

INSTYTUT INFORMATYKI TECHNICZNEJ
WYDZIAŁ INFORMATYKI I ZARZĄDZANIA
POLITECHNIKA WROCŁAWSKA

Raport serii PRE nr 01/2005

**Hybrydowa metoda rekomendacji
dokumentów w środowisku hipertekstowym**

(praca doktorska)

Maciej Kiewra

Wrocław 2005

Jestem ogromnie wdzięczny Promotorowi za nieocenioną opiekę merytoryczną i wskazówki, bez których nie powstałaby ta praca.

Składam także podziękowania prof. Adamowi Grzechowi za krytyczne uwagi i czas poświęcony tej pracy.

Szczególne dowody uznania należą się również mojej Żonie Ewie oraz Ojcu za tak potrzebne wsparcie i wiarę.

Spis treści:

Ważniejsze oznaczenia.....	4
1. Wstęp.....	6
2. Wybrane elementy teorii wyszukiwania informacji.....	11
2.1. Wyszukiwanie informacji – pojęcia ogólne.....	11
2.2. Badanie jakości systemów wyszukiwania.....	14
2.3. Podstawowe zagadnienia związane z hipertekstem.....	16
3. Przegląd metod rekomendacji.....	18
3.1. Sposoby klasyfikacji metod rekomendacji w środowisku hipertekstowym.....	18
3.2. Zunifikowane kryteria klasyfikacji systemów rekomendujących.....	22
3.3. Analiza systemów rekomendujących.....	24
4. Hybrydowa metoda rekomendacji.....	41
4.1. Wstępne definicje.....	41
4.2. Definicja metod składowych.....	47
4.3. Sformułowanie problemu.....	52
4.4. Ogólny opis hybrydowej metody rekomendacji.....	53
4.5. Definicja hybrydowej metody rekomendacji.....	56
4.6. Algorytm rekomendacji.....	66
5. Formalne własności metody rekomendacji.....	69
5.1. Ogólne własności metody rekomendacji.....	69
5.2. Ocena zdolności adaptacyjnych metody rekomendacji.....	79
5.3. Efektywność metody rekomendacji.....	98
5.4. Własności metod składowych.....	156
5.5. Wpływ wartości parametrów na zachowanie się metody rekomendacji.....	166
6. Eksperymentalna weryfikacja metody rekomendacji.....	175
6.1. Symulacyjne badanie efektywności.....	175
6.2. Zachowanie się metody w rzeczywistych systemach hipertekstowych.....	183
7. Podsumowanie.....	191
8. Literatura.....	194

Ważniejsze oznaczenia

Oznaczenie	Definicja lub pierwsze wystąpienie
α - parametr regulujący wpływ wartości funkcji rekomendacji w poprzednim kroku na obecną wartość funkcji rekomendacji	Definicja 4.5.3
β - parametr regulujący wpływ współczynnika wzorców nawigacyjnych na wartość funkcji rekomendacji	Definicja 4.5.3
δ - parametr regulujący wpływ podobieństwa do odwiedzanych dokumentów na wartość funkcji rekomendacji	Definicja 4.5.3
γ - parametr, który określa na ile obniży się wartość funkcji rekomendacji dla dokumentów zignorowanych w bieżącej sesji	4.5.3. Wektor dokumentów zignorowanych
ρ_s - parametr, który określa na ile obniży się wartość funkcji rekomendacji dla dokumentów odwiedzonych w bieżącej sesji	4.5.2 Wektor dokumentów odwiedzonych
c - wzorzec nawigacyjny (wektor średni grupy historycznych sesji)	4.2.2. Metoda rekomendacji oparta na zachowaniu użytkowników
C - grupa sesji historycznych, których wektory historycznych sesji są do siebie podobnie	4.2.2. Metoda rekomendacji oparta na zachowaniu użytkowników
d - dokument hipertekstowy	Definicja 4.1.1
$d_{\max}^{(i)}$ - dokument maksymalnie relewantny od kroku i	Definicja 5.2.3
D - zbiór dokumentów hipertekstowych	Definicja 4.1.1
$D_R^{(i)}$ - zbiór dokumentów relewantnych do potrzeb użytkownika w kroku i	4.1. Wstępne definicje
$E(SR, s_b)$ - efektywność systemu rekomendującego SR w bieżącej sesji s_b	4.1. Wstępne definicje
f - funkcja rekomendacji	Definicja 4.1.4
$\mathbf{g}^{(i)}$ - wektor dokumentów zignorowanych do kroku i	Definicja 4.1.2
m - liczba dokumentów rekomendowanych w każdym kroku	4.1. Sformułowanie problemu
m_r - liczba dokumentów relewantnych	Rozdział 5.3
n - liczba dokumentów w systemie hipertekstowym (liczność zbioru D)	4.1. Sformułowanie problemu
n_c - liczba grup powstałych w wyniku grupowania zbioru historycznych sesji	4.5.2 Wyznaczanie wzorców nawigacyjnych
n_{s_b} - długość bieżącej sesji s_b	Definicja 4.1.2
$\mathbf{nv}^{(i)}$ - wektor współczynników wzorców nawigacyjnych w kroku i	Definicja 4.5.2
nv_{\max} - maksymalny współczynnik wzorców nawigacyjnych dla dokumentów nirelewantnych	Definicja 5.3.2
nv_{\min} - minimalny współczynnik wzorców nawigacyjnych dla dokumentów relewantnych	Definicja 5.3.1
$prec(SR, s_b)$ - dokładność systemu rekomendującego SR w bieżącej sesji s_b	Definicja 4.1.5
$q(d)$ - funkcja przydatności dokumentu d	Definicja 4.5.1
R - zbiór wszystkich dokumentów zarekomendowanych w jednej sesji	4.1 Wstępne definicje

Oznaczenie	Definicja lub pierwsze wystąpienie
$R^{(i)}$ - zbiór dokumentów rekomendowanych użytkownikowi w kroku i	Definicja 4.1.2
$rec(SR, s_b)$ – kompletność systemu rekomendującego SR w bieżącej sesji s_b	Definicja 4.1.6
\mathbf{s} – wektor historycznej sesji	Definicja 4.1.3
$s_b^{(i)}$ – bieżąca sesja w kroku i	Definicja 4.1.2
S – zbiór sesji historycznych	4.1 Wstępne definicje
s_b – ciąg stanów sesji we wszystkich jej krokach	Definicja 4.1.2
$sim(\mathbf{a}, \mathbf{b})$ – podobieństwo między wektorami \mathbf{a} i \mathbf{b}	Definicja 2.1.2
sim_{nmax} – maksymalne podobieństwo dokumentów nerelevantnych	Definicja 5.3.4
sim_{rmin} – minimalne podobieństwo dokumentów relevantnych	Definicja 5.3.3
SR – system rekomendujący	Definicja 4.1.4
\mathbf{v}^d - wektor deskryptorów	Definicja 4.1.1
$\mathbf{z}^{(i)}$ - wektor dokumentów odwiedzonych do kroku i	Definicja 4.1.2

1. Wstęp

System rekomendujący (*ang. recommender system*), parafrazując definicję podaną w (Rashid i inni, 2002), to system, który sugeruje użytkownikowi potencjalnie interesujące go elementy na podstawie wiedzy o tym użytkowniku oraz informacji na temat przestrzeni dostępnych elementów. Metodą rekomendacji nazywamy ustrukturyzowany zespół czynności użytych przez system do dokonania rekomendacji. Głównym celem każdej metody rekomendacji jest przedstawienie użytkownikowi jak największej liczby elementów go interesujących (określanych także mianem relewantnych), przy całkowitym pominięciu elementów nierelewantnych (nie związanych z potrzebami użytkownika).

Systemy rekomendujące działają najczęściej w środowisku hipertekstowym. Idea hipertekstu zrodziła się 60 lat temu w pracy (Bush, 1945). Zakłada ona, że w treści dokumentów (nazywanych także stronami) znajdują się odsyłacze hipertekstowe, które prowadzą do innych dokumentów, co pozwala na swobodne przechodzenie pomiędzy nimi. Zarządzaniem oraz dostępem do pewnego zbioru dokumentów hipertekstowych zajmuje się system hipertekstowy. Przykładem systemu hipertekstowego jest witryna WWW. Przedmiotem niniejszej pracy są systemy rekomendujące współpracujące z dokładnie jednym systemem hipertekstowym.

Nadmiar i nieuporządkowanie informacji występującej w sieci Internet, a szczególnie w środowisku hipertekstowym WWW (*ang. World Wide Web*), jest problemem, na który często powołują się autorzy prac dotyczących metod rekomendacji dokumentów hipertekstowych (Moukas i Amalthea, 1997), (Chen i Sycara, 1998), (Cooley i inni, 1999), (Shahabi i inni, 2001), (Montaner i inni, 2003a). Nie umniejszając znaczenia zaawansowanych algorytmów wykorzystywanych przez te metody oraz skomplikowanych reprezentacji formalnych pozwalających przedstawić potrzeby i preferencje użytkownika, należałoby postawić sobie elementarne pytanie: jaki jest sens prac nad systemami rekomendacji, skoro istnieją globalne wyszukiwarki pozwalające w kilkanaście sekund dotrzeć do szukanej informacji niezależnie od witryny, w której się znajduje?

Pytanie jest tym bardziej zasadne, że na przestrzeni ostatniej dekady łatwo dostrzec wyraźną poprawę jakości globalnych systemów wyszukiwania dokumentów w środowisku WWW. Współczesne wyszukiwarki nie tylko indeksują znacznie większą liczbę dokumentów, lepiej porządkują wyniki wyszukiwania, ale także oferują swoim użytkownikom wiele użytecznych funkcji mających na celu łatwiejsze dotarcie do pożądanej informacji (do najpopularniejszych z nich należy zaliczyć: ograniczenie wyszukiwania do określonej witryny lub wybranego języka naturalnego oraz znalezienie wszystkich stron odsyłających do danego dokumentu).

Jest jednak jedna rzecz, której nie jest i nigdy nie będzie w stanie wykonać żadna z wyszukiwarek, a mianowicie: żadna z nich nie będzie mogła znaleźć zbioru dokumentów relewantnych, jeśli użytkownik nie sformułuje odpowiedniego zapytania. Innymi słowy, wszystkie zaawansowane technologie usprawniające indeksowanie dokumentów lub pozwalające lepiej uporządkować zbiór dokumentów relewantnych nie mogą zostać użyte, jeśli użytkownik nie wpisze dwóch lub trzech słów kluczowych charakteryzujących jego potrzeby. Oczywiście, istnieje cały szereg technik pozwalających poprawić zapytanie zadane przez użytkownika (*ang. query reformulation*) np. (Rocchio, 1971), (Singhal i inni, 1999), ale także te techniki uzależnione są od sformułowania zapytania.

Niechęć użytkownika do zadawania zapytań może mieć różnorodne przyczyny. Pierwsza z nich, najbardziej oczywista, to brak umiejętności w posługiwaniu się wyszukiwarką. Druga to

lenistwo użytkownika, który woli mniej lub bardziej świadomie wybierać kolejne dokumenty zamiast sformułować zapytanie. Trzecia z nich to trudność ujęcia swojej potrzeby informacyjnej w formie kilku słów kluczowych. Czwarty powód to niewiedza dotycząca istnienia w systemie elementów, które odpowiadają określonym potrzebom.

Zdaniem autora właśnie ten ostatni powód jest najsilniejszym argumentem uzasadniającym potrzebę prac nad systemami rekomendującymi, gdyż pokazuje on, że rekomendacja nie tylko ułatwia dotarcie do szukanej informacji, ale także pełni funkcję informacyjną, szczególnie, gdy wiedza użytkownika dotycząca interesującego go tematu jest mocno ograniczona. Funkcja informacyjna systemu rekomendującego spełnia dwie zasadnicze role. Po pierwsze, rekomendując dany dokument (zasób) powiadamia się użytkownika o posiadaniu informacji na dany temat. Przykładowo, nowy użytkownik systemu hipertekstowego poświęconego edukacji, może dowiedzieć się, dzięki rekomendacji, że system ten posiada dokumenty dotyczące rankingu szkół wyższych. Druga funkcja informacyjna pozwala uzupełnić wiedzę użytkownika na interesujący go temat. Na przykład, użytkownikowi czytającemu dokument hipertekstowy dotyczący recenzji filmu „Podwójne życie Weroniki” Krzysztofa Kieślowskiego system rekomendujący może zaproponować recenzje mniej znanych filmów tego reżysera takich jak „Przypadek” czy „Bez końca”. Warto zauważyć, że w obydwu przypadkach użytkownik nie byłby w stanie poprawnie sformułować zapytania bądź to z braku wiedzy na temat istnienia danego dokumentu w systemie (przykład, rankingu szkół wyższych), bądź to z niewiedzy o istnieniu danego obiektu w świecie rzeczywistym (film „Przypadek”).

System rekomendujący posiada wiele cech systemu wyszukiwania. Do najważniejszych należy zaliczyć:

- Wspólny cel – w obydwu typach systemów należy zidentyfikować i przedstawić zbiór dokumentów relewantnych do potrzeb użytkownika.
- Użycie rankingu - w obu przypadkach wynikowy zbiór dokumentów może zostać uporządkowany według malejącej relewancji do potrzeb użytkownika.
- Sprzężenie zwrotne od użytkownika (*ang. feedback*) - zarówno wyniki wyszukiwania jak i rekomendacji mogą ulec zmianie na skutek interakcji z użytkownikiem. W przypadku wyszukiwania będzie to modyfikacja zapytania, a w przypadku rekomendacji odwiedzenie kolejnego dokumentu lub jego ocena.

Cechy te powodują, że część autorów traktuje wyszukiwarki jako szczególny przypadek systemów rekomendujących, przykładowo: (Burke, 2002). W niniejszej pracy przyjmuje się jednak, że systemy wyszukiwania nie mogą zostać uznane za systemy rekomendujące z następujących powodów:

- W systemach wyszukiwania charakterystyka dokumentów relewantnych podana jest przez użytkownika w sposób bezpośredni (np. wyświetl wszystkie dokumenty zawierające słowa kluczowe: *sejm, poseł, ustawa*). Systemy rekomendujące zaś określają relewancję dokumentu na podstawie obserwacji zachowań użytkownika.
- Systemy wyszukiwania, w odróżnieniu od rekomendujących, oczekują na sformułowanie zapytania. W przypadku systemów rekomendujących użytkownik zazwyczaj nie musi dokonywać żadnych dodatkowych czynności, aby otrzymać potencjalnie interesujące go dokumenty (najczęściej są one dołączane do dokumentu pobieranego przez użytkownika).
- Systemy wyszukiwania wyświetlają użytkownikowi wszystkie dokumenty relewantne do sformułowanego zapytania (najczęściej w postaci listy stronicowanej), w od-

różnieniu od systemów rekomendujących, które proponują kilka dokumentów najlepiej odpowiadających potrzebom użytkownika.

Z drugiej jednak strony zbieżności pomiędzy rekomendacją oraz wyszukiwaniem powodują, że do oceny jakości obydwu typów systemów można stosować te same miary. Najważniejszą z nich jest efektywność będącą kombinacją: dokładności i kompletności¹. Pierwsza z nich opisuje zdolność systemu do rekomendowania dokumentów relewantnych, a druga określa ile różnych dokumentów relewantnych zaproponował system rekomendujący.

Analizując efektywność systemów rekomendujących należy mieć na uwadze fakt, że użytkownik w obrębie oglądanego właśnie dokumentu najczęściej „otrzymuje” 2-3 odsyłacze do potencjalnie interesujących go zasobów (nie trudno sobie wyobrazić zdumienia użytkownika na widok 50 sugerowanych odsyłaczy). Z tego właśnie powodu do osiągnięcia maksymalnej dokładności nie jest konieczne określenie wszystkich dokumentów relewantnych. „Wystarczy” wybrać tylko kilka z nich, o których wiadomo, że na pewno odpowiadają oczekiwaniom użytkownika. Nie jest to wcale zadaniem łatwym. Dlatego też od połowy lat 90 pojawiają się coraz to nowe pomysły pozwalające lepiej zidentyfikować te „kilka” najlepszych dokumentów (patrz rozdział 3).

Przeglądając listę publikacji, dołączoną do niniejszej pracy trudno oprzeć się wrażeniu, że „dobra” metoda rekomendacji to taka, która proponuje dokumenty relewantne. Jest to tylko część prawdy, gdyż ciągłe rekomendowanie tych samych elementów doprowadzi niechybnie do zubożenia użytkownika, na rekomendowaną treść. Takie niepożądane zachowanie systemu rekomendującego będzie, co prawda, cechować się wysoką dokładnością (proponowane są dokumenty relewantne), ale będzie charakteryzować się niską kompletnością, co w konsekwencji uniemożliwi systemowi rekomendującemu osiągnięcie wysokiej efektywności.

Dostępne prace nie przedstawiają wielu sposobów na uzyskanie wysokiej kompletności. Jeden z nielicznych pomysłów opiera się na unikaniu rekomendacji dokumentów już odwiedzonych przez użytkownika. W żadnej ze znanych autorowi prac², nie przeanalizowano jak powinien zachować się system rekomendujący, gdy zaproponowane dokumenty zostały zignorowane przez użytkownika. Problem nie jest banalny, gdyż, tak naprawdę, nieznane są rzeczywiste przyczyny, dla których rekomendacja może spotkać się z brakiem zainteresowania. Pierwsza z domniemyanych przyczyn to nierelewanca rekomendacji (podpowiedziano dokumenty nieodpowiadające potrzebom użytkownika). Z drugiej jednak strony, jest możliwe, że odbiorca uważa rekomendację za relewantną, ale przeglądając bieżący dokument znalazł inne odsyłacze, które wzbudziły jego większe zainteresowanie. Po trzecie jest też całkiem prawdopodobne, że użytkownik nie zauważył proponowanej treści.

Zdaniem autora, decyzja o ponownym zarekomendowaniu tego samego dokumentu, powinna sprowadzać się do sprawdzenia czy istnieją inne niezarekomendowane jeszcze dokumenty, które są na tyle relewantne, by można je podpowiedzieć użytkownikowi. Tutaj pojawia się kolejny problem: w miarę jak będzie się powiększał zbiór zarekomendowanych już dokumentów, które prawie „na pewno” są relewantne do potrzeb użytkownika, odnalezienie nierekomendowanego jeszcze dokumentu relewantnego będzie coraz trudniejsze. Z jednej bowiem strony istnieje niebezpieczeństwo obniżenia dokładności przez zarekomendowanie dokumentu nierelewantnego (w celu zwiększenia kompletności proponowane będą dokumenty coraz

¹ W niniejsze pracy przyjmuje się, że efektywność jest średnią geometryczną kompletności i dokładności.

² W (Kiewra, 2005b) autor niniejszej pracy przedstawił koncepcję, która będzie wykorzystana w metodzie rekomendacji zdefiniowanej w rozdziale 4.

bardziej „odległe” od potrzeb użytkownika). Z drugiej zaś, pominięcie nawet kilku dokumentów relewantnych przyczyni się do znacznego spadku kompletności.

Reasumując, metoda rekomendacji cechująca się dużą efektywnością powinna zawrzeć „mądry kompromis” pomiędzy dużą różnorodnością (wysoką kompletnością) podpowiadanej treści a jej relewancją (wysoką dokładnością).

Uwzględnienie problemu kompletności nie jest jedynym ważnym aspektem rekomendacji, który nie został należycie przeanalizowany w literaturze. Drugim z nich jest zmienność preferencji użytkownika. Jest oczywiste, że w związku z brakiem zapytań przesyłanych do systemu, informacja na temat odwiedzanych dokumentów jest jednym z podstawowych źródeł pozwalających poznać potrzeby użytkownika. Większość systemów rekomendujących przy ocenie relewancji danego dokumentu bierze pod uwagę wszystkie dokumenty odwiedzone podczas bieżącej sesji (lub ewentualnie kilka ostatnich). Na tej podstawie budowany jest tzw. profil użytkownika. Określenie profilu jest bardzo korzystne, szczególnie gdy użytkownik odwiedzi dokument, dla którego trudno dokonać rekomendacji lub który jest na tyle różny od wcześniej odwiedzonych dokumentów, że należy się spodziewać, że został on odwiedzony przez pomyłkę. Z drugiej jednak strony, odmienna treść ostatnio odwiedzonych dokumentów może być przejawem zmiany potrzeb informacyjnych użytkownika. Reasumując, system rekomendujący nie tylko powinien określić, które z dostępnych dokumentów są relewantne, ale także musi przewidzieć, kiedy zidentyfikowane przez niego dokumenty relewantne stają się nirelewantnymi. Z tych właśnie powodów oryginalne definicje kompletności i dokładności rekomendacji sformułowane w rozdziale 4, będą uwzględniać zmienność preferencji użytkownika.

Dwa wyżej wymienione problemy są motywem powstania niniejszej pracy, w ramach której przedstawiona zostanie metoda rekomendacji o nazwie *AdaptRank*. Biorąc pod uwagę wyniki najnowszych badań wskazujących, że łączenie kilku technik rekomendacji w tzw. *metodę hybrydową* daje wyższą dokładność niż metody składowe, opracowana metoda będzie opierać się na dwóch składowych często wykorzystywanych w rekomendacji. Pierwsza z nich zakłada, że dokumenty relewantne i odwiedzane przez użytkownika posiadają dużą liczbę wspólnych deskryptorów (słów kluczowych).

Druga składowa wykorzystuje następującą intuicję: użytkownicy mający podobne potrzeby informacyjne zachowują się w podobny sposób. Tak więc dokumenty relewantne są to takie dokumenty, które były często odwiedzane przez innych użytkowników, którzy zachowują się podobnie do „właściciela” bieżącej sesji. Do mierzenia jak często dany dokument był odwiedzany przez użytkowników zachowujących się podobnie używany będzie *współczynnik wzorców nawigacyjnych*, którego oryginalna definicja znajduje się w rozdziale 4.

Celem niniejszej pracy jest opracowanie i zbadanie własności hybrydowej metody rekomendacji dokumentów opartej na podobieństwie między dokumentami oraz zachowaniu użytkowników, która w wielu przypadkach pozwoli uzyskać większą efektywność rekomendacji od efektywności jej składowych.

W szczególności, w pracy zostanie formalnie udowodnione, a następnie zilustrowane eksperymentami, że zdefiniowana metoda hybrydowa, będzie się cechować wyższą efektywnością, gdy:

- Podobieństwo dokumentów relewantnych do dokumentów odwiedzanych w bieżącej sesji jest odpowiednio duże i wyraźnie wyższe od podobieństwa dokumentów nirelewantnych do dokumentów odwiedzanych w bieżącej sesji.

- Wartości współczynników wzorców nawigacyjnych dokumentów relewantnych są odpowiednio duże i wyraźnie wyższe od współczynników wzorców nawigacyjnych dla dokumentów nirelewantnych.

Wyższa efektywność będzie uzyskana niezależnie od tego, czy użytkownik ignorował czy też korzystał z rekomendowanych przez system dokumentów. Warto zauważyć, że powyższe założenia odpowiadają przypadkom, w których obydwie metody składowe działają poprawnie. Nie ma bowiem większego sensu analizować czy metoda hybrydowa osiągnie większą efektywność, gdy jej składowe nie są w stanie poprawnie zidentyfikować dokumentów relewantnych.

Należy podkreślić, że uzyskanie wyższej efektywności było możliwe dzięki wyraźnie wyższej kompletności. Jest to konsekwencją wprowadzenia oryginalnych mechanizmów zwiększających kompletność, które jednocześnie umożliwiają zachowanie dużej dokładności. Dodatkowo pokazano, że zdefiniowana metoda jest w stanie bardzo szybko dostosować się do zmieniających się potrzeb użytkownika.

Pozostałą część niniejszej pracy została podzielona na następujące rozdziały:

- **W rozdziale drugim** przedstawiono pojęcia i definicje w zakresie teorii wyszukiwania informacji znane z literatury. Rozdział ten nie tylko umożliwi uściślenie często wieloznacznych pojęć (np. relewancja), ale także pozwoli na łatwiejsze odróżnienie definicji znanych z literatury, od oryginalnych definicji wprowadzonych przez autora w dalszej części pracy.
- **W rozdziale trzecim** opisano różne metody klasyfikacji systemów rekomendujących wraz z krytyczną analizą metod rekomendacji znanych z literatury.
- **Rozdział czwarty** zawiera formalny model rekomendacji będący usystematyzowaniem i uporządkowaniem pojęć związanych z rekomendacją pojawiających się w literaturze. Przy użyciu tego modelu określony został problem, a także hybrydowa metoda rekomendacji *AdaptRank*, będąca jego rozwiązaniem.
- **Rozdział piąty** poświęcony jest formalnej analizie metody zdefiniowanej w poprzednim rozdziale. Określono w nim, między innymi, warunki, przy których hybrydowa metoda *AdaptRank* osiąga wyższą dokładność, kompletność i efektywność od dokładności, kompletności i efektywności jej metod składowych. Rozdział zakończony jest dyskusją na temat wpływu wartości parametrów na zachowanie się metody.
- **Rozdział szósty** zawiera badania symulacyjne nad efektywnością metody *AdaptRank* oraz opis systemu będącego wynikiem implementacji tej metody. Dodatkowo przedstawione zostały eksperymenty przeprowadzone na dwóch rzeczywistych systemach hipertekstowych.
- **Ostatni rozdział** poświęcony jest podsumowaniu wyników pracy oraz nakreśleniu głównych kierunków przyszłych badań.

2. Wybrane elementy teorii wyszukiwania informacji

Celem rozdziału jest przedstawienie pojęć i definicji w zakresie teorii wyszukiwania informacji znanych z literatury. Rozdział ten nie tylko umożliwi uściślenie często wieloznacznych pojęć (np. relewancja), ale także pozwoli łatwiej odróżnić ogólnie znane definicje, od oryginalnych definicji wprowadzonych przez autora w dalszej części pracy.

2.1. Wyszukiwanie informacji - pojęcia ogólne

Wyszukiwanie informacji (*ang. Information Retrieval*) zajmuje się wydobywaniem informacji, poszukiwaniem metadanych zawartych w dokumentach oraz znajdowaniem dokumentów spełniających określone warunki. Do odnajdywania dokumentów z pewnego zbioru dokumentów $D=\{d_1, d_2, \dots, d_n\}$ na podstawie kryteriów zadanych przez użytkownika służą systemy wyszukiwania. Kryteria te najczęściej nazywane są zapytaniem (*ang. query*). Przez P oznaczamy zbiór zapytań użytkowników systemu.

Do określania czy dany dokument spełnia oczekiwania użytkownika używa się terminu relewancji. Chociaż relewancja jest jednym z najważniejszych pojęć w teorii wyszukiwania informacji istnieją spore problemy z jej formalną definicją (Mizzaro, 1997), (Froehlich, 1994), (Schamber, 1990). Wyczerpujący przegląd koncepcji relewancji biorący pod uwagę 130 różnych publikacji związanych z relewancją można znaleźć w (Mizzaro, 1997). Na potrzeby niniejszej pracy przyjmuje się za (Bookstein, 1979) podział relewancji na subiektywną relewancję użytkownika oraz szacowaną relewancję systemu wyszukiwania określaną także skrótem RSV (*ang. Retrieval Status Value*). Pierwsza z nich jest pojęciem abstrakcyjnym oznaczającym subiektywne odczucie użytkownika co do przydatności danego dokumentu względem jego potrzeb informacyjnych. Przez $D_p \subseteq D$ oznaczany będzie podzbiór dokumentów relewantnych (w odczuciu użytkownika) do zadanego zapytania p .

Relewancja dokumentu do danego zapytania podana przez system wyszukiwania jest liczbą rzeczywistą obliczaną na podstawie funkcji wyszukiwania¹.

Definicja 2.1.1

Funkcją wyszukiwania $f_w: D \times P \rightarrow \mathfrak{R}$ nazywamy funkcję, która dla każdego dokumentu $d \in D$ i każdego zapytania $p \in P$ przyporządkowuje liczbę rzeczywistą określającą stopień relewancji dokumentu d do zadanego zapytania p (zakłada się, że im wartość funkcji jest większa tym dokument jest bardziej relewantny do zadanego zapytania).

Dokumenty spełniające kryteria wyszukiwania wyświetlane są użytkownikowi w postaci listy zwanej rankingiem dokumentów. Jest on uporządkowany malejąco według wartości funkcji wyszukiwania obliczonej dla zadanego zapytania i danego dokumentu. Przez $R_p \subseteq D$ oznaczać

¹ Funkcja ta w literaturze angielskiej określana jest także jako *rank function* (Zhang i Dong, 2000), *relevance function* (Bodoff, 2004), *matching score* (Sparck i inni, 2000), *retrieval status value* (Bookstein, 1979), *relevance score*. W niniejszej pracy przyjmuje się tłumaczenie dostępne w (Kłopotek, 2001), gdyż pozwoli ono na łatwe odróżnienie funkcji wyszukiwania (będącej funkcją relewancji używaną w wyszukiwaniu) od funkcji rekomendacji, która może zostać określona jako funkcja relewancji używana w rekomendacji.

będziemy podzbiór zbioru dokumentów D zawierający dokumenty wyświetlone użytkownikowi w odpowiedzi na zapytanie p .

Model wyszukiwania informacji określa sposób reprezentacji dokumentów, zapytań użytkownika oraz funkcji wyszukiwania. Na przestrzeni ostatnich 30 lat powstało wiele różnych modeli wyszukiwania informacji np. model logiczny (Lalmas, 1998) czy probabilistyczny (Rijsbergen, 1979), (Sparck, 2000).

W niniejszej pracy do zdefiniowania problemu rekomendacji użyty został najbardziej popularny model zwany modelem wektorowym, który został wprowadzony przez Saltona (Salton, 1983). Model ten zakłada, że dla zbioru D istnieje zbiór deskryptorów (słów kluczowych) w nich występujących, oznaczany przez $W = \{w_1, w_2, \dots, w_{n_w}\}$. Dokument d reprezentowany jest przez wektor deskryptorów $\mathbf{v}^d = (v_1^d, v_2^d, \dots, v_{n_w}^d)$, którego współrzędna v_j^d określa wagę deskryptora w_j w dokumencie d dla $j=1, 2, \dots, n_w$.

Współrzędna v_j^d wektora deskryptorów dokumentu d odpowiada istotności deskryptora w_j w dokumencie d . W obliczaniu wartości v_j^d stosuje się znaną formułę Saltona, częstość terminu – odwrotna częstość wystąpień w dokumentach (*ang. term frequency - inverse document frequency*) (Salton, 1989):

$$v_j^d = wf_j^d \cdot \log\left(\frac{\text{card}(D)}{n_d}\right) \quad (2.1.1)$$

gdzie: wf_j^d oznacza częstość deskryptora w_j w dokumencie d , a n_d to liczba dokumentów ze zbioru D , które zawierają deskryptor w_j . Formuła Saltona „promuje” deskryptory, które występują z dużą częstotliwością w małej liczbie dokumentów (deskryptory występujące w dużej liczbie dokumentów najczęściej niosą ze sobą szum informacyjny i nie ułatwiają wydobycia unikatowych cech dokumentów).

W modelu wektorowym także zapytanie reprezentowane jest jako wektor deskryptorów $\mathbf{v}^p = (v_1^p, v_2^p, \dots, v_{n_w}^p)$, którego współrzędna v_j^p określa wagę deskryptora w_j w zapytaniu p .

Do obliczania wartości funkcji wyszukiwania w modelu wektorowym używa się podobieństwa dwóch wektorów. W niniejszej pracy będzie się używać cosinusowej miary podobieństwa (Salton, 1989):

Definicja 2.1.2

Podobieństwem między wektorami $\mathbf{a}=(a_1, a_2, \dots, a_m)$ i $\mathbf{b}=(b_1, b_2, \dots, b_m)$, których współrzędne są liczbami rzeczywistymi, nazywany jest cosinus kąta utworzonego przez wektory \mathbf{a} i \mathbf{b}

$$\text{sim}(\mathbf{a}, \mathbf{b}) = \cos(\mathbf{a}, \mathbf{b}) = \frac{\sum_{j=1}^m a_j \cdot b_j}{\sqrt{\sum_{j=1}^m a_j^2} \cdot \sqrt{\sum_{j=1}^m b_j^2}}$$

O wybraniu cosinusowej miary podobieństwa zdecydowały następujące czynniki:

- Niezależność wielkości podobieństwa od długości wektorów (jedynie od ich kierunku i zwrotu)

- Zerowe wartości tych samych współrzędnych nie powodują wzrostu podobieństwa między wektorami. Wektory, pomiędzy którymi liczone będzie podobieństwo (wektory deskryptorów oraz wektory historycznych sesji) posiadają dużą liczbę współrzędnych o wartościach równych 0, wskutek czego użycie np. funkcji odległości euklidesowej jako miary „niepodobieństwa” spowodowałoby, że podobieństwo między wektorami zawierającymi dużą liczbę zerowych współrzędnych mogłoby być niewspółmiernie zawyżone.

Miara przedstawiona w definicji 2.1.2 będzie używana do określania podobieństwa między dokumentami oraz do obliczania podobieństwa między wektorem dokumentów odwiedzonych do kroku i (definicja 4.1.2) a wzorcami nawigacyjnymi przedstawionymi w definicji 4.5.2. Dodatkowo podobieństwo wykorzystywane będzie do grupowania wektorów historycznych sesji (proces ten jest opisany w punkcie 4.2.2).

Sprzężenie zwrotne relewancji (*ang. relevance feedback*) (Rocchio, 1971) jest często używaną techniką pozwalającą na automatyczne przededefiniowanie zapytania na podstawie ocen relewancji dokonanych przez użytkownika, a dotyczących dokumentów zwróconych przez system wyszukiwania. Załóżmy, że użytkownik po sformułowaniu zapytania w postaci wektora terminów $\mathbf{v}^p = (v_1^p, v_2^p, \dots, v_{n_v}^p)$ przejrzy kilka dokumentów ze zbioru dokumentów $R_p \subseteq D$ wyświetlonych przez system i wskaże przynajmniej jeden dokument relewantny i przynajmniej jeden nirelewantny. Wtedy współrzędna $v_j^{p'}$ wektora terminów $\mathbf{v}^{p'}$ odpowiadającego zmodyfikowanemu zapytaniu przyjmie postać:

$$v_j^{p'} = \alpha v_j^p + \beta \sum_{d_j \in D_p^R} \frac{\mathbf{v}^{d_j}}{\text{card}(D_p^R)} - \gamma \sum_{d_j \in D_p^N} \frac{\mathbf{v}^{d_j}}{\text{card}(D_p^N)} \quad (2.1.2)$$

gdzie: $D_p^R \subseteq R_p$ to zbiór dokumentów odwiedzonych przez użytkownika i uznanych za relewantne, a $D_p^N \subseteq R_p$ to zbiór dokumentów odwiedzonych przez użytkownika i uznanych za nirelewantne, α , β i γ to parametry regulujące odpowiednio wpływ poprzedniego zapytania, dokumentów ocenionych jako relewantne i dokumentów ocenionych jako nirelewantne.

Warto też nadmienić, że składnik $\sum_{d_j \in D_p^R} \frac{\mathbf{v}^{d_j}}{\text{card}(D_p^R)}$ nazywany jest sprzężeniem zwrotnym do-

datnim (*ang. positive relevance feedback*), a składnik $\sum_{d_j \in D_p^N} \frac{\mathbf{v}^{d_j}}{\text{card}(D_p^N)}$ nazywany jest sprzę-

żeniem zwrotnym ujemnym (*ang. negative relevance feedback*). Odpowiednio zmodyfikowane sprzężenie zwrotne relewancji będzie używane w metodzie rekomendacji przedstawionej w rozdziale 4.

Współczesne systemy wyszukiwania przy porządkowaniu rankingu dokumentów relewantnych oprócz podobieństwa między zapytaniem i danym dokumentem wykorzystują także jakość dokumentu, która nie zależy od zapytania. Do oszacowania jakości dokumentu używa się bądź to informacji na temat odsyłaczy wskazujących na ten dokument (Brin i Page, 1998), (Kleinberg, 1999) (im więcej dobrych dokumentów wskazuje na dany dokument tym musi być on lepszy) bądź statystyk odwiedzin (WWW, 2005b), (Zhu i Gauh, 2000) (im częściej dany dokument jest odwiedzany tym większa jest jego wartość).

2.2. Badanie jakości systemów wyszukiwania

Subiektywna miara relewancji danego dokumentu d w odczuciu użytkownika może różnić się znacznie od wartości funkcji wyszukiwania, dlatego w wyszukiwaniu informacji pojawiły się wielkości pozwalające mierzyć jakość systemów wyszukiwawczych. Najczęściej stosowana jest dokładność oraz kompletność (Rijsbergen, 1979):

Definicja 2.2.1

Dokładnością wyszukiwania (*ang. precision*) nazywamy stosunek liczby wyszukanych dokumentów relewancych do liczby wszystkich wyszukanych dokumentów:

$$prec(p) = \frac{card(R_p \cap D_p)}{card(R_p)}$$

Definicja 2.2.2

Kompletnością wyszukiwania (*ang. recall*) nazywany jest stosunek liczby wyszukanych dokumentów relewancych do liczby dokumentów relewancych znajdujących się w systemie:

$$rec(p) = \left(\frac{card(R_p \cap D_p)}{card(D_p)} \right)$$

Kompletność i dokładność często są reprezentowane przez ujednoczoną miarę zwaną efektywnością. W literaturze istnieje kilka sposobów wyznaczania efektywności (Heine, 1973), (Meadow, 1992), (Rijsbergen, 1979), (Voiskunskii, 1997), (Lewis, 1994). Najważniejsze z nich zostały zebrane w poniższej tabeli.

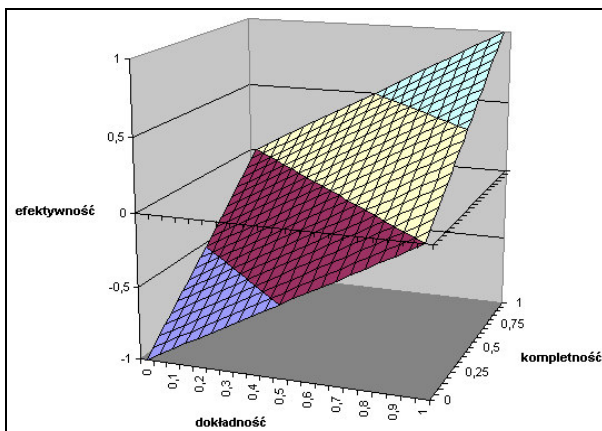
Autor	Miara	Uwagi
Borko	$E_B = rec + prec - 1$	Najprostsza z przedstawionych miar.
Meadow	$E_M = 1 - \frac{\sqrt{(1 - prec^2) + (1 - rec^2)}}{\sqrt{2}}$	Rosnąca funkcja wklęsła.
Heine	$E_H = 1 - \frac{1}{\frac{1}{rec} + \frac{1}{prec} - 1}$	Im większa wartość E_H tym mniejsza efektywność.
Vickery	$E_{Vi} = 1 - \frac{1}{\frac{2}{rec} + \frac{2}{prec} - 3}$	Podobna do miary wprowadzonej przez Heine. Ze względu na różnice w mianowniku wartości efektywności są trochę większe niż w przypadku miary wprowadzonej przez Heine.
Rijsbergen	$E_R = \frac{1}{\alpha \cdot \frac{1}{prec} + \frac{1}{rec} (1 - \alpha)}$	Uogólnienie miar zaprezentowanych przez Vickery i Heine. Współczynnik α mierzy relatywną wartość dokładności w porównaniu do kompletności.
Voiskunski	$E_V = \sqrt{prec \cdot rec}$	Średnia geometryczna dokładności i kompletności.
Lewis	$E_L = \frac{(b^2 + 1) \cdot prec \cdot rec}{b^2 \cdot prec + rec}$	Parametr b kontroluje wpływ kompletności i dokładności (wraz ze wzrostem b rośnie znaczenie kompletności; dla $b=1$ kompletność i dokładność są tak samo ważne)

Tabela 2.2.1 Najważniejsze miary określające efektywność systemów wyszukiwania.

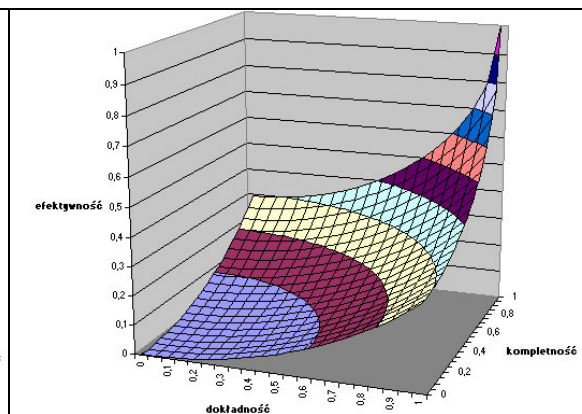
Zmodyfikowane definicje kompletności i dokładności będą używane do oceny rekomendacji (patrz rozdział 4). Na potrzeby niniejszej pracy przyjmuje się definicje efektywności zaproponowaną przez Voiskunskiego. Na jej wybór złożyły się następujące czynniki:

- Niskie wartości dla rekomendacji, w których kompletność rośnie kosztem dokładności lub dokładność wzrasta przy wyraźnym spadku kompletności. W pierwszym przypadku można mieć podejrzenie, że rekomendowana jest losowo dobierana treść. W drugim zaś należy domniemywać, że podpowiadany jest stale jeden lub kilka „sprawdzonych” dokumentów. Może to doprowadzić do szybkiego zubożnienia użytkownika, który w konsekwencji przestanie reagować na rekomendowaną zawartość.
- Dla dwóch rekomendacji, których suma dokładności i kompletności jest taka sama, wyższą efektywnością odznaczać się będzie rekomendacja, w której wartości kompletności i dokładności są zbliżone. W ten sposób promowane będą systemy, które potrafią „znaleźć” rozsądny kompromis pomiędzy podpowiadaniem wszystkiego, co w jakikolwiek sposób odnosi się do interesującego zagadnienia a rekomendowaniem tylko kilku elementów, o których wiadomo, że na pewno są związane z szukaną treścią.
- Prostota obliczeń.

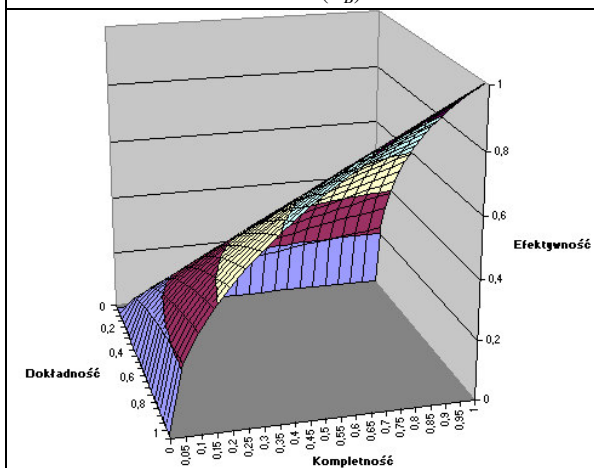
Na rysunku 2.2.1 przedstawiono wykresy efektywności w zależności od dokładności i kompletności.



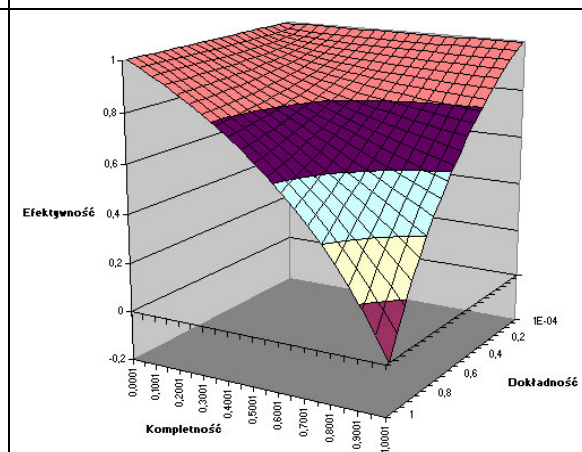
Rysunek 2.2.1a Wykres miary wprowadzonej przez Borko (E_B)



Rysunek 2.2.1b Wykres miary wprowadzonej przez Meadow (E_M)



Rysunek 2.2.1c Wykres miary, którą wprowadził Voiskunskii (E_V)

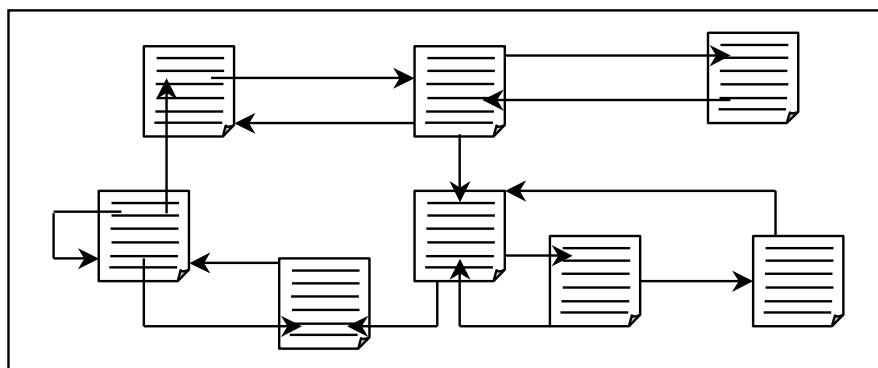


Rysunek 2.2.1d Wykres miary, którą wprowadził Heine (E_H)

2.3. Podstawowe zagadnienia związane z hipertekstem

Jak już zostało wspomniane w poprzednim rozdziale, idea hipertekstu zrodziła się 60 lat temu w pracy (Bush, 1945). Zakłada ona, że dokumenty zawierają w swojej treści odsyłacze do innych dokumentów, co pozwala na swobodne przechodzenie pomiędzy nimi. Zazwyczaj tych połączeń jest na tyle dużo, że w ramach pewnego systemu hipertekstowego (rozumianego tutaj jako grupa dokumentów hipertekstowych) możliwe jest „dotarcie” z dowolnego dokumentu do dowolnego innego dokumentu korzystając tylko i wyłącznie z odsyłaczy. Należy jednak zauważyć, że wybieranie odsyłaczy nie jest jedyną dostępną metodą poruszania się po współczesnych systemach hipertekstowych (np. użytkownik może przejść do kolejnego dokumentu poprzez wyszukiwarkę lub podając adres dokumentu z listy ulubionych).

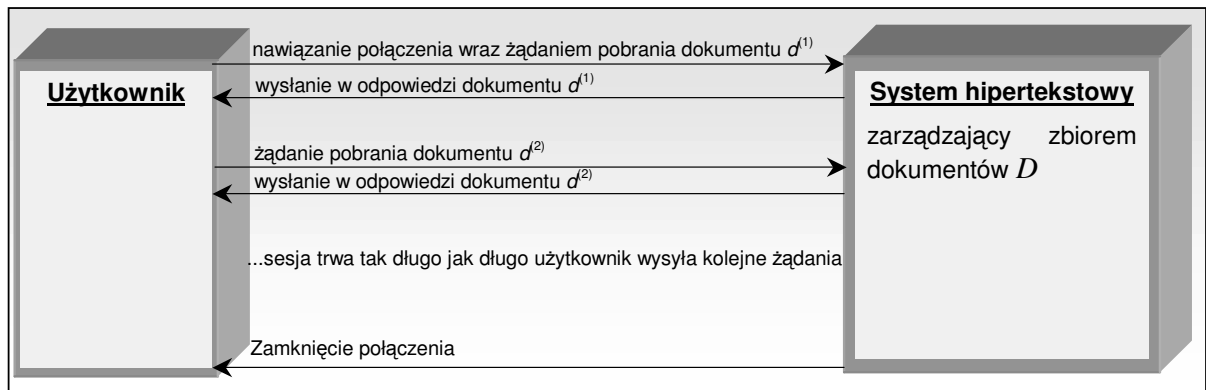
Parafrazując definicję podaną w (Nielsen, 1990), hipertekst jest sposobem reprezentacji informacji za pomocą grafu skierowanego, którego węzły (dokumenty hipertekstowe) zawierają pewne porcje informacji (najczęściej w postaci tekstu), a krawędzie nazywane odsyłaczami hipertekstowymi pozwalają na swobodne przemieszczanie się pomiędzy węzłami (rysunek 2.2.2). Każdy odsyłacz łączy dokument źródłowy (*ang. source node*) z dokumentem docelowym (*ang. target node*).



Rysunek 2.2.2 Graficzna reprezentacja hipertekstu.

Zarządzaniem oraz dostępem do zbioru dokumentów hipertekstowych D wraz z odsyłaczami zajmuje się system hipertekstowy. Typowym przykładem systemu hipertekstowego jest witryna WWW. Dokumenty wchodzące w skład danego systemu hipertekstowego mogą zawierać odsyłacze hipertekstowe do dokumentów znajdujących się w innych systemach hipertekstowych.

Użytkownik komunikuje się z systemem hipertekstowym przesyłając zapytania pobrania dokumentów. Jedno zapytanie dotyczy dokładnie jednego dokumentu $d \in D$ i zawiera identyfikator jednoznacznie determinujący dokument d . W odpowiedzi na zapytanie system hipertekstowy zwraca stosowny dokument. Zakłada się, że przed przesłaniem pierwszego zapytania użytkownik musi uzyskać połączenie z systemem hipertekstowym (w przypadku witryny WWW nawiązanie połączenia jest równoważne uruchomieniu przeglądarki i wpisaniu adresu URL (WWW, 2005a) witryny z ewentualną ścieżką dostępu do konkretnego dokumentu). Po zakończeniu pracy z systemem hipertekstowym następuje zamknięcie połączenia (w przypadku witryny WWW połączenie zostaje zakończone z chwilą zamknięcia przeglądarki). Sekwencja zapytań pobrania dokumentów przesłanych w ramach jednego połączenia nazywana będzie sesją (patrz rysunek 2.2.3). Każde pobranie dokumentu nazywane będzie krokiem sesji.



Rysunek 2.2.3. Schemat komunikacji użytkownika z systemem hipertekstowym.

Należy zauważyć, że system hipertekstowy może posiadać dodatkowe mechanizmy usprawniające komunikację z użytkownikiem. Przykładowo, jeśli system hipertekstowy zawiera wbudowany system wyszukiwania, to oprócz przesyłania żądań pobrania dokumentów, możliwe jest także przesyłanie zapytań wspomnianych w podrozdziale 2.1. W tym przypadku w odpowiedzi na zapytanie system hipertekstowy, a ściślej mówiąc system wyszukiwawczy, przedstawia wynik wyszukiwania w postaci dokumentu zawierającego ranking odsyłaczy do dokumentów, które są relewantne do zadanego zapytania. Warto podkreślić, że wyszukiwanie jest procesem pomocniczym w stosunku do tradycyjnej komunikacji żądanie – odpowiedź, a jego głównym celem jest przedstawienie listy identyfikatorów (podanych w postaci odsyłaczy) pozwalających użytkownikowi na wysłanie kolejnych żądań pobrania dokumentów.

3. Przegląd metod rekomendacji

Celem rozdziału 3 jest przedstawienie obecnego stanu wiedzy na temat systemów rekomendujących działających w środowisku hipertekstowym. W celu uporządkowania pojęć używanych przy okazji charakteryzowania systemów rekomendujących, dokonano krótkiego przeglądu metod klasyfikacji tych systemów. W dalszej części rozdziału scharakteryzowano kryteria pozwalające określić niemal każdy aspekt działania systemu rekomendującego. Krytyczna analiza rozwiązań znanych z literatury przeprowadzona została właśnie pod kątem tych kryteriów.

3.1. Sposoby klasyfikacji metod rekomendacji w środowisku hipertekstowym

Pierwsze systemy rekomendujące pojawiły się w latach dziewięćdziesiątych wraz z upowszechnieniem się Internetu w środowisku akademickim np. (Goldberg i inni, 1992), (Lieberman, 1995). Różnorodność metod i systemów rekomendujących opisywanych rokrocznie w literaturze fachowej zaowocował publikacjami o charakterze porównawczym, których głównym celem było przedstawienie sposobów klasyfikacji systemów rekomendujących oraz zdefiniowanie ich podstawowych cech. W (Montaner i inni, 2003a) zaproponowano 8 podstawowych kryteriów, przy użyciu których scharakteryzowano 37 różnych systemów rekomendujących. Do cech tych zaliczono:

- **Dziedzinę** (*ang. domain*) – określa rodzaj rekomendowanych elementów (np. płyty w sklepie internetowym, dokumenty, najnowsze informacje itp.).
- **Reprezentację profilu użytkownika** (*ang. user profile representation*) - charakteryzuje struktury danych, jakimi posługuje się system do przechowywania informacji na temat profilu użytkownika (wektor binarny, historia zakupów, zbiór atrybutów itp.).
- **Stan początkowy profilu** (*ang. initial profile generation*) – określa jaka jest zawartość nowego profilu użytkownika. Może być on pusty, wygenerowany na podstawie wypełnionej ankiety albo utworzony w oparciu o dane dotyczące istniejących profili.
- **Generowanie profilu użytkownika** (*ang. profile learning technique*) – charakteryzuje sposób, w jaki budowany jest profil (grupowanie, indeksowanie, selekcja terminów, sieć neuronowa itp.)
- **Sprzężenie zwrotne** (*ang. feedback*) - determinuje, w jaki sposób użytkownik komunikuje się z systemem. Komunikację można podzielić na bezpośrednią (wypełnianie ankiet, wyrażanie opinii o widzianych elementach) lub pośrednią (system analizując zachowanie użytkownika buduje jego profil).
- **Adaptację profilu** (*ang. profile adaptation technique*) - pozwala prześledzić jak formalna reprezentacja profilu dostosowuje się do zmian preferencji użytkownika. W najbardziej trywialnym przypadku zmiany wprowadzone są na bezpośrednie życzenie użytkownika. Bardziej zaawansowanym rozwiązaniem jest aktualizacja profilu poprzez dodanie informacji pochodzących z kolejnych interakcji z systemem.

- **Filtrowanie** (*ang. filtering*) określa rodzaj danych, które służą do budowania profilu. Autorzy wyróżnili cztery rodzaje filtrowania: filtrowanie zawartości, filtrowanie użytkowników ze względu na ich cechy demograficzne lub zbudowany profil oraz filtrowanie hybrydowe. Szczegółowy opis każdego z nich znajduje się w punkcie 3.3.2.
- **Dopasowywanie rekomendowanych elementów** – (*ang. matching*) opisuje sposób, w jaki system rekomendujący wybiera elementy, które będą zarekomendowane (podobieństwo cosinusowe, klasyfikacja w oparciu o sieci neuronowe, porównywanie grafów itp.).

Klasyfikacja podana w (Montaner i inni, 2003a) nie jest jedyną. (Burke, 2002) oprócz podania typów systemów hybrydowych (patrz podpunkt 3.3.2.3) dzieli systemy rekomendujące ze względu na:

- **Stosowaną technikę** – autor określa tym terminem rodzaj danych pozwalających zbudować profil użytkownika. Cecha ta jest odpowiednikiem filtrowania zdefiniowanego w (Montaner i inni, 2003a) z tą tylko różnicą, że autor definiuje dwie dodatkowe techniki. Pierwsza z nich oparta jest na **użyteczności dla użytkownika**. Użyteczność ta pozwala na zarekomendowanie elementów bazując na atrybutach, które nie opisują danego elementu w sposób bezpośredni (przykładowo reputacja sprzedawcy lub dostępność w magazynie). Druga z technik opiera się na **wiedzy** i zakłada, że system rekomendujący posiada dodatkowe źródło informacji. Do tej grupy autor zakwalifikował techniki wnoszące na bazie przykładów (*ang. case based reasoning*) oraz systemy wyszukiwania, gdyż posiadają one dodatkową wiedzę na temat potrzeb użytkownika w postaci zapytania.
- **Dane posiadane przed rekomendacją** nazwane przez autora tłem (*ang. background*).
- **Dane wejściowe wprowadzane przez bieżącego użytkownika** – są one odpowiednikiem sprzężenia zwrotnego w (Montaner i inni, 2003a).
- **Proces rekomendacji** – definiuje algorytm użyty w rekomendacji

W (Resnick i Varian, 1997) podzielono wszystkie cechy systemów rekomendujących na trzy grupy. Pierwsza z nich zawiera następujące charakterystyki projektowe:

- **Rodzaj danych otrzymywanych od użytkownika** - przykładowo 7 stopniowa skala ocen, fakt odwiedzenia danego dokumentu, zakup produktu itp.
- **Sposób pobierania danych** – odpowiednik sprzężenia zwrotnego wyróżnionego przez Montanera (bezpośredni lub pośredni).
- **Anonimowość użytkownika** – określa, w jaki sposób identyfikowany jest użytkownik (anonimowy identyfikator, pseudonim, numer dowodu tożsamości itp.).
- **Agregacja profilu użytkownika** – determinuje, w jaki sposób dane użytkownika zgromadzone przez system są zamieniane w profil (np. integracja zawartości tekstowej z ocenami najbliższych profili za pomocą wag itp.).
- **Wykorzystanie rekomendacji** – określa, w jaki sposób wykorzystywany jest zbudowany profil. Przykładowo, ocenienie przydatności danych elementów dla użytkownika może zaowocować przedstawieniem listy najbardziej relewantnych pozycji lub wyeliminowaniu pozycji nirelewantnych.

Druga grupa cech wymienionych w (Resnick i Varian, 1997) związana jest z charakterystyką rekomendowanych elementów, a wyróżniono w niej:

- **Przedmiot rekomendacji** - (produkty, dokumenty, najnowsze wiadomości, a nawet ludzie).
- **Liczba rekomendowanych dokumentów** – jest to ważne kryterium, gdyż innymi charakterystykami powinien odznaczać się system rekomendujący restauracje w danym mieście (zakłada się, że zbiór restauracji posiada od kilkudziesięciu do kilkuset elementów), a innymi system podpowiadający interesujące dokumenty w portalu internetowym zawierającym 100 000 zaindeksowanych stron.
- **Szybkość zmian** – niektóre systemy cechują się dużą zmiennością (przykładowo rekomendacja najnowszych wiadomości). W innych systemach nowe pozycje pojawiają się stosunkowo rzadko (rekomendacja kin lub teatrów).
- **Struktura kosztu** – próbuje ocenić jaki jest koszt niezarekomendowania relewantnego dokumentu w stosunku do zarekomendowania nierelewantnego. Autorzy mają tu na myśli problem ustalenia priorytetów pomiędzy dużą dokładnością kosztem małej kompletności oraz dużą kompletnością przy małej dokładności. Warto nadmienić, że problem ten został już zauważony w wyszukiwaniu informacji (Rijsbergen, 1979).

Ostania grupa cech podana przez autorów (Resnick i Varian, 1997) odnosi się do charakterystyki użytkowników. Wyróżniono w niej następujące kryteria:

- **Dostawcy rekomendacji** – określa rodzaj użytkowników, których gusta lub aktywność wykorzystywane są w bieżącym procesie rekomendacji.
- **Gęstość rekomendacji** – determinuje jak dużo użytkowników wpływa na przeciętną rekomendację systemu.
- **Rodzaj konsumentów rekomendacji** – identyfikuje odbiorców rekomendacji.
- **Zmienność gustów** – określa, z jaką częstotliwością dany użytkownik zmienia swoje zainteresowania (nigdy, rzadko, kilkakrotnie podczas jeden sesji).

Kończąc przegląd metod klasyfikacji systemów rekomendujących nie sposób nie wspomnieć o klasyfikacji systemów rekomendujących produkty w handlu internetowym zaproponowanej w (Schafer i inni, 2001). Chociaż praca ta nie uwzględnia rekomendacji dokumentów, to zaproponowane kryteria klasyfikujące mogą być z powodzeniem wykorzystane w tej dziedzinie rekomendacji. Jej autorzy zaproponowali metodę klasyfikacji opartą na wejściu i wyjściu systemów z osobnym uwzględnieniem sposobu rekomendacji i aspektów projektowych. Właśnie sposób rekomendacji zasługuje na szczególną uwagę z racji jego oryginalności. Kryterium to dzieli wszystkie metody rekomendacji na 7 następujących kategorii:

- **Proste pozyskiwanie informacji** (*ang. raw retrieval*) – używane jest w systemach, które komunikują się z użytkownikiem przez interfejs wyszukiwawczy. Przykładowo, aby uzyskać listę „rekomendowanych” albumów danego zespołu muzycznego należy sformułować zapytanie zawierające nazwę zespołu. Ze względu na charakter tych systemów autorzy proponują zamienną nazwę „rekomendacja zerowa” (*ang. null recommendation*).
- **Rekomendacja nieautomatyczna** – wykorzystywana jest w systemach, które korzystają z preferencji ekspertów danej społeczności. Dobrym przykładem tego rodzaju rekomendacji jest lista filmów stulecia opracowana przez znanego krytyka filmowego.

Warto nadmienić, że metoda ta nie wykorzystuje żadnych algorytmów ani nie wymaga żadnych dodatkowych obliczeń.

- **Podsumowania statystyczne** – metoda ta spotykana jest w systemach, które nie dostarczają żadnego rodzaju personalizacji. Systemy te rekomendują elementy bazując na sporządzonych statystykach, które są takie same dla wszystkich użytkowników (np. 10 najlepiej sprzedających się książek).
- **Rekomendacja oparta na atrybutach** – charakteryzuje systemy rekomendujące elementy o atrybutach podobnych do atrybutów elementów, o których wiadomo, że interesują danego użytkownika (np. zważywszy na fakt, że dany klient przegląda w sklepie internetowym sekcje poświęconą muzyce poważnej można zaproponować mu kupno zebranych dzieł Bacha).
- **Korelacja między produktami** – stosowana jest w systemach, które podpowiadają elementy będące w ścisłym powiązaniu z elementami, którymi zainteresowany jest użytkownik. Przykładowo, jeśli aparaty fotograficzne są często kupowane razem ze statywami, to jest wskazane zaproponować kupno statywu osobie, która w swoim wirtualnym koszyku umieściła aparat fotograficzny (podobnie, jeśli użytkownik ogląda dokument A, wskazanym byłoby zarekomendować mu dokument często odwiedzany z dokumentem A).
- **Korelacja między użytkownikami** – przeznaczona jest dla systemów, w których rekomendowane są elementy, które zostały pozytywnie ocenione przez użytkowników o profilach podobnych do profilu, dla którego przeprowadza się rekomendację. W tej kategorii zawierają się metody oparte na filtrowaniu kolaboratywnym oraz demograficznym (Montaner i inni, 2003a).
- **Metody hybrydowe** – odnosi się do systemów, które wykorzystują kilka z powyżej opisanych metod.

Biorąc pod uwagę cechy wejścia i wyjścia systemów rekomendujących zaproponowanych w (Schafer i inni, 2001) możemy wyróżnić następujące charakterystyki:

- **Dane wejściowe użytkownika docelowego** – określają rodzaj danych uzyskiwanych od użytkownika, dla którego przeprowadzana będzie rekomendacja (ankieta, ocena rekomendowanej treści, odwiedzane odsyłacze, historia zakupów itp.).
- **Dane wejściowe społeczności użytkowników** – charakteryzuje dane, jakie otrzymuje system rekomendujący od całej społeczności użytkowników. Najczęściej są to te same dane, które pobierane są od użytkownika docelowego
- **Dane wyjściowe (rezultat rekomendacji)** – determinuje, w jaki sposób rekomendowane elementy są przedstawiane użytkownikowi (sugestia wpleciona w prezentowaną treść, spersonalizowana reklama przy użyciu poczty elektronicznej itp.).

Na zakończenie opisu klasyfikacji dostępnej w (Schafer i inni, 2001) przedstawione zostaną kryteria charakteryzujące architekturę systemów rekomendujących:

- **Sposób dokonywania obliczeń** – określa, kiedy dokonywane są obliczenia związane z rekomendacją. W ramach tego kryterium można wyróżnić trzy rodzaje systemów. W pierwszym z nich wszystkie wielkości obliczane są w czasie rzeczywistym podczas pracy z użytkownikiem docelowym (*ang. on-line*). W drugim rodzaju systemów część obliczeń dokonywana jest w czasie rzeczywistym, a część z nich jest wykonywana periodycznie w czasie zmniejszonej aktywności systemu (*ang. off-line*). Trzecia nieco

hipotetyczna możliwość zakłada, że całokształt obliczeń dokonywany jest w czasie zmniejszonej aktywności systemu (może być przydatne tylko i wyłącznie dla systemów, które prezentują rekomendacje niespersonalizowane np. lista 10 najlepiej sprzedających się książek).

- **Stopień personalizowania** – determinuje jak bardzo zindywidualizowana jest przedstawiana rekomendacja. Autorzy proponują wyznaczyć 3 stopnie spersonalizowania systemów rekomendujących: **systemy niespersonalizowane** obejmują wszystkie te systemy, w których rekomendacja nie zależy ani od zachowania ani od charakterystyki użytkownika; w **personalizacji ulotnej** rekomendacja zależy wyłącznie od zachowania użytkownika (przykładowo, historia zakupów lub analiza razem odwiedzonych dokumentów); w **personalizacji trwałej** rekomendowana treść zależy zarówno od zachowania jak i charakterystyki użytkownika. Przykładowo, dwóch użytkowników, którzy obejrżeli te same elementy uzyskają różne rekomendacje, jeśli ich dane osobowe są różne (płeć, wykształcenie, wiek).
- **Dostarczanie** – określa czy rekomendowana treść dostarczana jest, gdy użytkownik dokonuje interakcji z systemem (*ang. pull technologies*) czy też gdy użytkownik jest nieaktywny (*ang. push technologies*).

Zunifikowana lista kryteriów klasyfikujących systemy rekomendujące pozwala na przeprowadzenie przejrzystej dyskusji na temat niemalże każdego aspektu ich funkcjonowania. Należy jednak zauważyć, że żadna z wyżej wspomnianych metod klasyfikacji nie została przygotowana tylko i wyłącznie dla systemów rekomendujących dokumenty hipertekstowe. Dodatkowo każda z klasyfikacji proponuje kryteria pomijane w konkurencyjnych podejściach. Z tych właśnie powodów następujący podrozdział poświęcony będzie przedstawieniu zunifikowanych kryteriów klasyfikacji systemów rekomendujących, który pozwoli na przeprowadzenie krytycznej analizy metod rekomendacji dostępnych w literaturze.

3.2. Zunifikowane kryteria klasyfikacji systemów rekomendujących.

W niniejszym podrozdziale zaproponowane zostaną zunifikowane kryteria klasyfikacji systemów rekomendujących. Autor sugeruje podzielić wszystkie kryteria na trzy grupy dotyczące: ogólnych cech systemu rekomendacji, zarządzania profilem użytkownika oraz architektury systemu.

Do pierwszej grupy (ogólne cechy systemu rekomendacji) zaliczamy:

- **Dziedzinę** – określa, jakie elementy są przedmiotem rekomendacji (dokumenty hipertekstowe, produkty, najnowsze wiadomości, osoby).
- **Technikę rekomendacji** – determinuje rodzaj danych używanych do rekomendacji i sposób ich użycia (filtrowanie użytkowników, filtrowanie treści, metoda hybrydowa itp.).
- **Zastosowane algorytmy uczące się** – w ramach jednej techniki (np. filtrowania treści) jest możliwe użycie różnych algorytmów uczących się (np. sieci neuronowe, drzewa decyzyjne, grupowanie itp.).
- **Sposób pobierania danych** – określa sposób, w jaki pobierane są dane od użytkowników (przykładowo, ocena oglądanych dokumentów, obserwacja kupowanych produktów itp.).

- **Różnorodność rekomendacji** – nowe kryterium, pozwalające podzielić systemy rekomendacji ze względu na mechanizmy mające na celu zwiększenie różnorodności rekomendowanej treści.

W drugiej grupie (zarządzanie profilem) znajdują się następujące kryteria:

- **Reprezentacja profilu użytkownika** – charakteryzuje struktury danych, jakimi posługuje się system do przechowywania informacji na temat profilu użytkownika (wektor binarny, historia zakupów, zbiór atrybutów itp.).
- **Profil początkowy** – określa, jaka jest zawartość nowego profilu użytkownika. Może być on pusty, wygenerowany na podstawie wypełnionej ankiety albo utworzony w oparciu o dane dotyczące istniejących profili.
- **Adaptacja profilu** – pozwala prześledzić jak formalna reprezentacja profilu dostosowuje się do zmian preferencji użytkownika.
- **Prywatność** – określa sposób, w jaki identyfikowany jest użytkownik (anonimowy identyfikator sesji, numer dowodu tożsamości, pseudonim itp.).

Do ostatniej grupy (architektura systemu) zaliczono:

- **Czasową dystrybucję obliczeń** – determinuje, kiedy dokonywane są obliczenia związane z rekomendacją. Najczęściej spotykane rozwiązania to obliczanie wszystkich niezbędnych wartości w czasie rzeczywistym (*ang. on-line*) lub ograniczenie operacji dokonywanych w czasie rzeczywistym na rzecz okresów zmniejszonej aktywności systemu.
- **Miejsce instalacji** – nowe kryterium pozwalające ustalić gdzie uruchamia się program zawierający implementację systemu rekomendującego (serwer WWW, przeglądarka klienta itp.).
- **Prezentację rekomendacji** – dotyczy sposobu, w jaki prezentowana jest rekomendowana treść.

Poniższa tabela prezentuje odpowiedniki zunifikowanych kryteriów klasyfikacji w opracowaniach wymienionych powyżej.

Lp.	Kryterium klasyfikacji systemów rekomendujących	Montaner 2003	Burke 2002	Resnick 1997	Schafer 2001
1	Dziedzina	Dziedzina	Brak	Przedmiot rekomendacji	Brak
2	Technika rekomendacji	Filtrowanie	Technika rekomendacji	Agregacja profilu	Sposób rekomendacji
3	Zastosowane algorytmy uczące się	Generowanie profilu użytkownika	Proces rekomendacji	Brak	Brak
4	Sposób pobierania danych	Sprzężenie zwrotne	Dane wejściowe	Sposób pobierania danych	Dane wejściowe użytkownika docelowego (częściowo)
5	Różnorodność rekomendacji	Brak	Brak	Częściowo zmienność gustów	Brak

Lp.	Kryterium klasyfikacji systemów rekomendujących	Montaner 2003	Burke 2002	Resnick 1997	Schafer 2001
6	Reprezentacja profilu użytkownika	Reprezentacja profilu użytkownika	Brak	Częściowo reprezentacja profilu	Brak
7	Profil początkowy	Profil początkowy	Brak	Brak	Brak
8	Adaptacja profilu	Adaptacja profilu	Brak	Częściowo zmienność gustów	Brak
9	Prywatność	Brak	Brak	Anonimowość użytkownika	Brak
10	Czasowa dystrybucja obliczeń	Brak	Brak	Brak	Sposób dokonywania obliczeń
11	Miejsce instalacji	Brak	Brak	Brak	Brak
12	Prezentacja rekomendacji	Brak	Brak	Wykorzystanie rekomendacji	Dane wyjściowe

Tabela 3.2.1 Odpowiedniki zunifikowanych kryteriów klasyfikacji w opracowaniach: (Montaner, 2003), (Burke, 2002), (Resnick, 1997) oraz (Schafer, 2001).

3.3. Analiza systemów rekomendujących

Zunifikowane kryteria klasyfikacji podane w poprzednim podrozdziale posłużą do analizy systemów rekomendujących dostępnych w literaturze.

3.3.1. Dziedzina

Kilkanaście lat badań nad systemami rekomendującymi zaowocowało różnorodnością obiektów będących przedmiotami rekomendacji, począwszy od filtrowania poczty elektronicznej (Goldberg i inni, 1992), poprzez rekomendowanie najnowszych wiadomości (Billsus i Pazzani, 1999), (Mobasher i inni, 2000a) czy produktów w handlu elektronicznym (Krulwich, 1997), a skończywszy na podpowiadaniu osób (Shah, 1997), dowcipów (Goldebrg i inni, 2001) oraz przepisów gastronomicznych (Svensson, 2000).

Z punktu widzenia tematyki niniejszej pracy najbardziej interesujące są metody rekomendacji dokumentów, w obrębie których można rozróżnić dwie grupy: rekomendację lokalną i globalną. Metody pochodzące z pierwszej grupy proponują dokumenty należące do jednego systemu hipertekstowego (pojedynczej witryny) (Mobasher i inni, 2000b), (Shahabi i inni, 2001). Druga grupa metod nie wprowadza ograniczeń co do pochodzenia podpowiadanego dokumentu. W tym przypadku zbiór dokumentów do rekomendowania tworzony jest przez formułowanie zapytań do wyszukiwarek (Chen i Sykara, 1998) lub poprzez analizę odsyłaczy znajdujących się w oglądanych dokumentach (Pazzani i inni, 1996).

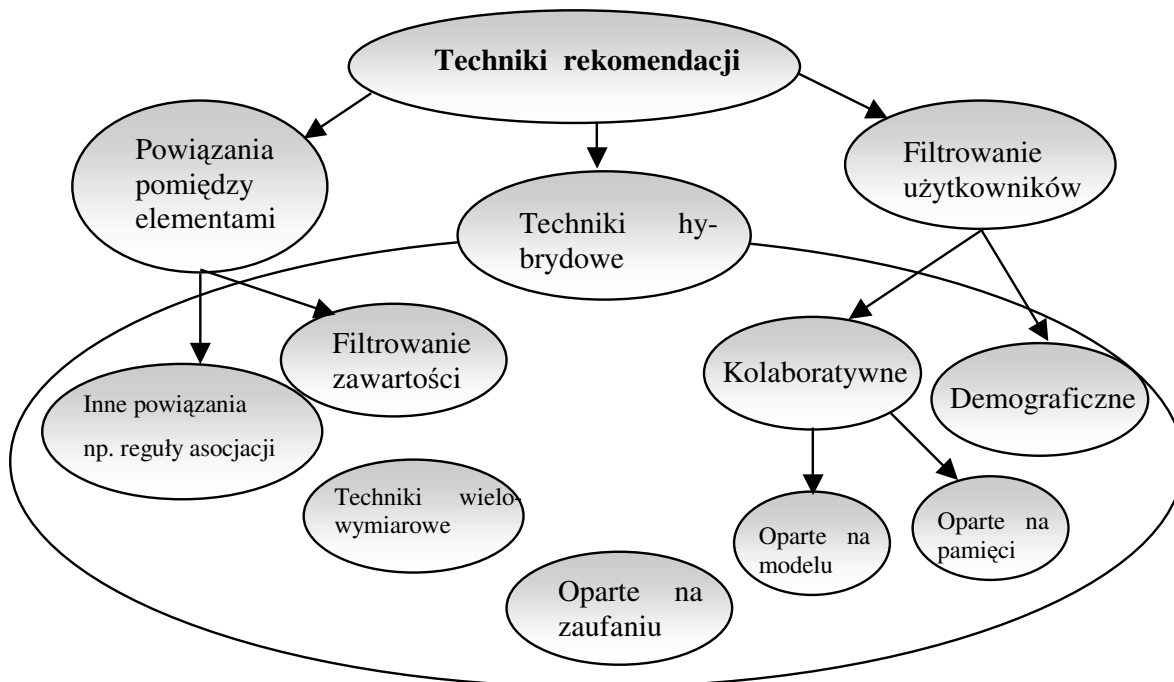
3.3.2. Technika rekomendacji

Technika rekomendacji jest najważniejszym, ze wszystkich opisywanych kryteriów. Została ona przyjęta za (Burke, 2002) i określa rodzaj danych i sposób ich użycia w metodzie reko-

mendacji. Dokładny podział i nazewnictwo poszczególnych technik różni się w zależności od autorów. W niniejszej pracy wyróżniono następujące kategorie:

- **Techniki oparte na powiązaniu między elementami** – obejmuje wszystkie systemy rekomendujące, w których podstawą do rekomendacji jest powiązanie, najczęściej przedstawiane w postaci podobieństwa cech – (*ang. content based filtering*), pomiędzy rekomendowanymi elementami, a elementami odwiedzionymi lub ocenionymi wcześniej (Lieberman, 1995), (Pazzani i inni, 1996), (Stefani i Strappavara, 1998).
- **Filtrowanie użytkowników** – systemy rekomendujące z tej grupy wykorzystują następującą intuicję: użytkownika będą interesowały elementy odwiedzone albo pozytywnie ocenione przez podobnych odbiorców. Reprezentacja profilu jest podstawą podziału filtrowania użytkowników na dwie podkategorie. Pierwsza z nich - filtrowanie demograficzne przedstawia każdą osobę jako zbiór cech demograficznych (płeć, wiek, wykształcenie itp.) (Krulwich, 1997), (Pazzani, 1999), druga zaś, znacznie częściej stosowana, nazywana filtrowaniem kolaboratywnym (*ang. collaborative filtering*) identyfikuje użytkownika za pomocą profilu powstałego z analizy jego zachowań (Goldberg i inni, 1992), (Goldberg i inni, 2001), (Sarwar i inni, 2001).
- **Techniki hybrydowe** – zawiera systemy rekomendujące, które bądź to wykorzystują obydwie wyżej wymienione techniki (Balabanovic i Shoham, 1997), (Pazzani i inni, 1999), (Mobasher i inni, 2005), bądź jedną z technik wzbogaconą o inne oryginalne podejścia. Przykładowo, w pracy (Adomavicius i Tuzhilin, 2004) zaproponowano, aby standardowy system rekomendujący oparty o filtrowanie wykorzystywał do rekomendacji informację kontekstową np.: pora roku, godzina dzień tygodnia itp.

Rysunek 3.3.1 obrazuje podział technik rekomendacji przyjętych w tej pracy. Do technik hybrydowych należy zaliczyć wszystkie metody, które wykorzystują jednocześnie kilka innych technik (przykładowo filtrowanie użytkowników, powiązania między dokumentami lub inne podejścia). Dodatkowo należy zauważyć, że podany podział zawiera w sobie wiele podtechnik stawianych przez innych autorów na równi z filtrowaniem użytkowników i powiązaniem między elementami. Przykładowo, (Burke, 2002) wyodrębnia techniki oparte na wiedzy oraz użyteczności dla użytkownika (patrz podrozdział 3.1). Warto podkreślić, że zarówno techniki oparte na wiedzy jak i na użyteczności nie mogą działać bez techniki opartej na powiązaniach między elementami lub filtrowaniu użytkowników. Podobne rozumowanie można przeprowadzić w stosunku do metod rekomendacji działających w oparciu o eksplorację danych (*ang. data mining*) w środowiskach hipertekstowych (Mobasher i inni, 2000a), (Mobasher i inni, 2000b), (Mobasher, 2001). Jeśli oparte są one wyłącznie na analizie danych dotyczących zawartości elementów (*ang. content mining*) albo na analizie danych dotyczących zachowań użytkowników, to będą zaliczane odpowiednio do technik opartych o powiązania między elementami bądź do filtrowania użytkowników (Mobasher i inni, 2000a), (Mobasher i inni, 2001). Podgrupa tych metod działających w oparciu o integrację zawartości i aktywności użytkownika zaliczana będzie do technik hybrydowych (Mobasher i inni, 2000b), (Kazienko i Kiewra, 2003a), (Kazienko i Kiewra, 2004b).



Rysunek 3.3.1 Klasyfikacja technik rekomendacji.

3.3.2.1. Techniki oparte na powiązaniach między elementami

Techniki oparte na powiązaniach między elementami rekomendują dokumenty lub produkty związane z elementami, które zostały odwiedzone lub pozytywnie ocenione przez użytkownika docelowego. Gdy powiązanie to odnosi się do treści mamy do czynienia z *filtrowaniem zawartości* (ang. *content based filtering*) (Pazzani i inni, 1996), (Stefani i Strappavara, 1998), (Chen i Sycara, 1998). Przykładowo, w (Pazzani i inni, 1996) na podstawie ocen dokumentów odwiedzonych przez użytkownika obliczane jest prawdopodobieństwo, z jakim dany deskryptor może wystąpić w dokumencie relewantnym. Następnie, przy użyciu naiwnego klasyfikatora Bayesa obliczane jest prawdopodobieństwo, z jakim nieodwiedzony dokument będzie zakwalifikowany do zbioru dokumentów relewantnych.

Bardzo często stosowane jest też podobieństwo pomiędzy wektorami dokumentów. Poszczególne cechy dokumentów – najczęściej deskryptory odpowiadają kolejnym współrzędnym wektora. (Chen i Sycara, 1998). W metodach tych zakłada się, że jeśli dokument *A* widziany przez użytkownika był relewantny do jego potrzeb, to dokument *B* bardzo podobny do dokumentu *A* także będzie dokumentem relewantnym. Jako miarę podobieństwa stosuje się najczęściej cosinus kąta między dwoma wektorami cech (patrz definicja 2.1.2).

Podobieństwo ze względu na posiadane cechy nie jest jedynym możliwym kryterium służącym do rekomendowania dokumentów wzajemnie powiązanych. Przykładowo w (Mobasher i inni, 2001) przedstawiono metodę bazującą na dokumentach często odwiedzanych razem przez innych użytkowników. W tym przypadku wykorzystano wyszukiwanie reguł asocjacji w dużych wolumenach danych.

Rekomendacja oparta tylko i wyłącznie na powiązaniu między elementami obarczona jest poważnymi wadami szeroko opisanymi w literaturze np. (Montaner i inni, 2003). Pierwsza z nich związana jest z ograniczonymi możliwościami analizy. Najlepiej zilustrować tę niedo-

godność następującym przykładem: założmy, że każdy dokument reprezentowany jest przez zbiór deskryptorów będący podzbiorem zbioru wszystkich terminów występujących w kolekcji dokumentów. W takim przypadku może pojawić się pewien dokument d_n , w którym nie będzie występować żaden z deskryptorów typowych dla całej kolekcji (przykładowo, gdy dokument d_n będzie zawierać wyłącznie zdjęcia lub rysunki). Dokument d_n nie będzie, więc podobny do żadnego innego elementu kolekcji, przez co nie zostanie nigdy zarekomendowany nawet wtedy, gdy będzie odznaczał się dużą relewancją w stosunku do potrzeb użytkownika.

Druga wada występująca w systemach opartych tylko i wyłącznie na powiązaniach między elementami to nadspecjalizacja. Uwidacznia się ona szczególnie w systemach korzystających z podobieństwa między dokumentami. Należy bowiem zauważyć, że najbardziej podobna do dokumentu obecnie odwiedzanego jest jego kopia (a w przypadku wyeliminowania kopii dokument niemalże identyczny). Powoduje to, że rekomendowane elementy będą dotyczyć pewnej bardzo wąskiej dziedziny. Dziedzina ta może nie zawierać wszystkich elementów relewantnych, co w konsekwencji spowoduje obniżenie kompletności rekomendacji. Pewnym sposobem wyeliminowania nadspecjalizacji jest wyłączenie z rekomendacji elementów bardzo podobnych do obecnie oglądanego (Pazzani i inni, 2000), lecz należy stwierdzić, że rozwiązanie to jest bardzo kontrowersyjne, gdyż trudno ustalić, gdzie znajduje się granica pomiędzy duplikatem lub „prawie duplikatem”, a dokumentem wysoce relewantnym.

Nadspecjalizacja uwidacznia się także w technikach rekomendacji opartych na dokumentach odwiedzonych razem - wspomnianych powyżej (Mobasher i inni, 2001). Techniki te rekomendują dokumenty, które są bardzo często odwiedzane z dokumentami widzianymi przez użytkownika w bieżącej sesji. W wielu przypadkach może budzić to podejrzenie, że przyczyną ich wspólnego pobierania związana jest ze strukturą nawigacyjną serwisu. Przykładowo, większość użytkowników witryny ZSI opisanej w eksperymentach (patrz rozdział 6) wchodzi do witryny przez stronę główną, co przy użyciu wyżej wspomnianej metody rekomendacji może powodować, że właśnie strona główna będzie często rekomendowana niezależnie od jej rzeczywistej relewancji do oglądanej treści.

Na zakończenie warto wspomnieć o braku weryfikacji jakości dokumentów rekomendowanych tylko i wyłącznie na podstawie podobieństwa wektora deskryptorów. W innych technikach weryfikacji tej dokonują użytkownicy (strony o niskiej jakości czy relewancji nie są odwiedzane lub są oceniane negatywnie). W (Shardanand i Maes, 1995) bardzo słusznie zauważono, że metody bazujące wyłącznie na podobieństwie dokumentów tak samo określają relewancję dokumentu bardzo dobrze zredagowanego (przejrzystego, bez błędów stylistycznych ani logicznych) i dokumentu o bardzo niskiej jakości, jeśli obydwa z nich zawierają te same deskryptory.

3.3.2.2. Filtrowanie użytkowników

Drugą dużą grupą technik rekomendacji jest filtrowanie użytkowników. W metodach należących do tej grupy wychodzi się z założenia, że podobni użytkownicy mają podobne gusta. Kryteria służące do określenia podobieństwa między użytkownikami są podstawą do wyodrębnienia dwóch głównych kategorii w obrębie tej grupy. Pierwsza z nich - filtrowanie demograficzne zakłada, że dwóch użytkowników jest podobnych, jeśli podobne są ich dane demograficzne (wiek, wykształcenie, płeć) (Krulwich, 1997), (Pazzani, 1999). Chociaż filtrowanie demograficzne może dawać dobre rezultaty w rekomendacji produktów, to jednak w przypadku podpowiadania dokumentów jego użycie jest bardzo dyskusyjne. Po pierwsze istnieje spora grupa witryn, w których dane personalne nie mają żadnego znaczenia. Trudno, na przykład, oczekiwać, że w witrynie poświęconej historii starożytnej Grecji potrzeby informacyjne

młodych kobiet będą diametralnie różne od potrzeb mężczyzn w średnim wieku. Po drugie użytkownicy bardzo niechętnie udostępniają swoje dane osobowe serwisom internetowym, a co gorsza, podane informacje bardzo trudno zweryfikować.

Z powyższych powodów filtrowanie kolaboratywne - druga kategoria technik z tej grupy wydaje się być bardziej adekwatna w przypadku rekomendowania dokumentów. W filtrowaniu kolaboratywnym dwóch użytkowników uważa się za podobnych, jeśli obydwoj odwiedzają te same elementy lub je podobnie oceniają. W (Breese i inni, 1998) podzielono wszystkie metody oparte na filtrowaniu kolaboratywnym na bazujące na pamięci (*ang. memory based*) oraz bazujące na modelu (*ang. model based*).

W technikach bazujących na pamięci, profil aktualnego użytkownika porównywany jest ze wszystkimi profilami użytkowników dostępnymi w systemie (Resnick i inni, 1994), (Shardanand i Maes, 1995). Podstawową zaletą tego podejścia jest łatwa adaptacja metody do zmian preferencji odbiorców. Przykładowo, załóżmy, że w sklepie muzycznym pojawiły się płyty będące prekursorami nowo powstałego nurtu muzycznego. W tym przypadku rekomendacja dokonywana przez metodę opartą na pamięci bezbłędnie zarekomenduje wyżej wspomniane płyty pod warunkiem posiadania kilku bądź kilkunastu ocen innych użytkowników. W odróżnieniu od systemu bazującego na modelu, który nie będzie w stanie dokonać poprawnej rekomendacji aż do momentu zaktualizowania modelu. Główną wadą filtrowania kolaboratywnego opartego na pamięci jest jego niska wydajność dla systemów cechujących się dużą grupą użytkowników. Aby wyeliminować tą wadę w (Yu i inni, 2004) zaproponowano porównywanie profilu bieżącego użytkownika tylko i wyłącznie z profilami dobrze reprezentującymi całą społeczność. Usprawnienie to sprawia jednak, że metoda zaproponowana w (Yu i inni, 2004) powinna być zakwalifikowana do technik bazujących na modelu.

Ciekawym rozszerzeniem filtrowania użytkowników bazującego na pamięci jest tzw. lista inwersyjna użytkowników (*ang. inverse user frequency*), która korzystając z osiągnięć wyszukiwania informacji (Salton, 1983) zakłada, że elementy dobrze oceniane przez bardzo wielu użytkowników są „szumem informacyjnym” i nie powinny być używane do określania podobieństwa między dwoma użytkownikami. Jest to oczywista analogia do terminów, występujących w wielu dokumentach, które są automatycznie wykluczane ze zbioru deskryptorów.

W pracy (Breese, 1998) dotyczącej filtrowania kolaboratywnego opartego na modelu zaproponowano model probabilistyczny. Zakładał on, że można oszacować prawdopodobieństwo poszczególnych ocen wystawionych przez użytkownika dla niewidzianego przez niego elementu. W szacowaniu tym bierze się pod uwagę oceny tego użytkownika dotyczące innych elementów oraz oceny innych użytkowników. W rezultacie szacowana ocena dla niewidzianego produktu będzie wartością oczekiwaną. W (Breese, 1998) podano dwa alternatywne modele probabilistyczne: pierwszy oparty na grupowaniu, a drugi na sieci Bayesa.

W literaturze znane są także inne przykłady filtrowania oparte na modelu. Przykładowo w (Aggarwal i inni, 1999) zaproponowano opisać oceny społeczności użytkowników za pomocą grafu skierowanego, którego węzły oznaczają użytkowników, a krawędzie stopień podobieństwa między nimi. Przewidzenie oceny, jaką dany użytkownik przypisałby niewidzianemu elementowi sprowadza się do znalezienia bezpośredniej drogi od bieżącego użytkownika do użytkownika, który już ocenił rozpatrywany element.

Stosunkowo często stosowane jest też grupowanie podobnych użytkowników na postawie ocen lub odwiedzin dokumentów (Goldberg i inni, 2001), (Mobasher i inni, 2000b), (Hoffman, 2003). Warto także wspomnieć o wykorzystaniu probabilistycznego modelu relacyjnego do filtrowania (Getoor i Sahami, 1999). Probabilistyczny model relacyjny pozwala określić prawdopodobieństwo wystąpienia konkretnej relacji między obiektami będącymi

instancjami danych encji. W przypadku filtrowania użytkowników można mówić o relacji (kupowania bądź oglądania) pomiędzy obiektami będącymi instancjami dwóch encji: produkt i użytkownik. Według autorów powyższej pracy, probabilistyczny model relacyjny jest łatwiejszy do interpretacji niż sieci Bayesa i pozwala na lepsze modelowanie złożonych relacji występujących między encjami.

Techniki oparte tylko i wyłącznie na filtrowaniu użytkowników posiadają także szereg wad szczegółowo opisanych w literaturze (Balabanovic i Shoham, 1997), (Montaner i inni, 2003a):

- Problem nowego elementu –do czasu zebrania odpowiedniej liczby odwiedzin (ocen) system rekomendujący nie jest w stanie zarekomendować nowego elementu żadnemu użytkownikowi. Dodatkowo dokonanie rekomendacji dla użytkownika, który odwiedził (ocenił) tylko jeden nowy dokument może odznaczać się niską dokładnością.
- Rozrzedzenie danych – gdy liczba użytkowników jest znacznie mniejsza od liczby elementów lub dostępne elementy są usuwane z systemu po krótkim okresie czasu a w ich miejsce wprowadzane są nowe, łatwo może dojść do sytuacji, w której pojawi się spora grupa elementów, które nie są często odwiedzane razem, chociaż są relewantne dla użytkowników mających podobne preferencje. Jeśli dwa elementy relewantne nie są odwiedzane razem, to odwiedzenie jednego z nich nie spowoduje zarekomendowania drugiego, co może doprowadzić do obniżenia się kompletności.
- Nietypowe preferencje i potrzeby – techniki rekomendacji oparte na filtrowaniu użytkowników mogą odznaczać się bardzo niską dokładnością w przypadku użytkowników, których potrzeby i zainteresowania odbiegają znacznie od typowych wzorców. Problem ten dotyczy głównie filtrowania kolaboratywnego opartego na modelu, który siłą rzeczy dokonuje pewnych uogólnień i uproszczeń.
- Kłopoty z osiągnięciem masy krytycznej – aby systemy rekomendujące bazujące na filtrowaniu użytkowników mogły poprawnie działać potrzebna jest duża liczba użytkowników, którzy odwiedzili dużą liczbę dokumentów, co może być trudne do osiągnięcia w systemach, które posiadają stosunkowo małą liczbę użytkowników lub, w których elementy zmieniają się bardzo szybko (np. witryna poświęcona najnowszym wiadomościom).

3.3.2.3. Techniki hybrydowe

Techniki hybrydowe łączą w sobie kilka innych technik. Metody rekomendacji działające w oparciu o techniki hybrydowe nazywane są metodami hybrydowymi, a systemy rekomendacji stosujące te metody systemami hybrydowymi. Już w (Balabanovic i Shoham, 1997) oraz (Pazzani, 1999) pokazano za pomocą eksperymentów, że systemy hybrydowe integrujące filtrowanie użytkowników z technikami opartymi na powiązaniach między elementami dają lepszą dokładność niż systemy rekomendujące w oparciu o jedną z dwóch wyżej wymienionych technik. W (Burke, 2002) wyróżniono 7 sposobów integracji technik w jedną metodę hybrydową. Tabela 3.3.1 prezentuje krótką charakterystykę każdego ze sposobów.

Sposób integracji	Opis	Przykład publikacji
Wagowa	Wartości numeryczne przypisane elementom w różnych metodach składowych łączone są w jedną wartość przy pomocy ważonej sumy.	(Claypool i inni, 1999), (Pazzani, 1999), (Mobasher i inni, 2000b)
Przełączająca	System rekomendacji w każdym kroku decyduje, którą metodę wybrać do zarekomendowania elementów.	(Tran i Cohen, 2000)
Mieszana	Użytkownikowi przedstawiane są równoległe rezultaty pochodzące z kilku metod.	(Smyth i Cotter, 2000)
Integracja cech	W ramach jednej metody rekomendacji używa się dane źródłowe pochodzące z kilku metod	(Basu i inni, 1998), (Kazienko i Kiewra, 2004a)
Kaskadowa	Jedna metoda rekomendacji jest używana do „filtrowania” rezultatów drugiej	(Burke, 2002)
Powiększanie cech	Dane wyjściowe jednej metody traktowane są jako dane wejściowe drugiej	(Mooney i Roy, 1999)
Meta-integracja	Model utworzony przez jedną metodę używany jest do utworzenia modelu używanego przez drugą metodę	(Balabanovic i Shoham, 1997), (Pazzani, 1999), (Mobasher, 2005)

Tabela 3.3.1 Rodzaje integracji w metodach hybrydowych podane za (Burke, 2002)

Przedstawiona tabela wymaga kilka wyjaśnień. Po pierwsze **integracja wagowa** wydaje się być najbardziej naturalnym rozwiązaniem, gdyż na wynik rekomendacji mają wpływ wszystkie metody składowe. Wpływ ten jest regulowany odpowiednimi wagami, które nie muszą być stałe. Przykładowo, w (Claypool i inni, 1999) wagi zmieniają się ze względu na zachowanie użytkownika (jeśli rekomendowane elementy są odrzucane przez użytkownika, to obniża się waga metody, która miała większe znaczenie dla dokonania tej konkretnej rekomendacji). W (Mobasher i inni, 2000b) wagi określające wpływ współczynnika treściowego i nawigacyjnego uzależnione są od podobieństwa do obecnej sesji użytkownika.

Mimo niezaprzeczalnych zalet podejścia wagowego (łatwość implementacji i interpretacji oraz duża elastyczność pozwalająca dostosować wartość wag do charakterystyki systemu), w publikacji (Burke, 2002) kwestionuje się jego przydatność, gdy poszczególne metody rekomendacji dają wyniki niezuniformowane. Zarzut ten jest mocno dyskusyjny, gdyż niezuniformowanie może zostać rozwiązane za pomocą normalizacji wartości pochodzących z różnych metod składowych.

W **integracji mieszanej i przełączającej** występuje problem braku wzajemnego wzmocnienia się metod składowych. Przykładowo załóżmy, że w systemie hybrydowym istnieją dwie metody rekomendacji, posiadające odpowiednio funkcję relewancji r_1 oraz r_2 , które dla par użytkownik - dokument przypisują nieujemną liczbę rzeczywistą nie większą, niż 1, która określa relewancję elementu do potrzeb użytkownika. Załóżmy, że dla użytkownika u mamy następujący zbiór elementów $\{e_1, e_2, e_3\}$ oraz, że:

$$r_1(u, e_1)=0,8, r_1(u, e_2)=0,79, r_1(u, e_3)=0$$

$$r_2(u, e_1)=0, r_2(u, e_2)=0,77, r_2(u, e_3)=0,78$$

Jeśli w integracji przełączającej należałoby zarekomendować jeden element, byłby to e_1 . Natomiast w mieszanej zaproponowany zostałby element e_1 oraz e_3 (każda z metod składowych dostarcza jeden element). Jest to oczywiście sprzeczne z intuicją, która nakazywałaby podpowiedzieć element e_2 .

Przy **integrowaniu cech**, podobnie jak w przypadku integracji wagowej, wszystkie „metody składowe mają wpływ na ostateczny wynik, ale trudno jest mówić o dużej elastyczności (zmiana znaczenia jednej ze składowych wiąże się najczęściej z przebudową modelu).

Integracja kaskadowa pozwala poprawić wydajność hybrydowego systemu rekomendującego, ponieważ tylko jedna metoda składowa wykonywana jest dla całego zbioru elementów. Najważniejszym mankamentem tej metody jest „zacieranie” różnic między relewancją elementów wyznaczoną przez metody nie będące ostatnie w kaskadzie. Powracając do wyżej przytoczonego przykładu założmy, że tym razem funkcje rekomendacji dają niższe wartości dla użytkownika u oraz zbioru elementów $\{e_1, e_2, e_3\}$:

$$r_1(u, e_1)=0,99, r_1(u, e_2)=0,1, r_1(u, e_3)=0$$

$$r_2(u, e_1)=0,78, r_2(u, e_2)=0,79, r_2(u, e_3)=0$$

Jeśli metoda z funkcją r_2 jest „drugą w kaskadzie” (filtruje i sortuje zbiór kandydatów uzyskanych przy pomocy funkcji r_1) to przy rekomendowaniu jednego elementu podpowiedziany zostanie element e_2 , choć intuicja kazałaby podpowiedzieć element e_1 .

Powiększanie cech jest bardzo ciekawym rozwiązaniem, szczególnie jeśli celem jest poprawienie efektywności gotowego systemu rekomendującego, którego z różnych powodów nie można zmodyfikować (brak kodu źródłowego, środków lub umiejętności). Istnieje tu jednak niebezpieczeństwo wadliwego działania systemu, jeśli pierwsza z zastosowanych metod nie działa w sposób prawidłowy (problem ten nie występuje w integracji wagowej, gdyż możliwe jest obniżenie wagi metody składowej, co do której istnieją podejrzenia, że podpowiada błędne elementy).

Meta-integracja jest bardzo popularnym sposobem integracji kolaboratywnego filtrowania użytkowników z filtrowaniem zawartości. Jest on podobny do powiększania cech z tą tylko różnicą, że w meta-integracji pierwsza metoda nie jest traktowana jako „czarna skrzynka”. Potrzebna jest więc znajomość pierwszej metody rekomendacji co najmniej na poziomie modelu. Co z jednej strony pozwala przynajmniej częściowo „zareagować”, gdy istnieje podejrzenie, że model stworzony przez pierwszą metodę jest obciążony błędami z drugiej jednak strony potrzebna jest większa wiedza dotyczącą technik wykorzystywanych przez pierwszą metodę.

Należy zauważyć, że nie mniej ważna od sposobu integracji kilku technik rekomendacji jest ich analiza pod kątem rozwiązywalności problemów, z jakimi borykają się ich składowe. Nie jest prawdą, że techniki hybrydowe pozwalają zawsze wyeliminować wady występujące w technikach składowych. Dobrą ilustracją tej tezy jest integracja dwóch technik rekomendacji opartych odpowiednio na grupowaniu sesji użytkowników i grupowaniu podobnych dokumentów przedstawionych w (Mobasher, 2000b). Pierwsza z nich zakłada, że użytkownikowi powinny być podpowiadane dokumenty z grupy tematycznej, do której została zakwalifikowana jego sesja. Druga zaś, rekomenduje dokumenty, które były odwiedzane w uśrednionej sesji grupy użytkowników, którzy zachowują się podobnie. Podejście to eliminuje, co prawda, problem nowych dokumentów, ale nie eliminuje zbyt dużego uogólnienia typowego dla rekomendacji opartej na grupowaniu. Ten mankament uwidacznia się szczególnie, gdy użytkownik znajduje się na krawędzi kilku grup. (patrz punkt 3.3.3 oraz rysunek 3.3.2).

Tak jak to zostało zasugerowane na rysunku 3.3.1, systemy hybrydowe nie muszą tylko i wyłącznie wykorzystywać technik opartych na filtrowaniu użytkowników lub związkach między elementami. Ostatnie lata zaowocowały badaniami nad nowymi technikami, które mogą zostać zastosowane w systemach hybrydowych (także w tej pracy użyto przydatności dokumentu, która nie może być zakwalifikowana do żadnej z technik przedstawionych powyżej – patrz podpunkt 4.5.1). Przykładowo, w (Adomavicius i Tuzhilin, 2003) zaproponowano rekomen-

dację filmów opartą na filtrowaniu kolaboratywnym wzbogaconym o informacje kontekstowe. W tym przypadku do informacji kontekstowej należy zaliczyć osobę, z którą użytkownik ma zamiar obejrzeć film (żona, przyjaciel, dziecko) i czas, w którym film będzie emitowany (dzień roboczy, koniec tygodnia itp.) W (Adomavicius, 2005) zaproponowano, aby standardową dwuwymiarową przestrzeń poszukiwań (*użytkownik x produkt*) zastąpić wielowymiarową, *użytkownik x produkt x kontekst*, gdzie kontekst także może składać się z wielu wymiarów.. Głównym problemem występującym w tego rodzaju systemach, jest trudność w reprezentowaniu, pobieraniu i aktualizowaniu wiedzy kontekstowej. Przytoczmy za (Adomavicius i Tuzhilin, 2003) przykład rekomendacji wycieczek, w którym autorzy postulują, aby systemy rekomendujące przy sugerowaniu wycieczek brały pod uwagę porę roku, długość dni wolnych, którymi dysponuje klient oraz osoby towarzyszące. W tym przypadku tylko pora roku może zostać określona automatycznie na podstawie zegara systemowego, natomiast osoby towarzyszące oraz długość wakacji muszą być podane bezpośrednio przez użytkownika. Dodatkowo należy zauważyć, że informacje kontekstowe, chociaż bardzo przydatne w rekomendacji produktów wydają się mniej ważne przy okazji podpowiadania dokumentów relevantnych witryny. Trudno sobie wyobrazić, aby potrzeby użytkowników witryny dotyczącej historii starożytnej Grecji zmieniały się w zależności od pory roku czy dnia tygodnia.

Również w pracy (Burke, 2002) zostało zaproponowane użycie technik opartych na wiedzy (jeśli użytkownik jest jarozsem, nie powinna być mu proponowana restauracja, w której królują potrawy mięsne). Z praktycznego punktu widzenia, techniki oparte na wiedzy opisywane przez (Burke, 2002) są równoważne rekomendacji bazującej na informacji kontekstowej opisanej w poprzednim akapicie.

Zupełnie innym rodzajem technik zajęto się w (Montaner i inni, 2004), gdzie zaproponowano wzbogacenie filtrowania kolaboratywnego o sieci społeczne (*ang. social networks*). W podejściu tym grupa użytkowników reprezentowana jest za pomocą grafu, którego węzłami są użytkownicy będący zrzeszeni w pewnej społeczności. Krawędzie grafu powinny opisywać relacje między nimi (np. przyjaźń lub wzajemne zaufanie). W (Montaner, 2003b) opisano system *GenialChef* stosujący tą ideę. Dokonuje on rekomendacji restauracji w oparciu o opinie użytkowników, których bieżący użytkownik obdarza zaufaniem. Także i w tym przypadku można mieć wrażenie, że rozwiązanie to trudno zastosować w praktyce, z powodu problemów związanych z budową sieci powiązań między użytkownikami (dane dotyczące relacji interpersonalnych są bardzo poufne i trudno sobie wyobrazić, aby użytkownicy chętnie je udostępniali).

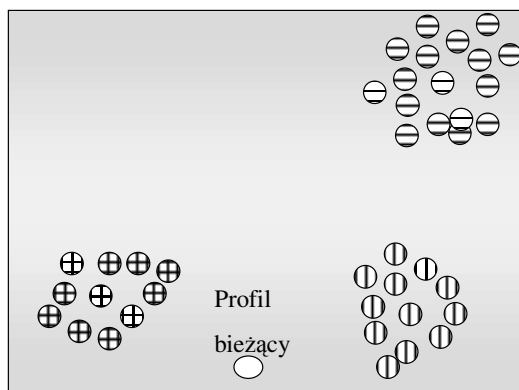
W (Li i Zaianem, 2004) zasugerowano wzbogacenie metody hybrydowej opartej na filtrowaniu kolaboratywnym i treści o informację na temat odsyłaczy między dokumentami. Autorzy proponują pogrupować wszystkie dokumenty danej witryny ze względu na ich odwiedzanie w podobnych sesjach. Następnie, do każdej grupy należy dodać dokumenty połączone odsyłaczami hipertekstowymi z dokumentami należącymi już do tej grupy. Rozwiązanie to jest umotywowane chęcią wzbogacenia każdej grupy nowymi dokumentami, które nie były jeszcze często odwiedzane, przez co nie mają szans znaleźć się w żadnej z grup, a w konsekwencji będą rzadko rekomendowane. Użycie odsyłaczy hipertekstowych w sposób zaproponowany w tej pracy jest bardzo dyskusyjny, gdyż w lokalnej witrynie dużo odsyłaczy ma charakter tylko i wyłącznie nawigacyjny. Przykładowo, wyobraźmy sobie, że w pewnej witrynie dotyczącej sportu większość stron w grupie poświęconej piłce nożnej posiada odnośnik do polityki prywatności stosowanej przez właściciela witryny.

3.3.3. Zastosowane algorytmy uczące się

Kilkanaście lat badań nad systemami rekomendującymi zaowocowało wykorzystaniem najrozmaitszych algorytmów uczących, do najpopularniejszych należy zaliczyć:

- **Grupowanie** – (Goldberg i inni, 2001), (Mobasher i inni, 2000b), (Hoffman, 2003) najczęściej grupowani są użytkownicy lub elementy podlegające rekomendacji. W pierwszym przypadku tworzone są grupy użytkowników odznaczających się podobnym zachowaniem, w drugim zaś generowane są kategorie dokumentów lub produktów. W obydwu przypadkach rekomendacja oparta jest na przyporządkowaniu bieżącego użytkownika do danej grupy i podpowiedzeniu mu niewidzianych przez niego elementów, które także do niej należą. Przyporządkowanie to sprowadza się najczęściej do określenia podobieństwa między „typowym przedstawicielem” każdej grupy, a formalną reprezentacją profilu.

Główną zaletą grupowania jest jego wydajność, gdyż profil bieżącego użytkownika nie jest porównywany ze wszystkimi elementami podlegającymi rekomendacji bądź wszystkimi użytkownikami, lecz jedynie z pewnymi typowymi reprezentantami tych zbiorów. Wadą takiego podejścia jest spadek dokładności systemu rekomendującego, gdy profil użytkownika znajduje się pomiędzy kilkoma grupami, patrz rysunek 3.3.2. Na rysunku tym dokumenty zostały przedstawione jako punkty w przestrzeni dwuwymiarowej, których wypełnienie identyfikuje grupę, do której przynależą. Punkt o białym wypełnieniu reprezentuje profil bieżącego użytkownika. Profil ten jest porównywany z „typowymi przedstawicielami” każdej grupy zaznaczonymi pogrubioną obwódką. Kiedy sesja bieżącego użytkownika nie da się zakwalifikować do żadnej z grup rekomendowane dokumenty mogą być „odległe” od potrzeb użytkownika.



Rysunek 3.3.2. Użycie grupowania może przekładać się na niską dokładność, gdy profil bieżącego użytkownika jest odległy od typowych grup.

- **Algorytm najbliższego sąsiada** – (Resnick i inni, 1994), (Shardanand i Maes, 1995) w przeciwieństwie do grupowania, w algorytmie najbliższego sąsiada formalna reprezentacja profilu użytkownika porównywana jest w zależności od zastosowanej techniki bądź to ze wszystkimi elementami podlegającymi rekomendacji bądź to ze wszystkimi użytkownikami. Podstawą do rekomendacji jest najbliższy element (bądź grupa elementów) z przeszukiwanego zbioru. Przykładowo, jeśli wiadomo, że bieżący użytkownik kupił w sklepie internetowym artykuły oznaczone przez $\{e_1, e_2, e_3, e_4, e_5\}$, to w celu zastosowania algorytmu najbliższego sąsiada dla filtrowania kolaratywnego należałoby znaleźć użytkownika, który nabył wszystkie (bądź większość) artykułów z powyższego zbioru i zarekomendować mu artykuły, których jeszcze nie kupił, a które zostały nabyte przez „najbliższego sąsiada”. Główną zaletą te-

go algorytmu jest jego lepsze zachowanie w przypadku użytkowników nietypowych. Algorytm najbliższego sąsiada, wykorzystując filtrowanie oparte na pamięci, odznacza się wadami typowymi dla tej techniki (niska wydajność systemu oraz nadmierna specjalizacja).

- **Naiwny klasyfikator Bayesa-** (Mladenic, 1996), (Pazzani i inni, 1996), (Breese i inni, 1998) w algorytmie tym określa się prawdopodobieństwo zajścia zdarzenia A – dokument d jest relewantny do potrzeb użytkownika, pod warunkiem zajścia iloczynu zdarzeń niezależnych $B_1 \cap B_2 \cap \dots \cap B_n$:

$$P(A|B_1 \cap B_2 \cap \dots \cap B_n) = P(A) \prod_{i=1}^n P(B_i)$$

Gdzie zdarzenie B_i ($i=1, 2, \dots, n$) oznacza wystąpienie deskryptora w_i w dokumencie d . Prawdopodobieństwa występujące po prawej stronie powyższej równości wyznaczone są na podstawie danych treningowych.

Oprócz najczęściej wykorzystywanych algorytmów przedstawionych powyżej, stosuje się także wiele innych. Przykładowo, w (Sarukkai, 2000) zaproponowano, a w (Zhu, 2001) usprawniono użycie łańcuchów Markova do rekomendacji dokumentów. W pracach tych tworzony jest skierowany graf z wagami, w którym dokumenty stanowią węzły grafu, a krawędź między węzłem A oraz B oznacza przejście z dokumentu A na dokument B . Waga jest przypisana do krawędzi i oznacza ile razy takie przejście nastąpiło (brane są pod uwagę wszystkie sesje użytkowników). Tak utworzony graf przekształcany jest w model Markova pozwalający określić, z jakim prawdopodobieństwem użytkownik przejdzie z dokumentu A na dokument B .

W (Montaner, 2003b) zastosowano wnioskowanie na bazie przykładu (*ang. case based reasoning*) do rekomendowania restauracji. W podejściu tym proces rekomendacji sprowadza się do znalezienia restauracji, które w przeszłości podobały się użytkownikowi i podpowiedzenia mu restauracji podobnych. W rekomendowaniu przez wnioskowanie na bazie przykładów wykorzystuje się typowe dla tej dziedziny fazy (wyszukiwanie przykładów, ich reutilizacja, weryfikacja oraz ewentualne zatrzymanie do przyszłego użycia).

Do algorytmów uczących się użytych w rekomendacji należy także zaliczyć wcześniej wspomniane wyszukiwanie reguł asocjacji (Mobasher i inni, 2001), sieci Bayesa (Breese i inni, 1998) oraz drzewa decyzyjne (Krulwich, 1996).

Na koniec należałoby wymienić zastosowanie algorytmów genetycznych. W (Moukas, 1997) zaproponowano by populacja agentów przeszukiwała zasoby Internetu. Agenci, którzy nie znaleźliby relewantnych treści powinni zostać wyeliminowani, zaś ci którzy znaleźli żądane dokumenty powinni wzajemnie krzyżować swoje cechy w celu utworzenia kolejnej populacji doskonalszych agentów. W (Sheth i Maes, 1993) zaproponowano użycie algorytmów genetycznych w celu wyeliminowania problemu nadspecjalizacji w filtrowaniu treści.

3.3.4. Sposób pobierania danych

Duża część technik używanych w systemach rekomendacji bazuje na ocenach wprowadzonych przez użytkownika. Przy rekomendacji produktów częstokroć rezygnuje się z obligowania użytkownika do bezpośredniego oceniania relewancji elementów, gdyż sam fakt zakupu danego produktu lub włożenie go do wirtualnego koszyka traktowany jest jako pozytywna ocena danego towaru.

W przypadku dokumentów jest niezmiernie trudno o taką „pośrednią ocenę”, więc pojawiły się pomysły, aby za pozytywnie ocenione uznać dokumenty, które użytkownik zapisał na swoim dysku lokalnym bądź, które zostały wydrukowane (Goldberg i inni, 1992), (Kamba i inni, 1995). Propozycja ta są bardzo dyskusyjne. Po pierwsze mogą być wykorzystane tylko w systemach działających po stronie klienta (w punkcie 3.3.13 opisano wady takich systemów). Dodatkowo, fakt niewydrukowania danego dokumentu nie musi się wiązać z jego małą relewancją, lecz po prostu z brakiem drukarki lub jej awarią.

Innym pomysłem przedstawionym w (Resnick, 1994) było założenie, że dokumentom relewantnym poświęca się więcej czasu. Ilość czasu poświęcona danemu dokumentowi może być obliczona poprzez różnicę czasu, jaki upłynął pomiędzy przesłaniem dwóch kolejnych żądań. Także ten pomysł wydaje się być kontrowersyjnym. Po pierwsze, nie jest możliwe określenie czasu poświęconego ostatniemu dokumentowi z sesji. Po drugie, użytkownik mógł zażądać kilku dokumentów niemalże jednocześnie w celu ich późniejszego przeczytania. Po trzecie nie ma żadnej metody pozwalającej na sprawdzenie czy w czasie, jaki upłynął pomiędzy pojawieniem się dwóch kolejnych żądań użytkownik w rzeczywistości zajmował się czytaniem a nie na przykład rozmową przez telefon.

Dlatego też twórcy systemów rekomendujących dokumenty częstokroć czekają na bezpośrednią ocenę podaną przez użytkownika (Pazzani, 1996), (Chen i Sycara, 1998). Jednak już w (Carroll i Rosson, 1987) zauważono, że użytkownicy są bardzo niechętni w dostarczaniu bezpośredniej informacji na temat oglądanych elementów. Ciągła prośba o ocenianie widzianej treści może być dla użytkownika bardzo uciążliwa i w zależności od jej nachalności może spotkać się z całkowitym zignorowaniem lub nawet zrezygnowaniem z używania danego systemu.

Problem ignorowania ocen został bardzo ciekawie rozwiązany w pracy (Goldberg i inni, 2001) dotyczącej systemu rekomendującego dowcipy, w którym użytkownik musi ocenić obecnie czytany dowcip, aby zobaczyć następny. Rozwiązanie to, choć pomysłowe nie jest, oczywiście, w stanie poradzić sobie z użytkownikami, którzy będą nadawać losowe oceny, aby jak najszybciej przejść do następnej porcji informacji.

Z podanych wyżej powodów odchodzi się od pobierania od użytkownika ocen dotyczących oglądanej treści, przyjmując, że odwiedzenie danego dokumentu jest przejawem zainteresowania jego treścią. W ten sposób, system rekomendujący zdobywa informację na temat dokumentów odwiedzonych w sposób niewidoczny dla użytkownika (Stefani i Strappavara, 1998), (Kazienko i Kiewra, 2004b).

Nie sposób też nie wspomnieć o systemach, które pobierają dane od użytkownika hybrydowo (Kamba i inni, 1995), (Resnick, 1994), (Montaner, 2003b). W tym przypadku najczęściej zakłada się, że dane wprowadzone bezpośrednio przez użytkownika są ważniejsze od informacji zdobytych pośrednio.

3.3.5. Różnorodność rekomendacji

Różnorodność rekomendacji to oryginalna cecha systemu rekomendującego wprowadzona przez autora tej pracy. Określa ona, jakie mechanizmy są używane przez system rekomendujący do zapewnienia zróżnicowanej treści. Wprowadzenie tego kryterium ma na celu podkreślenie znaczenia kompletności dla procesu rekomendacji. Tak jak to zostało zaznaczone we wstępie, autorzy prac dotyczących rekomendacji zdają się zapominać, że wysoka dokładność nie jest jedynym warunkiem koniecznym do osiągnięcia wysokiej efektywności metody rekomendacji. Drugim warunkiem jest, właśnie, wysoka kompletność.

W przypadku większości autorów kwestia różnorodności rekomendacji sprowadza się jedynie do nierekomendowania elementów już odwiedzonych (nie jest rozpatrywany problem rekomendowanych elementów, które nie wzbudziły zainteresowania użytkownika). Przykładowo, w kolaboratywnym filtrowaniu użytkowników zakłada się otwarcie, że selekcja elementów do rekomendacji odbywa się ze zbioru elementów jeszcze niewidzianych bądź niekupionych. Do nielicznych wyjątków należy zaliczyć publikację (Goker i Thompson, 2000), w której zaproponowano specjalny mechanizm zliczający liczbę razy, jaką dany element był już odbiorcy pokazywany w celu ograniczenia dokonywania tej samej rekomendacji wielokrotnie. Zastanawiająca jest postawa autora (Montaner, 2003b), który opisując mechanizm użyty w (Goker i Thompson, 2000) poddał go krytyce stwierdzając, że podejście to, chociaż, innowacyjne może doprowadzić do „podejrzanych zachowań systemu”, który dla tych samych danych wejściowych może zaprezentować różne rezultaty.

Problem różnorodności rekomendacji dokumentów nie jest wcale banalny, szczególnie gdy podpowiadane dokumenty spotykają się z ignorancją użytkownika. Z jednej bowiem strony użytkownik może być nieusatsfakcjonowany podpowiadaną treścią, z drugiej jednak zignorowanie rekomendacji może wynikać z jej słabej widoczności lub obecności w treści dokumentu innych „statycznych odsyłaczy”, które z punktu widzenia użytkownika są ciekawsze niż podpowiadane elementy.

Dodatkowo system rekomendujący, posiadający mechanizmy zapewniające osiągnięcie wysokiej różnorodności prezentowanej treści, powinien podpowiadać w każdym kroku bieżącej sesji jeszcze nierekomendowane elementy relewantne, a ponowne zaprezentowanie tego samego elementu powinno nastąpić dopiero po wyczerpaniu się relewantnych elementów, które nie były jeszcze rekomendowane. Nie jest to zadaniem prostym, gdyż w kolejnych krokach bieżącej sesji proponowane będą elementy coraz bardziej „odległe” od potrzeb użytkownika (co wiąże się z ryzykiem zarekomendowania elementów nierelewantnych, a w konsekwencji ze spadkiem dokładności).

3.3.6. Reprezentacja profilu użytkownika

Profil użytkownika pozwala przedstawić w sposób formalny jego gusta i preferencje. Najpowszechniejszą formą reprezentacji jest wektor lub zbiór atrybutów (Lieberman, 1995), (Moukas, 1997), (Balabanovic i Shoham 1997), (Mobasher i inni, 2000), (Mobasher i inni, 2001). Dodatkowo, w systemach komercyjnych użytkownik jest bardzo często reprezentowany przez historię swoich zakupów (WWW, 2005d), (WWW, 2005e). Choć model wektorowy ze względu na swoją intuicyjność oraz łatwy i przejrzysty sposób liczenia podobieństwa między elementami jest niemalże standardem, to jednak użycie specyficznych algorytmów uczących się (patrz podpunkt 3.3.3) może wymusić inną reprezentację profilu użytkownika.

Klasycznym przykładem innych podejść jest publikacja (Krulwich, 1996), gdzie zastosowanie drzew decyzyjnych sprawiło, że to właśnie drzewo decyzyjne reprezentuje profil użytkownika. Podobnie w (Stefani i Strappavara, 1998) zaproponowano sieć semantyczną jako sposób reprezentacji danych o użytkowniku. Sieć semantyczna w tym przypadku jest grafem, którego węzły reprezentują terminy kluczowe (deskrytory), a krawędzie odpowiadają współwystępowaniu tych terminów w dokumentach.

3.3.7. Profil początkowy

W wielu pracach przyjmuje się, że profil początkowy jest pusty (WWW, 2005d), (WWW, 2005e), (Balabanovic i Shoham, 1997). W formalnej reprezentacji profilu przekłada się to najczęściej na wektor zerowy lub zbiór pusty (Chen i Sycara, 1998).

Duża część systemów z lat dziewięćdziesiątych tworzyła profil początkowy na podstawie próbnych danych ocenianych przez użytkownika lub na podstawie wypełnianej ankiety (Pazzani i inni, 1996). Obydwa rozwiązania obarczone są pewnymi niedogodnościami. Przypisywanie użytkownikowi pustego profilu powoduje, że do czasu zbudowania chociażby prowizorycznego profilu użytkownika należy zaniechać jakiegokolwiek rekomendacji lub też należy liczyć się z niebezpieczeństwem podpowiadania dokumentów zupełnie nerelevantnych. Z drugiej strony zadawanie pytań lub prośba o wypełnienie ankiety może być dla użytkownika bardzo uciążliwe (patrz podpunkt 3.3.5), szczególnie gdy wizyta w witrynie jest anonimowa.

3.3.8. Adaptacja profilu

Dostosowywanie profilu do zmieniających się potrzeb oraz gustów użytkownika nie jest zadaniem łatwym i może prowadzić do problemu, który w (Burke, 2002) określony został jako „plastyczność kontra stabilność”. Szybkie dostosowywanie się do zmian gustów użytkownika może być zaletą systemu (warto przypomnieć, że zdolność adaptacji do nowych warunków jest jedną z definicji inteligencji), ale z drugiej strony może być niekorzystna, gdyż w przypadku punktowych zmian potrzeb użytkownika może zostać stracony jego poprawnie zbudowany profil. Przykładowo, załóżmy, że stały klient księgarni internetowej lubujący się w kryminałach postanowił zakupić w niniejszej księgarni kilka tomików poezji dla jednego ze swych bliskich. Zapamiętanie przez system faktu kupna tomików poezji może przełożyć się na zarekomendowanie dzieł zebranych Zbigniewa Herberta przy kolejnej wizycie tego klienta w systemie (nawet gdy podczas następnej wizyty użytkownik będzie chciał nabyć książkę odpowiadającą własnym zainteresowaniom). Dlatego w przypadku rekomendacji produktów wydaje się być w miarę rozsądnym rozwiązaniem przechowywanie zagregowanego stałego profilu użytkownika pozwalającego tuż po pojawieniu się użytkownika w systemie dokonać rekomendacji w oparciu o „stare i sprawdzone kryteria”. Jednocześnie w momencie zauważenia zachowań odbiegających od zapamiętanych schematów wskazanym byłoby szybkie dostosowanie się do nowych potrzeb.

W przypadku rekomendacji dokumentów potrzeba utrzymywania długofalowego profilu nie jest tak bardzo uzasadniona szczególnie ze względu na rosnącą popularność „sesji anonimowych” (patrz kolejny podpunkt). Najprostszym sposobem adaptacji profilu do nowych potrzeb użytkownika jest dodanie nowych informacji przykładowo: (Pazzani i inni, 1996) oraz (Chen i Sycara, 1998). Metody w ten sposób aktualizujące profil nazywane są agregacyjnymi. W podejściu tym pojawia się jednak problem równoważności pomiędzy obecnym zachowaniem użytkownika, a jego zachowaniem w przeszłości. Przykładowo, rozpatrzmy przypadek, w którym użytkownik łączy się z daną witryną w poszukiwaniu informacji na temat hoteli położonych w pobliżu renomowanych wyciągów narciarskich. Załóżmy, że ten sam użytkownik 6 miesięcy wcześniej interesował się pobytem nad morzem. Z punktu widzenia metody agregacyjnej bieżąca potrzeba zdobycia informacji dotyczącej hoteli w górach jest tak samo ważna jak informacja dotycząca nadmorskich kurortów.

Dlatego też wiele metod rekomendujących uwzględnia czynnik czasowy bądź to w formie postulatu (Stefani i Strappavara, 1998) bądź konkretnej formuły. Większość metod bazuje na pojęciu stopniowego zapominania (*ang. gradual forgetting*) wprowadzonej w (Web i Kuzmecz, 1996), a wykorzystywanej chociażby w (Stefani i Strappavara, 1998), (Mobasher i inni, 2000b), (Montaner, 2003b). Przykładowo w (Mobasher i inni, 2000b) zaproponowano, aby znaczenie dokumentów odwiedzonych w przeszłości malało w tempie geometrycznym. Alternatywą dla stopniowego zapominania jest uwzględnianie tylko kilku ostatnio odwiedzonych elementów (Mobasher i inni, 2001).

3.3.9. Prywatność

W systemach rekomendujących instalujących się po stronie klienta (Pazzani i inni, 1996) nie istniał problem poufności danych użytkownika, gdyż dane te znajdowały się w jego komputerze osobistym (w przypadku tych systemów zakładano, że dane dotyczące zachowania użytkownika nie były przekazywane na inne serwery). Wiele prac w zupełności pomija problem prawa użytkownika do prywatności i poufności danych o nim przechowywanych (Chen i Sycara, 1998), (Stefani i Strappavara, 1998).

Warto podkreślić, że przechowywanie danych o użytkowniku jest w pełni zrozumiałe i potrzebne w przypadku sklepów internetowych. Jednakże przechowywanie danych osobowych w przypadku rekomendowania dokumentów hipertekstowych jest bardzo dyskusyjne. Należy zdać sobie sprawę, że przeciętny użytkownik każdego dnia może odwiedzić nawet do kilkudziesięciu witryn. Dlatego też rejestrowanie się w witrynie, nawet jeśli wiąże się tylko i wyłącznie z podaniem pseudonimu postrzegane jest jako niemiłe utrudnienie. Wiele współczesnych systemów rekomendujących zakłada, że sesja użytkownika jest anonimowa (Mobasher i inni, 2001b), (Kazienko i Kiewra, 2003a). W sesji anonimowej użytkownik po połączeniu się z systemem otrzymuje anonimowy identyfikator, który jest ważny tylko podczas bieżącej sesji (po zamknięciu połączenia identyfikator jest automatycznie usuwany z systemu). Należy zauważyć, że w systemach używających anonimowych sesji nie można mówić o podobnych użytkownikach, a raczej o podobnych sesjach.

3.3.10. Czasowa dystrybucja obliczeń

Duża część systemów powstałych w latach dziewięćdziesiątych, szczególnie tych instalujących się po stronie klienta, dokonywała wszystkich obliczeń podczas pracy użytkownika z systemem (*ang. on-line*) (Pazzani i inni, 1996). Szybko jednak zdano sobie sprawę, że wykonywanie dużej liczby operacji w czasie rzeczywistym będzie oddziaływało negatywnie na czas odpowiedzi systemu, a pośrednio na skalowalność metody. Dlatego, wielu autorów sugeruje, wzorem systemów wyszukujących, wykonywać wszystkie operacje niezależne od docelowego użytkownika w czasie najniższej aktywności systemu - najczęściej nocą lub nad ranem. Przykładowo, (Chen i Sycara, 1998), (Stefani i Strappavara, 1998), (Mobasher i inni, 2000b), (Kazienko i Kiewra, 2003a). Tak jak już wspomniano w podrozdziale 3.2 istnieje też grupa systemów bazujących tylko i wyłącznie na rekomendacji statystycznej (Schafer i inni, 2001), w których wszystkie obliczenia mogą być dokonane w czasie najniższej aktywności systemu.

3.3.11. Miejsce instalacji

W latach dziewięćdziesiątych przeważały systemy instalujące się po stronie klienta (Stefani i Strappavara, 1998), (Pazzani, 1996), (Chen i Sycara, 1998). Przykładowo w (Pazzani i inni, 1996) system rekomendujący dokonywał obliczeń na dokumentach uprzednio skopiowanych na komputer użytkownika. System WebMate zaproponowany w (Chen i Sycara, 1998) nie tylko instalował specjalny program integrujący się z przeglądarką, ale także wymagał, aby wszystkie żądania HTTP przechodziły przez odpowiedni serwer proxy¹.

Upowszechnienie się języków skryptowych wykonujących się po stronie serwera WWW przełożyło się na pojawienie systemów rekomendujących uruchamiających się po stronie

¹Serwer proxy jest programem działającym w architekturze klient-serwer, który pośredniczy pomiędzy rzeczywistym klientem (dla którego jest serwerem) i rzeczywistym serwerem (dla którego jest klientem).

serwera (Shahabi i inni, 2001), (Mobasher i inni, 2001), (Kazienko i Kiewra, 2003a), (Montaner i inni, 2003a). W systemach tych maszyna użytkownika ograniczona jest do roli „cienkiego klienta” mającego za zadanie przedstawić wyniki rekomendacji.

Niemniej ważnym powodem spadku popularności systemów rekomendujących instalujących się po stronie klienta jest kwestia bezpieczeństwa. Użytkownicy niechętnie instalują systemy rekomendujące w obawie, że programy te pod pozorem rekomendacji „szpiegują” ich poczynania. Dodatkowo od drugiej połowy lat dziewięćdziesiątych obserwuje się wzrost programów bombardujących użytkownika nachalną reklamą (*ang. adware*). Programy te integrują się w sposób automatyczny z przeglądarką i przedstawiają treść o charakterze ściśle komercyjnym za każdym razem, gdy użytkownik prześle żądanie pobrania danego dokumentu (niezależnie od witryny, do której żądanie to jest skierowane).

3.3.12. Prezentacja rekomendacji

Ostatnim z opisywanych kryteriów jest prezentacja rekomendacji. Najczęściej rekomendowane elementy dołączane są do dokumentu zażadane przez użytkownika w postaci listy odsyłaaczy hipertekstowych wzbogaconych o krótki opis (WWW, 2005d), (WWW, 2005e), (Kiewra, 2005b). Istnieje też cały szereg oryginalnych rozwiązań. Przykładowo, w (Pazzani, 1996) zaproponowano oznaczać wszystkie odsyłaacze znajdujące się na bieżącej stronie symbolami graficznymi ilustrującymi relewancję dokumentów docelowych do potrzeb użytkownika. System *SiteIF* (Stefani i Strappavara, 1998) rekomenduje użytkownikowi wszystkie dokumenty relewantne na początku pracy z systemem.

Dość oryginalny pomysł został zaproponowany w pracy (Goldberg, 2001) dotyczącej rekomendacji dowcipów. System opisany w tej publikacji przedstawia użytkownikowi jeden dowcip należący do wybranej grupy. Tak jak już to zostało wspomniane w punkcie 3.3.5 następny element tej grupy pojawia się po oceniu obecnie oglądanej treści. Gdy użytkownik zobaczy już wszystkie elementy z danej grupy, przedstawiane będą dowcipy rekomendowane przez system.

3.3.13. Podsumowanie analizy systemów rekomendujących.

Z przedstawionej analizy wynika, że najsłabszą stroną współczesnych systemów rekomendujących jest brak mechanizmów zapewniających różnorodność rekomendowanej treści, co może prowadzić do niskiej kompletności rekomendacji, a w konsekwencji przyczynić się do obniżenia efektywności.

W przypadku technik rekomendacji, wady filtrowania użytkowników i technik opartych na powiązaniach między elementami powodują coraz większe zainteresowanie technikami hybrydowymi, w obrębie których integracja wagowa wydaje się być najbardziej obiecująca ze względu na brak niedogodności istniejących w innych rozwiązaniach, łatwość interpretacji i implementacji oraz dużą elastyczność pozwalającą dostosować wartość wag do charakterystyki systemu. Jednocześnie tak jak zauważono w podpunkcie 3.3.2.2 nie każda integracja technik opartych na filtrowaniu użytkowników i technik opartych na powiązaniach między dokumentami powoduje wyeliminowanie wszystkich wad występujących w tych technikach. W szczególności pożądanym jest integrowanie technik obarczonych problemem nadspecjalizacji z technikami cechującymi się zbyt dużym uogólnieniem rekomendacji.

Metoda *AdpatRank* przedstawiona w następnym rozdziale nie tylko posiada mechanizmy zapewniające dużą różnorodność rekomendowanej treści, ale także jest metodą hybrydową opartą na integracji wagowej, w której łączone jest podobieństwo między dokumentami (problemem nadspecjalizacji) z grupowaniem sesji użytkowników (nadmierne uogólnienie). Dodat-

kowo należy podkreślić, że przy definiowaniu metody *AdaptRank* starano się uniknąć wszystkich niedogodności występujących w istniejących systemach rekomendujących opisanych przy okazji charakteryzowania zunifikowanych kryteriów klasyfikacji systemów rekomendujących. Przykładowo, w przypadku adaptacji profilu, biorąc pod uwagę małą elastyczność metod agregacyjnych, zastosowano mechanizm stopniowego zapominania. Podobnie mając na uwadze niechęć użytkownika do bezpośredniej oceny odwiedzanych dokumentów metoda *AdaptRank* analizuje zachowanie użytkownika. Wykorzystuje się tu przede wszystkim informacji o dokumentach, jakie użytkownik odwiedził w bieżącej sesji. Dodatkowo brana jest pod uwagę informacja o rekomendowanych dokumentach, które nie wzbudziły zainteresowania użytkownika (zostały zignorowane) – jest to rozwiązanie nowatorskie, które według wiedzy autora nie było wykorzystane w żadnej innej pracy. Na zakończenie należy dodać, że pełna lista rozwiązań zastosowanych w metodzie *AdaptRank* znajduje się w tabeli 4.4.1 (patrz podrozdział 4.4 – Ogólny opis metody rekomendacji).

4. Hybrydowa metoda rekomendacji

Celem niniejszego rozdziału jest przedstawienie hybrydowej metody rekomendacji *AdaptRank* opartej na podobieństwie dokumentów i wzorcach nawigacyjnych. Metoda ta jest ujęta w formalnym modelu zdefiniowanym przez autora. Przetworzony model jest usystematyzowaniem i uporządkowaniem pojęć związanych z rekomendacją pojawiających się w literaturze. Pierwszy podrozdział zawiera definicje wykorzystywane w dalszej części pracy. W drugim zdefiniowano metody składowe, które będą integrowane w ramach metody *AdaptRank*. Następnie sformułowany zostanie problem rekomendacji. Kolejny podrozdział jest poświęcony definicji hybrydowej metody rekomendacji *AdaptRank* będącej rozwiązaniem uprzednio sformułowanego problemu. Na zakończenie rozdziału przedstawiony zostanie algorytm umożliwiający zaimplementowanie opisanej metody.

4.1. Wstępne definicje

Rekomendowaniem dokumentów zajmuje się system rekomendujący będący elementem systemu hipertekstowego opisanego w podrozdziale 2.3. Przed podaniem formalnej definicji systemu rekomendującego zdefiniowane zostaną wielkości wchodzące w jego skład. System rekomendujący rozważany w niniejszej pracy rekomenduje dokumenty hipertekstowe ze zbioru $D = \{d_1, d_2, \dots, d_n\}$. Do opisu treści dokumentów hipertekstowych ze zbioru D używamy skończonego i niepustego zbioru deskryptorów (słów kluczowych) $W = \{w_1, w_2, \dots, w_{n_w}\}$. Pierwsza definicja przedstawia elementy dokumentu należące do zbioru D .

Definicja 4.1.1

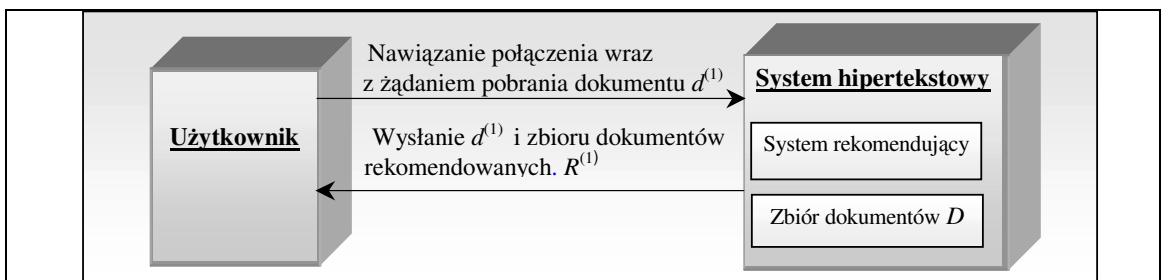
Hipertekstowym dokumentem $d \in D$ nazywamy uporządkowaną piątkę $d = (u^d, t_0^d, t_u^d, L^d, \mathbf{v}^d)$, w której:

- u^d - oznacza jednoznaczny identyfikator dokumentu d
- t_0^d - znacznik czasowy, w którym dokument d pojawił się w systemie hipertekstowym
- t_u^d - liczba jednostek czasu, podczas których dokument d był niedostępny
- L^d - zbiór odsyłaczy hipertekstowych umieszczonych wewnątrz dokumentu d
- $\mathbf{v}^d = (v_1^d, v_2^d, \dots, v_{n_w}^d)$ – wektor deskryptorów charakteryzujący treść dokumentu d o długości równej liczbie deskryptorów; współrzędna $v_j^d \in [0, 1]$ dla $j=1, \dots, n_w$ jest liczbą oznaczającą ważność deskryptora w_j w dokumencie d (dokładny sposób obliczania współrzędnych wektora \mathbf{v}^d podany jest w punkcie 4.5.3)

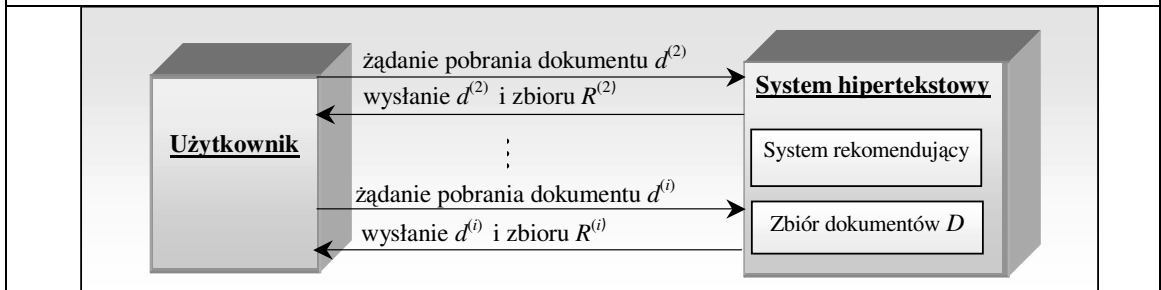
Znacznik czasowy t (*ang. timestamp*) to liczba jednostek czasu, które upłynęły od północy dnia pierwszego stycznia 1970 roku do daty i godziny, które ten znacznik oznacza. Przykła-

dowo, jeśli założymy, że jednostką czasu jest milisekunda to znacznik $t=1121000596827$ będzie określał godzinę 15, minut 3, sekund 16 oraz 827 milisekund dnia 10 lipca 2005 roku. Data, od której obliczana jest liczba jednostek czasu jest to tzw. data początkowa (*ang. epoch time*) używana w systemach UNIX oraz języku programowania Java. Wybór jednostki czasu uwarunkowany jest możliwościami platformy, w której implementowana jest metoda, najczęściej przyjmuje się sekundę lub milisekundę. Należy także zaznaczyć, że w przypadku witryny jednoznaczny identyfikator dokumentu u^d reprezentowany jest jako adres URL (WWW, 2005a).

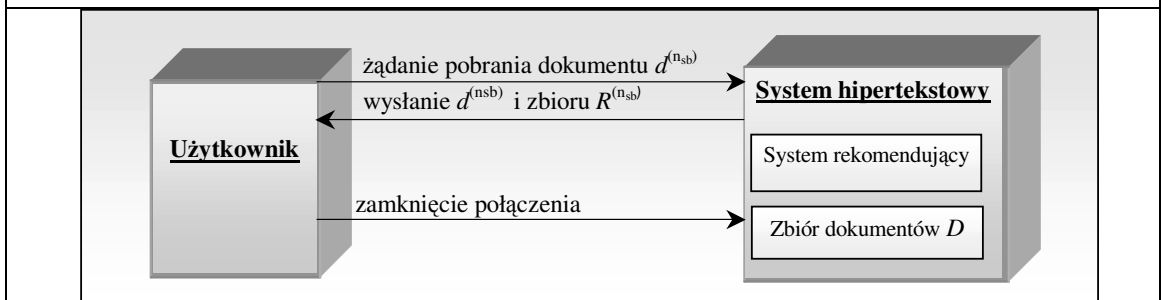
Użytkownik nie komunikuje się bezpośrednio z systemem rekomendującym, lecz z systemem hipertekstowym. Schemat komunikacji jest bardzo podobny do ogólnego schematu komunikacji przedstawionego w podrozdziale 2.3 (nawiązywane jest połączenie, a następnie użytkownik wysyła kolejne żądania pobrań dokumentów, zakończone zamknięciem połączenia). Sekwencja żądań pobrań dokumentów przesłanych w ramach jednego połączenia nazywana będzie sesją (patrz rysunek 4.1.1).



Rysunek 4.1.1a: Sesja użytkownika rozpoczyna się od nawiązania połączenia, po którym następuje przesłanie żądania pobrania pierwszego dokumentu. W odpowiedzi system hipertekstowy przesyła wybrany dokument dołączając pierwszy zbiór dokumentów rekomendowanych $R^{(1)}$



Rysunek 4.1.1b. W drugim kroku sesji przesyłane jest żądanie pobrania kolejnego dokumentu, który może, lecz nie musi należeć do zbioru $R^{(1)}$. W kolejnych krokach użytkownik przesyła żądania pobrania następnych dokumentów, a w odpowiedzi otrzymuje żądane dokumenty wraz ze zbiorami dokumentów rekomendowanych.



Rysunek 4.1.1c: Sesja użytkownika kończy się, gdy zamknięte zostanie połączenie z systemem hipertekstowym. W przypadku witryny WWW jest to równoważne z zamknięciem przeglądarki.

Każde pobranie dokumentu nazywane będzie krokiem sesji. Należy zauważyć, że w odpowiedzi na żądanie pobrania dokumentu wysłane w kroku i , oprócz dokumentu $d^{(i)}$ zwracany jest także zbiór $R^{(i)}$. Zawiera on odsyłacze do rekomendowanych dokumentów. Zbiór $R^{(i)}$ jest nowym elementem nie występującym w tradycyjnej komunikacji użytkownik - system hipertekstowy. W kolejnym kroku użytkownik może, ale nie musi zażądać przesłania dokumentu należącego do zbioru $R^{(i)}$. Jeśli dokument $d^{(i+1)}$ nie należy do zbioru $R^{(i)}$, to wszystkie dokumenty należące do zbioru $R^{(i)}$ uważane są za zignorowane w kroku i .

Sesja jest pojęciem fundamentalnym dla całej metody rekomendacji. W niniejszej pracy różniamy dwa rodzaje sesji: historyczną oraz bieżącą. Sesja historyczna to sesja, która miała miejsce w przeszłości i już została zakończona. Historyczne sesje służą do sporządzania wzorców nawigacyjnych (patrz definicja 4.5.2). Sesja bieżąca jest sesją właśnie trwającą, dla której dokonuje się rekomendacji. Poniżej zamieszczono jej formalną definicję.

Definicja 4.1.2

Bieżącą sesją w systemie hipertekstowym nazywamy ciąg jej stanów $s_b = (s_b^{(1)}, s_b^{(2)}, \dots, s_b^{(n_{sb})})$, gdzie n_{sb} oznacza długość sesji, natomiast $s_b^{(i)}$ jest stanem sesji s_b w kroku i ($i=1, \dots, n_{sb}$) będącym uporządkowaną trójką: $\langle R^{(i)}, \mathbf{z}^{(i)}, \mathbf{g}^{(i)} \rangle$ gdzie:

- $R^{(i)}$ to zbiór dokumentów zarekomendowanych użytkownikowi w kroku i ($R^{(i)} \subseteq D$)
- $\mathbf{z}^{(i)} = (z^{(i)}_1, z^{(i)}_2, \dots, z^{(i)}_n)$ to wektor dokumentów odwiedzonych do kroku i włącznie, którego współrzędna $z^{(i)}_j \in [0, 1]$ opisuje jak dawno dokument d_j był w tej sesji odwiedzony. Im wartość ta jest bliższa 0 tym dawniej dokument d_j był odwiedzony (0 – nie był odwiedzony, 1 – jest oglądany teraz). Dokładny sposób obliczenia tych współrzędnych podany jest w 4.5.1
- $\mathbf{g}^{(i)} = (g^{(i)}_1, g^{(i)}_2, \dots, g^{(i)}_n)$ to wektor dokumentów zignorowanych w kroku i , którego współrzędna $g^{(i)}_j \in [0, 1]$ opisuje jak dawno dokument d_j był w tej sesji zignorowany. Im wartość ta jest bliższa 0 tym dawniej dokument d_j był zignorowany (0 – nie był zignorowany, 1 – był zignorowany w poprzednim kroku). Dokładny sposób obliczenia tych współrzędnych podany jest w 4.5.2.

Niech $R = \bigcup_{i=1}^{n_{sb}} R^{(i)}$ oznacza zbiór wszystkich dokumentów zarekomendowanych w sesji s_b .

Warto zauważyć, że $\text{card}(R) \leq \sum_{i=1}^{n_{sb}} \text{card}(R^{(i)})$ gdyż jeden dokument może zostać podpowiedziany w kilku krokach tej samej sesji.

Długość sesji (tzn. liczba żądań pobrania dokumentu wysłanych do systemu hipertekstowego) zależy tylko i wyłącznie od użytkownika (każde żądanie przesłania dokumentu jest automatycznie wykonywane, a w odpowiedzi użytkownik otrzymuje żądany dokument wraz ze zbiorem dokumentów rekomendowanych). W praktyce jednak wprowadza się ograniczenie czasowe (*ang. timeout*), które powoduje, że jeśli użytkownik w określonej liczbie jednostek czasu nie prześle kolejnego żądania, to bieżącą sesję uznaje się za zakończoną (w przypadku witryn WWW liczba ta wynosi najczęściej 15-20 minut).

Sesja historyczna jest drugim rodzajem sesji służącym do wyznaczania typowych wzorców nawigacyjnych (patrz definicja 4.5.2). Poniżej znajduje się jej formalna definicja.

Definicja 4.1.3

Sesją historyczną nazywamy dwójkę (s, d_s) gdzie:

$s=(s_1, s_2, \dots, s_n)$ to wektor historycznej sesji, gdzie $s_j=1$, gdy dokument d_j był odwiedzony w tej sesji oraz $s_j=0$ w przeciwnym przypadku.

- d_s to dokument, który został odwiedzony w tej sesji jako pierwszy.

Przez S oznaczany będzie zbiór wszystkich historycznych sesji. Zakończona bieżąca sesja przekształcana jest w historyczną. Preferencje i zachowania użytkownika z ostatnio zakończonej sesji nie będą wpływać na dokumenty rekomendowane w innych sesjach, aż do momentu aktualizacji systemu. Dlatego też na potrzeby opisu metody rekomendacji można przyjąć, że zbiór S jest stały, a jego licznosc oznaczamy przez n_s .

Wektor s pozwala „zapamiętać”, jakie dokumenty były odwiedzane w danej historycznej sesji, a pierwszy dokument d_s jest niezbędny do otrzymania współczynnika otwarcia, którego sposób obliczenia przedstawiony zostanie w podrozdziale 4.5.

Przy przekształcaniu bieżącej sesji w sesję historyczną tracona jest spora ilość informacji (wektor dokumentów zignorowanych, zbiór dokumentów podpowiadanych R , kolejność odwiedzanych dokumentów). Jest to związane z praktycznym zastosowaniem metody (liczba historycznych sesji po kilku latach użytkowania systemu może być bardzo duża). Historyczne sesje wykorzystywane są do sporządzenia wzorców nawigacyjnych (patrz definicja 4.5.2), które są typowym przykładem filtrowania kolaboratywnego bazującego na modelu użytkowników (opisanego w punkcie 3.3.2.2). Warto zaznaczyć, że filtrowanie oparte na modelu użytkowników siłą rzeczy dokonuje pewnej generalizacji, która jest ceną, jaką trzeba zapłacić za skalowalność algorytmu rekomendacji.

Przyjmuje się, że użytkownik nawiązuje połączenie z systemem hipertekstowym w celu uzyskania konkretnych informacji. Warto przypomnieć, że użytkownik kontaktuje się z systemem hipertekstowym w ramach bieżącej sesji przedstawionej w definicji 4.1.2 oraz na rysunku 4.1.1. W każdym kroku sesji bieżącej użytkownik ma konkretne potrzeby informacyjne, które dzielą zbiór dokumentów D na 2 rozłączne podzbiory: zbiór $D_R^{(i)}$ dokumentów relevantnych oraz zbiór $D \setminus D_R^{(i)}$ dokumentów nierelwantnych. Pod wpływem odwiedzonych dokumentów preferencje użytkownika mogą się zmienić, więc jest możliwe, że w różnych krokach bieżącej sesji zbiory dokumentów relevantnych są różne lub nawet rozłączne. Przyjmuje się jednak, że w każdym kroku i zbiór $D_R^{(i)}$ jest niepusty. Rekomendowaniem dokumentów w ramach systemu hipertekstowego zajmować się będzie system rekomendujący, którego definicja znajduje się poniżej:

Definicja 4.1.4

Systemem rekomendującym SR nazywamy uporządkowaną trójkę $SR = \langle D, S, f \rangle$ gdzie: D to zbiór dokumentów wchodzących w skład systemu hipertekstowego, S to zbiór historycznych sesji, a f to funkcja rekomendacji, która w każdym kroku bieżącej sesji s_b każdemu dokumentowi ze zbioru D przyporządkowuje liczbę rzeczywistą określającą stopień relevantności danego dokumentu do potrzeb użytkownika w tym kroku.

Na przykład: $f(d, s_b^{(4)})=0,3$ oznacza, że w kroku 4 bieżącej sesji s_b wartość funkcji rekomendacji dla dokumentu d wynosić będzie 0,3. Z definicji 4.2.1 wynika, że system rekomendują-

cy SR posiada informacje na temat wszystkich dokumentów (czas życia, liczba nie działających odnośników wektor deskryptorów itp.), a także dysponuje danymi dotyczącymi dokumentów odwiedzanych przez użytkowników w każdej z historycznych sesji (zbiór S). Sesje te są anonimowe i nie są znane żadne dane osobowe użytkownika (płeć, wiek, stan cywilny itp.). Ponadto nie można stwierdzić, które z uprzednio zarejestrowanych sesji należą do użytkownika obecnie pracującego z systemem hipertekstowym¹.

Trzonem systemu rekomendującego jest funkcja rekomendacji f , która na podstawie zachowania użytkownika ocenia na ile użytkownik reprezentowany przez stan bieżącej sesji $s_b^{(i)}$ jest zainteresowany odwiedzeniem danego dokumentu d . Funkcja rekomendacji jest odpowiednikiem funkcji wyszukiwania występującej w systemie wyszukiwania przedstawionej w definicji 2.1.1. Choć obie funkcje są podobne to jednak różnią się co do dziedziny. W funkcji rekomendacji zapytanie użytkownika zostało zastąpione bieżącym stanem sesji w kroku i ($s_b^{(i)}$). Jest to związane bezpośrednio z różnicami między procesem rekomendacji a wyszukiwaniem. Rekomendacja w odróżnieniu od wyszukiwania nie oczekuje, aby użytkownik sformułował zapytanie, tylko proponuje mu dokumenty bazując na jego zachowaniu. Warto też podkreślić różnicę między metodą rekomendacji i funkcją rekomendacji. Metoda jest pojęciem szerszym i może być określona jako ustrukturyzowany zespół czynności użytych przez system do dokonania rekomendacji. Należy zauważyć, że samo określenie funkcji, która, oczywiście, jest najistotniejszym elementem metody rekomendacji, nie jest wystarczające do pełnego zdefiniowania metody rekomendacji (znaczy to, że znając postać funkcji nie można zaimplementować metody rekomendacji).

Nawiązując do rozważań na temat dwóch rodzajów rekomendacji zawartych w podrozdziale 2.1, należy stwierdzić, że zbiór $D_R^{(i)}$ wyraża relewancję subiektywną w odczuciu użytkownika w odróżnieniu od funkcji f , która określa relewancję szacowaną przez system rekomendujący.

Zakładamy, że system rekomendujący w każdym kroku bieżącej sesji podpowiada dokładnie m dokumentów ze zbioru D ($m \leq n$), a kryterium wyboru takich a nie innych dokumentów zawarte jest w funkcji rekomendacji f . Tak więc w kroku i bieżącej sesji określany jest zbiór dokumentów rekomendowanych $R^{(i)}$, gdzie $card(R^{(i)})=m$, takich że:

$$\forall_{d \in R^{(i)}, d' \in (D \setminus R^{(i)})} f(d, s_b^{(i)}) \geq f(d', s_b^{(i)}) \quad (4.1.1)$$

Proces selekcji dokumentów do zarekomendowania w kroku i może być porównany do sporządzenia rankingu dokumentów relewantnych do sformułowanego zapytania, z tą tylko różnicą, że w przypadku rekomendacji jedynie m pierwszych dokumentów zostanie przedstawionych użytkownikowi. Użycie rankingu w rekomendacji zwalnia system rekomendujący od obowiązku określenia całego zbioru dokumentów relewantnych w danym kroku $DR^{(i)}$. Wystarczy tylko podać m dokumentów, które na pewno są relewantne do potrzeb użytkownika.

Głównym zadaniem systemu SR jest maksymalizacja efektywności rekomendacji. Miara efektywności rekomendacji, analogicznie do efektywności wyszukiwania, jest średnią geometryczną wyznaczoną pomiędzy kompletnością i dokładnością.

$$E = \sqrt{prec \cdot rec} \quad (4.1.2)$$

¹ Problem rekomendacji byłby prostszy gdyby znane były zachowania użytkownika w przeszłości, biorąc jednak pod uwagę niechęć użytkowników do identyfikacji (opisaną w punkcie 2.3.9) zakłada się najgorszy scenariusz, w którym jedyne dane znane na temat danego użytkownika to jego zachowanie w bieżącej sesji.

Różnice pomiędzy wyszukiwaniem, a rekomendacją (a w szczególności założenie, że potrzeby użytkownika mogą zmieniać się podczas danej sesji) powodują, że jest niezbędne podanie odrębnych definicji kompletności i dokładności dla procesu rekomendacji.

Definicja 4.1.5

Dokładnością rekomendacji (*ang. precision*) systemu rekomendującego SR dla zakończonej bieżącej sesji s_b nazywamy stosunek liczby dokumentów relewantnych zarekomendowanych użytkownikowi podczas sesji s_b do liczby wszystkich dokumentów zarekomendowanych w tej sesji:

$$prec(SR, s_b) = \frac{\sum_{i=1}^{n_{s_b}} card(R^{(i)} \cap D_R^{(i)})}{\sum_{i=1}^{n_{s_b}} card(R^{(i)})}$$

Definicja 4.1.6

Kompletnością rekomendacji systemu rekomendującego SR w zakończonej bieżącej sesji s_b (*ang. recall*) nazywamy stosunek liczby różnych dokumentów relewantnych zarekomendowanych podczas tej sesji przez system SR , do liczby n_r oznaczającej największą liczbę różnych dokumentów relewantnych, które mogły być zarekomendowane użytkownikowi w tej sesji zakładając, że w każdym kroku rekomendowane jest maksymalnie m dokumentów:

$$rec(SR, s_b) = \frac{card\left(\bigcup_{i=1}^{n_{s_b}} (D_R^{(i)} \cap R^{(i)})\right)}{n_r}$$

Liczba n_r jest wyznaczana ze wzoru:

$$n_r = m + \sum_{i=2}^{n_{s_b}} \min\left(m, card\left(D_R^{(i)} \setminus \bigcup_{j=1}^{i-1} R^{(j)}\right)\right)$$

Główna różnica w definicjach dokładności i kompletności rekomendacji w porównaniu z ich odpowiednikami znanymi z wyszukiwania (patrz definicje 2.2.1 oraz 2.2.2) uwidacznia się w licznikach odpowiednich wyrażeń. Wprowadzenie zmienności potrzeb informacyjnych użytkownika w ramach tej samej bieżącej sesji ograniczyło zbiór dokumentów relewantnych tylko do tych zarekomendowanych dokumentów, które w danym kroku były relewantne do potrzeb użytkownika.

Przykładowo, jeśli $D_R^{(1)}=\{d_1, d_2, d_5\}$, $R^{(1)}=\{d_3\}$, $D_R^{(2)}=\{d_1, d_2, d_4\}$, $R^{(2)}=\{d_5\}$ oraz $D_R^{(3)}=\{d_3\}$, $R^{(3)}=\{d_1\}$ to kompletność i dokładność takiego systemu będzie wynosiła 0, chociaż $card(R \cap \left(\bigcup_{i=1}^3 D_R^{(i)}\right))=3$.

Dodatkowo, w przypadku kompletności zmieniała się wyraźnie postać mianownika. Jest to związane z charakterystyką procesu rekomendacji, w którym do kolejno odwiedzanych dokumentów dołączanych jest tylko m dokumentów uważanych przez system za relewantne. Należy zauważyć, że przyjęcie definicji kompletności znanej z wyszukiwania uzależniłoby wartość kompletności od długości bieżącej sesji. Co więcej, biorąc pod uwagę możliwość zmiany preferencji użytkownika, jest mało prawdopodobne, żeby bieżąca sesja była na tyle długa, aby wszystkie dokumenty relewantne z różnych zbiorów $D_R^{(i)}$ miały szanse być wypowiedziane. Przykładowo założmy, że podczas bieżącej sesji użytkownik 4 razy zmienił swoje preferencje i że dla każdego nowych preferencji użytkownika istniało 15 różnych dokumentów relewantnych. Przy założeniu, że w jednym kroku rekomendowane są 2 dokumenty potrzeba odwiedzenia 30 dokumentów, by wszystkie dokumenty relewantne mogły być zarekomendowane.

Kompletność z definicji 4.1.6 reprezentuje różnorodność rekomendowanych dokumentów relewantnych i nie dyskryminuje ani sesji zbyt krótkich, w których z racji ich długości liczba wypowiedzianych dokumentów relewantnych do liczby wszystkich dokumentów relewantnych musi być znikoma, ani zbyt długich, w których liczba zarekomendowanych dokumentów jest większa niż całkowita liczba dokumentów relewantnych, a co za tym idzie powtórzenia są nieuniknione.

4.2. Definicja metod składowych

Hybrydowa metoda rekomendacji *AdaptRank* oparta jest na dwóch metodach składowych, które zostaną sformalizowane w tym podrozdziale w celu łatwego porównania ich działania z działaniem metody *AdaptRank*. Każda z nich opiera się na rozwiązaniach znanych z literatury zaadaptowanych przez autora do modelu formalnego zaprezentowanego w niniejszej pracy. Należy dodać, że metody zdefiniowane w tym podrozdziale stanowiły punkt wyjścia do sformułowania hybrydowej metody *AdaptRank*, która nie tylko wprowadza nowy sposób integracji obydwu metod, ale także istotnie je modyfikuje w celu uzyskania wyższej efektywności. Do opisanie tych modyfikacji niezbędne jest uprzednie zdefiniowanie funkcji rekomendacji *AdaptRank* i jej odpowiednie przekształcenie, dlatego też w podrozdziale tym ograniczymy się do opisanie metod składowych tak jak podane zostały one w literaturze. Szczegółowa dyskusja na temat wprowadzanych modyfikacji przeprowadzona zostanie w podrozdziale 5.4.

4.2.1. Metoda rekomendacji oparta na podobieństwie tekstowym

Pierwsza z metod składowych rekomenduje dokumenty hipertekstowe, podobne do dokumentów odwiedzanych przez użytkownika. Do określania podobieństwa między dokumentami wykorzystano wektory deskryptorów $\mathbf{v}^d = (v_1^d, v_2^d, \dots, v_{n_w}^d)$ przedstawione w definicji 4.1.2.

Współrzędna v_j^d wektora deskryptorów dla dokumentu d odpowiada istotności deskryptora w_j w dokumencie d . W obliczaniu wartości v_j^d zastosowano formułę Saltona przytoczoną w podrozdziale 2.1, która jest tu powtórzona dla przypomnienia:

$$\mathbf{v}_j^d = wf_j^d \cdot \log\left(\frac{n}{n_j}\right) \quad (4.2.1)$$

gdzie wf_j^d oznacza częstość terminu w_j w dokumencie d , a n_j to liczba dokumentów, w których występuje termin w_j . Do wyznaczania podobieństwa między dokumentami użyte zostanie podobieństwo cosinusowe przedstawione w definicji 2.1.2. Oznacza to, że podobieństwo tekstowe między dokumentem d_j oraz d_k jest równe cosinusowi katów utworzonych przez wektory \mathbf{v}^{d_j} i \mathbf{v}^{d_k} tzn.:

$$\text{sim}(d_j, d_k) = \cos(\mathbf{v}^{d_j}, \mathbf{v}^{d_k}) \quad (4.2.2)$$

Rekomendowanie dokumentów podobnych do tych, które były już odwiedzone jest jedną z podstawowych technik rekomendacji (patrz punkt 2.3.2.1). Technika ta jest często wykorzystywana w praktycznych zastosowaniach. Przykładowo, elektroniczna biblioteka publikacji naukowych *CiteSeer* (WWW, 2005c) przy pobieraniu danego artykułu proponuje dokumenty do niego podobne. Analogicznie zachowuje się wiele sklepów internetowych - przykładowo *Amazon* (WWW, 2005d), *CD-NOW* (WWW, 2005e). Na uwagę zasługuje też fakt, że profesjonalne narzędzia do zarządzania dokumentami i nieuporządkowaną informacją w środowisku hipertekstowym takie jak *Autonomy* (WWW, 2005f) używają podobieństwa tekstowego do automatycznego generowania odsyłaczy wskazujących na dokumenty treściowo związane z czytany dokumentem (*ang. related documents*). Rozwiązania *Autonomy* zostały wykorzystane, między innymi, przez NASA, BBC ON-LINE czy agencję Reuters.

Rozpatrywana w niniejszej pracy metoda rekomendacji oparta na podobieństwie tekstowym jest adaptacją rozwiązania podanego w (Sarwar i inni, 2001) przez grupę badawczą *GroupLens* (WWW, 2005g), która od 1992 roku zajmuje się problemami rekomendacji i wyszukiwania informacji. Według autorów (Sarwar i inni, 2001) zaproponowana metoda odznacza się wyraźnie wyższą jakością od innych rozwiązań opartych na filtrowaniu kolaboratywnym lub najbliższym sąsiedztwie. Jakość w tym przypadku została określona za pomocą średniego błędu bezwzględnego (*ang. mean absolute error*). Na potrzeby niniejszej pracy zdefiniujemy funkcję rekomendacji f_i odpowiadającą zaadaptowanej metodzie:

Definicja 4.2.1

Wartość funkcji rekomendacji f_i dla dokumentu $d_j \in D$ w kroku i ($i=1, \dots, n_{sb}$), bieżącej sesji s_b jest równa sumie podobieństw dokumentu d_j do wszystkich dokumentów odwiedzonych w tej sesji podzielonej przez i :

$$f_i(d_j, s_b^{(i)}) = \frac{1}{i} \cdot \sum_{k=1}^i \text{sim}(\mathbf{v}^{d_j}, \mathbf{v}^{d^{(k)}})$$

gdzie. ($j=1, \dots, n$).

Dodatkowo zakłada się, że składowa metoda rekomendacji oparta na funkcji f_i w każdym kroku bieżącej sesji rekomenduje m dokumentów o najwyższych wartościach funkcji f .

4.2.2. Metoda rekomendacji oparta na zachowaniu użytkowników

Druga z metod składowych opiera się na wzorcach nawigacyjnych. Wzorce nawigacyjne odpowiadają typowym rodzajom zainteresowań użytkowników. Przykładowo na skutek grupowania sesji historycznych w systemie hipertekstowym dotyczącym sportu może się okazać, że wśród wzorców nawigacyjnych znalazł się jeden, w którym najczęściej są odwiedzane strony poświęcone koszykówce.

Wyznaczanie wzorców nawigacyjnych odbywa się poprzez grupowanie (*ang. clustering*) wektorów historycznych sesji (określonych w definicji 4.1.4). Kryterium grupowania jest podobieństwo między parami wektorów historycznych sesji, liczone jako cosinus kąta między nimi (patrz definicja 2.1.2). Rezultatem grupowania jest podział zbioru wszystkich sesji historycznych na n_c+1 niepustych grup (klastrów) takich, że:

$$S = \left(\bigcup_{j=0}^{n_c} C_j \right) \text{ oraz } C_0 \cap C_1 \dots \cap C_{n_c} = \emptyset. \quad (4.2.3)$$

Na uwagę zasługuje grupa C_0 , która powstaje przed właściwym procesem grupowania i zawiera wszystkie historyczne sesje, które zostały wyłączone z grupowania, z racji zbyt małej lub zbyt dużej liczby dokumentów w nich odwiedzonych. Grupa C_0 nie będzie brała udziału w rekomendacji (nie będzie dla niej wyznaczany wektor średni ani nie będzie określone podobieństwo do wektora dokumentów odwiedzonych do kroku i). Utworzenia grupy C_0 ma za zadanie wyłączenie z procesu grupowania tych historycznych sesji, dla których:

- ze względu na zbyt małą liczbę odwiedzonych dokumentów nie można ustalić preferencji użytkownika
- ze względu na zbyt dużą liczbę odwiedzonych dokumentów należy domniemywać, że użytkownik nie miał sprecyzowanych zainteresowań

Każda niepusta grupa C_j ($n_c \geq j > 0$) będzie reprezentowana przez wektor średni \mathbf{c}_j (centroid), zwany także wzorcem nawigacyjnym, którego formalna definicja znajduje się poniżej.

Definicja 4.2.2

Wzorcem nawigacyjnym (lub centroidem) grupy C_j ($j=1, \dots, n_c$) nazywamy wektor średni $\mathbf{c}_j = (c_{j1}, c_{j2}, \dots, c_{jn})$ utworzony z wektorów historycznych sesji należących do grupy C_j w następujący sposób:

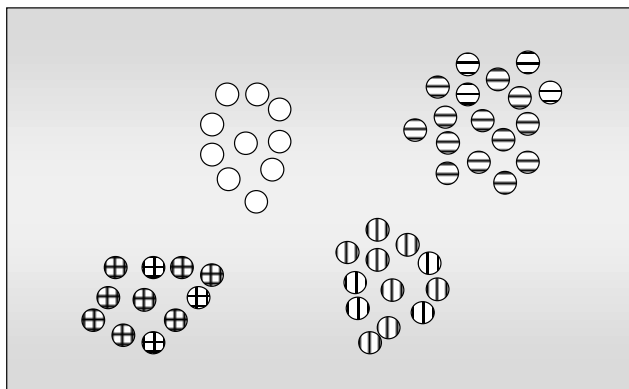
$$c_{jk} = \frac{1}{\text{card}(C_j)} \sum_{s \in C_j} s_k$$

dla $k = 1, 2, \dots, n$.

Tak wyznaczony wektor średni nazywa się wzorcem nawigacyjnym, gdyż odpowiada wzorcowej historycznej sesji dla pewnego zbioru użytkowników, których sesje należą do grupy C_j . W celu lepszego zobrazowania procesu grupowania można sprawdzić podobieństwo między wektorami historycznych sesji do odwrotności odległości pomiędzy punktami w dwuwymiarowej przestrzeni wektorowej (patrz rysunek 4.2.1). Wypełnienie punktu oznacza przynależność do konkretnej grupy. Centroidy poszczególnych grup zaznaczono pogrubioną linią.

Każda współrzędna wektora \mathbf{c}_j określa jak często odpowiadający jej dokument był odwiedzany w sesjach należących do C_j (0 – w ogóle nie był odwiedzany, 1 – był odwiedzany we

wszystkich sesjach). Wartości współrzędnych można także potraktować jako stopień, w jakim dany dokument jest reprezentantem danego wzorca nawigacyjnego. Załóżmy, że dla $n=8$ centroid pewnej grupy j jest równy $c_j=(0,05, 0,96, 0,98, 0, 0,8, 0,001, 0, 0,7)$. Analizując powyższy centroid łatwo zauważyć, że dokumenty d_2 i d_3 są jego dobrymi reprezentantami, gdyż zostały odwiedzone odpowiednio w 96% i 98% wszystkich historycznych sesji należących do grupy C_j .

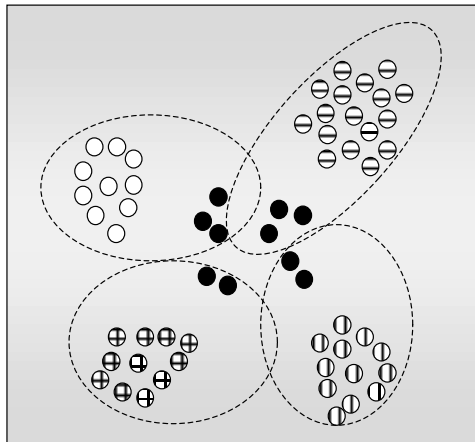


Rysunek 4.2.1 Ilustracja procesu wyznaczania wzorców nawigacyjnych.

W (Mobasher i inni, 2000b) zaproponowano użycie centroidów do klasyfikacji bieżącej sesji do jednego z wzorców nawigacyjnych w celu rekomendowania dokumentów będących „dobrymi przedstawicielami” tego wzorca nawigacyjnego. Przykładowo, wracając do wyżej przedstawionego centroidu $c_i=(0,05, 0,96, 0,98, 0, 0,8, 0,001, 0, 0,7)$ można stwierdzić, że użytkownik, który widział podczas bieżącej sesji dokumenty d_2 , d_5 oraz d_8 będzie potencjalnie zainteresowany dokumentem d_3 .

W metodzie *AdaptRank* zastąpiono klasyfikację do najbliższego wzorca nawigacyjnego obliczaniem wektora współczynników wzorców nawigacyjnych będącego kombinacją liniową dostatecznie bliskich wzorców nawigacyjnych, gdzie podobieństwo między danym wzorcem nawigacyjnym a wektorem dokumentów odwiedzonych do kroku i ($\mathbf{z}^{(i)}$) ma charakter wagi. Więcej informacji na temat klasyfikacji bieżącej sesji do wzorców nawigacyjnych można znaleźć w punkcie 4.5.4.

Osobnym zagadnieniem jest wybór algorytmu grupowania, który należy użyć w celu wyznaczenia wzorców nawigacyjnych. Dynamiczny rozwój nauki zajmującej się eksploracją danych (*ang. data mining*) przyczynił się do pojawienia wielu metod grupowania. Autor pracy sugeruje użycie algorytmu, w którym nie jest konieczne bezpośrednie określenie liczby grup, na które powinien zostać podzielony zbiór S , gdyż w przeciwnym przypadku może dojść do pogrupowania bardzo odległego od naturalnego podziału zbioru (patrz rysunek 4.2.2). Na rysunku zostało obrane $n_c=4$, choć należałoby raczej oczekiwać pięciu różnych grup. W eksperymentach zastosowano metodę hierarchiczną grupowania o nazwie HACM (Rasmussen, 1992).



Rysunek 4.2.2. Odgórne określenie liczby grup może doprowadzić do pogrupowania nie mającego nic wspólnego z realnym podziałem zbioru

Metoda składowa zaadaptowana w tej pracy jest częścią rozwiązania podanego w (Mobasher i inni, 2001b) pozwalającego zintegrować rekomendacje bazujące na wzorcach nawigacyjnych oraz grupach tematycznych. W niniejszej pracy użyto jedynie części metody opartej na wzorcach nawigacyjnych. Składowa ta nie jest tak często używana w praktyce jak poprzednia z uwagi na dość skomplikowaną implementację. W rozwiązaniach praktycznych częściej używa się filtrowania kolobaratywnego opartego na pamięci (patrz punkt 2.3.2). Należy jednak zauważyć, że tak jak to opisano w rozdziale 2 zarówno filtrowanie oparte na pamięci jak i rekomendacja bazująca na podobieństwie obarczone są błędem „nadspecjalizacji”. Dlatego też użycie metody opartej na wzorcach nawigacyjnych będącej typowym przykładem filtrowania opartego na modelu użytkowników jest bardziej uzasadnione. Filtrowanie oparte na modelu użytkowników pozwala bowiem uzyskać rekomendację cechującą się dość dużym uogólnieniem (prezentowani są typowi reprezentanci danej grupy zamiast dokumentów najbardziej podobnych do już odwiedzonych). Dodatkowo filtrowanie oparte na modelu użytkowników, w porównaniu do filtrowania opartego na pamięci, ma mniejsze wymagania co do mocy obliczeniowej i zasobów sprzętowych, co przekłada się na lepszą wydajność w systemach hipertekstowych z dużą liczbą historycznych sesji. Poniżej przedstawiona została funkcja rekomendacji f_c odpowiadająca zaadaptowanej metodzie. Podobnie jak w przypadku metody składowej wykorzystującej funkcję f_i , zakłada się, że składowa metoda rekomendacji oparta na funkcji f_c w każdym kroku bieżącej sesji rekomenduje m dokumentów o najwyższych wartościach funkcji f_c .

Definicja 4.2.3

Wartość funkcji rekomendacji f_c dla dokumentu $d_j \in D$ ($j=1, \dots, n_{sb}$) w kroku i ($i=1, \dots, n_{sb}$) bieżącej sesji s_b jest równa maksymalnej wartości wyznaczonej z iloczynów obliczonych pomiędzy j -tą współrzędną wzorców nawigacyjnych a podobieństwami pomiędzy wektorem dokumentów odwiedzonych do kroku i oraz wektorami wzorców nawigacyjnych:

$$f_c(d_j, s_b^{(i)}) = \max(c_{1j} \cdot \text{sim}(\mathbf{c}_1, \mathbf{z}^{(i)}), c_{2j} \cdot \text{sim}(\mathbf{c}_2, \mathbf{z}^{(i)}), \dots, c_{n_c j} \cdot \text{sim}(\mathbf{c}_{n_c}, \mathbf{z}^{(i)})),$$

4.3. Sformułowanie problemu**Definicja problemu:**Dane:

- Zbiór dokumentów D i zbiór historycznych sesji S .
- Miary jakości (dokładność, kompletność i efektywność).
- Bieżąca sesja s_b
- Składowa metoda rekomendacji oparta na funkcji f_t .
- Składowa metoda rekomendacji oparta na funkcji f_c .

Zadanie:

Zdefiniować hybrydową metodę rekomendacji dokumentów opartą na danych metodach rekomendacji, która w wielu przypadkach pozwoli uzyskać większą efektywność rekomendacji od efektywności rekomendacji jej metod składowych.

Ograniczenia:

1. Zbiór dokumentów relevantnych do potrzeb użytkownika w kroku i bieżącej sesji s_b jest niepusty $\forall_{0 < i \leq n_{sb}} (\text{card}(D_R^{(i)}) > 0)$.
2. Rekomendowane są dokumenty tylko ze zbioru D , a liczba dokumentów rekomendowanych jest mniejsza od liczby dokumentów w zbiorze D , czyli $n > m$.
3. W każdym kroku sesji bieżącej system nie rekomenduje dokumentu, który jest aktualnie odwiedzany, czyli $\forall_{0 < i \leq n_{sb}} (d^{(i)} \notin R^{(i)})$.

Tak jak już to zostało nadmienione w podrozdziale 2.3.5, przy okazji omawiania zmienności rekomendacji, większość systemów rekomendacji skupia się na podaniu mechanizmów pozwalających maksymalizować dokładność. Dotyczy to zarówno rozwiązań prostych jak i hybrydowych. Zdaniem autora rozwiązanie problemu postawionego powyżej wymaga opracowania metody rekomendacji, która oprócz zapewnienia mechanizmów zwiększających dokładność, będzie także zawierała techniki pozwalające zwiększyć kompletność rekomendacji.

4.4. Ogólny opis hybrydowej metody rekomendacji

W poprzednich podrozdziałach zdefiniowano system rekomendujący *SR* i sformułowano problem rekomendacji. W dalszej części tego rozdziału zaproponowane zostanie rozwiązanie tego problemu za pomocą hybrydowej metody rekomendacji nazwanej *AdaptRank*. Metoda ta bazuje na modelu wektorowym dokumentów i wykorzystuje zmodyfikowane sprzężenie relewancji znane z wyszukiwania (patrz podrozdział 2.1).

Tabela 4.4.1 zawiera podstawowe cechy systemu rekomendującego opartego o metodę *AdaptRank*. Cechy te odpowiadają zunifikowanym kryteriom klasyfikacji systemów rekomendujących opisanych w podrozdziale 2.2.

Cecha systemu	Wartość cechy w systemie implementującym <i>AdaptRank</i>
Cechy ogólne systemu	
Dziedzina	Dokumenty
Technika	Hybrydowa
Zastosowane algorytmu uczące się	Grupowanie
Sposób pobierania danych	Pośredni (obserwowane są żądane dokumenty i rekomendowane dokumenty, które zostały zignorowane)
Różnorodność rekomendacji	Zmniejszenie relewancji zarekomendowanych dokumentów (niezależnie czy zostały zignorowane, czy też odwiedzone)
Zarządzanie profilem użytkownika	
Reprezentacja profilu	Model wektorowy
Profil początkowy	Przydatność dokumentu określony za pomocą funkcji przedstawionej w definicji 4.5.1.
Adaptacja profilu	Stopniowe zapominanie
Prywatność	Użytkownik jest identyfikowany za pomocą anonimowego identyfikatora, który przestaje być ważny po zakończeniu sesji.
Architektura systemu	
Czasowa dystrybucja obliczeń	W czasie rzeczywistym obliczane są tylko wielkości zależne od bieżącego zachowania użytkownika. Większość czasochłonnych operacji wykonywanych jest okresowo w okresie zmniejszonej aktywności systemu.
Miejsce instalacji	Zależne od implementacji. System ROSA zawierający implementację metody działa po stronie serwera WWW (więcej informacji znajduje się w podrozdziale 6.2).
Prezentacja rekomendacji.	Zależna od implementacji. Wyżej wspomniany system ROSA prezentuje rekomendowane dokumenty za pomocą odsyłacza hipertekstowego z tytułem i krótkim streszczeniem rekomendowanego dokumentu.

Tabela 4.4.1 Cechy systemu rekomendującego implementującego metodę *AdaptRank*

Na wartość funkcji rekomendacji f dla dokumentu d w kroku i bieżącej sesji s_b składają się następujące elementy:

- Podobieństwo tekstowe dokumentu d do dokumentów odwiedzonych w tej sesji. Im dokument d jest bardziej podobny, tym wartość funkcji f jest większa.

- Liczba odwiedzin dokumentu d w historycznych sesjach podobnych do bieżącej sesji (im dokument d był częściej odwiedzany w tych sesjach, tym wartość funkcji f jest większa). Typowe sesje historyczne reprezentowane są przez wzorce nawigacyjne (patrz definicja 4.2.2). Do mierzenia jak często dany dokument był odwiedzany w sesjach podobnych do bieżącej używany będzie *współczynnik wzorców nawigacyjnych*, którego oryginalna definicja znajduje się w podrozdziale 4.5.4.
- Przydatność dokumentu d zdefiniowana jest jako kombinacja liniowa wielkości związanych z popularnością i jakością dokumentu (patrz podpunkt 4.5.1)
- Fakt zignorowania dokumentu. Jeśli dokument d był już rekomendowany w bieżącej sesji i został zignorowany, to fakt ten zmniejsza wartość funkcji f dla dokumentu d .
- Fakt odwiedzenia dokumentu d . Jeśli dokument d był już odwiedzony w bieżącej sesji to fakt ten zmniejsza wartość funkcji rekomendacji dla tego dokumentu.

Warto podkreślić, że metoda *AdaptRank* łączy podobieństwo tekstowe z grupowaniem. Jak zostało już zauważone przy okazji krytyki dostępnych metod, pierwsza ze składowych ma tendencje do rekomendowania treści bardzo pokrewnej do obecnie oglądanej, zaś druga generuje rekomendacje bardzo ogólne. Integracja tych dwóch metod przyczyni się do przynajmniej częściowego wyeliminowania dwóch podstawowych wad występujących w tych metodach: nadspecjalizacji (metoda oparta na podobieństwie tekstowym) i zbyt dużego uogólnienia (metoda oparta na wzorcach nawigacyjnych).

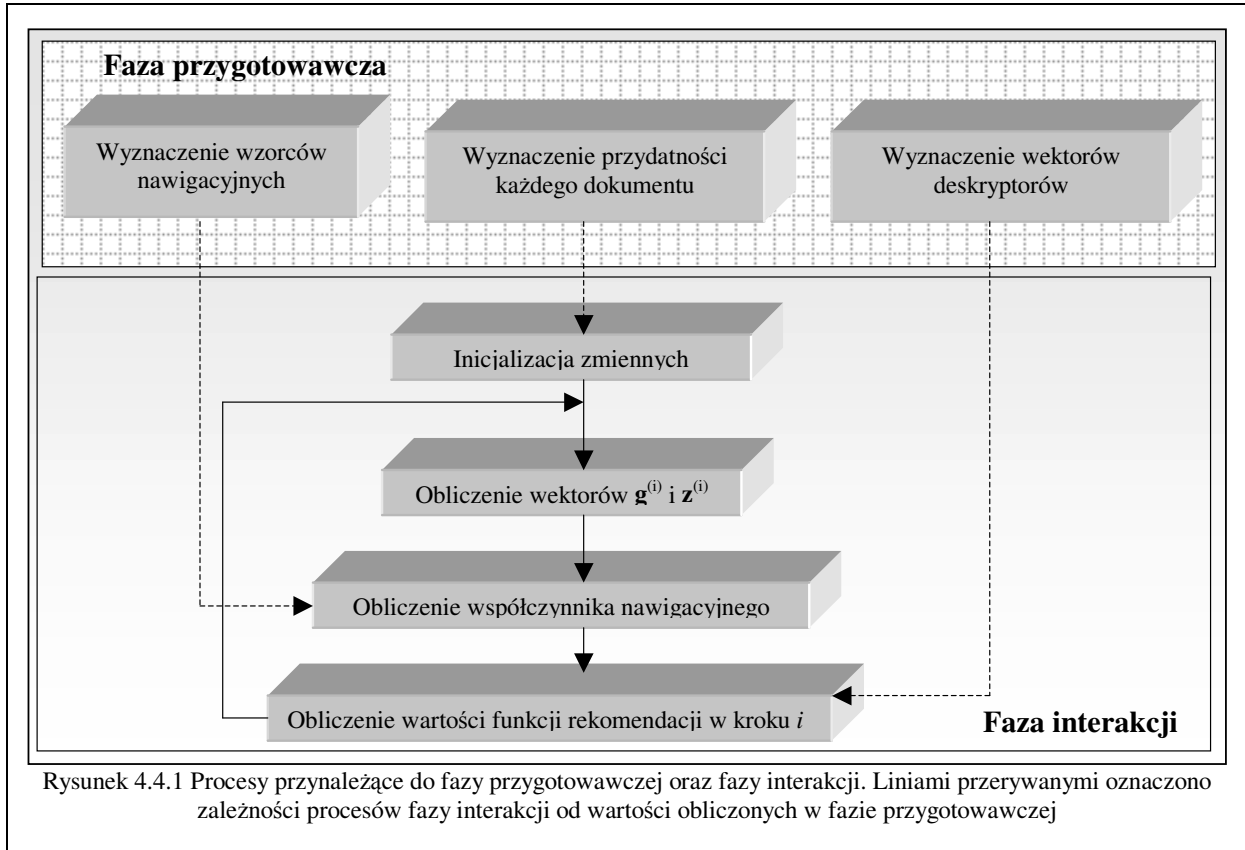
Ponadto, jednoczesne wykorzystanie 2 zupełnie różnych źródeł danych (zawartość tekstowa oraz zachowanie użytkowników) pozwala dokonać rekomendacji, nawet gdy dane pochodzące z jednego ze źródeł są niepełne bądź niedostępne. Przykładowo, dla nowych dokumentów istnieje spore prawdopodobieństwo braku danych dotyczących zachowania użytkowników (w tym przypadku z racji niskiej wartości współczynnika związanego z zachowaniem użytkowników, podobieństwo tekstowe będzie odgrywać dominującą rolę).

Nawet pobieżna analiza wyżej wymienionych czynników pozwala dostrzec ogrom obliczeń, które należy dokonać w celu określenia wartości funkcji rekomendacji. Z drugiej strony, z analizy charakterystyk systemu rekomendującego przedstawionych w punkcie 4.1 wynika, że rekomendowana treść powinna być przesyłana użytkownikowi wraz z żądanym dokumentem, (tak więc obliczenie wartości funkcji f nawet przy dużej liczbie paralelnych sesji nie może zajmować więcej niż kilka sekund). Z tych właśnie powodów proces obliczania wartości funkcji rekomendacji został podzielony na dwie fazy: przygotowawczą oraz interakcji. Podobne rozwiązania są szeroko rozpowszechnione w literaturze (Chen i Sykara, 1998), (Stefani i Strappavara, 1998), (Mobasher i inni, 2000b)¹.

Rysunek 4.4.1 przedstawia schemat metody *AdaptRank* z wyszczególnieniem wyżej wymienionych faz. Faza przygotowawcza zawiera czasochłonne operacje, które wymagają dużej ilości obliczeń obciążających system, a zarazem nie zależą od wartości bieżącej sesji. Procesy wchodzące w skład fazy przygotowawczej powinny być uruchamiane okresowo, gdyż dane posiadane przez system hipertekstowy ulegają ciągłym zmianom (przykładowo ciągle pojawiają się nowe sesje historyczne, a zbiór dokumentów D może ulec zmianie). Wskazaniem jest, by aktualizacja danych miała miejsce w okresie małej aktywności użytkowników systemu np. późną nocą lub nad ranem. Problem aktualizacji danych w procesie rekomendacji poruszony został w (Kazienko i Kiewra, 2003b). Wielkości obliczone w fazie przygotowawczej

¹ W literaturze anglojęzycznej faza przygotowawcza określana jest jako *off-line*, a faza interakcji jako *on-line*

traktowane są jako stałe podczas rekomendacji (np. zarejestrowanie nowej sesji historycznej nie powoduje automatycznej aktualizacji wzorców nawigacyjnych).



Na fazę przygotowawczą składają się następujące procesy:

- Wyznaczenie wektorów deskryptorów – dla każdego dokumentu d wyznaczany jest wektor \mathbf{v}^d (patrz definicja 4.1.2).
- Wyznaczenie przydatności dokumentu – dla każdego dokumentu określana jest przydatność definiowana jako liczba rzeczywista, która zależy od liczby odwiedzin, liczby nie działających odsyłaczy z niego wychodzących i innych czynników opisanych szczegółowo w punkcie 4.5.1.
- Wyznaczenie wzorców nawigacyjnych – historyczne sesje należące do zbioru S są grupowane w celu wyznaczenia zbioru typowych zachowań użytkowników. Każda z wyznaczonych grup odpowiada jednej typowej sesji reprezentowanej przez wektor średni (centroid) nazywany wzorcem nawigacyjnym.

Faza interakcji ogranicza się do operacji zależnych od zachowania użytkownika w bieżącej sesji s_b . Ma ona charakter iteracyjny i powtarzana jest za każdym razem, gdy użytkownik odwiedza kolejny dokument. W fazie interakcji można wyszczególnić następujące operacje:

- Wyznaczenie wektora $\mathbf{z}^{(i)}$ (dokumentów odwiedzonych do kroku i) - na podstawie dokumentów odwiedzonych w bieżącej sesji tworzony jest wektor obrazujący obecne zainteresowania użytkownika.

- Wyznaczenie wektora $\mathbf{g}^{(i)}$ (dokumentów zignorowanych do kroku i) - na podstawie dokumentów zarekomendowanych oraz odrzuconych w bieżącej sesji tworzony jest wektor określający, jakie zarekomendowane dokumenty nie spotkały się z zainteresowaniem użytkownika.
- Obliczenie wektora współczynników nawigacyjnych - podobieństwa między wektorem $\mathbf{z}^{(i)}$ (dokumentów odwiedzonych do kroku i), a poszczególnymi wzorcami nawigacyjnymi służy do obliczenia współczynników nawigacyjnych, których głównym zadaniem jest promowania dokumentów, które były często odwiedzane w sesjach składających się na wzorce nawigacyjne podobne do wektora $\mathbf{z}^{(i)}$.
- Obliczenie wartości funkcji f w kroku i – dla każdego dokumentu określana jest wartość funkcji rekomendacji zgodnie z formułą zawartą w definicji 4.5.3.

4.5. Definicja hybrydowej metody rekomendacji

Po ogólnym opisie metody *AdaptRank* w niniejszym podrozdziale przedstawiona zostanie definicja funkcji rekomendacji f będącej podstawą metody rekomendacji *AdaptRank*. Zanim to jednak nastąpi wprowadzone zostaną wielkości, niezbędne do wyznaczenia wartości funkcji rekomendacji: funkcja przydatności dokumentów, wektor dokumentów odwiedzonych, wektor dokumentów zignorowanych oraz wektor współczynników wzorców nawigacyjnych.

4.5.1. Określanie przydatności dokumentów

W rozdziale 2 przy okazji opisywania systemów wyszukiwania nadmieniono, że jednym z elementów wpływających na pozycję dokumentu w rankingu jest jego jakość, określająca ogólne walory dokumentu niezależne od zadanego zapytania. W niniejszej pracy zamiast terminu „jakość” używane będzie szersze pojęcie przydatności, które w tym przypadku będzie obejmowało cechy związane z jakością oraz popularnością dokumentu.

Funkcja pozwalająca określić przydatność dokumentu może być bardzo użyteczna w procesie rekomendacji z dwóch powodów. Po pierwsze pozwala ona zmniejszyć wpływ treści o niskiej wartości. Po drugie „bezkontekstowa” przydatność dokumentu może być bardzo ważnym kryterium rekomendacji, gdy liczba stron odwiedzonych przez użytkownika w bieżącej sesji jest jeszcze mała, a w konsekwencji jego preferencje nie są znane. Innymi słowy, autor pracy postuluje, aby system rekomendujący, w przypadku posiadania małej ilości danych o preferencjach użytkownika, podpowiadał dokumenty o obiektywnie dużej przydatności. Oczywiście, wraz ze wzrostem liczby odwiedzonych stron przydatność dokumentu powinna odgrywać coraz mniejszą rolę wobec innych wielkości pozwalających określić relewancję danego dokumentu w bieżącej sesji

W (Kiewra i Nguyen, 2005) przedstawiono kilka najbardziej znaczących rankingów przydatności dla systemów wyszukiwawczych o charakterze globalnym. Dodatkowo, pokazano, że globalne rankingi mogą źle porządkować dokumenty w przypadku pojedynczego systemu hipertekstowego. Dlatego w niniejszym podrozdziale zaproponowana zostanie nowa funkcja określająca przydatność dokumentu d oznaczana jako $q(d)$. Pozwoli ona wyznaczyć początkowy ranking przydatności przy użyciu danych statystycznych specyficznych dla pojedynczego systemu hipertekstowego.

Wszystkie wielkości numeryczne wykorzystywane do obliczania funkcji przydatności mogą zostać podzielone na dwie następujące grupy:

- Współczynniki użytkowania – w skład tej grupy wchodzi: współczynnik otwarcia i liczba odwiedzin.
- Współczynniki jakości – do których zaliczamy dostępność, liczbę działających odsyłaczy, oraz współczynnik nowości.

Kolejne punkty przybliżą każdy z wyżej wymienionych współczynników wraz z podaniem jego intuicji i sposobu wyznaczania.

Dostępność dokumentu

Nawet najciekawszy i najbardziej relewantny dokument nie ma dużej wartości, jeżeli nie jest zawsze dostępny dla użytkownika. Dostępność dokumentu „faworyzuje” dokumenty „bezwaryjne”, których żądanie przesłania nigdy lub prawie nigdy nie kończy się błędem. Dostępność dokumentu d wyznaczana jest poprzez odjęcie od jedności liczby powstałej w wyniku podzielenia całkowitego czasu, w którym dokument był niedostępny przez czas życia dokumentu w systemie hipertekstowym:

$$av(d) = 1 - \frac{t_u^d}{t_c - t_0^d} \quad (4.5.1)$$

gdzie t_c oznacza znacznik czasowy chwili obecnej.

Współczynnik działających odsyłaczy

Jednym z typowych zadań administratora systemu hipertekstowego jest poszukiwanie martwych odsyłaczy (*ang. dead links*) – tzn. takich odsyłaczy hipertekstowych, w których dokument docelowy nie istnieje lub od dłuższego czasu jest niedostępny. Istnienie martwych odsyłaczy powinno obniżać przydatność dokumentu, gdyż odnośniki do nie działających zasobów nie tylko irytują użytkownika, ale także kwestionują profesjonalizm całego systemu hipertekstowego. Współczynnik działających odsyłaczy dokumentu d można formalnie zapisać jako stosunek liczby działających odsyłaczy do wszystkich odsyłaczy, w których dokument d jest dokumentem źródłowym:

$$gl(d) = \begin{cases} \frac{card(L_g^d)}{card(L^d)} & \text{dla } card(L^d) > 0 \\ 0 & \text{dla } card(L^d) = 0 \end{cases} \quad (4.5.2)$$

gdzie L_g^d oznacza zbiór odsyłaczy, wychodzących z dokumentu d , które działają.

Liczba odwiedzin

Liczba żądań przesłania danego dokumentu (zwana też liczbą odwiedzin) jest jedną z podstawowych wielkości używanych w statystykach aktywności użytkowników systemów hipertekstowych. Liczba odwiedzin będzie wykorzystywana do „promowania” dokumentów często przesyłanych użytkownikom.

Głównym mankamentem związanym z tym współczynnikiem jest ciągle wzmacnianie dokumentów odznaczających się dużą popularnością (ich częste odwiedzanie powoduje, że zajmują wysokie miejsce w rankingu, co w konsekwencji powoduje, że są one jeszcze częściej odwiedzane). Jednym ze sposobów na zmniejszenie tego negatywnego trendu jest obliczanie funkcji średnich odwiedzin na daną jednostkę czasu. Średnia liczba odwiedzin dokumentu d będzie wyznaczona ze wzoru:

$$tr(d) = \begin{cases} \frac{card(U^d)}{t_c - t_0^d} & \text{dla } t_c - t_0^d > 0 \\ 0 & \text{dla } t_c - t_0^d = 0 \end{cases} \quad (4.5.3)$$

gdzie t_c oznacza znacznik czasowy chwili obecnej.

W celu porównywalnego wpływu wszystkich współczynników na przydatność dokumentów każdy z nich musi przyjmować wartości z przedziału $[0, 1]$ (musi zostać znormalizowany do 1). Najłatwiejszą metodą normalizacji do jedności jest podzielenie przez maksymalną liczbę odwiedzin. W ten sposób otrzymujemy względną średnią liczbę odwiedzin:

$$wtr(d) = \begin{cases} \frac{tr(d)}{\max(tr)} & \text{dla } \max(tr) > 0 \\ 0 & \text{dla } \max(tr) = 0 \end{cases} \quad (4.5.4)$$

gdzie $\max(tr)$ to maksymalna wartość funkcji tr .

Współczynnik otwarcia

Informacja na temat dokumentów, które odwiedzane są w bieżących sesjach jako pierwsze jest bardzo często wykorzystywana w statystykach aktywności użytkowników systemów hipertekstowych. Dokumenty te mogą być uznawane za łączniki między systemem hipertekstowym a światem zewnętrznym i powinny cechować się wysoką przydatnością. Przekonuje o tym analiza głównych powodów, dla których dany dokument został odwiedzony jako pierwszy w bieżącej sesji:

- Użytkownik wysłał do systemu hipertekstowego identyfikator dokumentu podając go z pamięci, albo z pewnej listy zanotowanych identyfikatorów¹ (w obydwu przypadkach można przypuszczać, że dokument ten jest dla użytkownika stosunkowo ważny skoro zadał sobie trud, aby go zapamiętać lub zapisać)
- Użytkownik odwiedził dany dokument na skutek wybrania odsyłacza hipertekstowego umieszczonego w dokumencie należącym do innego systemu hipertekstowego (przynajmniej autor dokumentu źródłowego musiał uważać docelowy dokument za ważny)
- Dokument został „odnaleziony” za pomocą globalnego systemu wyszukiwania. W tym przypadku możemy domniemywać, że użytkownik bazując na streszczeniu uważa, iż dokument ten jest relewantny do sformułowanego zapytania.

¹ Taka lista w przypadku systemu WWW nazywa się „listą ulubionych” i jest ona implementowana przez większość przeglądarek.

Współczynnik otwarcia dokumentu d jest to iloraz liczby historycznych sesji, w których dokument d był odwiedzony jako pierwszy do liczby wszystkich historycznych sesji, w których dokument d był odwiedzony:

$$or(d) = \begin{cases} \frac{card(S_1^d)}{card(S^d)} & \text{dla } card(S^d) > 0 \\ 0 & \text{dla } card(S^d) = 0 \end{cases} \quad (4.5.5)$$

gdzie $S^d \subseteq S$ oznacza zbiór wszystkich historycznych sesji, w których dokument d ($d \in D$) został odwiedzony, a $S_1^d \subseteq S^d \subseteq S$ to podzbiór historycznych sesji z S^d , w których dokument d był odwiedzony jako pierwszy, co formalnie można zapisać jako:

$$S_1^d = \{(s, d_s) \in S^d : d_s = d\} \quad (4.5.6)$$

Współczynnik nowości

Nowe dokumenty nie posiadają wysokiego miejsca w rankingach systemów wyszukiwania wykorzystujących jakość dokumentu z dwóch zasadniczych powodów. Po pierwsze, nie ma dostatecznie dużej liczby dokumentów, które by do nich odsyłały, co implikuje niską jakość liczoną w oparciu o odsyłacze (Brin i Page, 1998), (Kleinberg, 1999). Po drugie, mało użytkowników odwiedziło te dokumenty, co przekłada się na niewysoką jakość w rankingach typu *DirectHIT* (WWW, 2005b). Dlatego też, aby poprawić wartość jakości dla nowych dokumentów wprowadzony został współczynnik nowości

Współczynnik nowości dla danego dokumentu d może zostać obliczony ze wzoru:

$$nr(d) = \begin{cases} \rho_t^{day(t_c) - day(t_0^d)} & \text{dla } day(t_c) - day(t_0^d) > \varepsilon \\ 0 & \text{dla } day(t_c) - day(t_0^d) \leq \varepsilon \end{cases} \quad (4.5.7)$$

gdzie day – jest funkcją zamieniającą jednostkę czasu używaną w reprezentacji znacznika czasowego na liczbę dni, a $\rho_t \in (0,1)$ jest parametrem determinującym jak szybko dany dokument przestaje być uważany za nowy. Ze względu na ograniczenie kosztownych operacji zmiennoprzecinkowych wprowadzono wartość progową ε , poniżej której wartość funkcji nr jest równa zero.

Definicja funkcji przydatności

Definicja 4.5.1

Funkcją przydatności dokumentu d nazywamy znormalizowaną kombinację liniową współczynnika otwarcia dokumentu d , dostępności w systemie, liczby działających odsyłaczy, które on zawiera oraz wartości maksymalnej wyznaczonej pomiędzy jego współczynnikiem nowości, a średnią liczbą odwiedzin:

$$q(d) = \frac{\mu \cdot \max(nr(d), wtr(d)) + \lambda \cdot or(d) + \chi \cdot gl(d) + \phi \cdot av(d)}{\mu + \lambda + \phi + \chi}$$

gdzie μ , λ , χ , ϕ są parametrami z przedziału $[0, 1]$ ($\mu + \lambda + \phi + \chi > 0$) regulującymi wpływ poszczególnych składników na wartość funkcji przydatności..

Wartości parametrów μ , λ , χ , ϕ są uzależnione od charakterystyki systemu hipertekstowego. Przykładowo, jeśli dany system jest stosunkowo nowy i nie posiada jeszcze dużej ilości danych na temat zachowań użytkowników to wartości parametrów μ -i λ powinny być relatywnie mniejsze od wartości pozostałych parametrów. Suma występująca w mianowniku ma za zadanie znormalizowanie wartości funkcji przydatności do jedności.

4.5.2. Wektor dokumentów odwiedzonych

W definicji 4.1.3 ustalono, że wektor $\mathbf{z}^{(i)} = (z^{(i)}_1, z^{(i)}_2, \dots, z^{(i)}_n)$ dokumentów odwiedzonych do kroku i jest elementem stanu bieżącej sesji s_b w kroku i . Jego współrzędne określają jak dawno dany dokument d_j był odwiedzony w bieżącej sesji. Poniżej podany jest sposób wyznaczania współrzędnej $z^{(i)}_j$ stosowany w metodzie *AdaptRank*.

(4.5.8)

$$z_j^{(1)} = \begin{cases} 0 & \text{dla } d^{(1)} \neq d_j \\ 1 & \text{dla } d^{(1)} = d_j \end{cases}$$

$$z_j^{(i+1)} = \begin{cases} \rho_s \cdot z_j^{(i)} & \text{dla } d^{(i)} \neq d_j \\ 1 & \text{dla } d^{(i)} = d_j \end{cases}$$

gdzie $d^{(i)}$ to dokument odwiedzony w kroku i oraz $\rho_s \in (0, 1)$ to parametr, który określa jak szybko maleje znaczenie dokumentów odwiedzonych podczas bieżącej sesji.

Przyglądając się formule 4.5.8 łatwo zauważyć, że przed pierwszym krokiem wektor dokumentów odwiedzonych jest wektorem zerowym oraz że w kolejnych krokach współrzędna dokumentu właśnie odwiedzonych przyjmuje wartość jeden, a wszystkie pozostałe, które są niezerowe będą zmniejszać się w tempie geometrycznym.

Wektor dokumentów odwiedzonych służy do formalnego zapisu aktualnych preferencji ujawnionych w bieżącej sesji. W celu lepszego zilustrowania zmian wektora $\mathbf{z}^{(i)}$ można posłużyć

się następującym przykładem: system w kolejnych krokach przywiązuje coraz mniejsze znaczenie do faktu odwiedzenia danego dokumentu by w końcu o tym fakcie „zapomnieć”.

W procesie obliczania wartości funkcji rekomendacji f wektor dokumentów odwiedzonych spełnia dwa zasadnicze zadania:

- Określa podobieństwo obecnych zainteresowań użytkownika do wzorców nawigacyjnych określonych w fazie przygotowawczej. Posiada to następującą intuicję: system powinien rekomendować dokumenty nieodwiedzone przez użytkownika a będące często odwiedzane w historycznych sesjach składających się na wzorce nawigacyjne podobne do wektora $\mathbf{z}^{(i)}$
- Obniża wartość funkcji rekomendacji dla dokumentów ostatnio odwiedzonych

4.5.3. Wektor dokumentów zignorowanych

W definicji 4.1.3 ustalono, że wektor $\mathbf{g}^{(i)} = (g^{(i)}_1, g^{(i)}_2, \dots, g^{(i)}_n)$ dokumentów zignorowanych do kroku i jest elementem stanu bieżącej sesji s_b w kroku i . Jego współrzędne określają jak dawno dany dokument d_j był zignorowany w bieżącej sesji. Poniżej podany jest sposób wyznaczania współrzędnej $z^{(i)}_j$ stosowany w metodzie *AdaptRank*.

$$g_j^{(1)} = 0 \quad (4.5.9)$$

$$g_j^{(i+1)} = \begin{cases} 1 - \alpha & \text{dla } d^{(i)} \neq d_j \wedge g_j^{(i)} = 1 \\ \gamma \cdot g_j^{(i)} & \text{dla } d^{(i)} \neq d_j \wedge g_j^{(i)} \neq 1 \\ 1 & \text{dla } d^{(i)} = d_j \end{cases}$$

gdzie $d^{(i)}$ to dokument odwiedzony w kroku i oraz $\gamma \in (\alpha, 1)$ to parametr, który określa jak szybko maleje znaczenie zignorowania dokumentu podczas bieżącej sesji. Parametr $\alpha \in (0, 1)$ będzie określał, jaki wpływ na wartość funkcji rekomendacji w bieżącym kroku, ma wartość funkcji rekomendacji z kroku poprzedniego (patrz definicja 4.5.3).

Przyglądając się formule 4.5.9 łatwo zauważyć, że przed pierwszym krokiem wektor dokumentów zignorowanych jest wektorem zerowym oraz że w kolejnych krokach współrzędna dokumentu właśnie zignorowanego przyjmuje wartość równą jedności, a wszystkie pozostałe są zmniejszane w tempie geometrycznym, oprócz współrzędnych zerowych oraz współrzędnej odpowiadającej dokumentowi, który został zignorowany w poprzednim kroku. Powyższe odstępstwo gwarantuje, że w kolejnych krokach fakt zignorowania danego dokumentu będzie miał coraz mniejszy wpływ na wartość funkcji rekomendacji f (patrz twierdzenie 5.1.6).

4.5.4. Współczynnik wzorców nawigacyjnych

Wektor współczynników wzorców nawigacyjnych jest wektorem, którego każda współrzędna odpowiada jednemu dokumentowi ze zbioru D , a jej wartość z przedziału $[0, 1]$ określa zainteresowanie użytkownika określonym dokumentem. Zainteresowanie to obliczane jest na podstawie zachowań innych użytkowników reprezentowanych przez wzorce nawigacyjne. Poniżej znajduje się definicja wektora współczynników wzorców nawigacyjnych.

Definicja 4.5.2

Wektorem współczynników wzorców nawigacyjnych w kroku i ($i=1, \dots, n_{sb}$) nazywamy wektor $\mathbf{nv}^{(i)} = (nv^{(i)}_1, nv^{(i)}_2, \dots, nv^{(i)}_n)$, którego współrzędna $nv^{(i)}_j$ ($j=1, \dots, n$) jest średnią arytmetyczną otrzymaną z iloczynów wyznaczonych pomiędzy odpowiednimi współrzędnymi wzorców nawigacyjnych a podobieństwem wektora dokumentów odwiedzonych $\mathbf{z}^{(i)}$ do odpowiedniego wzorca nawigacyjnego, przy czym brane są pod uwagę tylko te wzorce nawigacyjne, których podobieństwo do wektora $\mathbf{z}^{(i)}$ jest odpowiednio duże, tj.:

$$nv_j^{(i)} = \frac{1}{n_b} \sum_{k=1}^{n_c} b_k \cdot c_{kj}$$

gdzie n_b oznacza liczbę wzorców, dla których $b_k > 0$, n_c – odpowiada liczbie utworzonych wzorców, a b_k jest wagą dla wzorca nawigacyjnego c_k i wyraża się wzorem:

$$b_k = \begin{cases} \text{sim}(\mathbf{z}^{(i)}, \mathbf{c}_k) & \text{gdym } \text{sim}(\mathbf{z}^{(i)}, \mathbf{c}_k) \geq \varepsilon_b \\ 0 & \text{gdym } \text{sim}(\mathbf{z}^{(i)}, \mathbf{c}_k) < \varepsilon_b \end{cases}$$

Parametr $\varepsilon_b \in (0, 1)$ określa próg, jaki musi przekroczyć podobieństwo między wzorcem nawigacyjnym \mathbf{c}_k a wektorem $\mathbf{z}^{(i)}$ aby wzorec ten miał wpływ na $\mathbf{nv}^{(i)}$.

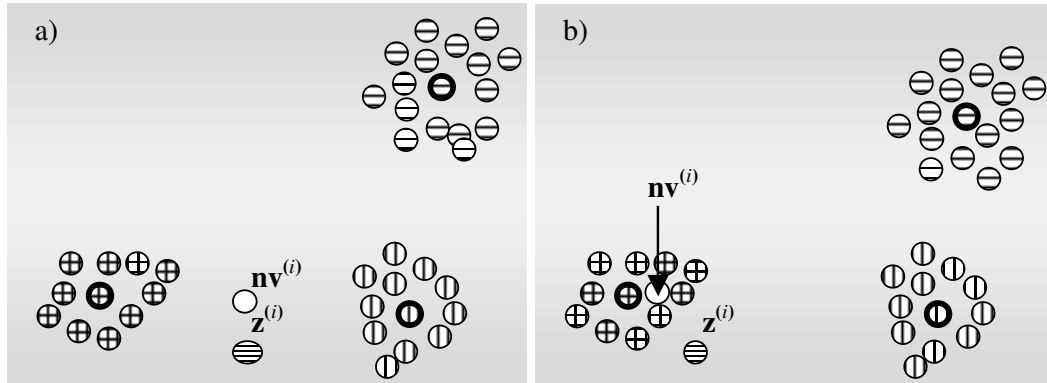
Współrzędną wektora współczynników wzorców nawigacyjnych odpowiadającą dokumentowi d będziemy w skrócie określać mianem współczynnika wzorców nawigacyjnych dokumentu d , a wektor współczynników wzorców nawigacyjnych, po prostu, współczynnikiem wzorców nawigacyjnych.

Rekomendacja w oparciu o współczynnik wzorców nawigacyjnych jest rozszerzeniem metody podanej w (Mobasher i inni, 2000b) z tą tylko różnicą, że na ostateczną wartość współczynnika wzorców nawigacyjnych wpływ mają wszystkie wzorce nawigacyjne, których podobieństwo do wektora dokumentów odwiedzonych jest dostatecznie duże. Wpływ ten jest właśnie określony przez podobieństwo między wektorem dokumentów odwiedzonych, a wzorcem nawigacyjnym (Kiewra, 2005c).

Warto podkreślić, że jeśli wektor $\mathbf{z}^{(i)}$ jest bardzo podobny do jednego ze wzorców i wyraźnie różny od pozostałych, to współczynnik wzorców nawigacyjnych upodabnia się do najbliższego wzorca nawigacyjnego (patrz rysunek 4.5.1b). Na rysunku tym wzorce nawigacyjne zaznaczone zostały za pomocą czarnej obwódki, a wektor współczynników wzorców nawigacyjnych jest punktem z białym wypełnieniem.

Jeżeli zaś istnieje kilka grup nawigacyjnych stosunkowo podobnych do wektora $\mathbf{z}^{(i)}$, to wektor $\mathbf{nv}^{(i)}$ będzie bardzo podobny do średniego wektora wyznaczonego z odpowiednich centroidów (patrz rysunek 4.5.1a). Oznacza to, że jeżeli zainteresowania właściciela bieżącej sesji odpowiadają jednemu ze wzorców nawigacyjnych (np. dotyczącego koszykówki), to dokumenty poświęcone koszykówce będą miały wysokie współrzędne w wektorze współczynników wzorców nawigacyjnych (rysunek 4.5.1b). Jeśli zaś zainteresowania użytkownika będą dotyczyły kilku wzorców nawigacyjnych (np. piłka nożna, koszykówka i siatkówka), to dokumen-

ty dotyczące tych 3 gier zespołowych będą miały największe współrzędne w wektorze $\mathbf{nv}^{(i)}$ (rysunek 4.5.1a).



Rysunek 4.5.1a-b „Położenie” wektora współczynników wzorców nawigacyjnych, gdy wektor dokumentów odwiedzonych jest tak samo oddalony od kilku najbliższych wzorców nawigacyjnych (rysunek a) i gdy wektor sesji jest wyraźnie bliższy jednemu z wzorców (rysunek b)

Na zakończenie opisu współczynnika wzorców nawigacyjnych należałoby przyjrzeć się roli parametru ε_b . Parametr ten eliminuje wpływ wzorców o niskim podobieństwie do wektora dokumentów odwiedzonych $\mathbf{z}^{(i)}$.

Nieuwzględnienie parametru ε_b doprowadziłoby do sytuacji, w której dla niektórych bieżących sesji wartość współczynnika nawigacyjnego dla relewantnych dokumentów mogłaby znacznie się zmniejszyć. Ilustruje to następujący przykład:

Przykład 4.5.1

Założmy, że $\varepsilon_b=0,5$, $n_c=3$, $\mathbf{c}_1=(1, 0,9, 0, 0, 0,8)$, $\mathbf{c}_2=(0, 0, 1, 1, 0)$, $\mathbf{c}_3=(0, 1, 0, 0, 1, 0)$ oraz $\mathbf{z}^{(2)}=(1, 0,9, 0, 0, 0)$. Z definicji 2.1.2 łatwo obliczyć podobieństwo między $\mathbf{z}^{(2)}$ a odpowiednimi wzorcami nawigacyjnymi (podane wyniki zaokrąglono do dwóch miejsc po przecinku)

$$\text{sim}(\mathbf{z}^{(2)}, \mathbf{c}_1)=0,86, \text{sim}(\mathbf{z}^{(2)}, \mathbf{c}_2)=0, \text{sim}(\mathbf{z}^{(2)}, \mathbf{c}_3)=0,07$$

Tylko podobieństwo pomiędzy \mathbf{c}_1 i $\mathbf{z}^{(2)}$ przekracza wartość progową określoną przez ε_b , więc stosując wzór na współczynnik wzorców nawigacyjnych podany w definicji 4.5.1 otrzymujemy:

$$\mathbf{nv}^{(2)} = 0,86 \cdot (1, 0,9, 0, 0, 0,8) = (0,86, 0,801, 0, 0, 0,688)$$

Założmy przez moment, że parametr ε_b nie jest używany, a wzór na współczynnik nawigacyjny w kroku i ma następującą postać:

$$nv_j^{(i)} = \frac{1}{n_c} \sum_{k=1}^{n_c} \text{sim}(\mathbf{z}^{(i)}, \mathbf{c}_k) \cdot c_{kj} \quad (4.5.10)$$

Wtedy współczynnik nawigacyjny z poruszanego przykładu byłby równy:

$$\begin{aligned} \mathbf{nv}'^{(2)} &= 0,33 \cdot [(1, 0,9, 0, 0, 0,8) \cdot 0,86 + 0,07 \cdot (0,1, 0, 0, 1, 0) + 0] = \\ &= 0,33[(0,86, 0,801, 0, 0, 0,688) + (0,007, 0, 0, 0,07, 0)] = (0,289, 0,267, 0, 0,023, 0,115) \end{aligned}$$

Łatwo zauważyć $\mathbf{nv}'^{(2)}$ ma współrzędne znacznie niższe od $\mathbf{nv}^{(2)}$. Gdyby podobieństwo $\text{sim}(\mathbf{z}^{(2)}, \mathbf{c}_2)$ było minimalnie większe od 0 współczynnik nawigacyjny $\mathbf{nv}'^{(2)}$ stanowiłby w

pewnym przybliżeniu trzecią część oryginalnego współczynnika. Jeśli istniałoby dziesięć wzorców o podobieństwie względnie $\mathbf{z}^{(2)}$ minimalnie większym od 0, współczynnik nawigacyjny $\mathbf{nv}^{(2)}$ stanowiłby w pewnym przybliżeniu dziesiątą część oryginalnego współczynnika i miałby bardzo mały wpływ na wartość funkcji rekomendacji (patrz podrozdział 4.5.3).

4.5.5. Definicja funkcji rekomendacji

Definicja 4.5.3

Wartość funkcji rekomendacji f dla dokumentu $d_j \in D$ w kroku i ($i = 1, \dots, n_{sb}$) bieżącej sesji s_b jest wyznaczana w następujący sposób:

$$f(d_j, s_b^{(1)}) = (0,5 \cdot (\beta + \delta) \cdot \alpha \cdot q(d_j) + \beta \cdot nv_j^{(1)} + \delta \cdot \text{sim}(\mathbf{v}^{d^{(1)}}, \mathbf{v}^{d_j})) \cdot (1 - z_j^{(1)})$$

$$f(d_j, s_b^{(i+1)}) = (\alpha \cdot f(d_j, s_b^{(i)}) + \beta \cdot nv_j^{(i+1)} + \delta \cdot \text{sim}(\mathbf{v}^{d^{(i+1)}}, \mathbf{v}^{d_j}) - g_j^{(i+1)}) \cdot (1 - z_j^{(i+1)})$$

gdzie $\beta, \delta \in (0, 1]$ oraz $\alpha \in (0; 1)$ to parametry regulujące wpływ poszczególnych elementów na wartości funkcji, a $d^{(i)}$ to dokument odwiedzony w kroku i .

Definicja 4.5.3 posiada następującą intuicję: gdy użytkownik wysłał żądanie dotyczące pierwszego dokumentu, wartość funkcji f dla konkretnego dokumentu jest kombinacją liniową przydatności tego dokumentu, podobieństwa tekstowego do dokumentu obecnie oglądanego oraz odpowiedniej współrzędnej współczynnika wzorców nawigacyjnych. Wyrażenie $(1 - z_j^{(i+1)})$ przyjmuje wartość 1 dla wszystkich dokumentów oprócz obecnie odwiedzanego, więc można stwierdzić, że nie wpływa ono na wartość funkcji f w pierwszym kroku.

W kolejnych krokach wartość funkcji f dla danego dokumentu zależy także od wartości funkcji w kroku poprzednim oraz odpowiedniej współrzędnej wektora dokumentów zignorowanych. Dodatkowo wyrażenie $(1 - z_j^{(i+1)})$ obniża wartości funkcji dla dokumentów, które były już odwiedzane w bieżącej sesji.

Przyglądając się definicji funkcji rekomendacji f trudno nie dostrzec podobieństwa ze sprzężeniem zwrotnym relewancji przedstawionym w podrozdziale 2.1. Podobieństwo to nie jest przypadkowe, gdyż zdaniem autora użycie zmodyfikowanego mechanizmu sprzężenia zwrotnego relewancji w rekomendacji pozwoli osiągnąć wysoką efektywność unikając jednocześnie typowych niedogodności sprzężenia zwrotnego relewancji znanych z publikacji dotyczących wyszukiwania. Po pierwsze, sprzężenie zwrotne relewancji obliuguje użytkownika do określania relewancji wyszukanych dokumentów (tak jak już to zostało opisane w rozdziale 2 użytkownicy bardzo niechętnie oceniają pokazywaną im treść). Metoda rekomendacji *AdaptRank* zamiast zmuszać użytkownika do określania relewancji dokumentów, zakłada, że relewantne do potrzeb użytkownika są dokumenty, które są:

- podobne do dokumentów ostatnio odwiedzonych
- często odwiedzane w sesjach, w których użytkownicy zachowują się podobnie do zachowania ujawniającego się bieżącej sesji.

Innymi słowy, współczynnik wzorców nawigacyjnych oraz podobieństwo danego dokumentu do obecnie odwiedzanego mogą być potraktowane jako sprzężenie zwrotne dodatnie. Jedno-

częście zignorowanie zarekomendowanego dokumentu może być potraktowane jako sprzężenie zwrotne ujemne. Dodatkowo, należy zauważyć, że w przypadku rekomendacji nie mamy do czynienia z zapytaniami, dlatego zamiast zwiększać bądź zmniejszać szacowaną relewancję terminu w zapytaniu nastąpi bezpośrednio modyfikowanie stopnia relewancji dokumentu.

Druga trudność związana z wykorzystaniem sprzężenia zwrotnego relewancji w wyszukiwaniu przejawia się w iteracyjnym charakterze procesu (w pierwszym kroku użytkownik zadaje zapytanie – otrzymuje listę dokumentów ocenia relewancję kilku z nich, otrzymuje kolejną listę, ocenia relewancję kilku dokumentów, otrzymuje kolejną listę itd.). Przekłada się to na wydłużenie czasu potrzebnego do uzyskania zbioru dokumentów relewantnych do zadanego zapytania. Warto podkreślić, że problem ten nie występuje w przypadku rekomendacji, gdyż preferencje użytkownika określane są iteracyjnie w kolejnych krokach bieżącej sesji.

Wpływ poszczególnych elementów na wartość funkcji jest regulowany za pomocą użytych parametrów. W tabeli 4.5.1 omówiono ich znaczenie wraz z proponowanymi wartościami dobranymi na podstawie analizy funkcji f przedstawionej w następnym rozdziale.

Nazwa parametru	Opis	Proponowana wartość
α	Określa jaki wpływ na obecną wartość funkcji f ma wartość funkcji f z poprzedniego kroku. W sposób pośredni α reguluje wpływ przydatności dokumentu na wynik rekomendacji	0,25
β	Określa w jakim stopniu współczynnik wzorców nawigacyjnych wpływa na wartości funkcji rekomendacji f	1
δ	Określa wpływ podobieństwa tekstowego na wartości funkcji rekomendacji f	1
γ	Określa na ile obniży się wartość funkcji rekomendacji dla dokumentów zignorowanych.	0,9
ρ_s	Określa na ile obniży się wartość funkcji rekomendacji dla dokumentów wcześniej widzianych. Parametr ten jest „ukryty” w wektorze dokumentów odwiedzonych $\mathbf{z}^{(i)}$.	0,75

Tabela 4.5.1 Parametry regulujące wpływ poszczególnych elementów składowych na funkcję f .

Proponowane wartości parametrów nie są ostateczne i niezmiennie, a powinny raczej zależeć od charakterystyki systemu hipertekstowego. Na przykład, gdy dane dotyczące zachowań użytkowników nie są dostępne w wystarczającej ilości, wartość parametru β powinna być relatywnie mniejsza od pozostałych parametrów. Szczegółowa dyskusja nad wartościami parametrów przedstawiona została w podrozdziale 5.5.

4.6. Algorytm rekomendacji

Po zdefiniowaniu funkcji rekomendacji możliwe jest przedstawienie algorytmu rekomendacji. Zakłada się, że wszystkie wartości, które powinny być obliczone w fazie wstępnej są dostępne. W celu jednoczesnego wyznaczenia wartości funkcji rekomendacji f dla wszystkich dokumentów ze zbioru d wprowadzone zostanie pomocnicze pojęcie wektora rekomendacji:

Definicja 4.6.1

Wektorem rekomendacji w kroku i ($i = 1, \dots, n_{sb}$) nazywamy wektor $\mathbf{r}^{(i)} = (r^{(i)}_1, r^{(i)}_2, \dots, r^{(i)}_n)$, którego współrzędna $r^{(i)}_j$ równa jest wartości funkcji f dla dokumentu d_j w kroku $i > 0$ bieżącej sesji s_b :

$$r^{(i)}_j = f(d_j, s^{(i)}_b)$$

dla $j = 1, 2, \dots, n$. Dodatkowo wektor $\mathbf{r}^{(0)} = (r^{(0)}_1, r^{(0)}_2, \dots, r^{(0)}_n)$ będzie nazywany wektorem początkowym rekomendacji, którego współrzędna $r^{(0)}_j$ równa:

$$r^{(0)}_j = 0,5 \cdot \alpha \cdot (\beta + \delta) \cdot q(d_j)$$

Innymi słowy współrzędna j wektora rekomendacji określa, w jakim stopniu użytkownik byłby skłonny zobaczyć dokument d_j biorąc pod uwagę jego zachowanie podczas bieżącej sesji.

W celu jednoczesnego wyznaczenia podobieństwa wszystkich dokumentów ze zbioru d do dokumentu $d^{(i)}$ właśnie odwiedzanego wprowadzone zostanie pomocnicze pojęcie wektora podobieństwa tekstowego. Wyznaczenie wartości tego wektora w fazie przygotowawczej pozwoli na szybkie i łatwe odnalezienie dokumentów podobnych do obecnie odwiedzanego bez potrzeby obliczania podobieństwa między odpowiednimi wektorami deskryptorów w fazie interakcji.

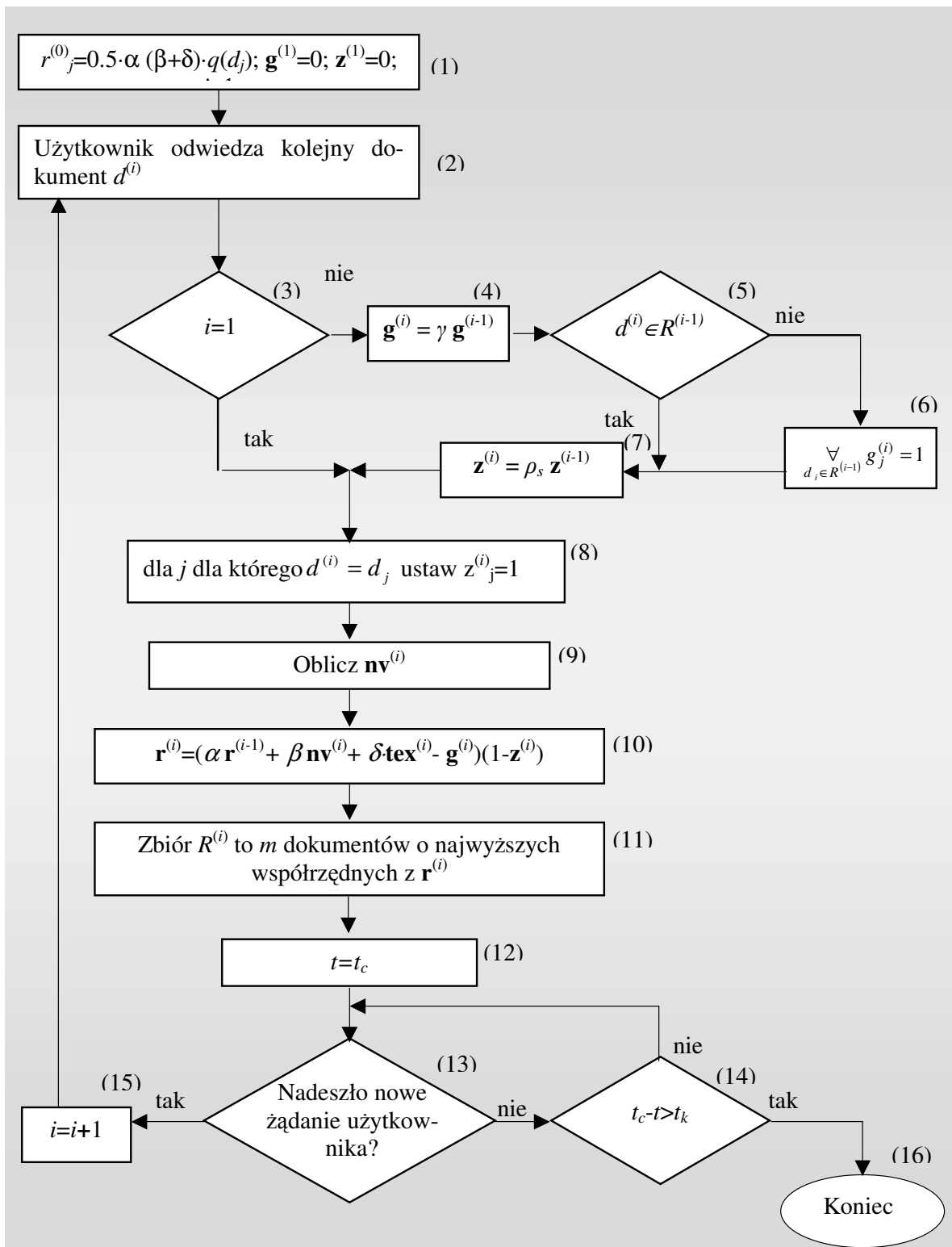
Definicja 4.6.2

Wektorem podobieństwa tekstowego dokumentu d_j ($j = 1, 2, \dots, n$) nazywamy wektor $\mathbf{tex}^{d_j} = (tex^{d_j}_1, tex^{d_j}_2, \dots, tex^{d_j}_n)$, którego współrzędna $tex^{d_j}_k$ (dla $k = 1, 2, \dots, n$) jest równa podobieństwu między wektorami deskryptorów dokumentów d_j oraz d_k :

$$tex^{d_j}_k = sim(\mathbf{v}^{d_j}, \mathbf{v}^{d_k})$$

W implementacji w celu zaoszczędzenia przestrzeni dyskowej można przekształcić zbiór wektorów podobieństwa w trójkątną macierz podobieństwa. Jednak na potrzeby algorytmu użyte będą wektory podobieństwa tekstowego zdefiniowane powyżej.

Rysunek 4.6.1 przedstawia schemat blokowy algorytmu rekomendacji przy użyciu metody *AdaptRank*.

Rysunek 4.6.1 Algorytm rekomendacji *AdaptRank*.

Każdy blok na rysunku 4.6.1 oznaczono numerem w celu łatwiejszego opisu algorytmu. Proces rekomendacji rozpoczyna się, gdy użytkownik połączy się z systemem. Jak można łatwo zauważyć na początku przypisywane są wartości wyjściowe składające się na profil użytkownika (blok 1). Po wysłaniu pierwszego żądania przesłania dokumentu algorytm przychodzi do

pierwszego kroku ($i=1$). W pierwszym kroku użytkownik nie zignorował jeszcze żadnego rekomendowanego dokumentu, gdyż żaden dokument nie został mu jeszcze podpowiedziany ani żadne dokumenty nie zostały wcześniej odwiedzone, wskutek czego pomijana jest część algorytmu dotycząca wyznaczenia wektora dokumentów zignorowanych. Podobnie nie zmniejszane są wartości współrzędnych dokumentów wcześniej odwiedzonych.

Algorytm przechodzi więc bezpośrednio do bloku 8, w którym przypisuje się wartość 1 współrzędnej wektora dokumentów odwiedzonych odpowiadającej właśnie odwiedzanemu dokumentowi. W bloku 9 obliczany jest współczynnik wzorców nawigacyjnych według formuły podanej w definicji 4.5.2. Następnie wyznaczany jest wektor rekomendacji dla pierwszego kroku (blok 10) i wybierane jest m rekomendowanych dokumentów, dla których wektor rekomendacji ma największe współrzędne (blok 11) i niniejszym kończy się krok 1.

Bloki oznaczone numerami 12, 13 i 14 służą do sprawdzenia czy bieżącą sesję należy uznać już za ukończoną. W bloku 11 pod zmienną t przypisywany jest znacznik chwili obecnej. Następnie przez t_k jednostek czasowych algorytm oczekuje na kolejne żądanie użytkownika (blok 13 i 14). Jeśli w czasie t_k żadne żądanie nie nadejdzie proces rekomendacji można uznać za zakończony. W przeciwnym przypadku algorytm przechodzi do następnego kroku. W kolejnych krokach przed powtórzeniem wszystkich operacji z kroku 1 należy zmniejszyć wartości współrzędnych wektora dokumentów odwiedzonych (blok 7) oraz wyznaczyć wektor dokumentów zignorowanych (blok 4, 5, 6). Zakłada się, że jeśli obecnie odwiedzany dokument należał do zbioru dokumentów rekomendowanych w poprzednim kroku, to żaden z zarekomendowanych dokumentów nie może zostać uznany za zignorowany.

5. Formalne własności metody rekomendacji

Celem rozdziału piątego jest przedstawienie formalnych własności metody rekomendacji *AdaptRank*. Można w nim wyróżnić następujące podrozdziały:

- *Ogólne własności metody rekomendacji* – poświęcony jest sformułowaniu ogólnych twierdzeń i wniosków, które będą przydatne w kolejnych podrozdziałach.
- *Ocena zdolności adaptacyjnych metody rekomendacji* – pokazuje, że niezależnie od doboru parametrów metoda *AdaptRank* jest w stanie dostosować się do nowych zainteresowań użytkownika. Dodatkowo wyprowadzony został wzór na maksymalną liczbę kroków potrzebnych do „zaadaptowania się” metody do nowych potrzeb użytkownika. Na zakończenie podrozdziału przeprowadzono dyskusję na temat wpływu parametrów na wyżej wspomnianą liczbę kroków.
- *Efektywności metody rekomendacji* - pokazuje, przy jakich założeniach metoda osiąga maksymalną dokładność, a przy jakich założeniach metoda osiąga maksymalną kompletność. Dodatkowo przeprowadzono dyskusję na temat wpływu wartości parametrów na osiągnięcia maksymalnej kompletności i dokładności.
- *Własności elementów składowych metody AdaptRank* – w podrozdziale tym przeanalizowano własności funkcji składowych funkcji *AdaptRank*.
- *Wpływ wartości parametrów na zachowanie się metody rekomendacji* – ostatni podrozdział poświęcony jest zebraniu informacji na temat wpływu wartości parametrów na zachowanie się metody.

5.1. Ogólne własności metody rekomendacji

W definicji 4.5.3 funkcja rekomendacji f w danym kroku determinowana jest przez wartość funkcji rekomendacji w kroku poprzednim. Biorąc pod uwagę iteracyjny charakter procesu rekomendacji, łatwo zauważyć, że formuła ta ułatwia podanie algorytmu rekomendacji. Z drugiej jednak strony nie jest korzystna dla analizy własności metody. Twierdzenie 5.1.1 podaje równoważną formułę, która ułatwi analizę metody *AdaptRank*.

Twierdzenie 5.1.1

Wartość funkcji rekomendacji f dla dokumentu d_j ($j=1,2,\dots,n$) w kroku i ($i=1,\dots,n_{sb}$) bieżącej sesji s_b może być wyznaczona za pomocą następującego wzoru:

$$f(d_j, s_b^{(i)}) = \alpha^i \cdot 0,5 \cdot (\beta + \delta) \cdot q(d_j) \cdot \prod_{k=1}^i (1 - z_j^{(k)}) + \beta \cdot \sum_{k=1}^i \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) + \delta \cdot \sum_{k=1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) - \sum_{k=1}^i \left(\alpha^{i-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right)$$

Dowód:

Przez $X(i)$ oznaczymy wielkość po prawej stronie równania w twierdzeniu. W celu udowodnienia powyższego twierdzenia zastosowana zostanie zasada indukcji matematycznej. Na początku pokazane zostanie, że powyższe twierdzenie jest prawdziwe dla $i=1$. Podstawiając $i=1$ otrzymujemy:

$$X(1) = \alpha \cdot \left(0,5 \cdot (\beta + \delta) \cdot q(d_j) \cdot (1 - z_j^{(1)})\right) + \beta \cdot nv_j^{(1)} \cdot (1 - z_j^{(1)}) + \delta \cdot \text{sim}(\mathbf{v}^{d^{(1)}}, \mathbf{v}^{d_j}) \cdot (1 - z_j^{(1)}) - g_j^{(1)} \cdot (1 - z_j^{(1)}) \quad (5.1.1)$$

Z formuły 4.5.9 wynika, że $g_j^{(1)} = 0$, co po wyłączeniu $(1 - z_j^{(1)})$ przed nawias daje:

$$X(1) = (1 - z_j^{(1)}) \cdot \left(\alpha \cdot 0,5 \cdot (\beta + \delta) \cdot q(d_j) + \beta \cdot nv_j^{(1)} + \delta \cdot \text{sim}(\mathbf{v}^{d^{(1)}}, \mathbf{v}^{d_j})\right) = f(d_j, s_b^{(1)}) \quad (5.1.2)$$

Tak, więc dla $i=1$ dowiedziona została prawdziwość twierdzenia, ponieważ otrzymaliśmy wyrażenie tożsame z definicją 4.5.3.

Zakładając, że twierdzenie jest prawdziwe dla i , pokażemy, że jest ono prawdziwe dla $i+1$.

Podstawmy najpierw $i+1$ do formuły X :

$$X(i+1) = \alpha^{i+1} \cdot \left(0,5 \cdot (\beta + \delta) \cdot q(d_j) \cdot \prod_{k=1}^{i+1} (1 - z_j^{(k)})\right) + \beta \cdot \sum_{k=1}^{i+1} \left(\alpha^{i+1-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^{i+1} (1 - z_j^{(l)})\right) + \delta \cdot \sum_{k=1}^{i+1} \left(\alpha^{i+1-k} \cdot \text{sim}(\mathbf{v}^{d^{(k+1)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^{i+1} (1 - z_j^{(l)})\right) - \sum_{k=1}^{i+1} \left(\alpha^{i+1-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^{i+1} (1 - z_j^{(l)})\right) \quad (5.1.3)$$

Po wyłączeniu $(1 - z_j^{(i+1)})$ przed nawias otrzymujemy:

$$X(i+1) = (1 - z_j^{(i+1)}) \left(\alpha^{i+1} \cdot \left(0,5 \cdot (\beta + \delta) \cdot q(d_j) \cdot \prod_{k=1}^i (1 - z_j^{(k)})\right) + \beta \cdot \sum_{k=1}^{i+1} \left(\alpha^{i+1-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)})\right) + \delta \cdot \sum_{k=1}^{i+1} \left(\alpha^{i+1-k} \cdot \text{sim}(\mathbf{v}^{d^{(k+1)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)})\right) - \sum_{k=1}^{i+1} \left(\alpha^{i+1-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)})\right) \right) \quad (5.1.4)$$

Następnie wyłączone zostaną ostatni składnik z trzech następujących sum: sumy wzorców nawigacyjnych, sumy podobieństw dokumentów i sumy współrzędnych wektora dokumentów zignorowanych:

$$\begin{aligned}
X(i+1) &= (1 - z_j^{(i+1)}) \left(\beta \cdot nv_j^{(i+1)} + \delta \cdot \text{sim}(\mathbf{v}^{d^{(i+1)}}, \mathbf{v}^{d_j}) - g_j^{(i+1)} \right) + \\
&\alpha^{i+1} \cdot \left(0,5 \cdot (\beta + \delta) \cdot q(d_j) \cdot \prod_{k=1}^i (1 - z_j^{(k)}) \right) + \beta \cdot \sum_{k=1}^i \left(\alpha^{i+1-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) + \\
&+ \delta \cdot \sum_{k=1}^i \left(\alpha^{i+1-k} \cdot \text{sim}(\mathbf{v}^{d^{(k+1)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) - \sum_{k=1}^i \left(\alpha^{i+1-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \quad (5.1.5)
\end{aligned}$$

Z wszystkich składników sumy znajdujących się w drugiej i trzeciej linii wyrażenia 5.1.5 wyłączone zostanie α :

$$\begin{aligned}
X(i+1) &= (1 - z_j^{(i+1)}) \left(\beta \cdot nv_j^{(i+1)} + \delta \cdot \text{sim}(\mathbf{v}^{d^{(i+1)}}, \mathbf{v}^{d_j}) - g_j^{(i+1)} \right) + \\
&\alpha \cdot \left[\alpha^i \cdot \left(0,5 \cdot (\beta + \delta) \cdot q(d_j) \cdot \prod_{k=1}^i (1 - z_j^{(k)}) \right) + \beta \cdot \sum_{k=1}^i \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) + \right. \\
&\left. + \delta \cdot \sum_{k=1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k+1)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) - \sum_{k=1}^i \left(\alpha^{i-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \right] \quad (5.1.6)
\end{aligned}$$

Łatwo zauważyć, że wyrażenia znajdujące się w nawiasie kwadratowym formuły 5.1.6 jest równoważne wartości funkcji rekomendacji $f(d_j, s_b^{(i)})$ z założenia indukcyjnego oraz z definicji 4.5.3 mamy, więc:

$$\begin{aligned}
X(i+1) &= (1 - z_j^{(i+1)}) \left(\beta \cdot nv_j^{(i+1)} + \delta \cdot \text{sim}(\mathbf{v}^{d^{(i+1)}}, \mathbf{v}^{d_j}) - g_j^{(i+1)} + \alpha \cdot f(d_j, s_b^{(i)}) \right) = \\
&= f(d_j, s_b^{(i+1)}) \quad (5.1.7)
\end{aligned}$$

Powyższa równość oznacza, że twierdzenie jest prawdziwe dla $i+1$. Wynika stąd, że jest prawdziwe dla każdego i .

□

Dokumenty nieodwiedzone i nieignorowane w bieżącej sesji są szczególnie istotne dla formalnej analizy zachowań metody rekomendacji. Zważywszy na fakt, że wartość współrzędnej wektora dokumentów odwiedzonych odpowiadająca nieodwiedzonemu dokumentowi jest zawsze równa 0, oraz że wartość współrzędnej wektora dokumentów zignorowanych odpowiadająca niezignorowanemu dokumentowi jest także równa 0 można sformułować następujące twierdzenie:

Twierdzenia 5.1.2

Wartość funkcji rekomendacji f dla dokumentu d_j ($j=1,2,\dots,n$) nieodwiedzzonego i niezignorowanego do kroku i ($i=1,\dots,n_{sb}$) bieżącej sesji s_b jest równa:

$$f(d_j, s_b^{(i)}) = \alpha^i \cdot \left(0,5 \cdot (\beta + \delta) \cdot q(d_j) \right) + \beta \cdot \sum_{k=1}^i \left(\alpha^{i-k} \cdot nv_j^{(k)} \right) + \delta \cdot \sum_{k=1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \right)$$

Dowód:

Z definicji 4.1.3 wiadomo, że jeśli dokument d_j nie był zignorowany w bieżącej sesji do kroku i to:

$$\forall_{i \geq i > 0} g_j^{(i)} = 0$$

tak więc wartość funkcji f dla dokumentu d_j niezignorowanego do kroku i bieżącej sesji s_b jest równa:

$$f(d_j, s_b^{(i)}) = \alpha^i \cdot \left(0,5 \cdot (\beta + \delta) \cdot q(d_j) \cdot \prod_{k=1}^i (1 - z_j^{(k)}) \right) + \beta \cdot \sum_{k=1}^i \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \quad (5.1.8)$$

$$+ \delta \cdot \sum_{k=1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right)$$

Także z definicji 4.1.3 wiadomo, że jeśli dokument d_j nie był odwiedzony w bieżącej sesji s_b do kroku i to:

$$\forall_{i \geq i > 0} z_j^{(i)} = 0$$

Tak więc wartość funkcji f dla dokumentu d_j niezignorowanego a zarazem nieodwiedzonego do kroku i sesji s_b jest równa:

$$f(d_j, s_b^{(i)}) = \alpha^i \cdot (0,5 \cdot (\beta + \delta) \cdot q(d_j)) + \beta \cdot \sum_{k=1}^i (\alpha^{i-k} \cdot nv_j^{(k)}) + \delta \cdot \sum_{k=1}^i (\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j})) \quad (5.1.9)$$

□

Kolejne twierdzenie (wyprowadzone bezpośrednio z pierwszego) determinuje maksymalną wartość funkcji rekomendacji dla nieodwiedzonego i niezignorowanego dokumentu (korzysta się tu z faktu, że maksymalna wartość podobieństwa dwóch wektorów oraz współrzędnych wzorca nawigacyjnego jest równa 1, a odpowiednia współrzędna wektora dokumentów zignorowanych jest równa 0).

Twierdzenie 5.1.3

Maksymalna wartość funkcji rekomendacji f dla dokumentu d_j ($j=1, 2, \dots, n$) nieodwiedzonego i niezignorowanego do kroku i ($i=1, 2, \dots, n_{s_b}$) bieżącej sesji s_b jest równa:

$$f(d_j, s_b^{(i)}) = \alpha^i \cdot 0,5 \cdot (\beta + \delta) + \beta \cdot \sum_{k=1}^i \alpha^{i-k} + \delta \cdot \sum_{k=1}^i \alpha^{i-k}$$

Dowód:

Z twierdzenia 5.1.2 znana jest wartość funkcji dla nieodwiedzonego i niezignorowanego dokumentu:

$$f(d_j, s_b^{(i)}) = \alpha^i \cdot (0,5 \cdot (\beta + \delta) \cdot q(d_j)) + \beta \cdot \sum_{k=1}^i (\alpha^{i-k} \cdot nv_j^{(k)}) + \delta \cdot \sum_{k=1}^i (\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j})) \quad (5.1.10)$$

Maksymalna wartość przydatności dokumentu d_j oraz maksymalna wartość współczynnika wzorców nawigacyjnych są równe jedności, również maksymalne podobieństwo między dokumentem d_j , a dokumentem $d^{(i)}$ odwiedzionym w kroku i jest równe jedności. Wziąwszy pod uwagę, że nie ma żadnych okoliczności, które wykluczałyby jednoczesne osiągnięcie wartości maksymalnej dla przydatności, współczynnika wzorców nawigacyjnych dokumentu d_j oraz jego podobieństwa do odwiedzanych dokumentów, możemy podstawić do wyrażenia 5.1.10 następujące wartości: $q(d_j) = 1$, $\forall_{i \geq k > 0} nv_j^{(k)} = 1$ oraz $\forall_{i \geq k > 0} sim(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) = 1$ co ostatecznie daje nam:

$$f(d_j, s_b^{(i)}) = \alpha^i \cdot 0,5 \cdot (\beta + \delta) + \beta \cdot \sum_{k=1}^i \alpha^{i-k} + \delta \cdot \sum_{k=1}^i \alpha^{i-k} \quad (5.1.11)$$

□

Biorąc pod uwagę twierdzenie 5.1.2 oraz następujące fakty: $q(d_j) \geq 0$, $\forall_{n_{s_b} \geq i > 0} nv_j^{(i)} \geq 0$, $\forall_{n_{s_b} \geq i > 0} sim(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \geq 0$, można sformułować następujący wniosek

Wniosek 5.1.1

Minimalna wartość funkcji rekomendacji f dla dokumentu d_j ($j=1, 2, \dots, n$) nieodwiedzonego i niezignorowanego do kroku i ($i=1, 2, \dots, n_{s_b}$) bieżącej sesji s_b jest równa 0, tj.:

$$f(d_j, s_b^{(i)}) \geq 0$$

Kolejne twierdzenie pokazuje o ile może maksymalnie wzrosnąć wartość funkcji rekomendacji w porównaniu do poprzedniego kroku. Twierdzenie to będzie przydatne, do uzasadnienia, dlaczego zbiór wartości funkcji rekomendacji powinien posiadać kres dolny oraz kres górny.

Twierdzenie 5.1.4

Maksymalny wzrost wartości funkcji rekomendacji dla dokumentu d_j ($j=1, 2, \dots, n$) nieodwiedzonego i niezignorowanego w bieżącej sesji s_b w stosunku do poprzedniego kroku wynosi $\beta + \delta$, tj.:

$$f(d_j, s_b^{(i+1)}) - f(d_j, s_b^{(i)}) \leq \beta + \delta$$

Dowód:

Należy wyznaczyć maksymalną różnicę między wartościami funkcji rekomendacji dla tego samego dokumentu w dwóch kolejnych krokach. Korzystając z definicji 4.5.3 można zapisać:

$$f(d_j, s_b^{(i+1)}) - f(d_j, s_b^{(i)}) = \left(\alpha \cdot f(d_j, s_b^{(i)}) + \beta \cdot nv_j^{(i+1)} + \delta \cdot sim(\mathbf{v}^{d^{(i+1)}}, \mathbf{v}^{d_j}) - g_j^{(i+1)} \right) \cdot (1 - z_j^{(i+1)}) - f(d_j, s_b^{(i)}) \quad (5.1.12)$$

Z twierdzenia wiadomo, że dokument d_j nie był ani odwiedzony ani zignorowany, więc $g_j^{(i+1)} = 0$ oraz $z_j^{(i+1)} = 0$. Podstawiając te wielkości do wzoru 5.1.12 otrzymujemy:

$$f(d_j, s_b^{(i+1)}) - f(d_j, s_b^{(i)}) = \left(\alpha \cdot f(d_j, s_b^{(i)}) + \beta \cdot nv_j^{(i+1)} + \delta \cdot \text{sim}(\mathbf{v}^{d^{(k+1)}}, \mathbf{v}^{d_j}) \right) - f(d_j, s_b^{(i)}) \quad (5.1.13)$$

Aby wzrost wartości funkcji był maksymalny, wartość funkcji w kroku i musi być minimalna, a wartość funkcji w kroku $i+1$ musi być maksymalna. Z wniosku 5.1.4 wiadomo, że wartość minimalna wartości funkcji rekomendacji f dla nieodwiedzonego i niezignorowanego dokumentu wynosi 0, czyli korzystając z 5.1.13 otrzymujemy:

$$f(d_j, s_b^{(i+1)}) - f(d_j, s_b^{(i)}) \leq \beta \cdot nv_j^{(i+1)} + \delta \cdot \text{sim}(\mathbf{v}^{d^{(k+1)}}, \mathbf{v}^{d_j}) \quad (5.1.14)$$

Największe wartości, jakie może osiągnąć współczynnik wzorców nawigacyjnych oraz podobieństwo między dwoma dokumentami są równe jednności. Podstawiając do 5.1.14 otrzymujemy: $nv_j^{(i+1)} = 1$ oraz $\text{sim}(\mathbf{v}^{d^{(k+1)}}, \mathbf{v}^{d_j}) = 1$

$$f(d_j, s_b^{(i+1)}) - f(d_j, s_b^{(i)}) \leq \beta + \delta \quad (5.1.15)$$

□

Kolejne twierdzenie ma charakter pomocniczy w stosunku do twierdzeń z rozdziału 5.3 i dotyczy wartości funkcji rekomendacji dla dokumentu odwiedzonego w jednym z poprzednich kroków bieżącej sesji.

Twierdzenie 5.1.5

Wartość funkcji rekomendacji w kroku i bieżącej sesji s_b dla dokumentu d_j ($j=1, 2, \dots, n$) odwiedzonego w kroku i' ($i' < i$) jest równa:

$$f(d_j, s_b^{(i)}) = \beta \cdot \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i'}) \right) + \delta \cdot \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - \rho_s^{l-i'}) \right) - \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i'}) \right)$$

dla $j=1, 2, \dots, n$

Dowód:

Ogólny wzór na wartość funkcji rekomendacji w kroku i bieżącej sesji s_b dla dokumentu d_j , na mocy twierdzenia 5.1.1 jest równy:

$$f(d_j, s_b^{(i)}) = \alpha^i \cdot \left(0,5 \cdot (\beta + \delta) \cdot q(d_j) \cdot \prod_{k=1}^i (1 - z_j^{(k)}) \right) + \beta \cdot \sum_{k=1}^i \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) + \delta \cdot \sum_{k=1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) - \sum_{k=1}^i \left(\alpha^{i-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \quad (5.1.16)$$

Jest prawdą, że:

$$\begin{aligned} \sum_{k=1}^i \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) &= \sum_{k=1}^{i'} \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^{i'} (1 - z_j^{(l)}) \right) + \\ &+ \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \end{aligned} \quad (5.1.17)$$

analogicznie:

$$\begin{aligned} \sum_{k=1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) &= \sum_{k=1}^{i'} \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^{i'} (1 - z_j^{(l)}) \right) + \\ &+ \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \end{aligned} \quad (5.1.18)$$

oraz:

$$\begin{aligned} \sum_{k=1}^i \left(\alpha^{i-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) &= \sum_{k=1}^{i'} \left(\alpha^{i-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^{i'} (1 - z_j^{(l)}) \right) + \\ &+ \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \end{aligned} \quad (5.1.19)$$

Biorąc pod uwagę fakt, jeśli dokument d_j był odwiedzony w kroku i' to $z_j^{(i')}=1$ co w konsekwencji daje:

$$\begin{aligned} \sum_{k=1}^{i'} \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^{i'} (1 - z_j^{(l)}) \right) &= 0 \quad \text{oraz} \quad \sum_{k=1}^{i'} \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^{i'} (1 - z_j^{(l)}) \right) = 0 \quad \text{oraz} \\ \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) &= 0 \quad \text{oraz} \quad \alpha^i \cdot \left(0,5 \cdot (\beta + \delta) \cdot q(d_j) \cdot \prod_{k=1}^i (1 - z_j^{(k)}) \right) = 0 \end{aligned}$$

Po uwzględnieniu tego faktu i podstawieniu 5.1.17, 5.1.18 oraz 5.1.19 do 5.1.16 otrzymujemy:

$$\begin{aligned} f(d_j, s_b^{(i)}) &= \beta \cdot \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) + \\ &+ \delta \cdot \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) - \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \end{aligned} \quad (5.1.20)$$

Jeśli dokument d_j był odwiedzony tylko kroku i' to $\forall_{i>i'} z_j^{(i)} = \rho_s^{i-i'}$ co pozwala zapisać:

$$\begin{aligned} f(d_j, s_b^{(i)}) &= \beta \cdot \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i'}) \right) + \\ &+ \delta \cdot \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - \rho_s^{l-i'}) \right) - \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i'}) \right) \end{aligned} \quad (5.1.21)$$

□

Ostatnie twierdzenie, podobnie do twierdzenia 5.1.5, ma charakter pomocniczy w stosunku do twierdzeń z rozdziału 5.3 i dotyczy wartości funkcji rekomendacji dla dokumentu zignorowanego w jednym z poprzednich kroków bieżącej sesji.

Twierdzenie 5.1.6

Jeśli dokument d_j ($j=1, \dots, n$) w bieżącej sesji s_b był zarekomendowany jedynie w kroku i' oraz został zignorowany przez użytkownika, to wartość funkcji rekomendacji dla dokumentu d w kroku i ($i > i'$) wyrazi się wzorem:

$$f(d_j, s_b^{(i)}) = \alpha^i \cdot (0,5 \cdot (\beta + \delta) \cdot q(d_j)) + \beta \cdot \sum_{k=1}^i (\alpha^{i-k} \cdot nv_j^{(k)}) + \delta \cdot \sum_{k=1}^i (\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j})) - \gamma^{i-i'-1}$$

Dowód:

Z treści twierdzenia wiadomo, że dokument d_j nie był odwiedzany w bieżącej sesji, co oznacza, że wartość funkcji rekomendacji w kroku i dla tego dokumentu jest równa:

$$f(d_j, s_b^{(i)}) = \alpha^i \cdot (0,5 \cdot (\beta + \delta) \cdot q(d_j)) + \beta \cdot \sum_{k=1}^i (\alpha^{i-k} \cdot nv_j^{(k)}) + \delta \cdot \sum_{k=1}^i (\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j})) - \sum_{k=1}^i (\alpha^{i-k} \cdot g_j^{(k)}) \quad (5.1.22)$$

Dokument był rekomendowany w kroku i' (wcześniej nie był rekomendowany), więc:

$$\forall_{i \geq i' > 0} g_j^{(i)} = 0$$

Tak więc można zapisać:

$$f(d_j, s_b^{(i)}) = \alpha^i \cdot (0,5 \cdot (\beta + \delta) \cdot q(d_j)) + \beta \cdot \sum_{k=1}^i (\alpha^{i-k} \cdot nv_j^{(k)}) + \delta \cdot \sum_{k=1}^i (\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j})) - \sum_{k=i'+1}^i (\alpha^{i-k} \cdot g_j^{(k)}) \quad (5.1.23)$$

Do udowodnienia twierdzenia wystarczy pokazać, że $\sum_{k=i'+1}^i (\alpha^{i-k} \cdot g_j^{(k)}) = \gamma^{i-i'-1}$

Wprowadźmy zmienną pomocniczą $i'' = i - i'$, tak więc do udowodnienia twierdzenia należy pokazać, że:

$$\sum_{k=i'+1}^{i''+i'} (\alpha^{i''+i'-k} \cdot g_j^{(k)}) = \gamma^{i''-1} \quad (5.1.24)$$

Dodatkowo wprowadźmy oznaczenie $X(i'') = \sum_{k=i'+1}^{i''+i'} (\alpha^{i''+i'-k} \cdot g_j^{(k)})$. Do udowodnienia 5.1.24 zastosowana zostanie zasada indukcji matematycznej. Dla $i'' = 1$ otrzymujemy:

$$X(1) = \sum_{k=i'+1}^{i'+1} (\alpha^{i''+i'-k} \cdot g_j^{(k)}) = \alpha^0 \cdot g_j^{(i'+1)} = g_j^{(i'+1)} \quad (5.1.25)$$

Gdy $i''=1$ to $i'-i=1$ czyli dokument był zarekomendowany w poprzednim kroku. Z formuły 4.5.2 wiadomo, że jeśli w kroku $i'+1$ zignorowany został dokument zaproponowany w kroku i' to $g_j^{(i'+1)} = 1$, czyli:

$$X(1) = 1 = \gamma^0 \quad (5.1.26)$$

Tak więc 5.1.24 jest prawdziwe dla $i''=1$. Wziąwszy pod uwagę, że zgodnie z formułą 4.5.2 wartości $g_j^{(i'+1)}$ oraz $g_j^{(i'+2)}$ są określoną odrębnym wyrażeniem podstawmy także $i''=2$:

$$X(2) = \sum_{k=i'+1}^{i'+2} (\alpha^{2+i'-k} \cdot g_j^{(k)}) = \alpha \cdot g_j^{(i'+1)} + g_j^{(i'+2)} \quad (5.1.27)$$

Zgodnie z formułą 4.5.2 $g_j^{(i'+1)} = 1$ oraz $g_j^{(i'+2)} = \gamma - \alpha$, co pozwala otrzymać:

$$X(2) = \alpha + \gamma - \alpha = \gamma \quad (5.1.28)$$

Tak więc 5.1.24 jest także prawdziwe dla $i''=2$. Następnie pokażemy, że jeśli 5.1.24 jest prawdziwe dla $i''>1$ to jest także prawdziwe dla $i''+1$ (powyżej pokazano, że gdy $i''=1$ to 5.1.24 jest także prawdziwe dla $i''=2$).

$$X(i''+1) = \sum_{k=i'+1}^{i''+1+i'} (\alpha^{i''+1+i'-k} \cdot g_j^{(k)}) \quad (5.1.29)$$

Z pod znaku sumy z formuły 5.1.29 wyłączony zostanie ostatnie składnik:

$$X(i''+1) = \alpha^{i''+1+i'-i''-1-i'} \cdot g_j^{(i''+1+i')} + \sum_{k=i'+1}^{i''+i'} (\alpha^{i''+1+i'-k} \cdot g_j^{(k)}) \quad (5.1.30)$$

Łatwo zauważyć, że $\sum_{k=i'+1}^{i''+i'} (\alpha^{i''+1+i'-k} \cdot g_j^{(k)}) = \alpha \sum_{k=i'+1}^{i''+i'} (\alpha^{i''+i'-k} \cdot g_j^{(k)})$. Dodatkowo z formuły 4.5.2 wiadomo, że dla $i''-i>1$ $g_j^{(i''+1+i')} = \gamma \cdot g_j^{(i''+i')}$:

$$X(i''+1) = \gamma \cdot g_j^{(i)} + \alpha \sum_{k=i'+1}^{i''+i'} (\alpha^{i''+i'-k} \cdot g_j^{(k)}) \quad (5.1.31)$$

Z założenia indukcyjnego wiadomo, że: $\sum_{k=i'+1}^{i''+i'} (\alpha^{i''+i'-k} \cdot g_j^{(k)}) = \gamma^{i''-1}$

$$X(i''+1) = \gamma \cdot g_j^{(i)} + \alpha \cdot \gamma^{i''-1} \quad (5.1.32)$$

Z formuły 4.5.2 wiadomo, że dla $i-i' > 2$:

$$g_j^{(i)} = (\gamma - \alpha)\gamma^{j-i-2} \quad (5.1.33)$$

Biorąc pod uwagę, że $i-i' = i''$ i podstawiając 5.1.33 do 5.1.32 otrzymujemy:

$$X(i''+1) = \gamma \cdot (\gamma - \alpha) \cdot \gamma^{i-i'-2} + \alpha \cdot \gamma^{i''-1} = \gamma^{i-i'} - \alpha\gamma^{i''-1} + \alpha \cdot \gamma^{i''-1} = \gamma^{i-i'} \quad (5.1.34)$$

Tak więc jeśli 5.1.24 jest prawdziwe dla $X(i''+1)$ jeśli jest prawdziwe dla $X(i'')$, więc można wnioskować, że 5.1.24 jest prawdziwe, co dowodzi prawdziwości całego twierdzenia. □

5.2. Ocena własności adaptacyjnych metody rekomendacji

Problem rekomendacji sformalizowany w rozdziale 4 zakłada, że użytkownik w kolejnych krokach bieżącej sesji może zmienić swoje zainteresowania, co przekłada się na zmianę zawartości zbioru dokumentów relewantnych. System rekomendacyjny powinien dostosowywać się do nowych potrzeb informacyjnych użytkownika nawet wtedy, gdy nowy zbiór dokumentów relewantnych jest podzbiorem zbioru dokumentów nierelevantnych z poprzedniego kroku ($D_R^{(i)} \subseteq D \setminus D_R^{(i-1)}$).

Poniżej przedstawione zostaną własności związane ze zdolnościami adaptacyjnymi metody *AdaptRank*. Pierwsze twierdzenie dotyczy kresu górnego zbioru wartości funkcji rekomendacji f . Twierdzenie to jest istotne dla pokazania, że metoda *AdaptRank* jest w stanie dostosować się do zmieniających się potrzeb użytkownika. Gdyby wartości funkcji f w kolejnych krokach rosły nieograniczenie, zgodnie ze wzorem podanym w definicji 4.5.3 znaczenie zachowań użytkownika w długich sesjach byłoby coraz mniejsze, co stoi w sprzeczności z intuicją, zakładającą, że dokumenty odwiedzone ostatnio powinny mieć największy wpływ na wyniki rekomendacji.

Dla lepszego zilustrowania tego problemu założymy, że po oglądnięciu i dokumentów użytkownik przestał się interesować dotychczasowo pogłębianym tematem, którego jednym z reprezentantów był nieodwiedzony i niezignorowany jeszcze dokument d_j i zajął się innym zagadnieniem, którego typowym przedstawicielem jest dokument d_k , który nie był w bieżącej sesji ani zignorowany, ani odwiedzony. Jeśli zbiór wartości funkcji f rósłby nieograniczenie, to dla odpowiednio dużego i można by oczekiwać $f(d_j, s_b^{(i)}) > 20000$. Założymy jednocześnie, że wartość rekomendacji dla dokumentu d_k w kroku i przyjmuje wartość minimalną $f(d_k, s_b^{(i)}) = 0$ (patrz wniosek 5.1.4). W tym przypadku można oszacować minimalną liczbę kroków, po których wartość funkcji rekomendacji dla dokumentu d_k będzie większa od funkcji d_j .

Aby oszacować tę wartość założymy, że w każdym kroku większym od i wartość funkcji rekomendacji dla dokumentu d_k przyrasta maksymalnie¹ (patrz twierdzenie 5.1.5), a jednocześnie wartość współczynnika wzorców nawigacyjnych dla dokument d_j , a także jego podobieństwo do kolejnych dokumentów odwiedzanych po kroku i są równe 0 (są to w obydwu przypadkach wartości minimalne), wtedy aby w pewnym kroku i' prawdą było: $f(d_k, s_b^{(i')}) > f(d_j, s_b^{(i)})$ musiałyby zostać spełnione:

$$(i'-i) \cdot (\beta + \delta) > \alpha^{(i'-i)} \cdot 100000 + 0 + 0$$

Przy założeniu, że $\beta + \delta = 1$ oraz $\alpha = 0,8$ powyższa nierówność jest prawdziwa dla $i'-i > 35$ co oznacza, że system rekomendujący potrzebowałby 36 kroków do przystosowania się do nowych potrzeb użytkownika. Tak duża liczba potrzebnych kroków przeczyłaby zdolności adaptacji algorytmu i miałaby negatywny wpływ na efektywność systemu rekomendującego.

¹ W rzeczywistości nie jest możliwe, aby we wszystkich krokach wartość funkcji przyrastała maksymalnie. Podany przykład jest sporym przejaskrawieniem mającym na celu pokazanie, co by było, gdyby zbiór wartości funkcji f nie miał kresu górnego.

Twierdzenie 5.2.1

Zbiór wartości funkcji f posiada kres górny równy:

$$\sup(f(d_j, s_b^{(i)})) = \frac{\beta + \delta}{1 - \alpha}$$

Dowód:

Wziąwszy pod uwagę to, że $\forall_{i>0} \left(0 \leq \prod_{l=k}^i (1 - z_j^{(l)}) \right)$ mamy:

$$\forall_{i>0} \left(\sum_{k=1}^i \left(\alpha^{i-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \geq 0 \right).$$

Co biorąc pod uwagę twierdzenia 5.1.1 pozwala stwierdzić, że następująca nierówność jest prawdziwa:

$$\begin{aligned} f(d_j, s_b^{(i)}) \leq & \alpha^i \cdot \left(0,5 \cdot (\beta + \delta) \cdot q(d_j) \cdot \prod_{k=1}^i (1 - z_j^{(k)}) \right) + \beta \cdot \sum_{k=1}^i \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) + \\ & + \delta \cdot \sum_{k=1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \end{aligned} \quad (5.2.1)$$

Następnie, wziąwszy pod uwagę, że:

$$\forall_{i>0} \left(0 \leq \prod_{l=k}^i (1 - z_j^{(l)}) \leq 1 \right)$$

prawdziwa jest nierówność:

$$\begin{aligned} f(d_j, s_b^{(i)}) \leq & \alpha^i \cdot \left(0,5 \cdot (\beta + \delta) \cdot q(d_j) \cdot \prod_{k=1}^i (1 - z_j^{(k)}) \right) + \beta \cdot \sum_{k=1}^i \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) + \\ & + \delta \cdot \sum_{k=1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \\ \leq & \alpha^i \cdot (0,5 \cdot (\beta + \delta) \cdot q(d_j)) + \beta \cdot \sum_{k=1}^i (\alpha^{i-k} \cdot nv_j^{(k)}) + \delta \cdot \sum_{k=1}^i (\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j})) \end{aligned} \quad (5.2.2)$$

Nierówności 5.2.1 oraz 5.2.2 pozwalają stwierdzić, że nieodwiedzone i niezignorowane dokumenty mają większe wartości funkcji rekomendacji od dokumentów odwiedzonych bądź zignorowanych przy takim samym podobieństwie do odwiedzanych dokumentów i takich samych współczynnikach wzorców nawigacyjnych.

Oznacza to, że maksymalną wartość funkcji rekomendacji f jest osiągnięta przez nieodwiedzone i niezignorowane dokumenty, co dodatkowo na mocy wniosku 5.1.1 pozwala zapisać:

$$\sup(f(d_j, s_b^{(i)})) = \sup \left(\alpha^i \cdot 0,5 \cdot (\beta + \delta) + \beta \cdot \sum_{k=1}^i \alpha^{i-k} + \delta \cdot \sum_{k=1}^i \alpha^{i-k} \right) \quad (5.2.3)$$

Łatwo zaważyć, że:

$$\sum_{k=1}^i \alpha^{i-k} = \sum_{k=0}^{i-1} \alpha^k$$

co pozwala przekształcić wyrażenie 5.2.3 do następującej postaci:

$$\sup(f(d_j, s_b^{(i)})) = \sup\left(\alpha^i \cdot 0,5 \cdot (\beta + \delta) + \beta \cdot \sum_{k=0}^{i-1} \alpha^k + \delta \cdot \sum_{k=0}^{i-1} \alpha^k\right) \quad (5.2.4)$$

Po prawej stronie wyrażenia otrzymano ciąg liczbowy zależny od i . Zbadajmy kiedy ten ciąg jest rosnący poprzez rozwiązanie następującej nierówności:

$$0,5 \cdot (\beta + \delta) \cdot \alpha^{i+1} + \beta \sum_{k=0}^i \alpha^k + \delta \sum_{k=0}^i \alpha^k - \left(0,5 \cdot (\beta + \delta) \cdot \alpha^i + \beta \sum_{k=0}^{i-1} \alpha^k + \delta \sum_{k=0}^{i-1} \alpha^k\right) > 0 \quad (5.2.5)$$

Po wyłączeniu $\sum_{k=0}^i \alpha^k$ oraz $\sum_{k=0}^{i-1} \alpha^k$ przed odpowiednie nawiasy otrzymujemy:

$$0,5 \cdot (\beta + \delta) \cdot \alpha^{i+1} + (\beta + \delta) \cdot \sum_{k=0}^i \alpha^k - \left(0,5 \cdot (\beta + \delta) \cdot \alpha^i + (\beta + \delta) \cdot \sum_{k=0}^{i-1} \alpha^k\right) > 0 \quad (5.2.6)$$

Pogrupowanie wyrażen pozwala zapisać:

$$0,5 \cdot (\beta + \delta) \cdot \alpha^{i+1} - 0,5 \cdot (\beta + \delta) \cdot \alpha^i + (\beta + \delta) \cdot \sum_{k=0}^i \alpha^k - (\beta + \delta) \cdot \sum_{k=0}^{i-1} \alpha^k > 0 \quad (5.2.7)$$

Wyłączając $\beta + \delta$ oraz $0,5 \cdot (\beta + \delta) \cdot \alpha^i$ przed nawias otrzymujemy:

$$0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot (\alpha - 1) + (\beta + \delta) \cdot \left(\sum_{k=0}^i \alpha^k - \sum_{k=0}^{i-1} \alpha^k\right) > 0 \quad (5.2.8)$$

Korzystając z $\sum_{k=0}^i \alpha^k - \sum_{k=0}^{i-1} \alpha^k = \alpha^i$ możemy napisać:

$$0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot (\alpha - 1) + (\beta + \delta) \cdot \alpha^i > 0 \quad (5.2.9)$$

Po wyłączeniu $(\beta + \delta) \cdot \alpha^i$ przed nawias otrzymujemy:

$$((\beta + \delta) \cdot \alpha^i) \cdot (0,5 \cdot (\alpha - 1) + 1) > 0 \quad (5.2.10)$$

Jako, że $(\beta + \delta) \cdot \alpha^i > 0$ dla dowolnego $i \in \mathbb{N}$, $\alpha \in (0, 1)$, $\beta \in (0, 1)$, $\delta \in (0, 1)$; cały iloczyn będzie dodatni gdy:

$$0,5 \cdot (\alpha - 1) + 1 > 0 \quad (5.2.11)$$

Co ostatecznie daje:

$$\alpha > -1 \quad (5.2.12)$$

Tak więc dla $\alpha \in (0, 1)$ ciąg z wyrażenia 5.2.4 jest zawsze rosnący.

Do wyznaczenia kresu górnego wystarczy, zatem policzyć następującą granicę:

$$\sup(f(d_j, s_b^{(i)})) = \lim_{i \rightarrow \infty} \left(\alpha^i \cdot 0,5 \cdot (\beta + \delta) + \beta \cdot \sum_{k=0}^{i-1} \alpha^k + \delta \cdot \sum_{k=0}^{i-1} \alpha^k \right) \quad (5.2.13)$$

Granica sumy jest równa sumie granic, więc możemy napisać:

$$\sup(f(d_j, s_b^{(i)})) = \lim_{i \rightarrow \infty} (\alpha^i \cdot 0,5 \cdot (\beta + \delta)) + \lim_{i \rightarrow \infty} \left(\beta \cdot \sum_{k=0}^{i-1} \alpha^k \right) + \lim_{i \rightarrow \infty} \left(\delta \cdot \sum_{k=0}^{i-1} \alpha^k \right) \quad (5.2.14)$$

Korzystając z faktu $\lim_{i \rightarrow \infty} c \cdot a_i = c \cdot \lim_{i \rightarrow \infty} a_i$ gdzie c to stała a a_i to ciąg liczbowy można zapisać:

$$\sup(f(d_j, s_b^{(i)})) = 0,5 \cdot (\beta + \delta) \cdot \lim_{i \rightarrow \infty} (\alpha^i) + \beta \cdot \lim_{i \rightarrow \infty} \left(\sum_{k=0}^{i-1} \alpha^k \right) + \delta \cdot \lim_{i \rightarrow \infty} \left(\sum_{k=0}^{i-1} \alpha^k \right) \quad (5.2.15)$$

Po policzeniu granic cząstkowych $\lim_{i \rightarrow \infty} (\alpha^i) = 0$ oraz $\lim_{i \rightarrow \infty} \left(\sum_{k=0}^{i-1} \alpha^k \right) = \frac{1}{1 - \alpha}$ otrzymujemy:

$$\sup(f(d_j, s_b^{(i)})) = \frac{\beta + \delta}{1 - \alpha} \quad (5.2.16)$$

□

Warto dodać, że dla wartości parametrów proponowanych w tabeli 4.5.1 kres górny zbioru wartości funkcji rekomendacji wynosi $\frac{8}{3}$.

Kolejne twierdzenie dotyczy kresu dolnego zbioru wartości funkcji rekomendacji f . Twierdzenie to jest również bardzo istotne dla pokazania własności adaptacyjnych metody. Brak kresu dolnego oznaczałoby, że wartości funkcji f dla pewnego zignorowanego dokumentu d miałyby nieograniczenie (z racji na ujemne sprzężenie zwrotne relewancji). Jeśli po zmianie zainteresowań dokument d stałby się relewantnym to hipotetyczny brak kresu dolnego przyczyniłby się do znacznego wydłużenia liczby kroków potrzebnych do zarekomendowania dokumentu d a w konsekwencji zdolności adaptacyjne metody *AdaptRank* byłyby podważone.

Twierdzenie 5.2.2

Zbiór wartości funkcji f posiada kres dolny równy:

$$\inf(f(d_j, s_b^{(i)})) = -\frac{1}{1 - \alpha}$$

Dowód:

Korzystając z faktu, że minimalne wartości współczynnika wzorców nawigacyjnych, funkcji przydatności dokumentu oraz podobieństwa między dwoma dokumentami są równe 0, można stwierdzić, że:

$$\sum_{k=1}^i \left(\alpha^{i-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \leq f(d_j, s_b^{(i)}) \quad (5.2.17)$$

A w konsekwencji wzięwszy pod uwagę, że:

$$\forall_{i>0} \left(0 \leq \prod_{l=k}^i (1 - z_j^{(l)}) \leq 1 \right)$$

można stwierdzić prawdziwość nierówności:

$$-\sum_{k=1}^i (\alpha^{i-k} \cdot g_j^{(k)}) \leq -\sum_{k=1}^i \left(\alpha^{i-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \leq f(d_j, s_b^{(i)}) \quad (5.2.18)$$

Biorąc pod uwagę, że $\forall_{i>0} (g_j^{(i)} \leq 1)$ można zapisać:

$$-\sum_{k=1}^i (\alpha^{i-k}) \leq -\sum_{k=1}^i (\alpha^{i-k} \cdot g_j^{(k)}) \leq f(d_j, s_b^{(i)}) \quad (5.2.19)$$

Wyrażenie 5.2.19 oznacza, że funkcja rekomendacji f osiąga najniższe wartości dla nieodwiedzonych i zignorowanych dokumentów, których przydatność, podobieństwo do dokumentów odwiedzonych oraz wartość współczynnika wzorców nawigacyjnych jest równa 0. Łatwo zaważyć, że:

$$\sum_{k=1}^i \alpha^{i-k} = \sum_{k=0}^{i-1} \alpha^k$$

co pozwala przekształcić wyrażenie 5.2.19 do następującej postaci:

$$-\sum_{k=0}^{i-1} (\alpha^k) \leq f(d_j, s_b^{(i)}) \quad (5.2.20)$$

Biorąc pod uwagę fakt, że dla $\alpha \in (0, 1)$ szereg: $-\sum_{k=1}^{i-1} (\alpha^k)$ jest malejący i dodatkowo:

$$\lim_{i \rightarrow \infty} \left(-\sum_{k=0}^{i-1} (\alpha^k) \right) = -\frac{1}{1-\alpha}$$

można stwierdzić:

$$\inf(f(d_j, s_b^{(i)})) = -\frac{1}{1-\alpha} \quad (5.2.21)$$

□

Warto dodać, że dla wartości parametrów proponowanych w tabeli 4.5.1 kres dolny zbioru wartości funkcji rekomendacji wynosi $-\frac{4}{3}$.

Należy zauważyć, że osiągnięcie kresu dolnego przez jakiś dokument jest mało prawdopodobne w praktyce, gdyż warunkiem koniecznym, aby wartość funkcji dla dokumentu d_j osiągnęła kres dolny jest rekomendowanie dokumentu d_j we wszystkich krokach bieżącej sesji i jego ciągle ignorowanie przez użytkownika.

Zanim przedstawione zostanie następane twierdzenie ilustrujące zdolności adaptacyjne metody *AdaptRank* wprowadzone zostanie pojęcie dokumentu maksymalnie relewantnego od kroku i .

Definicja 5.2.3

Dokumentem maksymalnie relewantnym od kroku i ($i=1, 2, \dots, n_{sb}-1$) w ramach bieżącej sesji nazywamy dokument $d_{\max}^{(i)}$, który począwszy od kroku $i+1$ jest najbardziej podobny do wszystkich odwiedzanych dokumentów oraz osiąga najwyższe wartości współczynnika wzorców nawigacyjnych (oznaczone są one jako $nv_{\max}^{(i)}$), tj.:

$$\forall_{n_{sp} \geq i > i} \left(\forall_{d_j \in D \setminus \{d_{\max}^{(i)}\}} \left(nv_{\max}^{(i)} \geq nv_j^{(i)} \wedge \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_{\max}^{(i)}}) \geq \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j}) \right) \right)$$

Dokument maksymalnie relewantny od kroku i jest to więc dokument, który od kroku $i+1$ odpowiada najlepiej potrzebom użytkownika przy założeniu, że dokumenty relewantne cechują się większym podobieństwem do dokumentów odwiedzonych niż dokumenty nirelewantne oraz że dokumenty relewantne osiągają wyższe wartości współczynnika wzorców nawigacyjnych.

Twierdzenia 5.2.4 – 5.2.8 pozwalają zmierzyć zdolności adaptacyjne metody *AdaptRank* za pomocą liczby oznaczającej maksymalną liczbę kroków potrzebnych do dostosowania się do nowych potrzeb użytkownika.

Twierdzenie 5.2.4 pokazuje, że dokument maksymalnie relewantny od kroku i zostanie zarekomendowany w skończonej liczbie kroków, nawet jeśli przed krokiem i dokument ten był zupełnie nirelewantny do potrzeb użytkownika.

Twierdzenie 5.2.4

Jeśli dokument maksymalnie relewantny od kroku i $d_{\max}^{(i)}$ nie był odwiedzony ani zignorowany w bieżącej sesji s_b , to zostanie on zarekomendowany, po co najwyżej $step$ krokach, gdzie $step$ to najmniejsza liczba całkowita spełniająca nierówność:

$$step > \log_{\alpha} \left(\frac{c}{c + 1 - 0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right)$$

a c to największa dodatnia liczba rzeczywista spełniająca warunek:

$$\forall_{i>i} \left(\forall_{d_j \in D \setminus \{d_{\max}^{(i)}\}} \left(nv_{\max}^{(i)} - nv_j^{(i)} \geq c \wedge sim(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_{\max}^{(i)}}) - sim(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j}) \geq c \right) \right)$$

Dowód:

Aby dokument $d_{\max}^{(i)}$ został zarekomendowany w kroku i' , niezależnie od wartości parametru m określającego ile dokumentów jest rekomendowanych w jednym kroku, musi być spełniona następująca nierówność:

$$\forall_{d_j \in D \setminus \{d_{\max}^{(i)}\}} \left(f(d_{\max}^{(i)}, s_b^{(i)}) > f(d_j, s_b^{(i)}) \right) \quad (5.2.22)$$

a liczba kroków potrzebnych do zarekomendowania dokumentu $d_{\max}^{(i)}$ wyrazi się wzorem:

$$step = i' - i \quad (5.2.23)$$

Zauważmy, że wartość funkcji f w kroku i zależy od jej wartości w kroku $i-1$, tak więc liczba kroków potrzebna do zarekomendowania dokumentu $d_{\max}^{(i)}$ będzie największa, gdy w kroku i wartość funkcji f dla dokumentu $d_{\max}^{(i)}$ będzie minimalna i jednocześnie będzie istnieć m dokumentów dla których wartość funkcji f w kroku i będzie maksymalna.

Z wniosku 5.1.1 wiadomo, że minimalna wartość funkcji f dla niezignorowanego i nieodwiedzanego dokumentu jest równa, 0 czyli:

$$f(d_{\max}^{(i)}, s_b^{(i)}) = 0 \quad (5.2.24)$$

Korzystając z twierdzenia 5.1.1 i przy założeniu, że do kroku i wartość funkcji rekomendacji dla dokumentu $d_{\max}^{(i)}$ była równa 0 można określić minimalną wartość funkcji f dla dokumentu $d_{\max}^{(i)}$ w kroku i' :

$$f(d_{\max}^{(i)}, s_b^{(i)}) = 0 + \beta \cdot \sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot nv_{\max}^{(k)}) + \delta \cdot \sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot sim(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_{\max}^{(i)}})) \quad (5.2.25)$$

Z drugiej strony z twierdzenia 5.1.3 wiadomo, że maksymalna wartość funkcji f w kroku i dla dokumentu d_j nie będącego maksymalnie relewantnym jest równa:

$$f(d_j, s_b^{(i)}) = 0,5 \cdot (\beta + \delta) \cdot \alpha^i + \beta \cdot \sum_{k=1}^i (\alpha^{i-k}) + \delta \cdot \sum_{k=1}^i (\alpha^{i-k}) \quad (5.2.26)$$

Zważywszy na fakt, że od kroku $i+1$ współczynnik wzorców nawigacyjnych dokumentu d_j oraz jego podobieństwo do dokumentów odwiedzonych muszą spełniać założenia twierdzenia, maksymalna wartość funkcji rekomendacji f dla dokumentu d_j w kroku i wyniesie:

$$\begin{aligned} f(d_j, s_b^{(i)}) = & 0,5 \cdot (\beta + \delta) \cdot \alpha^i + \beta \cdot \sum_{k=1}^i (\alpha^{i-k}) + \delta \cdot \sum_{k=1}^i (\alpha^{i-k}) + \\ & \beta \cdot \sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot nv_j^{(k)}) + \delta \cdot \sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot sim(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j})) \end{aligned} \quad (5.2.27)$$

Podstawiając pod 5.2.22 wartości funkcji f w kroku i' podane w 5.2.25 oraz 5.2.27 można zapisać:

$$\begin{aligned} \beta \cdot \sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot nv_{\max}^{(k)}) + \delta \cdot \sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot sim(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_{\max}^{(i)}})) > 0,5 \cdot (\beta + \delta) \cdot \alpha^i + \beta \cdot \sum_{k=1}^i (\alpha^{i-k}) + \\ + \delta \cdot \sum_{k=1}^i (\alpha^{i-k}) + \beta \cdot \sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot nv_j^{(k)}) + \delta \cdot \sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot sim(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j})) \end{aligned} \quad (5.2.28)$$

Po odjęciu od obu stron wyrażenia znajdującego się po prawej stronie nierówności i pogrupowaniu otrzymujemy:

$$\begin{aligned} \beta \cdot \sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot nv_{\max}^{(k)}) - \beta \cdot \sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot nv_j^{(k)}) + \delta \cdot \sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot sim(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_{\max}^{(i)}})) \\ - \delta \cdot \sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot sim(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j})) - 0,5 \cdot (\beta + \delta) \cdot \alpha^i - \beta \cdot \sum_{k=1}^i (\alpha^{i-k}) - \delta \cdot \sum_{k=1}^i (\alpha^{i-k}) > 0 \end{aligned} \quad (5.2.29)$$

Wyłączając β z dwóch pierwszych wyrażen oraz δ z dwóch następnych możemy zapisać:

$$\begin{aligned} \beta \cdot \left(\sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot nv_{\max}^{(k)}) - \sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot nv_j^{(k)}) \right) + \\ + \delta \cdot \left(\sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot sim(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_{\max}^{(i)}})) - \sum_{k=i+1}^{i'} (\alpha^{i-k} \cdot sim(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j})) \right) - 0,5 \cdot (\beta + \delta) \cdot \alpha^i \\ - \beta \cdot \sum_{k=1}^i (\alpha^{i-k}) - \delta \cdot \sum_{k=1}^i (\alpha^{i-k}) > 0 \end{aligned} \quad (5.2.30)$$

Korzystając z faktu: $\sum_{k=i+1}^{i'} a_k - \sum_{k=i+1}^{i'} b_k = \sum_{k=i+1}^{i'} (a_k - b_k)$ oraz wyłączając $\sum_{k=1}^i (\alpha^{i-k})$ z dwóch ostatnich wyrażen możemy zapisać:

$$\begin{aligned}
& \beta \cdot \left(\sum_{k=i+1}^i (\alpha^{i-k} \cdot nv_{\max}^{(k)} - \alpha^{i-k} \cdot nv_j^{(k)}) \right) + \tag{5.2.31} \\
& + \delta \cdot \left(\sum_{k=i+1}^i (\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_{\max}^{(i)}}) - \alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j})) \right) - \\
& - 0,5 \cdot (\beta + \delta) \cdot \alpha^i - (\beta + \delta) \cdot \sum_{k=1}^i (\alpha^{i-k}) > 0
\end{aligned}$$

Wyłączając α^{i-k} z dwóch pierwszych różnic otrzymujemy:

$$\begin{aligned}
& \beta \cdot \left(\sum_{k=i+1}^i (\alpha^{i-k} \cdot (nv_{\max}^{(k)} - nv_j^{(k)})) \right) + \delta \cdot \left(\sum_{k=i+1}^i (\alpha^{i-k} \cdot (\text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_{\max}^{(i)}}) - \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j})) \right) - \tag{5.2.32} \\
& - 0,5 \cdot (\beta + \delta) \cdot \alpha^i - (\beta + \delta) \cdot \sum_{k=1}^i (\alpha^{i-k}) > 0
\end{aligned}$$

Z założenia twierdzenia wiadomo, że

$$\forall_{i>i} \left(\bigvee_{d_j \in D \setminus \{d_{\max}^{(i)}\}} \left(nv_{\max}^{(i)} - nv_j^{(i)} \geq c \wedge \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_{\max}^{(i)}}) - \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j}) \geq c \right) \right) \tag{5.2.33}$$

Dlatego korzystając z przechodniości relacji większości możemy stwierdzić, że 5.2.32 będzie prawdziwe gdy prawdziwe będzie:

$$\beta \cdot \left(\sum_{k=i+1}^i (\alpha^{i-k} \cdot c) \right) + \delta \cdot \left(\sum_{k=i+1}^i (\alpha^{i-k} \cdot c) \right) - 0,5 \cdot (\beta + \delta) \cdot \alpha^i - (\beta + \delta) \cdot \sum_{k=1}^i (\alpha^{i-k}) > 0 \tag{5.2.34}$$

Po wyłączeniu $\sum_{k=i+1}^i (\alpha^{i-k} \cdot c)$ z dwóch pierwszych wyrażeń otrzymujemy:

$$(\beta + \delta) \cdot \sum_{k=i+1}^i (\alpha^{i-k} \cdot c) - 0,5 \cdot (\beta + \delta) \cdot \alpha^i - (\beta + \delta) \cdot \sum_{k=1}^i (\alpha^{i-k}) > 0 \tag{5.2.35}$$

Z definicji 4.7.1 wiadomo, że $\beta + \delta > 0$ więc po podzieleniu obu stron nierówności przez $\beta + \delta$ można zapisać:

$$\sum_{k=i+1}^i (\alpha^{i-k} \cdot c) - 0,5 \cdot \alpha^i - \sum_{k=1}^i (\alpha^{i-k}) > 0 \tag{5.2.36}$$

Korzystając z faktu $\sum_{k=i+1}^i (\alpha^{i-k} \cdot c) = c \cdot \sum_{k=i+1}^i \alpha^{i-k}$ otrzymujemy:

$$c \cdot \sum_{k=i+1}^i (\alpha^{i-k}) - 0,5 \cdot \alpha^i - \sum_{k=1}^i (\alpha^{i-k}) > 0 \tag{5.2.37}$$

$\sum_{k=i+1}^i \alpha^{i-k}$ to suma $(i'-i)$ wyrazów ciągu geometrycznego o pierwszym wyrazie równym 1 i stałym ilorazie równym α , zaś $\sum_{k=1}^i \alpha^{i-k}$ to suma i wyrazów ciągu arytmetycznego o pierwszym wyrazie równym α^{i-i} i o stałym ilorazie równym α ($0 < \alpha < 1$). Korzystając z tego faktu mamy:

$$c \cdot \frac{1 - \alpha^{i-i}}{1 - \alpha} - \alpha^{i-i} \cdot \frac{1 - \alpha^i}{1 - \alpha} - 0,5 \cdot \alpha^i > 0 \quad (5.2.38)$$

Z nierówności 5.2.38 wyznaczona zostanie wartość i' , dla większej przejrzystości obliczeń wprowadzona zostanie zmienna pomocnicza $x = \alpha^{i'}$ ($1 > x > 0$):

$$c \cdot \frac{1 - \frac{x}{\alpha^i}}{1 - \alpha} - \frac{x}{\alpha^i} \cdot \frac{1 - \alpha^i}{1 - \alpha} - 0,5 \cdot x > 0 \quad (5.2.39)$$

Mnożąc obie strony przez zawsze dodatnie wyrażenie $1 - \alpha$ otrzymujemy:

$$c - \frac{c \cdot x}{\alpha^i} - \frac{x \cdot (1 - \alpha^i)}{\alpha^i} - 0,5 \cdot x \cdot (1 - \alpha) > 0 \quad (5.2.40)$$

Odejmując od obu stron c oraz wyłączając x przed nawias można zapisać:

$$x \cdot \left(\frac{-c - 1 + \alpha^i}{\alpha^i} - 0,5 \cdot (1 - \alpha) \right) > -c \quad (5.2.41)$$

Po sprowadzeniu wyrażenia w nawiasie do wspólnego mianownika otrzymujemy:

$$x \cdot \left(\frac{-c - 1 + \alpha^i - 0,5 \cdot \alpha^i + 0,5 \cdot \alpha^{i+1}}{\alpha^i} \right) > -c \quad (5.2.42)$$

Grupując wyrażenia podobne i mnożąc obie strony nierówności przez -1 otrzymujemy:

$$x \cdot \left(\frac{c + 1 - 0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}}{\alpha^i} \right) < c \quad (5.2.43)$$

Jako, że ($0 < \alpha < 1$) to $\frac{c + 1 - 0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}}{\alpha^i} > 0$ (dla $i=1, \dots, n_{sb}$), tak więc mnożąc obie strony

nierówności 5.2.43 przez $\frac{\alpha^i}{c + 1 - 0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}}$ otrzymujemy:

$$x < \frac{c \cdot \alpha^i}{c + 1 - 0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \quad (5.2.44)$$

Podstawiając za $x = \alpha^{i'}$ mamy:

$$\alpha^{i'} < \frac{c \cdot \alpha^i}{c + 1 - 0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \quad (5.2.45)$$

Wyrażenie z prawej strony nierówności jest zawsze dodatnie dla $i = 1, \dots, n_{s_b}$ oraz $(0 < \alpha < 1)$ dodatkowo operacji logarytmowania nierówności dla podstawy logarytmu mniejszej od jedności obliuguje do zmiany znaku nierówności, co daje:

$$i' > \log_{\alpha} \left(\frac{c \cdot \alpha^i}{c + 1 - 0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right) \quad (5.2.45)$$

Nierówność 5.2.45 określa numer kroku (liczony od początku bieżącej sesji), do którego dokument $d_{\max}^{(i)}$ maksymalnie relewantny od kroku i na pewno zostanie zarekomendowany. Odejmując numer kroku, po którym użytkownik zmienił swoje zainteresowania można otrzymać maksymalną liczbę kroków potrzebnych do zarekomendowania dokument $d_{\max}^{(i)}$:

$$i' - i > \log_{\alpha} \left(\frac{c \cdot \alpha^i}{c + 1 - 0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right) - i \quad (5.2.46)$$

Korzystając z faktu, że dla $a > 0 \wedge a \neq 1 \wedge b > 0 \wedge c > 0$: $\log_a(b \cdot c) = \log_a b + \log_a c$ oraz z 5.2.23 otrzymujemy:

$$step > \log_{\alpha} \left(\frac{c}{c + 1 - 0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right) + \log_{\alpha}(\alpha^i) - i \quad (5.2.47)$$

Co biorąc pod uwagę, że $\log_{\alpha}(\alpha^i) = i$, ostatecznie daje:

$$step > \log_{\alpha} \left(\frac{c}{c + 1 - 0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right) \quad (5.2.48)$$

Co było do udowodnienia. □

Wyrażenie określające liczbę kroków potrzebnych do zarekomendowania dokumentu $d_{\max}^{(i)}$ maksymalnie relewantnego od kroku i warte jest dodatkowej dyskusji. Należy zauważyć, że maksymalna liczba kroków z twierdzenia 5.2.4 nie zależy od wartości parametrów β oraz δ , które to parametry regulują odpowiednio wpływ współczynników wzorców nawigacyjnych i podobieństwa do odwiedzonych dokumentów na wartość funkcji rekomendacji f .

Warto także podkreślić, że dość skomplikowana postać formuły określającej maksymalną liczbę kroków potrzebnych do zarekomendowania dokumentu maksymalnie relewantnego wymaga dodatkowej analizy. Najprostsza forma analizy oparta jest na podaniu kilku wartości $step$. Przykładowo, dla $\alpha = 0,5$, $c = 0,4$ oraz $i = 10$, $step \approx 1,81$ co oznacza, że w dwóch następujących krokach dokument maksymalnie relewantny zostanie na pewno podpowiedziany.

Obliczanie wartości wyrażenia $step$ dla różnych kombinacji c , α oraz i , choć pomaga zrozumieć sens twierdzenia, nie jest najlepszą metodą do badania zdolności adaptacyjnych metody. Dlatego też reszta tego podrozdziału poświęcona zostanie na przedstawienie twierdzeń, które pozwolą lepiej scharakteryzować wyżej wymienione zdolności.

Kolejne 3 twierdzenia pokazują jak dobór wartości parametru α wpływa na maksymalną liczbę kroków potrzebnych do zarekomendowania dokumentu maksymalnie relewantnego od kroku i . Zależność liczby kroków od parametru α może zostać wyrażona następującą funkcją:

$$step(\alpha) = \log_{\alpha} \left(\frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right)$$

gdzie: $\alpha \in (0,1)$, $c \in (0,1)$ oraz $i > 1$.

Twierdzenie 5.2.5

Funkcja $step(\alpha) = \log_{\alpha} \left(\frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right)$ gdzie $\alpha \in (0, 1)$ jest rosnącą w całej dziedzinie, gdzie i to numer kroku, od którego dokument $d_{\max}^{(i)}$ jest maksymalnie relewantny, a wartość stałej c została określona w ramach twierdzenia 5.2.4.

Dowód:

Najpierw pokażemy, że następująca funkcja jest rosnąca w dziedzinie $\alpha \in (0, 1)$:

$$g(\alpha) = \frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \quad (5.2.49)$$

Funkcja $g(\alpha)$ jest różniczkowalna do określenia jej monotoniczności obliczona zostanie jej pierwsza pochodna:

$$g'(\alpha) = \left(\frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right)' \quad (5.2.50)$$

Co daje dalej:

$$g'(\alpha) = -c \cdot \frac{1}{(c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1})^2} \cdot (c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1})' \quad (5.2.51)$$

i ostatecznie:

$$g'(\alpha) = c \cdot \frac{0,5 \cdot i \cdot \alpha^{i-1} + 0,5 \cdot (i+1) \cdot \alpha^i}{(c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1})^2} \quad (5.2.52)$$

Łatwo zauważyć, że dla $\alpha \in (0, 1)$, $i > 0$ oraz $c > 0$ jest spełniona następująca nierówność $0,5 \cdot i \cdot \alpha^{i-1} + 0,5 \cdot (i+1) \cdot \alpha^i > 0$, więc:

$$g'(\alpha) > 0 \quad (5.2.53)$$

Na podstawie 5.2.53 możemy twierdzić, że funkcja $g(\alpha)$ jest rosnąca w przedziale $\alpha \in (0, 1)$:

Następnie pokazane zostanie, że:

$$0 < \frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} < 1 \quad (5.2.54)$$

Łatwo zauważyć, że dla $\alpha \in (0, 1)$ oraz $c \in (0, 1]$ prawdą jest $\frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} > 0$

Mnożąc obie strony przez wartość $c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}$ dodatnią dla $\alpha \in (0, 1)$ oraz $c \in (0, 1]$ otrzymujemy:

$$c < c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1} \quad (5.2.55)$$

Dodanie do obu stron $(0,5 \cdot \alpha^i + 0,5 \cdot \alpha^{i+1} - c)$ pozwoli przekształcić 5.2.55 do następującej postaci:

$$0,5 \cdot \alpha^i + 0,5 \cdot \alpha^{i+1} < 1 \quad (5.2.56)$$

Łatwo zauważyć, że dla $\alpha \in (0, 1)$

$$0,5 \cdot \alpha^i + 0,5 \cdot \alpha^{i+1} \leq 0,5 \cdot \alpha + 0,5 \cdot \alpha < 1 \quad (5.2.57)$$

Więc dla $\alpha \in (0, 1)$ nierówność 5.2.54 jest prawdziwa.

Następnie pokażemy, że funkcja: $h(\alpha) = \log_{\alpha} b$ jest funkcją rosnącą dla $\alpha \in (0, 1)$ oraz $b \in (0, 1)$. Korzystając z faktu, że dla $a > 0 \wedge a \neq 1 \wedge x > 0 \wedge x \neq 1 \wedge y > 0$: $\log_x y = \frac{\log_a y}{\log_a x}$ zapiszemy funkcję $h(\alpha)$ w następujący sposób:

$$h(\alpha) = \frac{\ln b}{\ln \alpha} \quad (5.2.58)$$

Funkcja $h(\alpha)$ jest różniczkowalna, więc możemy policzyć jej pierwszą pochodną:

$$h'(\alpha) = \ln b \cdot \left(\frac{1}{\ln \alpha} \right)' \quad (5.2.59)$$

Co korzystając z wzoru na pochodną funkcji złożonej daje:

$$h'(\alpha) = -\ln b \cdot \frac{1}{(\ln \alpha)^2} \cdot \frac{1}{\alpha} \quad (5.2.60)$$

Aby funkcja $h(\alpha)$ była rosnąca musi być spełniony następujący warunek:

$$h'(\alpha) > 0 \quad (5.2.61)$$

Biorąc pod uwagę fakt, że $\ln b < 0$ dla $b \in (0, 1)$ a w konsekwencji $-\ln b > 0$ można stwierdzić, że:

$$-\ln b \cdot \frac{1}{(\ln \alpha)^2} \cdot \frac{1}{\alpha} > 0 \quad (5.2.62)$$

Aby udowodnić, że funkcja $step(\alpha)$ jest rosnąca pokażemy, że dla $1 > \alpha' > \alpha > 0$:

$$step(\alpha') - step(\alpha) > 0 \quad (5.2.63)$$

Co po podstawieniu $step(\alpha) = \log_{\alpha} \left(\frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right)$ daje:

$$\log_{\alpha'} \left(\frac{c}{c+1-0,5 \cdot \alpha'^i - 0,5 \cdot \alpha'^{i+1}} \right) - \log_{\alpha} \left(\frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right) > 0 \quad (5.2.64)$$

Na początku tego dowodu pokazano, że dla $\alpha \in (0, 1)$ funkcja

$g(\alpha) = \frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}}$ jest rosnąca, co sprawia, że prawdziwa jest nierówność:

$$\begin{aligned} \log_{\alpha'} \left(\frac{c}{c+1-0,5 \cdot \alpha'^i - 0,5 \cdot \alpha'^{i+1}} \right) - \log_{\alpha} \left(\frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right) > \\ \log_{\alpha'} \left(\frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right) - \log_{\alpha} \left(\frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right) \end{aligned} \quad (5.2.65)$$

Korzystając z przechodności relacji większości do udowodnienia 5.2.64 wystarczy pokazać, że dla $1 > \alpha' > \alpha > 0$:

$$\log_{\alpha'} \left(\frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right) - \log_{\alpha} \left(\frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right) > 0 \quad (5.2.66)$$

Podstawiając $b = \frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}}$ otrzymujemy:

$$\log_{\alpha'}(b) - \log_{\alpha}(b) > 0 \quad (5.2.67)$$

Powyżej pokazano, że $1 > \frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} > 0$, a także, że funkcja $h(\alpha) = \log_{\alpha} b$ jest funkcją rosnącą dla $\alpha \in (0, 1)$ oraz $b \in (0, 1)$ wskutek czego z własności funkcji rosnącej można twierdzić, że dla $1 > \alpha' > \alpha > 0$ nierówność 5.2.66 jest prawdziwa, a w konsekwencji prawdziwa jest też nierówność 5.2.63 i całe twierdzenie. □

Twierdzenie 5.2.5 jest zgodne z intuicją, gdyż tak jak zostało to określone w podrozdziale 4.5 parametr α określa jaki wpływ na obecną wartość funkcji f ma wartość funkcji f z poprzedniego kroku. Im wartość parametru α jest większa tym „trudniej” jest metodzie zaadoptować się do nowych potrzeb użytkownika.

Omawiając zdolności adaptacyjne metody w zależności od parametru α warto przeanalizować zachowanie się funkcji określającej maksymalną liczbę kroków, gdy parametr α będzie miał wartość zbliżoną do minimalnej (twierdzenie 5.2.6) lub zbliżoną do maksymalnej twierdzenie 5.2.7

Twierdzenie 5.2.6

Funkcja $step(\alpha) = \log_{\alpha} \left(\frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right)$ posiada w punkcie $\alpha=0$ granicę prawostronną równą 0: $\lim_{\alpha \rightarrow 0^+} step(\alpha) = 0$.

Dowód:

Korzystając z faktu, że dla $a>0 \wedge a \neq 1 \wedge x>0 \wedge x \neq 1 \wedge y>0$: $\log_x y = \frac{\log_a y}{\log_a x}$ otrzymujemy:

$$\lim_{\alpha \rightarrow 0^+} step(\alpha) = \frac{\ln \left(\frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right)}{\ln \alpha} \quad (5.2.68)$$

Gdy $\alpha=0$ wartość licznika jest określona możemy więc napisać:

$$\lim_{\alpha \rightarrow 0^+} step(\alpha) = \frac{\ln \left(\frac{c}{c+1} \right)}{\ln \alpha} \quad (5.2.69)$$

Licznik wyrażenia nie zależy od α , mianownik zaś jest rozbieżny do nieskończoności, wskutek czego możemy postulować prawdziwość twierdzenia 5.2.6

□

W przypadku bardzo małych wartości parametru α wartość funkcji $step(\alpha)$ zmierza do zera, co jest zgodne jest z intuicją: jeśli dokumenty odwiedzone w poprzednich krokach wpływają w bardzo małym stopniu na wartość funkcji rekomendacji w bieżącym kroku to zmiana charakteru odwiedzonych dokumentów przekłada się automatycznie na zmianę rekomendacji. Należy zaznaczyć, że przypisanie parametrowi α wartości bliskiej zera może spowodować drastyczny spadek dokładności rekomendacji, gdyż użytkownik szukając interesujących go treści może przesłać żądanie dotyczące zupełnie nierelevantnego dokumentu.

Kolejne twierdzenie pokazuje jak zmieniają się zdolności adaptacyjne metody, gdy parametr α jest bliski jedności. Tak jak zostanie to udowodnione maksymalna liczba kroków dąży w tym przypadku do pewnej stałej liczby, której wartość uzależniona jest od liczby dokumentów odwiedzonych w bieżącej sesji oraz stałej c .

Twierdzenie 5.2.7

Funkcja $step(\alpha) = \log_{\alpha} \left(\frac{c}{c+1-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right)$ posiada w punkcie $\alpha=1$ granicę lewostronną równą $\frac{i+0,5}{c}$, tj.: $\lim_{\alpha \rightarrow 1^-} step(\alpha) = \frac{i+0,5}{c}$

Dowód:

Korzystając z faktu, że dla $a > 0 \wedge a \neq 1 \wedge x > 0 \wedge x \neq 1 \wedge y > 0$: $\log_x y = \frac{\log_a y}{\log_a x}$ otrzymujemy:

$$\lim_{\alpha \rightarrow 1^-} \text{step}(\alpha) = \frac{\ln\left(\frac{c}{1+c-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}}\right)}{\ln \alpha} \quad (5.2.70)$$

Granica mianownika jest równa: $\lim_{\alpha \rightarrow 1^-} \ln(1) = 0$, co w połączeniu z granicą licznika:

$$\lim_{\alpha \rightarrow 1^-} \ln\left(\frac{c}{1+c-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}}\right) = \lim_{\alpha \rightarrow 1^-} \ln(1) = 0 \quad (5.2.71)$$

pozwala zastosować twierdzenie l'Hospitala o granicy ilorazu funkcji, których wartości granic są równe 0 :

$$\lim_{\alpha \rightarrow 1^-} \text{step}(\alpha) = \lim_{\alpha \rightarrow 1^-} \frac{\left(\ln\left(\frac{c}{1+c-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}}\right)\right)'}{(\ln \alpha)'} \quad (5.2.72)$$

Pochodna licznika jest równa:

$$\left(\ln\left(\frac{c}{1+c-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}}\right)\right)' = (\ln c)' - \left(\ln(1+c-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1})\right)' \quad (5.2.73)$$

Co daje dalej:

$$\begin{aligned} \left(\ln\left(\frac{c}{1+c-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}}\right)\right)' &= -\frac{-0,5 \cdot i \cdot \alpha^{i-1} - 0,5 \cdot (i+1) \cdot \alpha^i}{1+c-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} = \\ &= \frac{0,5 \cdot i \cdot \alpha^{i-1} + 0,5 \cdot (i+1) \cdot \alpha^i}{1+c-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \end{aligned} \quad (5.2.74)$$

Podstawiając pod 5.2.72 pochodną obliczoną w 5.2.74 i licząc pochodną mianownika otrzymujemy:

$$\lim_{\alpha \rightarrow 1^-} k(\alpha) = \lim_{\alpha \rightarrow 1^-} \frac{0,5 \cdot i \cdot \alpha^{i-1} + 0,5 \cdot (i+1) \cdot \alpha^i}{1+c-0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \cdot \frac{1}{\alpha} \quad (5.2.75)$$

Co ostatecznie pozwala zapisać:

$$\lim_{\alpha \rightarrow 1^-} k(\alpha) = \frac{0,5 \cdot i + 0,5 \cdot (i+1)}{1 + c - 0,5 - 0,5} = \frac{i + 0,5}{c} \quad (5.2.76)$$

□

Twierdzenie 5.2.7 pokazuje, że gdy dokumenty odwiedzone w poprzednich krokach mają bardzo duży wpływ na wartość funkcji rekomendacji (α zmierza do jedności), wówczas największa liczba kroków potrzebna do zarekomendowania dokumentu maksymalnie relewantnego jest „prawie” proporcjonalna do liczby odwiedzonych dokumentów (wartość granicy określona w 5.2.76 jest proporcjonalna do i). Dodatkowo, wartość ta jest odwrotnie proporcjonalna do stałej c określającej na ile dokument maksymalnie relewantny uzyskuje wyższe podobieństwo do dokumentów odwiedzanych oraz na ile jego współczynnik wzorców nawigacyjnych jest wyższy od współczynnika wzorców nawigacyjnych innych dokumentów.

Metody rekomendacji, w których funkcja rekomendacji zależy od dokumentów odwiedzonych w przeszłości cechują się mniejszą zdolnością adaptacji dla sesji, w których odwiedzona została duża liczba dokumentów. Jest to po części zrozumiałe, gdyż jeśli użytkownik wysłał już znaczną liczbę żądań przesłania dokumentu, to można przypuszczać, że jego preferencje są znane systemowi rekomendującemu. Z drugiej jednak strony bardzo duża liczba kroków potrzebnych do zarekomendowania dokumentu maksymalnie relewantnego w długich sesjach świadczyłaby o słabych zdolnościach adaptacyjnych metody. Do badania jak zmienia się liczby kroków potrzebnych do zarekomendowania dokumentu maksymalnie relewantnego w zależności od wzrostu liczby dokumentów odwiedzonych w danej sesji przed zmianą zainteresowań będziemy używać następującego ciągu:

$$step_i = \log_{\alpha} \left(\frac{c}{c + 1 - 0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right)$$

Twierdzenie 5.2.8 pokazuje, że jeśli krok i , w którym nastąpiła zmiana zainteresowań użytkownika jest odpowiednio duży, to dalszy wzrost wartości i nie będzie przekładał się na wzrost maksymalnej liczby kroków potrzebnych do zarekomendowania dokumentu maksymalnie relewantnego od kroku i . To znaczy, że maksymalna liczba kroków potrzebnych do zarekomendowania dokumentu maksymalnie relewantnego będzie taka sama niezależnie czy użytkownik zmienił swoje zainteresowanie po 30, 100 czy 1000 krokach.

Twierdzenie 5.2.8

Jeśli dany jest ciąg: $step_i = \log_{\alpha} \left(\frac{c}{c + 1 - 0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right)$

gdzie $i > 0$ oraz $i \in \mathbb{N}$ to: $\exists \forall \forall_{i' > i} |step_{i''} - step_i| < 1$

Dowód:

Na początku policzymy granicę ciągu $step_i$:

$$\lim_{i \rightarrow \infty} step_i = \lim_{i \rightarrow \infty} \log_{\alpha} \left(\frac{c}{c+1 - 0,5 \cdot \alpha^i - 0,5 \cdot \alpha^{i+1}} \right) = \log_{\alpha} \left(\frac{c}{c+1} \right) \quad (5.2.77)$$

Z definicji granicy ciągu oraz z 5.2.77 wiadomo, że:

$$\lim_{i \rightarrow \infty} step_i = \frac{c}{c+1} \Leftrightarrow \forall_{\varepsilon > 0} \left(\exists_{l > 0} \forall_{i > l} \left(\left| step_i - \frac{c}{c+1} \right| < \varepsilon \right) \right) \quad (5.2.78)$$

Na mocy 5.2.78 przypisując pod $\varepsilon=0,5$ można stwierdzić:

$$\exists_{l > 0} \forall_{i > l} \left(\left| step_i - \frac{c}{c+1} \right| < 0,5 \right) \quad (5.2.79)$$

Co pozwala napisać:

$$\exists_{l > 0} \forall_{i > l} \forall_{i' > i} \left(\left| step_{i'} - \frac{c}{c+1} \right| < 0,5 \wedge \left| step_i - \frac{c}{c+1} \right| < 0,5 \right) \quad (5.2.80)$$

$\left| step_{i'} - \frac{c}{c+1} \right| < 0,5 \wedge \left| step_i - \frac{c}{c+1} \right| < 0,5$ można przekształcić do układu nierówności:

$$\begin{cases} step_{i'} - \frac{c}{c+1} < 0,5 \\ step_{i'} - \frac{c}{c+1} > -0,5 \\ \frac{c}{c+1} - step_i < 0,5 \\ \frac{c}{c+1} - step_i > -0,5 \end{cases} \quad (5.2.81)$$

Dodając stronami nierówności: 1 z 3 oraz 2 z 4 otrzymujemy:

$$\begin{cases} step_{i'} - step_i < 1 \\ step_{i'} - step_i > -1 \end{cases} \quad (5.2.82)$$

Co jest równoważne:

$$\left| step_{i'} - step_i \right| < 1 \quad (5.2.83)$$

Podstawiając 5.2.83 do 5.2.80 otrzymujemy:

$$\bullet \exists_{l > 0} \forall_{i > l} \forall_{i' > i} \left(\left| step_{i'} - step_i \right| < 1 \right) \quad (5.2.84)$$

□

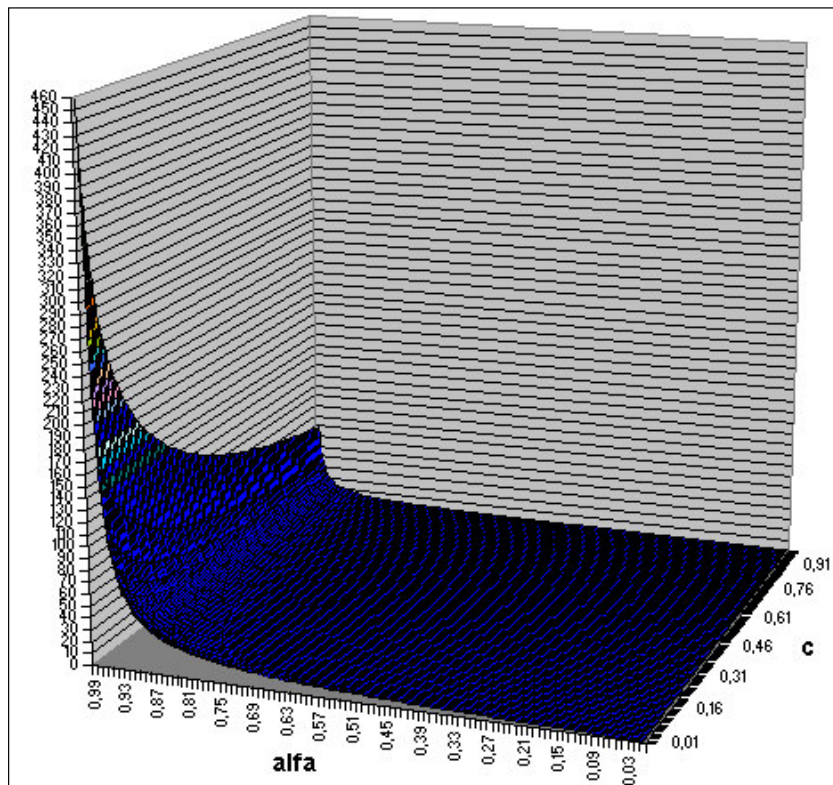
Twierdzenie 5.2.8 jest bardzo ważne, gdyż pokazuje, że dla dostatecznie długich bieżących sesji, dalszy wzrost długości sesji nie przekłada się na spadek zdolności adaptacyjnych metody.

Co więcej z formuły 5.2.77 wynika, że maksymalna liczba kroków dla długich sesji dąży do pewnej liczby zależnej od parametru α oraz od zmiennej c określającej na ile dokument maksymalnie relewantny jest bardziej podobny do odwiedzanych dokumentów oraz na ile dokument maksymalnie relewantny osiąga wyższe wartości współczynnika wzorców nawigacyjnych w stosunku do pozostałych dokumentów.

Niech $step_{i\max}(\alpha, c) = \log_{\alpha} \left(\frac{c}{c+1} \right)$ oznacza funkcję pozwalającą wyznaczyć wyżej wspomnianą liczbę. Na rysunku 5.2.1 zamieszczono wykres tej funkcji. Jak widać tylko dla dużych wartości parametru α oraz bardzo małych wartości c , funkcja ta przyjmuje bardzo duże wartości. Łatwo zauważyć, że dla $\alpha < 0,66$ oraz $c > 0,1$ maksymalna liczba kroków potrzebnych do zarekomendowania dokumentu maksymalnie relewantnego dla długich sesji nie przekroczy 6, gdyż:

$$\log_{0,66} \left(\frac{0,1}{0,1+1} \right) \approx 5,77 \quad (5.2.85)$$

Oznacza to, że dla powyższych wartości parametrów c oraz α liczba kroków potrzebnych do zarekomendowania dokumentu maksymalnie relewantnego od kroku i nigdy nie przekroczy 6, niezależnie od numeru kroku i , po którym użytkownik zmienił swoje zainteresowania.



Rysunek 5.2.1 Wykres funkcji $step_{i\max}(\alpha, c)$

5.3. Efektywność metody rekomendacji

W podrozdziale tym przedstawione są twierdzenia związane z efektywnością systemu rekomendującego używającego metody *AdaptRank*. W pierwszej części podrozdziału podane zostaną ogólne warunki, jakie musi spełniać dowolna funkcja rekomendacji, aby dokładność systemu rekomendującego odznaczała się wartością maksymalną. Dodatkowo określone będą sytuacje, w których metoda rekomendacji *AdaptRank* osiągnie maksymalną dokładność.

W drugiej części podrozdziału podane zostaną ogólne warunki, jakie musi spełniać dowolna funkcja rekomendacji, aby kompletność systemu rekomendującego osiągnęła wartość maksymalną. Dodatkowo określone będą założenia, przy których metoda rekomendacji *AdaptRank* odznacza się maksymalną kompletnością.

Trzecia część podrozdziału bazując na wynikach dwóch pierwszych, pozwala określić warunki zapewniające maksymalną efektywność metody *AdaptRank*. Jednocześnie zostanie pokazane, że wyżej wspomniane warunki nie są dostateczne, aby metody składowe osiągnęły maksymalną efektywność.

Każda z wyżej wymienionych części zawiera analizę wpływu wartości parametrów na dokładność, kompletność i efektywność metody.

5.3.1. Dokładność metody rekomendacji

Twierdzenie 5.3.1 określa ogólne warunki jakie musi spełniać dowolna funkcja rekomendacji, aby dokładność systemu rekomendującego w bieżącej sesji wyniosła wartość maksymalną. Twierdzenie to jest bardzo ważne, gdyż pokazuje, że do osiągnięcia maksymalnej dokładności nie jest wymagane, aby system rekomendujący określił cały zbiór dokumentów relewantnych (zbiór ten jest znany tylko użytkownikowi i wyznaczenie go za pomocą formalnej metody nie jest zadaniem łatwym). Do osiągnięcia maksymalnej dokładności wystarczy tylko, aby co najmniej m dokumentów relewantnych miało wartości funkcji rekomendacji większe od wartości tej funkcji dla dokumentów nirelewantnych.

Twierdzenie 5.3.1

System rekomendujący SR osiąga maksymalną dokładność w bieżącej sesji s_b , tj. $prec(SR, s_b) = 1$, jeśli w każdym kroku tej sesji istnieje co najmniej m dokumentów relewantnych, dla których wartości funkcji rekomendacji są większe od wartości tej funkcji dla dokumentów nirelewantnych, tj. gdy:

$$\forall_{n_{s_b} \geq i > 0} \exists D_R^{(i)} \subseteq D_R^{(i)} \forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D \setminus D_R^{(i)}} (f(d_r, s_b^{(i)}) > f(d_n, s_b^{(i)}) \wedge card(D_R^{(i)}) \geq m)$$

Dowód:

Zgodnie z definicją 4.1.5 system rekomendujący w bieżącej sesji s_b będzie odznaczał się maksymalną dokładnością, gdy:

$$\frac{\sum_{i=1}^{n_{sb}} \text{card}(R^{(i)} \cap D_R^{(i)})}{\sum_{i=1}^{n_{sb}} \text{card}(R^{(i)})} = 1 \quad (5.3.1)$$

W dowodzie twierdzenia wykorzystana zostanie zasada indukcji matematycznej ze względu na liczbę kroków w bieżącej sesji. Najpierw pokazane zostanie, że twierdzenie jest prawdziwe, gdy w bieżącej sesji s_b odwiedzony został tylko jeden dokument ($n_{sb}=1$). Gdy odwiedzony jest tylko jeden dokument wówczas dokładność rekomendacji jest równa:

$$\text{prec}(SR, s_b) = \frac{\text{card}(R^{(1)} \cap D_R^{(1)})}{\text{card}(R^{(1)})} \quad (5.3.2)$$

Wziąwszy pod uwagę, że w pojedynczym kroku rekomendowane jest m dokumentów można stwierdzić, że: $\text{card}(R^{(1)})=m$. Jednocześnie z założeń twierdzenia wiadomo, że istnieje co najmniej m dokumentów relewantnych, dla których wartość funkcji rekomendacji jest wyższa od wartości funkcji rekomendacji dla dokumentów nirelevantnych, co pozwala stwierdzić, że: $\text{card}(R^{(1)} \cap D_R^{(1)}) = m$ czyli:

$$\text{prec}(SR, s_b) = \frac{m}{m} = 1 \quad (5.3.3)$$

Udowodnione zostało, że twierdzenie 5.3.1 jest prawdziwe dla $n_{sb}=1$, następnie pokazane zostanie, że jeśli twierdzenie jest prawdziwe dla $n_{sb} = k$ to prawdziwe jest także dla $n_{sb} = k+1$.

Podstawiając pod wyrażenie z definicji 4.1.5 $n_{sb} = k + 1$ otrzymujemy:

$$\text{prec}(SR, s_b) = \frac{\sum_{i=1}^{k+1} \text{card}(R^{(i)} \cap D_R^{(i)})}{\sum_{i=1}^{k+1} \text{card}(R^{(i)})} \quad (5.3.4)$$

Wyłączając spod znaku sumy ostatni składnik otrzymujemy:

$$\text{prec}(SR, s_b) = \frac{\text{card}(R^{(k+1)} \cap D_R^{(k+1)}) + \sum_{i=1}^k \text{card}(R^{(i)} \cap D_R^{(i)})}{\text{card}(R^{(k+1)}) + \sum_{i=1}^k \text{card}(R^{(i)})} \quad (5.3.5)$$

W ostatnim kroku $k+1$ także jest zarekomendowane m dokumentów, więc: $\text{card}(R^{(k+1)}) = m$. Jednocześnie z założeń twierdzenia wiadomo, że w kroku $k+1$ istnieje m dokumentów relewantnych, dla których wartość funkcji rekomendacji jest wyższa od wartości

funkcji rekomendacji dla dokumentów nierelevantnych, co pozwala stwierdzić, że:
 $card(R^{(k+1)} \cap D_R^{(k+1)}) = m$ czyli:

$$prec(SR, s_b) = \frac{m + \sum_{i=1}^k card(R^{(i)} \cap D_R^{(i)})}{m + \sum_{i=1}^k card(R^{(i)})} \quad (5.3.6)$$

Z założenia indukcyjnego wiadomo, że: $\frac{\sum_{i=1}^k card(R^{(i)} \cap D_R^{(i)})}{\sum_{i=1}^k card(R^{(i)})} = 1$ więc:

$\sum_{i=1}^k card(R^{(i)} \cap D_R^{(i)}) = \sum_{i=1}^k card(R^{(i)})$ co pozwala stwierdzić, że twierdzenie jest prawdziwe dla $k + 1$, wynika stąd, że jest prawdziwe dla każdego $n_{sb} > 0$.

□

Kolejne twierdzenia będą wymagały zdefiniowania czterech dodatkowych wielkości.

Definicja 5.3.1

Minimalnym współczynnikiem wzorców nawigacyjnych nv_{rmin} dla dokumentów relevantnych nazywamy liczbę będącą najmniejszą wartością współczynnika wzorców nawigacyjnych osiągniętą przez dokumenty relevantne w bieżącej sesji s_b , tj.:

$$nv_{rmin} = \min \{nv_d^{(i)} : 0 < i \leq n_{sb} \text{ oraz } d \in D_R^{(i)}\}$$

Definicja 5.3.2

Maksymalnym współczynnikiem wzorców nawigacyjnych nv_{nmax} dla dokumentów nierelevantnych nazywamy liczbę będącą największą wartością współczynnika wzorców nawigacyjnych osiągniętą przez dokumenty nierelevantne w bieżącej sesji s_b , tj.:

$$nv_{nmax} = \max \{nv_d^{(i)} : 0 < i \leq n_{sb} \text{ oraz } d \in D \setminus D_R^{(i)}\}$$

Definicja 5.3.3

Minimalnym podobieństwem dokumentów relewantnych do dokumentów odwiedzonych w bieżącej sesji s_b nazywamy liczbę sim_{rmin} spełniającą następujący warunek:

$$sim_{rmin} = \min \{ sim(d, d^{(i)}) : 0 < i \leq n_{sb} \text{ oraz } d \in D_R^{(i)} \}$$

Definicja 5.3.4

Maksymalnym podobieństwem dokumentów nirelewantnych do dokumentów odwiedzonych w bieżącej sesji s_b nazywamy liczbę sim_{nmin} spełniającą następujący warunek:

$$sim_{nmin} = \max \{ sim(d, d^{(i)}) : 0 < i \leq n_{sb} \text{ oraz } d \in D \setminus D_R^{(i)} \}$$

Twierdzenie 5.3.2 pokazuje przy jakich warunkach dokładność metody rekomendacji *AdaptRank* będzie maksymalna (równa jedności).

Twierdzenie 5.3.2

Jeśli w każdym kroku i ($i=2, \dots, n_{sb}$) bieżącej sesji s_b , użytkownik odwiedza jeden z dokumentów rekomendowanych w kroku poprzednim i system hipertekstowy zawiera m_r ($m_r > m > 0$) dokumentów relewantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{sb})}$ oraz we wszystkich krokach bieżącej sesji spełnione są następujące warunki:

$$nv_{rmin} > \frac{0,5 \cdot \alpha}{1 - \rho_s^{[m_r/m]}} + \frac{nv_{nmax}}{(1 - \rho_s^{[m_r/m]}) \cdot (1 - \alpha)} \text{ oraz } sim_{rmin} > \frac{0,5 \cdot \alpha}{1 - \rho_s^{[m_r/m]}} + \frac{sim_{nmax}}{(1 - \rho_s^{[m_r/m]}) \cdot (1 - \alpha)},$$

to dokładność systemu rekomendującego, wykorzystującego metodę *AdaptRank*, w bieżącej sesji s_b będzie równa 1 niezależnie od liczby dokumentów odwiedzonych w tej sesji (gdzie $[m_r/m]$ to część całkowita ilorazu m_r/m).

Dowód:

Zacniemy od przekształcenia nierówności wyrażających warunki twierdzenia. Mnożąc obie strony pierwszej nierówności przez zawsze dodatnie $\beta \cdot (1 - \rho_s^{[m_r/m]})$ i drugiej przez zawsze dodatnie $\delta \cdot (1 - \rho_s^{[m_r/m]})$ otrzymujemy postacie równoważne:

$$\begin{aligned} (1 - \rho_s^{[m_r/m]}) \cdot \beta \cdot nv_{rmin} &> 0,5 \cdot \beta \cdot \alpha + \beta \cdot \frac{1}{(1 - \alpha)} nv_{nmax} \\ (1 - \rho_s^{[m_r/m]}) \cdot \delta \cdot sim_{rmin} &> 0,5 \cdot \delta \cdot \alpha + \delta \cdot \frac{1}{(1 - \alpha)} \cdot sim_{nmax} \end{aligned} \quad (5.3.7)$$

Dodając stronami nierówności z 5.3.7 otrzymujemy:

$$\begin{aligned} & (1 - \rho^{[m_r/m]}) \cdot \beta \cdot nv_{r_{\min}} + (1 - \rho^{[m_r/m]}) \cdot \delta \cdot sim_{r_{\min}} > \\ & 0,5 \cdot (\beta + \delta) \cdot \alpha + \left(\frac{1}{1 - \alpha} \right) \cdot (\beta \cdot nv_{n_{\max}} + \delta \cdot sim_{n_{\max}}) \end{aligned} \quad (5.3.8)$$

Z twierdzenia 5.3.1 wiadomo, że aby system rekomendujący osiągnął maksymalną dokładność wartości funkcji f dla dokumentów nierelevantnych musi być mniejsza niż dla m dokumentów relevantnych:

$$\forall_{d_n \in D^N, d_r \in D_R} f(d_r, s_b^{(i)}) > f(d_n, s_b^{(i)}) \quad (5.3.9)$$

W dowodzie zostanie pokazane, że zgodnie z założeniami twierdzenia 5.3.9 jest spełnione nawet gdy:

- Wszystkie dokumenty relevantne osiągają minimalne wartości funkcji rekomendacji, które są zgodne z założeniami twierdzenia.
- Wszystkie dokumenty nierelevantne osiągają maksymalne wartości funkcji rekomendacji, które są zgodne z założeniami twierdzenia.

Zgodnie z założeniami twierdzenia w najgorszym z przypadków wszystkie dokumenty nierelevantne będą uzyskiwały we wszystkich krokach bieżącej sesji wartości maksymalne, co nastąpi gdy:

$$\forall_{n_b \geq i > 0} \forall_{d_n \in D \setminus D_R^{(i)}} (q(d_n) = 1 \wedge sim(d_n, d^{(i)}) = sim_{n_{\max}} \wedge nv_{d_n}^{(i)} = nv_{n_{\max}}) \quad (5.3.10)$$

Tak więc wartość funkcji rekomendacji dla dokumentu nierelevantnego d_n spełniać będzie nierówność:

$$f(d_n, s_b^{(i)}) \leq 0,5 \cdot (\beta + \delta) \cdot \alpha^i + \beta \cdot nv_{n_{\max}} \cdot \sum_{k=1}^i (\alpha^{i-k}) + \delta \cdot sim_{n_{\max}} \cdot \sum_{k=1}^i (\alpha^{i-k}) \quad (5.3.11)$$

$\sum_{k=1}^i (\alpha^{i-k}) = \sum_{k=1}^i (\alpha^{k-1})$ to suma i wyrazów ciągu geometrycznego o pierwszym wyrazie równym 1 i stałym ilorazie równym α , więc:

$$f(d_n, s_b^{(i)}) \leq 0,5 \cdot (\beta + \delta) \cdot \alpha^i + \frac{1 - \alpha^i}{1 - \alpha} \cdot (\beta \cdot nv_{n_{\max}} + \delta \cdot sim_{n_{\max}}) \quad (5.3.12)$$

Łatwo zauważyć, że:

$$\begin{aligned}
0,5 \cdot (\beta + \delta) \cdot \alpha^i + \frac{1 - \alpha^i}{1 - \alpha} \cdot (\beta \cdot nv_{n_{\max}} + \delta \cdot sim_{n_{\max}}) &\leq & (5.3.13) \\
0,5 \cdot (\beta + \delta) \cdot \alpha + \frac{1}{1 - \alpha} \cdot (\beta \cdot nv_{n_{\max}} + \delta \cdot sim_{n_{\max}}) &
\end{aligned}$$

Tak więc, jeśli prawdziwe jest 5.3.12 to prawdziwe będzie także:

$$f(d_n, s_b^{(i)}) \leq 0,5 \cdot (\beta + \delta) \cdot \alpha + \frac{1}{1 - \alpha} \cdot (\beta \cdot nv_{n_{\max}} + \delta \cdot sim_{n_{\max}}) \quad (5.3.14)$$

Aby wyznaczyć minimalną wartość funkcji rekomendacji dla dokumentu relewantnego jest potrzebne wprowadzenie następującego lematu:

Lemat 1.

Gdy wszystkie dokumenty relewantne w danej sesji spełniają następujący warunek:

$$\forall_{n_{s_b} \geq i > 0} \forall_{d_r \in D_r^{(i)}} (q(d_r) = 0 \wedge sim(d_r, d^{(i)}) = sim_{r_{\min}} \wedge nv_{d_r}^{(i)} = nv_{r_{\min}}),$$

to maksymalna wartość funkcji rekomendacji dla dokumentu relewantnego $d_{r_{\max}}$ w kroku $i > [m_r/m]$ bieżącej sesji s_b będzie wynosić:

$$\begin{aligned}
f(d_{r_{\max}}, s_b^{(i)}) &= \beta \cdot \sum_{k=i+1-[m_r/m]}^i \left(\alpha^{i-k} \cdot nv_{r_{\min}} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i+[m_r/m]}) \right) + \\
&+ \delta \cdot \sum_{k=i+1-[m_r/m]}^i \left(\alpha^{i-k} \cdot sim_{r_{\min}} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i+[m_r/m]}) \right)
\end{aligned}$$

Dowód Lematu 1:

Po pierwsze, należy zauważyć, że po $[m_r/m]$ krokach wszystkie dokumenty relewantne zostaną zarekomendowane (równe współczynniki wzorców nawigacyjnych oraz podobieństwa do dokumentów odwiedzonych powodują, że dokumenty już zaproponowane w bieżącej sesji będą miały niższe wartości funkcji rekomendacji od dokumentów jeszcze nierekomendowanych). Dzięki równym współczynnikom w kroku $[m_r/m]+1$ najwyższe wartości funkcji rekomendacji będą posiadały dokumenty zarekomendowane w pierwszym kroku, gdyż zgodnie z twierdzeniem 5.1.5 na wartość funkcji rekomendacji dokumentu odwiedzonego w kroku i' mają wpływ współczynniki z kroków o numerze większym od i' . Podobnie w kroku $[m_r/m]+2$ najwyższe wartości funkcji rekomendacji będą miały dokumenty zarekomendowane w kroku 2. Tak więc uogólniając w kroku $i > [m_r/m]$ największymi wartościami funkcji rekomendacji będą odznaczały się dokumenty, które były rekomendowane w kroku $i - [m_r/m]$. Na mocy twierdzenia 5.1.5 maksymalna wartość funkcji rekomendacji w kroku $i > [m_r/m]$ wynosi:

$$\begin{aligned}
f(d_{r_{\max}}, s_b^{(i)}) &= \beta \cdot \sum_{k=i+1-\lfloor m_r/m \rfloor}^i \left(\alpha^{i-k} \cdot nv_{r_{\min}} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i+\lfloor m_r/m \rfloor}) \right) + \\
&+ \delta \cdot \sum_{k=i+1-\lfloor m_r/m \rfloor}^i \left(\alpha^{i-k} \cdot sim_{r_{\min}} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i+\lfloor m_r/m \rfloor}) \right)
\end{aligned} \tag{5.3.15}$$

Co dowodzi prawdziwości lematu 1.

Lemat 1 pozwolił określić najwyższą wartość funkcji rekomendacji dla dokumentów relevantnych zakładając, że dokumenty relevantne uzyskują minimalne wartości: współczynników wzorców nawigacyjnych oraz podobieństwa do dokumentów odwiedzanych przy spełnionych założeniach twierdzenia. Należy zauważyć, że w krokach mniejszych równych $\lfloor m_r/m \rfloor$ najwyższa wartość funkcji rekomendacji dla dokumentów relevantnych będzie większa od najwyższych wartości w krokach wyższych niż $\lfloor m_r/m \rfloor + 1$, dlatego w dowodzie twierdzenia wystarczy ograniczyć się do kroków $i > \lfloor m_r/m \rfloor$. Zgodnie z lematem 1 można, więc zapisać:

$$\begin{aligned}
f(d_r, s_b^{(i)}) &\geq \beta \cdot \sum_{k=i+1-\lfloor m_r/m \rfloor}^i \left(\alpha^{i-k} \cdot nv_{r_{\min}} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i+\lfloor m_r/m \rfloor}) \right) + \\
&+ \delta \cdot \sum_{k=i+1-\lfloor m_r/m \rfloor}^i \left(\alpha^{i-k} \cdot sim_{r_{\min}} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i+\lfloor m_r/m \rfloor}) \right)
\end{aligned} \tag{5.3.16}$$

Wyłączając ostatni składnik ($k=i$) przed znak każdej z sum otrzymujemy:

$$\begin{aligned}
f(d_r, s_b^{(i)}) &\geq (1 - \rho^{m_r/m}) \cdot \beta \cdot nv_{r_{\min}} + (1 - \rho^{m_r/m}) \cdot \delta \cdot sim_{r_{\min}} + \\
&\beta \cdot \sum_{k=i+1-\lfloor m_r/m \rfloor}^{i-1} \left(\alpha^{i-k} \cdot nv_{r_{\min}} \cdot \left(\prod_{l=k}^{i-1} (1 - \rho_s^{l-i+\lfloor m_r/m \rfloor}) \right) \right) + \\
&\delta \cdot \sum_{k=i+1-\lfloor m_r/m \rfloor}^{i-1} \left(sim_{r_{\min}} \cdot \alpha^{i-k} \cdot \left(\prod_{l=k}^{i-1} (1 - \rho_s^{l-i+\lfloor m_r/m \rfloor}) \right) \right)
\end{aligned} \tag{5.3.17}$$

Łatwo zauważyć, że:

$$\begin{aligned}
&(1 - \rho^{m_r/m}) \cdot \beta \cdot nv_{r_{\min}} + (1 - \rho^{m_r/m}) \cdot \delta \cdot sim_{r_{\min}} + \\
&\beta \cdot \sum_{k=i+1-\lfloor m_r/m \rfloor}^{i-1} \left(\alpha^{i-k} \cdot nv_{r_{\min}} \cdot \left(\prod_{l=k}^{i-1} (1 - \rho_s^{l-i+\lfloor m_r/m \rfloor}) \right) \right) + \\
&\delta \cdot \sum_{k=i+1-\lfloor m_r/m \rfloor}^{i-1} \left(sim_{r_{\min}} \cdot \alpha^{i-k} \cdot \left(\prod_{l=k}^{i-1} (1 - \rho_s^{l-i+\lfloor m_r/m \rfloor}) \right) \right) \geq \\
&(1 - \rho^{m_r/m}) \cdot \beta \cdot nv_{r_{\min}} + (1 - \rho^{m_r/m}) \cdot \delta \cdot sim_{r_{\min}}
\end{aligned} \tag{5.3.18}$$

Więc jeśli prawdziwa jest nierówność 5.3.17, to prawdziwa będzie także następująca nierówność

$$f(d_r, s_b^{(i)}) \geq (1 - \rho^{[m_r/m]}) \cdot \beta \cdot nv_{r_{\min}} + (1 - \rho^{[m_r/m]}) \cdot \delta \cdot sim_{r_{\min}} \quad (5.3.19)$$

Podstawiając 5.3.14 i 5.3.19 pod 5.3.9 możemy zapisać:

$$(1 - \rho^{[m_r/m]}) \cdot \beta \cdot nv_{r_{\min}} + (1 - \rho^{[m_r/m]}) \cdot \delta \cdot sim_{r_{\min}} > \quad (5.3.20)$$

$$0,5 \cdot (\beta + \delta) \cdot \alpha + \frac{1}{1 - \alpha} \cdot (\beta \cdot nv_{n_{\max}} + \delta \cdot sim_{n_{\max}})$$

Nierówność 5.3.20 jest równoważna postaci założeń twierdzenia przekształconych do formy zapisanej w 5.3.8, więc twierdzenie jest prawdziwe. □

Przyglądając się twierdzeniu 5.3.2 należy zauważyć, że metoda rekomendacji *AdaptRank* będzie odznaczać się maksymalną dokładnością, jeśli istnieje będzie wyraźna różnica między współczynnikami nawigacyjnymi dokumentów relewantnych i dokumentów nirelewantnych. To samo tyczy się podobieństwa do dokumentów odwiedzonych. W praktyce metoda *AdaptRank* często powinna odznaczać się maksymalną dokładnością, gdyż wyraźnie wyższe podobieństwo dokumentów relewantnych do odwiedzonych oraz uznanie za relewantne dokumentów należących do tego samego wzorca nawigacyjnego co dokumenty odwiedzane jest zgodne z intuicją.

Należy zauważyć, że parametry α , m oraz ρ_s , a także liczba dokumentów relewantnych m_r determinują, jaka musi być różnica między $nv_{r_{\min}}$ i $nv_{n_{\max}}$ oraz między $sim_{r_{\min}}$ i $sim_{n_{\max}}$, aby system rekomendujący mógł osiągnąć maksymalną dokładność. Przykładowo, dla wartości parametrów zaproponowanych w tabeli 4.5.1 ($\alpha=0,25$, $\beta=1$, $\delta=1$, $\gamma=0,9$, $\rho_s=0,75$, $m=2$) i przy założeniu, że istnieje $m_r=30$ dokumentów relewantnych do potrzeb użytkownika oraz że $nv_{r_{\max}}=0,1$ oraz $sim_{n_{\max}}=0,1$ do osiągnięcia maksymalnej dokładności wystarczy, aby minimalne podobieństwo dokumentów relewantnych do dokumentów odwiedzonych w bieżącej sesji oraz minimalna wartość współczynnika wzorców nawigacyjnych nie były mniejsze niż 0,54. Tak dobrane wartości parametrów nie wydają się być trudnymi do osiągnięcia w praktyce.

Analizując powyższe twierdzenie nasuwa się następujące pytanie: dlaczego warunek $nv_{r_{\min}} > nv_{n_{\max}} \wedge sim_{r_{\min}} > sim_{n_{\max}}$ nie jest wystarczający dla zapewnienia maksymalnej dokładności? Powód jest bardzo prosty: według definicji metody *AdaptRank* dokumenty odwiedzane przez użytkownika, w następnym kroku odnotowują spadek wartości funkcji f regulowany parametrem ρ_s i mogłoby się tak zdarzyć, że po odwiedzeniu wszystkich dokumentów relewantnych wartość funkcji rekomendacji dla pewnego dokumentu nirelewantnego byłaby największa.

Wprowadzenie parametru ρ_s jest związane z koniecznością uniknięcia ciągłego rekomendowania tych samych dokumentów (co w konsekwencji prowadziło do drastycznego spadku kompletności). Z drugiej strony, spełnienie nierówności przedstawionych w twierdzeniu 5.3.2 gwarantuje, że dokumenty nirelewantne nie będą podpowiadane nigdy – niezależnie od długości sesji (nawet jeśli wszystkie dokumenty relewantne zostały już odwiedzone w bieżącej sesji).

Twierdzenie 5.3.2 może okazać się bardzo przydatne w praktyce, gdyż szacując wartości, jakie powinny przybierać zmienne $nv_{n_{\max}}$, $sim_{n_{\max}}$, $nv_{r_{\min}}$, $sim_{r_{\min}}$ oraz m_r można tak dobrać parametry α , ρ_s , oraz m , aby dokumenty nirelewantne nigdy nie zostały zarekomendowane. Jeśli wartości określające minimalny współczynnik wzorców nawigacyjnych dla dokumentów

relewantnych oraz minimalne podobieństwo dokumentów relewantnych do dokumentów odwiedzanych wydają się zbyt wygórowane dla przeciętnej sesji użytkownika możliwe jest zmiana wartości parametrów. Przykładowo, aby w powyżej podanym przykładzie obniżyć nv_{rmin} oraz sim_{rmin} do 0,38, wystarczy dla parametru ρ_s przypisać wartość 0,6.

Oszacowanie wyżej wspomnianych wartości zmiennych nv_{nmax} , sim_{nmax} , nv_{rmin} , sim_{rmin} nie jest zadaniem łatwym. Założenie zbyt dużych wartości nv_{nmax} oraz sim_{nmax} może doprowadzić do sytuacji, w której tylko kilka dokumentów będzie stale podpowiadanych lub co gorsza żaden dokument relewantny nie spełni założeń twierdzenia (minimalna wartość współczynników nawigacyjnego i podobieństwa będzie bardzo wysoka). „Niedoszacowanie” tych wielkości niesie ze sobą niebezpieczeństwo podpowiadania dokumentów nierelewantnych (minimalna wartość współczynników nawigacyjnego i podobieństwa będzie bardzo niska).

Powracając do przykładu przedstawionego powyżej należy stwierdzić, że w bieżącej sesji s_b dokładność rekomendacji będzie maksymalna, jeśli rzeczywiście $nv_{nmax}=0,1$ oraz $sim_{nmax}=0,1$. Jeśli jednak wartości nv_{nmax} oraz sim_{nmax} będą wyższe, to mogą pojawić się dokumenty nierelewantne posiadające większe wartości współczynnika wzorców nawigacyjnych lub podobieństwa do dokumentów odwiedzanych, a w konsekwencji dokładność rekomendacji może się obniżyć.

Jeśli jednak nv_{nmax} oraz sim_{nmax} zostały oszacowane w sposób poprawny, to spadek wartości prawych stron nierówności z twierdzenia 5.3.2 będzie przyczyniać się do wzrostu dokładności systemu rekomendującego, gdyż maksymalna dokładność osiągnięta będzie nawet w bieżących sesjach, w których dokumenty relewantne nie będą się odznaczać bardzo wysokim podobieństwem do dokumentów odwiedzonych ani dużymi wartościami współczynników wzorców nawigacyjnych. Sformalizowaniem powyższych rozważań jest wniosek 5.3.1

Wniosek 5.3.1

Jeśli w każdym kroku i ($i=2, \dots, n_{sb}$) bieżącej sesji s_b , użytkownik odwiedza jeden z dokumentów rekomendowanych w kroku poprzednim i system hipertekstowy zawiera m_r ($m_r > m > 0$) dokumentów relewantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{sb})}$, to obniżenie wartości następujących wyrażeń:

$$\frac{0,5 \cdot \alpha}{1 - \rho_s^{[m_r, l/m]}} + \frac{nv_{nmax}}{(1 - \rho_s^{[m_r, l/m]}) \cdot (1 - \alpha)} \quad \text{oraz} \quad \frac{0,5 \cdot \alpha}{1 - \rho_s^{[m_r, l/m]}} + \frac{sim_{nmax}}{(1 - \rho_s^{[m_r, l/m]}) \cdot (1 - \alpha)}$$

będzie powodowało wzrost dokładności metody rekomendacji *AdaptRank*.

Biorąc pod uwagę wniosek 5.3.1 trudno nie postawić sobie pytania: dlaczego nie przypisać wyrażeniom z wniosku 5.3.1 wartości minimalnych lub zbliżonych do minimalnych? Może się to okazać niekorzystne z dwóch następujących powodów. Po pierwsze minimalne wartości współczynników dla dokumentów relewantnych będą się niebezpiecznie „zbliżać” do maksymalnych współczynników dla dokumentów nierelewantnych co może spowodować pojawienie się takiej sesji, w której granice między dokumentami relewantnymi i nierelewantnymi zatrą się zupełnie (ze względu na ewentualny błąd oszacowania maksymalnych współczynników nawigacyjnych dokumentów nierelewantnych). Po drugie obniżenie wartości wyrażeń z wniosku 5.3.1 może bardzo niekorzystnie wpłynąć na kompletność rekomendacji (patrz punkt 5.3.2).

Wniosek 5.3.1 jest bardzo ważny, gdyż pozwala przeanalizować jak będzie zmieniać się dokładność metody rekomendacji w zależności od doboru parametrów. Poniższe 3 wnioski wraz

z twierdzeniem 5.3.3 przedstawiają wpływ parametrów α , ρ_s , m oraz m_r na dokładność rekomendacji przy spełnionych założeniach wniosku 5.3.1.

Parametry $0 < \rho_s < 1$ oraz $m, m_r \in \mathbb{N}$ znajdują się w dodatnich mianownikach wyrażen z wniosku 5.3.1. Zważywszy na fakt, że im większa jest wartość $1 - \rho_s^{[m_r/m]}$ tym mniejsze są wyżej wspomniane wyrażenia, można sformułować następujące trzy wnioski:

Wniosek 5.3.2

Jeśli spełnione są warunki wniosku 5.3.1, to im większa jest wartość parametru ρ_s tym mniejsza jest dokładność metody rekomendacji *AdaptRank*.

Wniosek 5.3.2 jest zgodny z intuicją, gdyż parametr ρ_s określa jak „mocno” fakt odwiedzenia dokumentu wpływa negatywnie na wartość jego funkcji rekomendacji w kolejnych krokach bieżącej sesji. Jeśli odwiedzenie danego dokumentu wiąże się z drastycznym spadkiem wartości jego funkcji rekomendacji w przyszłych krokach, łatwo może dojść do sytuacji, w której wartość funkcji rekomendacji dla dokumentów nierelevantnych przewyższy wartość funkcji rekomendacji dla odwiedzonych dokumentów relevantnych, nawet gdy podobieństwo dokumentów nierelevantnych do dokumentów odwiedzonych oraz ich współczynniki nawigacyjne są znacznie niższe od analogicznych wartości uzyskanych przez dokumenty relevantne.

Wniosek 5.3.3

Jeśli spełnione są warunki wniosku 5.3.1, to im większa jest liczba m_r dokumentów relevantnych do potrzeb użytkownika, tym większa jest dokładność metody rekomendacji *AdaptRank*.

Także ten wniosek jest zgodny z intuicją, gdyż jednym z głównych powodów, dla którego mogą zostać zarekomendowane dokumenty nierelevantne jest związana z czasowym „osłabieniem” wartości funkcji rekomendacji dla odwiedzonych dokumentów relevantnych. Jeśli zbiór dokumentów relevantnych jest odpowiednio duży, dokumenty nierelevantne mają małe szanse zostać podpowiedziane z wyżej określonej przyczyny, gdyż w tym samym czasie, w którym ostatnio odwiedzane dokumenty mają bardzo niskie wartości funkcji relevantności, dokumenty rekomendowane na początku sesji przestają być „osłabiane” przez parametr ρ_s .

Wniosek 5.3.4

Jeśli spełnione są warunki wniosku 5.3.1, to im większa jest wartość parametru m (określającego liczbę dokumentów rekomendowanych w pojedynczym kroku bieżącej sesji), tym mniejsza jest dokładność metody rekomendacji *AdaptRank*.

Interpretacja tego wniosku jest następująca: im więcej dokumentów jest rekomendowanych w jednym kroku bieżącej sesji, tym więcej dokumentów relevantnych powinno osiągać wysokie wartości współczynnika wzorców nawigacyjnych oraz podobieństwa do dokumentów odwiedzonych, aby zapewnić maksymalną dokładność, czyli tym trudniejsze jest osiągnięcie maksymalnej dokładności.

Twierdzenie 5.3.3

Jeśli spełnione są warunki wniosku 5.3.1, to im większa jest wartość parametru α , tym mniejsza jest dokładność metody rekomendacji *AdaptRank*.

Dowód:

Pokażemy, że funkcja $prec_{nv}(\alpha) = \frac{0,5 \cdot \alpha}{1 - \rho_s^{[m_r/m]}} + \frac{nv_{nmax}}{(1 - \rho_s^{[m_r/m]}) \cdot (1 - \alpha)}$ gdzie $\alpha \in (0, 1)$ jest rosnąca w całej dziedzinie. Funkcja $prec_{nv}(\alpha)$ jest różniczkowalna, więc do określenia monotoniczności obliczona zostanie pierwsza pochodna funkcji $prec_{nv}(\alpha)$:

$$prec_{nv}'(\alpha) = \frac{0,5}{1 - \rho_s^{[m_r/m]}} + \frac{nv_{nmax}}{1 - \rho_s^{[m_r/m]}} \cdot \left(\frac{1}{1 - \alpha} \right)' \quad (5.3.21)$$

co pozwala otrzymać:

$$prec_{nv}'(\alpha) = \frac{0,5}{1 - \rho_s^{[m_r/m]}} - \frac{nv_{nmax}}{(1 - \rho_s^{[m_r/m]})} \cdot \frac{1}{(1 - \alpha)^2} \cdot -1 \quad (5.3.22)$$

i ostatecznie:

$$prec_{nv}'(\alpha) = \frac{0,5}{1 - \rho_s^{[m_r/m]}} + \frac{nv_{nmax}}{(1 - \rho_s^{[m_r/m]})} \cdot \frac{1}{(1 - \alpha)^2} \quad (5.3.23)$$

Łatwo zauważyć, że pierwsza pochodna funkcji $prec_{nv}(\alpha)$ jest dodatnia w całej dziedzinie (gdyż $\alpha \in (0, 1)$, $\rho_s^{[m_r/m]} \in (0,1)$ oraz $nv_{nmax} \in (0, 1)$) co pozwala stwierdzić, że funkcja $prec_{nv}(\alpha)$ jest rosnąca w całej dziedzinie. Łatwo też zauważyć, że jeśli funkcja $prec_{nv}(\alpha)$ jest rosnąca to rosnąca też będzie funkcja $prec_{sim}(\alpha) = \frac{0,5 \cdot \alpha}{1 - \rho_s^{[m_r/m]}} + \frac{sim_{nmax}}{(1 - \rho_s^{[m_r/m]}) \cdot (1 - \alpha)}$ gdzie $\alpha \in (0, 1)$ co na mocy wniosku 5.3.1 dowodzi prawdziwości twierdzenia (wraz ze wzrostem parametru α rośnie wartość $\frac{0,5 \cdot \alpha}{1 - \rho_s^{[m_r/m]}} + \frac{sim_{nmax}}{(1 - \rho_s^{[m_r/m]}) \cdot (1 - \alpha)}$, więc dokładność maleje).

□

Kolejne twierdzenie pokazuje, że system rekomendujący używający metody *AdaptRank* może uzyskać maksymalną dokładność nie tylko, gdy użytkownik odwiedza rekomendowane dokumenty relewantne, ale także gdy je konsekwentnie ignoruje. Warto podkreślić, że zignorowanie rekomendowanych dokumentów nie musi oznaczać, że podpowiadana treść nie jest relewantna do potrzeb użytkownika. Nieodwiedzenie proponowanego dokumentu może wynikać z niezauważenia rekomendacji lub ze znalezienia innego dokumentu, który lepiej odpowiada bieżącym potrzebom.

Twierdzenie 5.3.4

Jeśli w każdym kroku i ($i=2, \dots, n_{sb}$) bieżącej sesji s_b , użytkownik odwiedza dokument, który nie był rekomendowany w kroku poprzednim ($d^{(i)} \notin R^{(i-1)}$) i system hipertekstowy zawiera m_r ($m_r > m$) dokumentów relewantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{sb})}$ oraz we wszystkich krokach bieżącej sesji spełnione są następujące warunki:

$$nv_{r \min} > 0,5 \cdot \alpha + \frac{nv_{n \max}}{1 - \alpha} + \frac{0,5 \cdot \gamma^{[m_r/m]-1}}{\beta} \left(1 + \frac{\alpha}{1 - \alpha^{[m_r/m]}} \right) \text{ oraz}$$

$$sim_{r \min} > 0,5 \cdot \alpha + \frac{sim_{n \max}}{1 - \alpha} + \frac{0,5 \cdot \gamma^{[m_r/m]-1}}{\delta} \left(1 + \frac{\alpha}{1 - \alpha^{[m_r/m]}} \right)$$

to dokładność systemu rekomendującego, wykorzystującego metodę *AdaptRank*, w bieżącej sesji s_b będzie równa 1 niezależnie od liczby dokumentów odwiedzonych w tej sesji.

Dowód:

Zacniemy od przekształcenia nierówności wyrażających warunki twierdzenia. Mnożąc obie strony pierwszej nierówności przez zawsze dodatnie β i drugiej przez zawsze dodatnie δ otrzymujemy postacie równoważne:

$$\beta \cdot nv_{r \min} > 0,5 \cdot \beta \cdot \alpha + \frac{\beta \cdot nv_{n \max}}{1 - \alpha} + 0,5 \cdot \gamma^{[m_r/m]-1} \left(1 + \frac{\alpha}{1 - \alpha^{[m_r/m]}} \right) \quad (5.3.24)$$

$$\delta \cdot sim_{r \min} > 0,5 \cdot \delta \cdot \alpha + \frac{\delta \cdot sim_{n \max}}{1 - \alpha} + 0,5 \cdot \gamma^{[m_r/m]-1} \left(1 + \frac{\alpha}{1 - \alpha^{[m_r/m]}} \right)$$

Dodając stronami nierówności z 5.3.24 otrzymujemy:

$$\beta \cdot nv_{r \min} + \delta \cdot sim_{r \min} > \quad (5.3.25)$$

$$0,5 \cdot (\beta + \delta) \cdot \alpha + \left(\frac{1}{1 - \alpha} \right) \cdot (\beta \cdot nv_{n \max} + \delta \cdot sim_{n \max}) + \gamma^{[m_r/m]-1} \left(1 + \frac{\alpha}{1 - \alpha^{[m_r/m]}} \right)$$

Z twierdzenia 5.3.1 wiadomo, że aby system rekomendujący osiągnął maksymalną dokładność wartości funkcji f dla dokumentów nirelewantnych musi być mniejsza niż dla m dokumentów relewantnych, co na pewno będzie spełnione gdy:

$$\forall_{ns_b \geq i > 0} \forall_{d_{nj} \in D^N, d_{rj} \in D_R} f(d_r, s_b^{(i)}) > f(d_n, s_b^{(i)}) \quad (5.3.26)$$

Zgodnie z założeniami twierdzenia w najgorszym z przypadków wszystkie dokumenty nierelevantne będą uzyskiwały we wszystkich krokach bieżącej sesji wartości maksymalne, gdy:

$$\forall_{n_{sb} \geq i > 0} \forall_{d_n \in D \setminus D_R^{(i)}} (q(d_n) = 1 \wedge sim(d_n, d^{(i)}) = sim_{n_{max}} \wedge nv_{d_n}^{(i)} = nv_{n_{max}}) \quad (5.3.27)$$

Tak więc wartość funkcji rekomendacji dla dokumentu nierelevantnego d_n spełniać będzie nierówność:

$$f(d_n, s_b^{(i)}) \leq 0,5 \cdot (\beta + \delta) \cdot \alpha^i + \beta \cdot nv_{n_{max}} \cdot \sum_{k=1}^i (\alpha^{i-k}) + \delta \cdot sim_{n_{max}} \cdot \sum_{k=1}^i (\alpha^{i-k}) \quad (5.3.28)$$

$\sum_{k=1}^i (\alpha^{i-k}) = \sum_{k=1}^i (\alpha^{k-1})$ to suma i wyrazów ciągu geometrycznego o pierwszym wyrazie równym 1 i stałym ilorazie równym α , więc:

$$f(d_n, s_b^{(i)}) \leq 0,5 \cdot (\beta + \delta) \cdot \alpha^i + \left(\frac{1 - \alpha^i}{1 - \alpha}\right) \cdot (\beta \cdot nv_{n_{max}} + \delta \cdot sim_{n_{max}}) \quad (5.3.29)$$

Łatwo zauważyć, że:

$$\begin{aligned} 0,5 \cdot (\beta + \delta) \cdot \alpha^i + \left(\frac{1 - \alpha^i}{1 - \alpha}\right) \cdot (\beta \cdot nv_{n_{max}} + \delta \cdot sim_{n_{max}}) &\leq \\ 0,5 \cdot (\beta + \delta) \cdot \alpha + \frac{1}{1 - \alpha} \cdot (\beta \cdot nv_{n_{max}} + \delta \cdot sim_{n_{max}}) & \end{aligned} \quad (5.3.30)$$

Co na mocy przechodności relacji większości pozwala stwierdzić, że jeśli prawdziwe jest 5.3.29 to prawdziwe jest też:

$$f(d_n, s_b^{(i)}) \leq 0,5 \cdot (\beta + \delta) \cdot \alpha + \left(\frac{1}{1 - \alpha}\right) \cdot (\beta \cdot nv_{n_{max}} + \delta \cdot sim_{n_{max}}) \quad (5.3.31)$$

Z drugiej strony zgodnie z założeniami twierdzenia w najgorszym z przypadków wszystkie dokumenty relevantne będą uzyskiwały we wszystkich krokach bieżącej sesji wartości minimalne, co nastąpi gdy:

$$\forall_{n_{sb} \geq i > 0} \forall_{d_r \in D_R^{(i)}} (q(d_r) = 0 \wedge sim(d_r, d^{(i)}) = sim_{r_{min}} \wedge nv_{d_r}^{(i)} = nv_{r_{min}}) \quad (5.3.32)$$

Co oznacza, że wartość funkcji rekomendacji dla dokumentu relevantnego, będzie spełniać nierówność:

$$f(d_r, s_b^{(i)}) \geq \beta \cdot nv_{r \min} \cdot \sum_{k=1}^i (\alpha^{i-k}) + \delta \cdot sim_{r \min} \cdot \sum_{k=1}^i (\alpha^{i-k}) - \sum_{k=1}^i (\alpha^{i-k} \cdot g_{d_r}^{(k)}) \quad (5.3.33)$$

$\sum_{k=1}^i (\alpha^{i-k}) = \sum_{k=1}^i (\alpha^{k-1})$, to suma i wyrazów ciągu geometrycznego o pierwszym wyrazie równym 1 i stałym ilorazie równym α , więc:

$$f(d_r, s_b^{(i)}) \geq \beta \cdot nv_{r \min} \cdot \frac{1-\alpha^i}{1-\alpha} + \delta \cdot sim_{r \min} \cdot \frac{1-\alpha^i}{1-\alpha} - \sum_{k=1}^i (\alpha^{i-k} \cdot g_{d_r}^{(k)}) \quad (5.3.34)$$

Łatwo zauważyć, że dla $i > 0$ jest prawdą, że:

$$\beta \cdot nv_{r \min} + \delta \cdot sim_{r \min} - \sum_{k=1}^i (\alpha^{i-k} \cdot g_{d_r}^{(k)}) \leq \beta \cdot nv_{r \min} \cdot \frac{1-\alpha^i}{1-\alpha} + \delta \cdot sim_{r \min} \cdot \frac{1-\alpha^i}{1-\alpha} - \sum_{k=1}^i (\alpha^{i-k} \cdot g_{d_r}^{(k)}) \quad (5.3.35)$$

Co korzystając z relacji przechodniości pozwala stwierdzić, że jeśli prawdziwa jest nierówność 5.3.35, to prawdziwe będzie także:

$$f(d_r, s_b^{(i)}) \geq \beta \cdot nv_{r \min} + \delta \cdot sim_{r \min} - \sum_{k=1}^i (\alpha^{i-k} \cdot g_{d_r}^{(k)}) \quad (5.3.36)$$

Należy zauważyć, że w przypadku ignorowania dokumentów relewantnych po $[m_r/m]$ krokach wszystkie dokumenty relewantne zostaną zarekomendowane, gdyż równe wartości funkcji przydatności, współczynników wzorców nawigacyjnych oraz podobieństwa do dokumentów odwiedzonych powodują, że o różnicach w wartości funkcji f decydują wartości odpowiednich współrzędnych wektora $\mathbf{g}^{(i)}$, które są zawsze zerowe dla dokumentów jeszcze nierekomendowanych w tej sesji.

Jednocześnie dzięki wyżej wspomnianym równym współczynnikom, w kroku $[m_r/m]+1$ najwyższe wartości funkcji rekomendacji będą posiadały dokumenty zarekomendowane w pierwszym kroku, gdyż współrzędna wektora $\mathbf{g}^{(i)}$ jest dodatnia tylko dla dokumentów, które już były zignorowane w bieżącej sesji, a wpływ wartości odpowiedniej współrzędnej wektora $\mathbf{g}^{(i)}$ w kolejnych krokach maleje. Podobnie w kroku $[m_r/m]+2$ najwyższe wartości funkcji rekomendacji będą miały dokumenty zarekomendowane w kroku 2. Tak więc uogólniając w kroku $i > [m_r/m]$ największą wartością funkcji rekomendacji będą odznaczały się dokumenty, które były rekomendowane w kroku $i - [m_r/m]$.

Z rozumowania przeprowadzonego powyżej wynika, że w kroku i najwyższą wartością funkcji rekomendacji będzie odznaczał się dokument d_r , który był zarekomendowany w krokach

$i - [m_r/m], i - 2[m_r/m], \dots, i - \left\lfloor \frac{i}{[m_r/m]} \right\rfloor \cdot [m_r/m] > 0$. Co korzystając dodatkowo z twierdzenia 5.1.6 pozwala zapisać:

$$f(d_r, s_b^{(i)}) \geq \beta \cdot nv_{r \min} + \delta \cdot sim_{r \min} - \gamma^{i - (i - [m_r/m]) - 1} - \alpha^{i - 2[m_r/m] - 1} \cdot \gamma^{i - [m_r/m] - (i - 2[m_r/m]) - 1} - \dots - \alpha^{\left\lfloor \frac{i}{[m_r/m]} \right\rfloor [m_r/m] - 1} \cdot \gamma^{[m_r/m] - 1} \quad (5.3.37)$$

Co po prostych przekształceniach daje:

$$f(d_r, s_b^{(i)}) \geq \beta \cdot nv_{r \min} + \delta \cdot sim_{r \min} - \gamma^{[m_r/m] - 1} - \alpha^{i - 2[m_r/m] - 1} \cdot \gamma^{[m_r/m] - 1} - \dots - \alpha^{\left\lfloor \frac{i}{[m_r/m]} \right\rfloor [m_r/m] - 1} \cdot \gamma^{[m_r/m] - 1} \quad (5.3.38)$$

Po wyłączeniu $\gamma^{[m_r/m] - 1}$ przed nawias otrzymujemy:

$$f(d_r, s_b^{(i)}) \geq \beta \cdot nv_{r \min} \cdot \frac{1 - \alpha^i}{1 - \alpha} + \delta \cdot sim_{r \min} \cdot \frac{1 - \alpha^i}{1 - \alpha} - \gamma^{[m_r/m] - 1} - \gamma^{[m_r/m] - 1} \left(\alpha^{i - 2[m_r/m] - 1} + \dots + \alpha^{\left\lfloor \frac{i}{[m_r/m]} \right\rfloor [m_r/m] - 1} \right) \quad (5.3.39)$$

$\alpha^{i - 2[m_r/m] - 1} + \dots + \alpha^{\left\lfloor \frac{i}{[m_r/m]} \right\rfloor [m_r/m] - 1}$ to suma $\left\lfloor \frac{i}{[m_r/m]} \right\rfloor - 1$ wyrazów ciągu geometrycznego o pierwszym wyrazie równym $\alpha^{i - 2[m_r/m]}$ i stałym ilorazie równym $\alpha^{[m_r/m]}$:

$$f(d_r, s_b^{(i)}) \geq \beta \cdot nv_{r \min} \cdot \frac{1 - \alpha^i}{1 - \alpha} + \delta \cdot sim_{r \min} \cdot \frac{1 - \alpha^i}{1 - \alpha} - \gamma^{[m_r/m] - 1} - \gamma^{[m_r/m] - 1} \cdot \alpha^{i - 2[m_r/m]} \cdot \frac{1 - (\alpha^{[m_r/m]})^{\left\lfloor \frac{i}{[m_r/m]} \right\rfloor - 1}}{1 - \alpha^{[m_r/m]}} \quad (5.3.40)$$

Warto zauważyć, że:

$$\begin{aligned} & \beta \cdot nv_{r \min} \cdot \frac{1 - \alpha^i}{1 - \alpha} + \delta \cdot sim_{r \min} \cdot \frac{1 - \alpha^i}{1 - \alpha} - \\ & - \gamma^{[m_r/m] - 1} - \gamma^{[m_r/m] - 1} \cdot \alpha^{i - 2[m_r/m]} \cdot \frac{1 - (\alpha^{[m_r/m]})^{\left\lfloor \frac{i}{[m_r/m]} \right\rfloor - 1}}{1 - \alpha^{[m_r/m]}} \geq \\ & \beta \cdot nv_{r \min} + \delta \cdot sim_{r \min} - \gamma^{[m_r/m] - 1} - \gamma^{[m_r/m] - 1} \cdot \frac{\alpha}{1 - \alpha^{[m_r/m]}} \end{aligned} \quad (5.3.41)$$

Podstawiając 5.3.41 i 5.3.31 do 5.3.26 otrzymujemy:

$$\beta \cdot nv_{r_{\min}} + \delta \cdot sim_{r_{\min}} - \gamma^{[m_r, l_m]-1} \left(1 + \frac{\alpha}{1 - \alpha^{[m_r, l_m]}} \right) > \quad (5.3.42)$$

$$0,5 \cdot (\beta + \delta) \cdot \alpha + \left(\frac{1}{1 - \alpha} \right) \cdot (\beta \cdot nv_{n_{\max}} + \delta \cdot sim_{n_{\max}})$$

Dodając do obu stron $\gamma^{[m_r, l_m]-1} \left(1 + \frac{\alpha}{1 - \alpha^{[m_r, l_m]}} \right)$ otrzymujemy:

$$\beta \cdot nv_{r_{\min}} + \delta \cdot sim_{r_{\min}} > \quad (5.3.43)$$

$$0,5 \cdot (\beta + \delta) \cdot \alpha + \left(\frac{1}{1 - \alpha} \right) \cdot (\beta \cdot nv_{n_{\max}} + \delta \cdot sim_{n_{\max}}) + \gamma^{[m_r, l_m]-1} \left(1 + \frac{\alpha}{1 - \alpha^{[m_r, l_m]}} \right)$$

5.3.43 jest równe przekształceniu założeń zawartych w 5.3.25 co pozwala wnioskować, że twierdzenie jest prawdziwe. □

Twierdzenie 5.3.4 jest podobne do twierdzenia 5.3.2. Główna różnica polega na zmianie zachowania użytkownika, który zgodnie z założeniami twierdzenia 5.3.4 konsekwentnie ignoruje podpowiadane mu dokumenty. Także i w tym przypadku metoda rekomendacji *AdaptRank* będzie odznaczać się maksymalną dokładnością, jeśli istnieje wyraźna różnica między współczynnikami nawigacyjnymi dokumentów relewantnych i dokumentów nierelewantnych. To samo tyczy się podobieństwa do dokumentów odwiedzonych.

W praktyce metoda *AdaptRank* często powinna odznaczać się wysoką dokładnością, gdyż wyraźnie wyższe podobieństwo dokumentów relewantnych do odwiedzanych oraz uznanie za relewantne dokumentów należących do tego samego wzorca nawigacyjnego, co dokumenty odwiedzane jest zgodne z intuicją.

Należy zauważyć, że parametry α , m oraz γ , a także liczba dokumentów relewantnych m_r determinują, jaka musi być różnica między $nv_{r_{\min}}$ i $nv_{r_{\max}}$ oraz między $sim_{r_{\min}}$ i $sim_{r_{\max}}$, aby system rekomendujący mógł osiągnąć maksymalną dokładność. Przykładowo, dla wartości parametrów zaproponowanych w tabeli 4.5.1 ($\alpha=0,25$, $\beta=1$, $\delta=1$, $\gamma=0,9$, $\rho_s=0,75$, $m=2$) i przy założeniu, że istnieje $m_r=30$ dokumentów relewantnych do potrzeb użytkownika oraz że $nv_{r_{\max}}=0,1$ oraz $sim_{r_{\max}}=0,1$ do osiągnięcia maksymalnej dokładności wystarczy, aby minimalne podobieństwo dokumentów relewantnych do dokumentów odwiedzonych w bieżącej sesji oraz minimalna wartość współczynnika wzorców nawigacyjnych nie były mniejsze niż 0,45. Tak dobrane wartości parametrów nie wydają się być trudnymi do osiągnięcia w praktyce.

Za pomocą wyżej wymienionych parametrów można zwiększać i zmniejszać dokładność metody *AdaptRank*. Podobnie jak w przypadku twierdzenia 5.3.2 drastyczne obniżenie prawych stron nierówności zawartych w twierdzeniu 5.3.4 nie zawsze jest korzystne, gdyż może spowodować zarekomendowanie dokumentów nierelewantnych. Analogicznie, zbyt wysokie wartości wyżej wspomnianych współczynników spowoduje cykliczne rekomendowanie tych samych dokumentów, co przyczyni się do obniżenia kompletności rekomendacji.

Wpływ zmiany wyrażeń podanych w twierdzeniu 5.3.4 na dokładność przedstawiony został w następującym wniosku (jest to wniosek analogiczny do wniosku 5.3.1).

Wniosek 5.3.5

Jeśli w każdym kroku i ($i=2, \dots, n_{sb}$) bieżącej sesji s_b , użytkownik odwiedza dokument, który nie był rekomendowany w kroku poprzednim ($d^{(i)} \notin R^{(i-1)}$) i system hipertekstowy zawiera m_r ($m_r > m$) dokumentów relewantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{sb})}$, to obniżenie wartości następujących wyrażeń:

$$0,5 \cdot \alpha + \frac{nv_{n_{\max}}}{1 - \alpha} + \frac{0,5 \cdot \gamma^{[m_r/m]-1}}{\beta} \left(1 + \frac{\alpha}{1 - \alpha^{[m_r/m]}} \right)$$

$$0,5 \cdot \alpha + \frac{sim_{n_{\max}}}{1 - \alpha} + \frac{0,5 \cdot \gamma^{[m_r/m]-1}}{\delta} \left(1 + \frac{\alpha}{1 - \alpha^{[m_r/m]}} \right)$$

będzie powodowało wzrost dokładności metody rekomendacji *AdaptRank*.

Wniosek 5.3.5 pozwala przeanalizować jak będzie zmieniać się dokładność metody rekomendacji w zależności od doboru parametrów. Parametr $0 < \gamma < 1$ znajduje się w licznikach dodatnich wyrażeń z wniosku 5.3.5, co pozwala sformułować następujący wniosek.

Wniosek 5.3.6

Jeśli spełnione są warunki wniosku 5.3.5, to wzrost wartości parametru γ będzie powodował spadek dokładności metody rekomendacji *AdaptRank*.

Wniosek 5.3.6 jest zgodny z intuicją, gdyż parametr γ określa jak „mocno” fakt zignorowania danego dokumentu wpływa negatywnie na wartość jego funkcji rekomendacji w kolejnych krokach. Jeśli zignorowanie danego dokumentu wiąże się z drastycznym spadkiem wartości funkcji rekomendacji dla tego dokumentu w kolejnych krokach, to łatwo może dojść do sytuacji, w której po zignorowaniu przez użytkownika wszystkich dokumentów relewantnych jeden z dokumentów nierелеwantnych osiągnie najwyższą wartość funkcji rekomendacji, a w konsekwencji zostanie zarekomendowany.

Parametry $1 \geq \beta > 0$ oraz $1 \geq \delta > 0$ znajdują się w dodatnich mianownikach wyrażeń z wniosku 5.3.5 co pozwala na sformułowanie następującego wniosku:

Wniosek 5.3.7

Jeśli spełnione są warunki wniosku 5.3.5, to im większa jest wartość parametrów β oraz δ tym większa jest dokładność metody rekomendacji *AdaptRank*.

Duże wartości parametrów β oraz δ powodują, wzrost znaczenia współczynnika wzorców nawigacyjnych oraz podobieństwa do dokumentów odwiedzanych na wartości funkcji rekomendacji dla wszystkich dokumentów, wskutek czego względne obniżenie się wartości funkcji rekomendacji dla relewantnych dokumentów zignorowanych wobec dokumentów nierелеwantnych nie będzie tak znaczne, a w konsekwencji dokumentom nierелеwantnym trudniej

będzie osiągnąć wyższe wartości funkcji rekomendacji od wartości funkcji rekomendacji dla dokumentów relewantnych.

Parametry m_r i m występują w wyrażeniu $\gamma^{[m_r/m]-1}$ znajdującym się w dodatnich licznikach formuł z wniosku 5.3.5 oraz w wyrażeniu $1 - \alpha^{[m_r/m]}$ znajdującym się w dodatnich mianownikach formuł z wniosku 5.3.5, co pozwala na sformułowanie następujących dwóch wniosków:

Wniosek 5.3.8

Jeśli spełnione są warunki wniosku 5.3.5, to im większa jest liczba m_r dokumentów relewantnych do potrzeb użytkownika, tym większa jest dokładność metody rekomendacji *AdaptRank*.

Także ten wniosek jest zgodny z intuicją, gdyż jeden z głównych powodów, dla którego mogą zostać zarekomendowane dokumenty nerelewantne, jest związany z czasowym „osłabieniem” wartości zignorowanych dokumentów relewantnych. Jeśli zbiór dokumentów relewantnych jest odpowiednio duży, dokumenty nerelewantne mają małe szanse zostać podpowiedziane z wyżej określonej przyczyny, gdyż w tym samym czasie, w którym ostatnio zignorowane dokumenty będą odznaczały się niskimi wartościami funkcji relewancji, dokumenty rekomendowane na początku sesji przestaną być „osłabiane” przez wektor $\mathbf{g}^{(i)}$.

Wniosek 5.3.9

Jeśli spełnione są warunki wniosku 5.3.5, to im większa jest wartość parametru m (określającego liczbę dokumentów rekomendowanych w pojedynczym kroku bieżącej sesji) tym mniejsza jest dokładność metody rekomendacji *AdaptRank*.

Interpretacja tego wniosku jest następująca: im więcej dokumentów jest rekomendowanych w jednym kroku bieżącej sesji, tym więcej dokumentów relewantnych powinno osiągać wysokie wartości współczynnika wzorców nawigacyjnych oraz podobieństwa do dokumentów odwiedzonych, aby zapewnić maksymalną dokładność, czyli tym trudniejsze jest osiągnięcie maksymalnej dokładności.

Wzrost parametru α powoduje wzrost wartości wyrażen $0,5 \cdot \alpha, \frac{1}{1-\alpha}, 1 + \frac{\alpha}{1-\alpha^{[m_r/m]}}$, co umożliwia sformułowanie następującego wniosku:

Wniosek 5.3.10

Jeśli spełnione są warunki wniosku 5.3.5, to im większa jest wartość parametru α tym mniejsza jest dokładność metody rekomendacji *AdaptRank*.

5.3.2. Kompletność metody rekomendacji *AdaptRank*

Przyglądając się definicji 4.1.6 dotyczącej kompletności rekomendacji łatwo zauważyć, że zarekomendowanie wszystkich dokumentów relewantnych w bieżącej sesji pozwoli osiągnąć maksymalną kompletność. Oznacza to, że podobnie jak w wyszukiwaniu do osiągnięcia maksymalnej kompletności „wystarczy wyświetlić” wszystkie relewantne dokumenty znajdujące się w systemie.

Niestety, w odróżnieniu od wyszukiwania, w rekomendacji podpowiadane dokumenty prezentowane są w ograniczonej liczbie, a dodatkowo w ramach bieżącej sesji preferencje użytkownika mogą ulec zmianie. Dlatego też do zarekomendowania wszystkich dokumentów relewantnych potrzeba dużej liczby kroków. Definicja 4.1.6 będąca adaptacją do specyfiki rekomendacji, definicji kompletności znanej z wyszukiwania, pozwala osiągnąć maksymalną kompletność także sesjom, w których liczba dokumentów odwiedzonych jest wyraźnie mniejsza od liczby dokumentów relewantnych do potrzeb użytkownika.

Twierdzenie 5.3.5 określa warunki, jakie musi spełniać dowolna funkcja rekomendacji, aby kompletność systemu rekomendującego w bieżącej sesji wyniosła wartość maksymalną niezależnie od długości tej sesji. To twierdzenie jest bardzo ważne, gdyż pokazuje, że do osiągnięcia maksymalnej kompletności nie wymaga od się systemu rekomendującego wyznaczenia całego zbioru dokumentów relewantnych (zbiór ten jest znany tylko użytkownikowi i określenie go za pomocą formalnej metody nie jest zadaniem łatwym). W celu uzyskania maksymalnej kompletności wystarczy w każdym kroku bieżącej sesji „zidentyfikować” m dokumentów relewantnych nierekomendowanych jeszcze w tej sesji.

Twierdzenie 5.3.5

Niech $D_{RN}^{(i)} \subseteq D_R^{(i)}$ oznacza zbiór dokumentów relewantnych w kroku i , które nie były zarekomendowane użytkownikowi w bieżącej sesji s_b . System rekomendacji SR w bieżącej sesji s_b osiąga maksymalną kompletność, jeśli w każdym kroku i ($i=1, 2, \dots, n_{sb}$) tej sesji, co najmniej m_{aux} dokumentów należących do zbioru $D_{RN}^{(i)}$ osiąga wartości funkcji rekomendacji wyższe od wartości tej funkcji dla dokumentów nie należących do tego zbioru, tj. gdy:

$$\forall_{n_{sb} \geq i > 0} \exists D_{RN}^{(i)} \subseteq D_R^{(i)} \forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D \setminus D_{RN}^{(i)}} \left((f(d_r, s_b^{(i)}) > f(d_n, s_b^{(i)})) \wedge (card(D_{RN}^{(i)}) \geq m_{aux}) \right)$$

gdzie m_{aux} to mniejsza z dwóch następujących liczb: licznosc zbioru $D_{RN}^{(i)}$ oraz liczba dokumentów m podpowiadanych w jednym kroku.

Dowód:

Zgodnie z definicją 4.1.6 system rekomendujący w bieżącej sesji s_b będzie odznaczał się maksymalną kompletnością, gdy:

$$\frac{card\left(\bigcup_{i=1}^{n_{sb}} (D_R^{(i)} \cap R^{(i)})\right)}{n_r} = 1 \quad (5.3.44)$$

gdzie n_r oznacza największą liczbę różnych dokumentów relewantnych, które mogły być zarekomendowane użytkownikowi w tej sesji zakładając, że w każdym kroku rekomendowane jest dokładnie m dokumentów $n_r = m + \sum_{i=2}^{n_{sb}} \min\left(m, card\left(D_R^{(i)} \setminus \bigcup_{j=1}^{i-1} R^{(j)}\right)\right)$.

Po pomnożeniu obu stron przez n_r otrzymujemy:

$$\text{card}\left(\bigcup_{i=1}^{n_{sb}}(D_R^{(i)} \cap R^{(i)})\right) = n_r \quad (5.3.45)$$

i dalej:

$$\text{card}\left(\bigcup_{i=1}^{n_{sb}}(D_R^{(i)} \cap R^{(i)})\right) = m + \sum_{i=2}^{n_{sb}} \min(m, \text{card}\left(D_R^{(i)} \setminus \bigcup_{j=1}^{i-1} R^{(j)}\right)) \quad (5.3.46)$$

Za pomocą zasady indukcji matematycznej pokażemy, że 5.3.46 jest prawdziwe dla dowolnej długości bieżącej sesji określonego przez n_{sb} pod warunkiem, że spełnione są założenia twierdzenia, czyli:

$$\forall_{n_{sb} \geq i > 0} \exists_{D_{RN}^{(i)} \subseteq D_{RN}^{(i)}} \forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D \setminus D_{RN}^{(i)}} \left((f(d_r, s_b^{(i)}) > f(d_n, s_b^{(i)})) \wedge (\text{card}(D_{RN}^{(i)}) \geq m_{aux}) \right) \quad (5.3.47)$$

Dla $n_{sb}=1$ otrzymujemy:

$$\text{card}\left(\bigcup_{i=1}^1(D_R^{(i)} \cap R^{(i)})\right) = m \quad (5.3.48)$$

5.3.47 ma być prawdziwe, gdy prawdą jest:

$$\exists_{D_{RN}^{(1)} \subseteq D_{RN}^{(1)}} \forall_{d_r \in D_R^{(1)}} \forall_{d_n \in D \setminus D_{RN}^{(1)}} \left((f(d_r, s_b^{(1)}) > f(d_n, s_b^{(1)})) \wedge (\text{card}(D_{RN}^{(1)}) \geq m_{aux}) \right) \quad (5.3.49)$$

Łatwo zauważyć, że w pierwszym kroku: $D_R^{(1)} = D_{RN}^{(1)}$ (żaden dokument nie był wcześniej rekomendowany w tej bieżącej sesji) oraz $m_{aux}=m$, gdyż z definicji problemu wiadomo, że $D_R^{(1)} \geq m$. Tak więc 5.3.49 gwarantuje w tym przypadku, że co najmniej m najwyższych wartości funkcji rekomendacji będzie należeć do dokumentów ze zbioru $D_{RN}^{(1)}$, co oznacza, że w pierwszym kroku zostanie zarekomendowanych m dokumentów ze zbioru $D_{RN}^{(1)}$. Co dowodzi prawdziwości twierdzenia dla $n_{sb}=1$.

Następnie pokażemy, że gdy twierdzenie jest prawdziwe dla pewnego $n_{sb}>0$ to jest też prawdziwe dla $n_{sb}+1$. Podstawiając $n_{sb}+1$ do 5.3.46 otrzymujemy:

$$\text{card}\left(\bigcup_{i=1}^{n_{sb}+1}(D_R^{(i)} \cap R^{(i)})\right) = m + \sum_{i=2}^{n_{sb}+1} \min(m, \text{card}\left(D_R^{(i)} \setminus \bigcup_{j=1}^{i-1} R^{(j)}\right)) \quad (5.3.50)$$

Wyłączając ostatnie składniki z odpowiednich sum otrzymujemy:

$$\begin{aligned}
& \text{card}\left(D_R^{(n_{sb}+1)} \cap R^{(n_{sb}+1)}\right) + \text{card}\left(\bigcup_{i=1}^{n_{sb}} (D_R^{(i)} \cap R^{(i)})\right) \\
& - \text{card}\left(\left(D_R^{(n_{sb}+1)} \cap R^{(n_{sb}+1)}\right) \cap \bigcup_{i=1}^{n_{sb}+1} (D_R^{(i)} \cap R^{(i)})\right) = \\
& = m + \min\left(m, \text{card}\left(D_R^{(n_{sb}+1)} \setminus \bigcup_{j=1}^{n_{sb}-1} R^{(j)}\right)\right) + \sum_{i=2}^{n_{sb}} \min(m, \text{card}\left(D_R^{(i)} \setminus \bigcup_{j=1}^{i-1} R^{(j)}\right))
\end{aligned} \tag{5.3.51}$$

Należy zauważyć, że formuła 5.3.47 gwarantuje, że:

$$\begin{aligned}
& \text{card}\left(D_R^{(n_{sb}+1)} \cap R^{(n_{sb}+1)}\right) - \text{card}\left(\left(D_R^{(n_{sb}+1)} \cap R^{(n_{sb}+1)}\right) \cap \bigcup_{i=1}^{n_{sb}+1} (D_R^{(i)} \cap R^{(i)})\right) = \\
& = \min\left(m, \text{card}\left(D_R^{(n_{sb}+1)} \setminus \bigcup_{j=1}^{n_{sb}-1} R^{(j)}\right)\right)
\end{aligned} \tag{5.3.52}$$

więc:

$$\begin{aligned}
& \min\left(m, \text{card}\left(D_R^{(n_{sb}+1)} \setminus \bigcup_{j=1}^{n_{sb}-1} R^{(j)}\right)\right) + \text{card}\left(\bigcup_{i=1}^{n_{sb}} (D_R^{(i)} \cap R^{(i)})\right) = \\
& \min\left(m, \text{card}\left(D_R^{(n_{sb}+1)} \setminus \bigcup_{j=1}^{n_{sb}-1} R^{(j)}\right)\right) + m + \sum_{i=2}^{n_{sb}} \min(m, \text{card}\left(D_R^{(i)} \setminus \bigcup_{j=1}^{i-1} R^{(j)}\right))
\end{aligned} \tag{5.3.53}$$

i ostatecznie:

$$\text{card}\left(\bigcup_{i=1}^{n_{sb}} (D_R^{(i)} \cap R^{(i)})\right) = m + \sum_{i=2}^{n_{sb}} \min(m, \text{card}\left(D_R^{(i)} \setminus \bigcup_{j=1}^{i-1} R^{(j)}\right)) \tag{5.3.54}$$

Tak więc zostało pokazane, że 5.3.46 jest prawdziwe dla $n_{sb}+1$, jeśli jest prawdziwe dla n_{sb} , co pozwala stwierdzić, że całe twierdzenie jest prawdziwe. \square

Powyższe twierdzenie pokazuje, że do osiągnięcia maksymalnej kompletności wystarczy tylko, aby co najmniej m dokumentów relewantnych nierekomendowanych jeszcze w bieżącej sesji miało wartości funkcji rekomendacji większe zarówno od wartości tej funkcji dla dokumentów nierekomendowanych oraz dokumentów relewantnych już w tej sesji rekomendowanych.

Twierdzenie 5.3.6 określa warunki, przy jakich metoda *AdaptRank* pozwoli osiągnąć maksymalną kompletność pod warunkiem, że użytkownik odwiedza rekomendowane dokumenty relewantne.

Twierdzenie 5.3.6

Jeśli w każdym kroku i ($i=2, \dots, n_{sb}$) bieżącej sesji s_b , użytkownik odwiedza jeden z dokumentów rekomendowanych w kroku poprzednim ($d^{(i)} \in R^{(i-1)}$) i system hipertekstowy zawiera m_r ($m_r > m$) dokumentów relewantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{sb})}$ oraz w $[m_r/m]$ pierwszych krokach bieżącej sesji spełnione są następujące warunki:

$$\begin{aligned} nv_{r \min} &> 0,5 \cdot \alpha + \frac{nv_{n \max}}{1 - \alpha} \\ sim_{r \min} &> 0,5 \cdot \alpha + \frac{sim_{n \max}}{1 - \alpha} \\ nv_{r \min} &> (1 - \rho_s) \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} \\ sim_{r \min} &> (1 - \rho_s) \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} \end{aligned}$$

to kompletność systemu rekomendującego, używającego metody *AdaptRank*, w bieżącej sesji s_b będzie równa 1 niezależnie od liczby dokumentów odwiedzonych w tej sesji.

Dowód:

Przekształćmy najpierw 2 pierwsze nierówności z twierdzenia. Mnożąc pierwszą nierówność przez β , a drugą przez δ otrzymujemy:

$$\begin{aligned} \beta \cdot nv_{r \min} &> 0,5 \cdot \beta \cdot \alpha + \frac{\beta \cdot nv_{n \max}}{1 - \alpha} \\ \delta \cdot sim_{r \min} &> 0,5 \cdot \delta \cdot \alpha + \frac{\delta \cdot sim_{n \max}}{1 - \alpha} \end{aligned} \quad (5.3.55)$$

Sumując obydwie nierówności można zapisać:

$$\beta \cdot nv_{r \min} + \delta \cdot sim_{r \min} > 0,5 \cdot (\beta + \delta) \cdot \alpha + \frac{\beta \cdot nv_{n \max} + \delta \cdot sim_{n \max}}{1 - \alpha} \quad (5.3.56)$$

Następnie przekształćmy dwie ostatnie nierówności z twierdzenia. Mnożąc trzecią nierówność przez β , a czwartą przez δ otrzymujemy:

$$\begin{aligned} \beta \cdot nv_{r \min} &> (1 - \rho_s) \cdot \beta \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} \\ \delta \cdot sim_{r \min} &> (1 - \rho_s) \cdot \delta \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} \end{aligned} \quad (5.3.57)$$

Dodając nierówności stronami otrzymujemy:

$$\beta \cdot nv_{r \min} + \delta \cdot sim_{r \min} > (1 - \rho_s) \cdot \beta \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} + \delta \cdot (1 - \rho_s) \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} \quad (5.3.58)$$

Z twierdzenia 5.3.5 wiadomo, że aby kompletność była maksymalna to w każdym kroku bieżącej sesji, w której zbiór $D_{RN}^{(i)}$ jest niepusty, co najmniej m_{aux} dokumentów należących do zbioru $D_{RN}^{(i)}$ musi osiągać wartości funkcji rekomendacji wyższe od wartości funkcji rekomendacji dla dokumentów nie należących do zbioru $D_{RN}^{(i)}$. Łatwo zauważyć, że jeśli zbiór dokumentów relewantnych jest taki sam we wszystkich krokach bieżącej sesji to przy rekomendowaniu tylko i wyłącznie różnych dokumentów relewantnych po $[m_r/m]$ krokach kompletność osiągnie wartości maksymalną i już nie zmniejszy się niezależnie od długości bieżącej sesji.

To znaczy, że udowodnianie twierdzenia wystarczy pokazać, że dla każdego odwiedzonego dokumentu relewantnego $d_r \in D_R^{(i)} \setminus D_{RN}^{(i)}$ oraz dla każdego nirelewantnego dokumentu $d_n \in D \setminus D_R^{(i)}$ zajdzie:

$$\forall_{[m_r/m] \geq i > 0} f(d_m, s_b^{(i)}) > f(d_r, s_b^{(i)}) \wedge f(d_m, s_b^{(i)}) > f(d_n, s_b^{(i)}) \quad (5.3.59)$$

Dokumenty należące do zbioru $D_{RN}^{(i)}$ osiągną minimalne wartości funkcji rekomendacji zgodne z założeniami twierdzenia, czyli pod warunkiem, że:

$$\forall_{[m_r/m] \geq i > 0} \forall_{d_m \in D_R^{(i)}} (q(d_r) = 0 \wedge sim(d_m, d^{(i)}) = sim_{r \min} \wedge nv_{d_m}^{(i)} = nv_{r \min}) \quad (5.3.60)$$

Dokumenty należące do zbioru $D \setminus D_{RN}^{(i)}$ będą się odznaczać wartościami maksymalnymi funkcji rekomendacji zgodnymi z założeniami twierdzenia. Oznacza to, że dokumenty nirelewantne będą spełniać:

$$\forall_{[m_r/m] \geq i > 0} \forall_{d_n \in D \setminus D_R^{(i)}} (q(d_n) = 1 \wedge sim(d_n, d^{(i)}) = sim_{n \max} \wedge nv_{d_n}^{(i)} = nv_{n \max}) \quad (5.3.61)$$

Dokumenty relewantne odwiedzone w bieżącej sesji będą spełniać:

$$\forall_{[m_r/m] \geq i > 0} \forall_{d_r \in D \setminus D_R^{(i)}} (q(d_r) = 1 \wedge sim(d_r, d^{(i)}) = 1 \wedge nv_{d_r}^{(i)} = 1) \quad (5.3.62)$$

Z 5.3.60 można wywnioskować, że wartość funkcji rekomendacji dla dokumenty relewantnego nieodwiedzonego w bieżącej sesji będzie spełniać nierówność:

$$f(d_m, s_b^{(i)}) \geq \beta \cdot nv_{r \min} \cdot \sum_{k=1}^i (\alpha^{i-k}) + \delta \cdot sim_{r \min} \cdot \sum_{k=1}^i (\alpha^{i-k}) \quad (5.3.63)$$

$\sum_{k=1}^i (\alpha^{i-k}) = \sum_{k=1}^i (\alpha^{k-1})$ to suma i wyrazów ciągu geometrycznego o pierwszym wyrazie równym 1 i stałym ilorazie równym α , więc:

$$f(d_m, s_b^{(i)}) \geq \left(\frac{1-\alpha^i}{1-\alpha} \right) \cdot (\beta \cdot nv_{r \min} + \delta \cdot sim_{r \min}) \quad (5.3.64)$$

W lemacie 1 z twierdzenia 5.3.2 pokazano, że jeśli zarekomendowany i niezignorowany dokument relewantny d_r spełnia:

$$\forall_{n_{s_b} \geq i > 0} \forall_{d_r \in D_R^{(i)}} (q(d_r) = 0 \wedge sim(d_r, d^{(i)}) = sim_{r \min} \wedge nv_{d_r}^{(i)} = nv_{r \min}) \quad (5.3.65)$$

To maksymalna wartość funkcji rekomendacji dla dokumentu relewantnego wyniesie:

$$\begin{aligned} f(d_{r \max}, s_b^{(i)}) = & \beta \cdot \sum_{k=i+1-[\lfloor m_r / m \rfloor]}^i \left(\alpha^{i-k} \cdot nv_{r \min} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i+[\lfloor m_r / m \rfloor]}) \right) + \\ & + \delta \cdot \sum_{k=i+1-[\lfloor m_r / m \rfloor]}^i \left(\alpha^{i-k} \cdot sim_{r \min} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i+[\lfloor m_r / m \rfloor]}) \right) \end{aligned} \quad (5.3.66)$$

Łatwo zauważyć, że jeśli:

$$\forall_{n_{s_b} \geq i > 0} \forall_{d_r \in D_R^{(i)}} (sim(d_r, d^{(i)}) = 1 \wedge nv_{d_r}^{(i)} = 1) \quad (5.3.67)$$

to maksymalna wartość funkcji rekomendacji dla już odwiedzonego dokumentu relewantnego będzie spełniać nierówność:

$$\begin{aligned} f(d_r, s_b^{(i)}) \leq & \beta \cdot \sum_{k=i+1-[\lfloor m_r / m \rfloor]}^i \left(\alpha^{i-k} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i+[\lfloor m_r / m \rfloor]}) \right) + \\ & + \delta \cdot \sum_{k=i+1-[\lfloor m_r / m \rfloor]}^i \left(\alpha^{i-k} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i+[\lfloor m_r / m \rfloor]}) \right) \end{aligned} \quad (5.3.68)$$

W tym przypadku wartość $q(d_r)$ nie ma żadnego znaczenia, gdyż jak to pokazano w twierdzeniu 5.1.5 wartość funkcji rekomendacji dla odwiedzonego dokumentu d_r nie zależy od $q(d_r)$.

Łatwo zauważyć, że prawdziwa jest następująca nierówność:

$$\begin{aligned} \beta \cdot \sum_{k=i+1-[\lfloor m_r / m \rfloor]}^i (\alpha^{i-k} \cdot (1 - \rho_s)) + \delta \cdot \sum_{k=i+1-[\lfloor m_r / m \rfloor]}^i (\alpha^{i-k} \cdot (1 - \rho_s)) \geq \\ \beta \cdot \sum_{k=i+1-[\lfloor m_r / m \rfloor]}^i \left(\alpha^{i-k} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i+[\lfloor m_r / m \rfloor]}) \right) + \delta \cdot \sum_{k=i+1-[\lfloor m_r / m \rfloor]}^i \left(\alpha^{i-k} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i+[\lfloor m_r / m \rfloor]}) \right) \end{aligned} \quad (5.3.69)$$

Co oznacza, że jeśli prawdziwa jest 5.3.68 to prawdziwe będzie także:

$$f(d_r, s_b^{(i)}) \leq \beta \cdot \sum_{k=i+1-\lfloor m_r/m \rfloor}^i (\alpha^{i-k} \cdot (1 - \rho_s)) + \delta \cdot \sum_{k=i+1-\lfloor m_r/m \rfloor}^i (\alpha^{i-k} \cdot (1 - \rho_s)) \quad (5.3.70)$$

Wyłączając $1 - \rho_s$ przed obydwoma znakami sumy otrzymujemy:

$$f(d_r, s_b^{(i)}) \leq (1 - \rho_s) \cdot \beta \cdot \sum_{k=i+1-\lfloor m_r/m \rfloor}^i (\alpha^{i-k}) + \delta \cdot (1 - \rho_s) \cdot \sum_{k=i+1-\lfloor m_r/m \rfloor}^i (\alpha^{i-k}) \quad (5.3.71)$$

$\sum_{k=i+1-\lfloor m_r/m \rfloor}^i (\alpha^{i-k})$ jest równa sumie $\lfloor m_r/m \rfloor$ wyrazów ciągu geometrycznego o stałym ilorazie równym α i pierwszym wyrazie równym 1:

$$f(d_r, s_b^{(i)}) \leq (1 - \rho_s) \cdot \beta \cdot \frac{1 - \alpha^{\lfloor m_r/m \rfloor}}{1 - \alpha} + \delta \cdot (1 - \rho_s) \cdot \frac{1 - \alpha^{\lfloor m_r/m \rfloor}}{1 - \alpha} \quad (5.3.72)$$

Z drugiej strony w dowodzie twierdzenia 5.3.2 pokazano, że jeśli spełnione jest 5.3.61 to wartość funkcji f dla dokumentu nierelevantnego d_n spełnia nierówność:

$$f(d_n, s_b^{(i)}) \leq 0,5 \cdot (\beta + \delta) \cdot \alpha + \frac{1}{1 - \alpha} \cdot (\beta \cdot nv_{n \max} + \delta \cdot sim_{n \max}) \quad (5.3.73)$$

Korzystając z 5.3.59, 5.3.64 5.3.72 oraz 5.3.73 można stwierdzić, że do udowodnienia twierdzenia wystarczy pokazać, że: spełnione jest 5.3.74 oraz 5.3.75:

$$\begin{aligned} \frac{1 - \alpha^i}{1 - \alpha} \cdot (\beta \cdot nv_{r \min} + \delta \cdot sim_{r \min}) &> (1 - \rho_s^{\lfloor m_r/m \rfloor - 1}) \cdot \beta \cdot \frac{1 - \alpha^{\lfloor m_r/m \rfloor}}{1 - \alpha} + \\ &+ \delta \cdot (1 - \rho_s^{\lfloor m_r/m \rfloor - 1}) \cdot \frac{1 - \alpha^{\lfloor m_r/m \rfloor}}{1 - \alpha} \end{aligned} \quad (5.3.74)$$

$$\begin{aligned} \frac{1 - \alpha^i}{1 - \alpha} \cdot (\beta \cdot nv_{r \min} + \delta \cdot sim_{r \min}) &> \\ 0,5 \cdot (\beta + \delta) \cdot \alpha + \frac{1}{1 - \alpha} \cdot (\beta \cdot nv_{n \max} + \delta \cdot sim_{n \max}) \end{aligned} \quad (5.3.75)$$

Łatwo zauważyć, że dla $i \geq 2$ prawdziwa jest następująca nierówność:

$$\begin{aligned} \left(\frac{1 - \alpha^i}{1 - \alpha} \right) \cdot (\beta \cdot nv_{r \min} + \delta \cdot sim_{r \min}) &\geq (\beta \cdot nv_{r \min} + \delta \cdot sim_{r \min}) > \\ \left(\frac{1 - \alpha^2}{1 - \alpha} \right) \cdot (\beta \cdot nv_{r \min} + \delta \cdot sim_{r \min}) \end{aligned} \quad (5.3.76)$$

Nierówność 5.3.74 musi być spełniona dla $i \geq 2$, gdyż w pierwszym kroku zbiór dokumentów rekomendowanych w bieżącej sesji jest pusty, co biorąc pod uwagę 5.3.76 pozwala stwierdzić prawdziwość następującej nierówności:

$$\begin{aligned} \left(\frac{1-\alpha^2}{1-\alpha}\right) \cdot (\beta \cdot nv_{r\min} + \delta \cdot sim_{r\min}) &> (1-\rho_s^{[m_r/m]-1}) \cdot \beta \cdot \frac{1-\alpha^{[m_r/m]}}{1-\alpha} + \\ \delta \cdot (1-\rho_s^{[m_r/m]-1}) \cdot \frac{1-\alpha^{[m_r/m]}}{1-\alpha} & \end{aligned} \quad (5.3.77)$$

Dzieląc obie strony przez zawsze dodatnie $1+\alpha$ otrzymujemy:

$$\begin{aligned} \beta \cdot nv_{r\min} + \delta \cdot sim_{r\min} &> (1-\rho_s^{[m_r/m]-1}) \cdot \beta \cdot \frac{1-\alpha^{[m_r/m]}}{1-\alpha^2} + \\ + \delta \cdot (1-\rho_s^{[m_r/m]-1}) \cdot \frac{1-\alpha^{[m_r/m]}}{1-\alpha^2} & \end{aligned} \quad (5.3.78)$$

Podobnie biorąc pod uwagę 5.3.76 należy stwierdzić, że prawdziwe jest:

$$\begin{aligned} \beta \cdot nv_{r\min} + \delta \cdot sim_{r\min} &> \\ 0,5 \cdot (\beta + \delta) \cdot \alpha + \frac{1}{1-\alpha} \cdot (\beta \cdot nv_{n\max} + \delta \cdot sim_{n\max}) & \end{aligned} \quad (5.3.79)$$

5.3.78 i 5.3.79 są równe przekształceniom twierdzenia podanym w 5.3.56 i 5.3.58, co dowodzi prawdziwości twierdzenia. □

Interpretacja twierdzenie 5.3.6 jest następująca: jeśli wszystkie dokumenty relewantne mają dostatecznie wysokie współczynniki wzorców nawigacyjnych oraz odznaczają się odpowiednio dużym podobieństwem do dokumentów odwiedzonych, wtedy już odwiedzone dokumenty relewantne będą osiągały mniejsze wartości funkcji rekomendacji (gdyż będzie je osłabiał parametr ρ_s) i w ten sposób różne dokumenty relewantne będą odwiedzane w kolejnych krokach. Jednocześnie odpowiednia duża różnica między $nv_{n\max}$ i $nv_{r\min}$ oraz $sim_{n\max}$ i $sim_{r\min}$ spowoduje, że dokumenty nierelatywne nie będą w tej sesji rekomendowane.

Należy zauważyć, że podobnie jak w przypadku twierdzenia 5.3.2 parametry α , m oraz ρ_s , a także liczba dokumentów relewantnych m determinują, jaka musi być różnica między $nv_{r\min}$ i $nv_{r\max}$ oraz między $sim_{r\min}$ i $sim_{n\max}$, aby system rekomendujący nie podpowiadał nierelawantnych dokumentów (pierwsza i druga nierówność z twierdzenia). Dodatkowo nierówności trzecia i czwarta określają na ile wysokie muszą być $nv_{r\min}$ oraz $sim_{r\min}$, aby dokument już odwiedzony w tej sesji cechował się niższą wartością funkcji rekomendacji od wartości funkcji rekomendacji dla dokumentu relewantnego jeszcze w tej sesji nierekomendowanego.

Przykładowo, dla wartości parametrów zaproponowanych w tabeli 4.5.1 ($\alpha=0,25$, $\beta=1$, $\delta=1$, $\gamma=0,9$, $\rho_s=0,75$, $m=2$) i przy założeniu, że istnieje $m_r=30$ dokumentów relewantnych do potrzeb użytkownika oraz że $nv_{r\max}=0,1$ oraz $sim_{n\max}=0,1$, do osiągnięcia maksymalnej kompletności wystarczy, aby minimalne podobieństwo dokumentów relewantnych do dokumentów odwiedzonych w bieżącej sesji oraz minimalna wartość współczynnika wzorców nawigacyjnych nie były mniejsze niż 0,27. Tak dobrane wartości parametrów nie wydają się być trudnymi do osiągnięcia w praktyce.

Podobnie jak w przypadku twierdzenia 5.3.2 drastyczne obniżenie prawych stron nierówności zawartych w twierdzeniu 5.3.6 nie zawsze jest korzystne, gdyż może spowodować zarekomendowanie dokumentów nierelevantnych. Analogicznie zbyt wysokie wartości wyżej wspomnianych współczynników spowoduje obniżenie się kompletności.

Wpływ zmiany wyrażeń podanych w twierdzeniu 5.3.6 na kompletność przedstawiony został w następującym wniosku (jest to wniosek analogiczny do wniosków 5.3.1 i 5.3.5).

Wniosek 5.3.11

Jeśli w każdym kroku ($i=2, \dots, n_{sb}$) bieżącej sesji s_b , użytkownik odwiedza jeden z dokumentów rekomendowanych w kroku poprzednim ($d^{(i)} \in R^{(i-1)}$) i system hipertekstowy zawiera m_r ($m_r > m$) dokumentów relevantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{sb})}$ to spadek wartości następujących wyrażeń: $0,5 \cdot \alpha + \frac{nv_{n_{\max}}}{1-\alpha}$, $0,5 \cdot \alpha + \frac{sim_{n_{\max}}}{1-\alpha}$, $\frac{(1-\rho_s) \cdot (1-\alpha^{[m_r/m]})}{1-\alpha^2}$ oraz $\frac{(1-\rho_s) \cdot (1-\alpha^{[m_r/m]})}{1-\alpha^2}$ będzie powodował wzrost kompletności metody rekomendacji *AdaptRank*.

Wniosek 5.3.11 jest bardzo ważny, gdyż pozwala przeanalizować jak będzie zmieniać się kompletność metody rekomendacji w zależności od doboru parametrów. Poniższe 3 wnioski wraz twierdzeniem 5.3.7 przedstawiają wpływ parametrów α , ρ_s , m oraz m_r na kompletność rekomendacji, przy spełnionych założeniach wniosku 5.3.11.

Parametr $0 < \rho_s < 1$ znajduje się w licznikach wyrażeń z wniosku 5.3.11. Zważywszy na fakt, że im większa jest wartość $1 - \rho_s^{[m_r/m]}$, tym większe są wyżej wspomniane wyrażenia, można sformułować następujący wniosek.

Wniosek 5.3.12

Jeśli spełnione są warunki wniosku 5.3.11, to im większa jest wartość parametru ρ_s tym większa jest kompletność metody rekomendacji *AdaptRank*.

Wniosek 5.3.12 jest zgodny z intuicją – gdyż parametr ρ_s określa jak „mocno” fakt odwiedzenia dokumentu wpływa negatywnie na wartość jego funkcji rekomendacji w kolejnych krokach. Jeśli odwiedzenie danego dokumentu wiąże się z drastycznym spadkiem jego funkcji rekomendacji to łatwo dochodzi do sytuacji, w której wartość funkcji rekomendacji dla dokumentów relevantnych nieodwiedzonych w bieżącej sesji przewyższy wartość funkcji rekomendacji dla dokumentów odwiedzonych w bieżącej sesji.

Łatwo zauważyć, że wzrost wartości wyrażenia $1 - \alpha^{[m_r/m]}$ będzie powodował wzrost wyrażeń z wniosku 5.3.11, co pozwala na sformułowanie następujących dwóch wniosków.

Wniosek 5.3.13

Jeśli spełnione są warunki wniosku 5.3.11, to im większa jest liczba m_r dokumentów relewantnych do potrzeb użytkownika, tym mniejsza jest kompletność metody rekomendacji *AdaptRank*.

Także ten wniosek jest zgodny z intuicją, gdyż kluczem do osiągnięcia maksymalnej kompletności jest czasowe „osłabienie” wartości odwiedzonych dokumentów relewantnych. Jeśli zbiór dokumentów relewantnych jest zbyt duży odwiedzone dokumenty przestają być „osłabiane” przez parametr ρ_s a co za tym idzie nie jest możliwe zarekomendowanie wszystkich nierekomendowanych dokumentów relewantnych w $[m_r/m]$ pierwszych krokach bieżącej sesji.

Wniosek 5.3.14

Jeśli spełnione są warunki wniosku 5.3.11, to im większa jest wartość parametru m (określającego jak dużo dokumentów jest rekomendowanych w pojedynczym kroku bieżącej sesji) tym większa jest kompletność metody rekomendacji *AdaptRank*.

Wzrost wartości parametru m powoduje wzrost kompletności, gdyż zarekomendowanie dużej liczby dokumentów w jednym kroku zmniejsza liczbę kroków potrzebnych do zarekomendowania wszystkich dokumentów relewantnych. Zmniejszenie liczby kroków jest korzystne, gdyż z każdym odwiedzionymi krokami, fakt odwiedzenia dokumentu wpływa coraz słabiej na wartość funkcji f dla dokumentów odwiedzonych na początku bieżącej sesji, co powoduje, że wartości funkcji dla dokumentów zarekomendowanych na początku bieżącej sesji będą coraz większe i wypowiedzenie nierekomendowanego dokumentu relewantnego będzie coraz trudniejsze. Z drugiej jednak strony należy zauważyć, że duża liczba dokumentów podpowiadana w jednym kroku może powodować dezorientację użytkownika.

Twierdzenie 5.3.7

Jeśli spełnione są warunki wniosku 5.3.11 i $[m_r/m] > 2$ to im większa jest wartość parametru α , tym mniejsza jest kompletność metody rekomendacji *AdaptRank*.

Dowód:

Pokażemy, że funkcja $rec(\alpha) = \frac{(1 - \rho_s) \cdot (1 - \alpha^{[m_r/m]})}{1 - \alpha^2}$ gdzie $\alpha \in (0, 1)$ jest rosnąca w całej dziedzinie. W tym celu obliczona zostanie pierwsza pochodna funkcji $rec(\alpha)$:

$$rec'(\alpha) = (1 - \rho_s) \cdot \left(\frac{(1 - \alpha^{[m_r/m]})}{1 - \alpha^2} \right)' \quad (5.3.80)$$

Po zastosowaniu wzoru na pochodną ilorazu otrzymujemy:

$$rec'(\alpha) = (1 - \rho_s) \cdot \frac{(1 - \alpha^{[m_r/m]}) \cdot (1 - \alpha^2) - (1 - \alpha^{[m_r/m]}) \cdot (1 - \alpha^2)'}{(1 - \alpha^2)^2} \quad (5.3.81)$$

Funkcja $rec(\alpha)$ będzie rosnąca, gdy pierwsza pochodna będzie dodatnia, czyli:

$$(1 - \rho_s) \cdot \frac{(1 - \alpha^{[m_r/m]})' \cdot (1 - \alpha^2) - (1 - \alpha^{[m_r/m]}) \cdot (1 - \alpha^2)'}{(1 - \alpha^2)^2} > 0 \quad (5.3.82)$$

Mnożąc obie strony przez zawsze dodatnie $\frac{(1 - \alpha^2)^2}{1 - \rho_s}$ otrzymujemy:

$$(1 - \alpha^{[m_r/m]})' \cdot (1 - \alpha^2) - (1 - \alpha^{[m_r/m]}) \cdot (1 - \alpha^2)' > 0 \quad (5.3.83)$$

Po obliczeniu pochodnych cząstkowych otrzymujemy:

$$-[m_r/m] \cdot \alpha^{[m_r/m]-1} \cdot (1 - \alpha^2) + (1 - \alpha^{[m_r/m]}) \cdot 2\alpha > 0 \quad (5.3.84)$$

Dla łatwiejszego dokonania dalszych obliczeń i większej przejrzystości podstawmy $x=m_r/m$, $x \in N$ oraz $x > 2$

$$-x \cdot \alpha^{x-1} \cdot (1 - \alpha^2) + (1 - \alpha^x) \cdot 2\alpha > 0 \quad (5.3.85)$$

Po dokonaniu operacji mnożenia mamy:

$$-x \cdot \alpha^{x-1} + x \cdot \alpha^{x+1} + 2 \cdot \alpha - 2 \cdot \alpha^{x+1} > 0 \quad (5.3.86)$$

Dzieląc obie strony przez $\alpha > 0$:

$$-x \cdot \alpha^{x-2} + x \cdot \alpha^x + 2 - 2 \cdot \alpha^x > 0 \quad (5.3.87)$$

Podstawiając pod lewą stronę nierówności 5.3.72, $x=2$ otrzymujemy:

$$-2 + 2 \cdot \alpha^2 + 2 - 2 \cdot \alpha^2 = 0 \quad (5.3.88)$$

Żeby udowodnić, że 5.3.87 jest prawdziwe dla $x > 2$ wystarczy pokazać, że ciąg $c_x = -x \cdot \alpha^{x-2} + x \cdot \alpha^x + 2 - 2 \cdot \alpha^x$ jest rosnący, czyli że

$$c_{x+1} - c_x > 0 \quad (5.3.89)$$

Co daje nam:

$$-(x+1) \cdot \alpha^{x-1} + (x+1) \cdot \alpha^{x+1} + 2 - 2 \cdot \alpha^{x+1} + x \cdot \alpha^{x-2} - x \cdot \alpha^x - 2 + 2 \cdot \alpha^x > 0 \quad (5.3.90)$$

Po uporządkowaniu 5.3.90 otrzymujemy:

$$(x-1) \cdot \alpha^{x+1} - (x+1) \cdot \alpha^{x-1} - (x-2) \cdot \alpha^x + x \cdot \alpha^{x-2} > 0 \quad (5.3.91)$$

Po podzieleniu obu stron przez zawsze dodatnie α^{x-2} :

$$(x-1)\alpha^3 - (x+1)\alpha - (x-2)\alpha^2 + x > 0 \quad (5.3.92)$$

Podstawiając pod lewą stronę nierówności 5.3.92 $\alpha=1$ otrzymujemy

$$x-1-x-1-x+2+x=0 \quad (5.3.93)$$

Co pozwala zapisać 5.3.92 jako iloczyn następujących wielomianów:

$$((x-1)\alpha^2 + \alpha - x)(\alpha - 1) > 0 \quad (5.3.94)$$

Sprowadźmy do postaci iloczynu następujące wyrażenie $(x-1)\alpha^2 + \alpha - x$. W tym celu policzmy wyznacznik równania kwadratowego Δ :

$$\Delta = 1 + 4(x-1)x \quad (5.3.95)$$

I dalej:

$$\Delta = 1 - 4x + 4x^2 = (1-2x)^2 \quad (5.3.96)$$

Co pozwala zapisać 5.3.94 jako:

$$\left(\alpha - \frac{-1 + |1-2x|}{2(x-1)} \right) (\alpha - 1) \left(\alpha - \frac{-1 - |1-2x|}{2(x-1)} \right) > 0 \quad (5.3.97)$$

Dla $x > 0$ ($x \in \mathbb{N}$) $|1-2x|=2x-1$

$$\left(\alpha - \frac{-1 + 2x - 1}{2(x-1)} \right) (\alpha - 1) \left(\alpha - \frac{-1 - 2x + 1}{2(x-1)} \right) > 0 \quad (5.3.98)$$

Co ostatecznie daje:

$$(\alpha - 1)(\alpha - 1) \left(\alpha - \frac{-2x}{2(x-1)} \right) > 0 \quad (5.3.99)$$

Tak więc 5.3.92 będzie prawdziwe gdy:

$$\alpha > \frac{-2x}{2(x-1)} \quad (5.3.100)$$

Dla $x > 0$ prawa strona jest ujemna tak więc 5.3.92 jest prawdziwe dla każdego $0 < \alpha < 1$ a w konsekwencji ciąg c_n jest rosnący, co pozwala stwierdzić, że funkcja $rec(\alpha)$ jest rosnąca w całej dziedzinie. Dodatkowo w dowodzie twierdzenia 5.3.3 pokazano, że funkcje $prec_{sim}(\alpha) = \frac{0,5 \cdot \alpha}{1 - \rho_s^{[m_r/m]}} + \frac{sim_{n \max}}{(1 - \rho_s^{[m_r/m]}) \cdot (1 - \alpha)}$ i $prec_{nv}(\alpha) = \frac{0,5 \cdot \alpha}{1 - \rho_s^{[m_r/m]}} + \frac{nv_{n \max}}{(1 - \rho_s) \cdot (1 - \alpha)}$ są rosnące dla każdego $0 < \alpha < 1$. co oznacza, że rosnące będą także funkcje $rec_{sim}(\alpha) = 0,5 \cdot \alpha + \frac{sim_{n \max}}{1 - \alpha}$ oraz $rec_{nv}(\alpha) = 0,5 \cdot \alpha + \frac{nv_{n \max}}{1 - \alpha}$. Co dowodzi prawdziwości twierdzenia. □

Kolejne twierdzenie pokazuje, że metoda *AdaptRank* pozwala także osiągnąć maksymalną kompletność, gdy użytkownik ignoruje wszystkie dokumenty rekomendowane przez system.

Twierdzenie 5.3.8

Jeśli w każdym kroku i ($i=2, \dots, n_{sb}$) bieżącej sesji s_b , użytkownik odwiedza dokument nierekomendowany w kroku poprzednim ($d^{(i)} \notin R^{(i-1)}$) i system hipertekstowy zawiera m_r ($m_r > m$) dokumentów relewantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{sb})}$ oraz w $[m_r/m]$ pierwszych krokach bieżącej sesji spełnione są następujące warunki:

$$nv_{r \min} > \frac{\alpha^{[m_r/m]} \cdot 0,5}{1 + \alpha} + \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} - \frac{0,5 \cdot \gamma^{[m_r/m]-2}}{\beta \cdot (1 + \alpha)}$$

$$sim_{r \min} > \frac{\alpha^{[m_r/m]} \cdot 0,5}{1 + \alpha} + \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} - \frac{0,5 \cdot \gamma^{[m_r/m]-2}}{\delta \cdot (1 + \alpha)}$$

$$nv_{r \min} > 0,5 \cdot \alpha + \frac{nv_{n \max}}{1 - \alpha}$$

$$sim_{r \min} > 0,5 \cdot \alpha + \frac{sim_{n \max}}{1 - \alpha}$$

to kompletność systemu rekomendującego w bieżącej sesji s_b będzie równa 1 niezależnie od liczby dokumentów odwiedzonych w tej sesji.

Dowód:

Na początek przekształćmy pierwszą i drugą nierówność podaną w twierdzeniu. Obie strony pierwszej nierówności zostaną pomnożone przez β , a drugiej przez δ :

$$\begin{aligned} \beta \cdot nv_{r \min} &> \frac{\alpha^{[m_r/m]} \cdot 0,5 \cdot \beta - 0,5 \cdot \gamma^{[m_r/m]-2}}{1 + \alpha} + \beta \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} \\ \delta \cdot sim_{r \min} &> \frac{\alpha^{[m_r/m]} \cdot 0,5 \cdot \delta - 0,5 \cdot \gamma^{[m_r/m]-2}}{1 + \alpha} + \delta \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} \end{aligned} \quad (5.3.101)$$

Dodając obie nierówności stronami otrzymujemy:

$$\begin{aligned} \beta \cdot nv_{r \min} + \delta \cdot sim_{r \min} &> \\ \frac{\alpha^{[m_r/m]} \cdot 0,5 \cdot (\beta + \delta) - \gamma^{[m_r/m]-2}}{1 + \alpha} + (\beta + \delta) \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} \end{aligned} \quad (5.3.102)$$

Z twierdzenia 5.3.5 wiadomo, że aby kompletność była maksymalna to w każdym kroku $i > 0$ bieżącej sesji, w której zbiór $D_{RN}^{(i)}$ jest niepusty, co najmniej m_{aux} dokumentów należących do zbioru $D_{RN}^{(i)}$ musi osiągać wartości funkcji rekomendacji wyższe od wartości funkcji rekomendacji dla dokumentów nie należących do zbioru $D_{RN}^{(i)}$. Łatwo zauważyć, że jeśli w każdym kroku do kroku $[m_r/m]$ włącznie rekomendowano różne dokumenty relewantne to kompletność w tej sesji będzie maksymalna i już nie zmniejszy się niezależnie od długości bieżącej sesji, pod warunkiem, że zbiór dokumentów relewantnych będzie taki sam we wszystkich krokach bieżącej sesji (co zgodne jest z założeniami twierdzenia).

To znaczy, że do udowodnienia twierdzenia wystarczy pokazać, że dla każdego zignorowanego dokumentu relewantnego $d_r \in D_R^{(i)} \setminus D_{RN}^{(i)}$ i dla każdego dokumentu relewantnego nierekomendowanego w bieżącej sesji $d_{rn} \in D_{RN}^{(i)}$ oraz dla każdego nierekomendowanego dokumentu $d_n \in D \setminus D_R^{(i)}$ zajdzie:

$$\forall_{[m_r/m] \geq i > 0} f(d_m, s_b^{(i)}) > f(d_r, s_b^{(i)}) \wedge f(d_m, s_b^{(i)}) > f(d_n, s_b^{(i)}) \quad (5.3.103)$$

Dokumenty należące do zbioru $D_{RN}^{(i)}$ osiągną wartości minimalne funkcji rekomendacji zgodne z założeniami twierdzenia pod warunkiem, że:

$$\forall_{[m_r/m] \geq i > 0} \forall_{d_m \in D_{RN}^{(i)}} (q(d_m) = 0 \wedge sim(d_m, d^{(i)}) = sim_{r \min} \wedge nv_{d_m}^{(i)} = nv_{r \min}) \quad (5.3.104)$$

Należy zauważyć, że warunek, dla którego zachodzi $f(d_m, s_b^{(i)}) > f(d_n, s_b^{(i)})$ został określony w dowodzie twierdzenia: 5.3.6 jest to

$$\begin{aligned} nv_{r \min} &> 0,5 \cdot \alpha + \frac{nv_{n \max}}{1 - \alpha} \\ sim_{r \min} &> 0,5 \cdot \alpha + \frac{sim_{n \max}}{1 - \alpha} \end{aligned}$$

W dowodzie niniejszego twierdzenia określić należy jedynie warunki, dla których:

$$\forall_{[m_r/m] \geq i > 0} f(d_m, s_b^{(i)}) > f(d_r, s_b^{(i)}) \quad (5.3.105)$$

Po pierwsze należy zauważyć, że dla $i=1$ wyrażenie 5.3.105 jest zawsze prawdziwe bo żaden z dokumentów nie był jeszcze rekomendowany co oznacza, że do udowodnienia 5.3.104 wystarczy udowodnić:

$$\forall_{[m_r/m] \geq i > 1} f(d_m, s_b^{(i)}) > f(d_r, s_b^{(i)}) \quad (5.3.106)$$

Rozpatrzmy najbardziej pesymistyczny przypadek, w którym wszystkie dokumenty relewantne, które zostały już zarekomendowane w bieżącej sesji cechują się maksymalnymi wartościami odpowiednich współczynników. Oznacza to, że spełniona jest następująca formuła:

$$\forall_{[m_r/m] \geq i > 0} \forall_{d_r \in D_R^{(i)}} (q(d_r) = 1 \wedge sim(d_r, d^{(i)}) = 1 \wedge nv_{d_r}^{(i)} = 1) \quad (5.3.107)$$

Jeśli 5.3.107 jest prawdziwe, to korzystając z twierdzenia 5.1.6 możemy stwierdzić, że wartość funkcji rekomendacji w kroku i dla relewantnych dokumentów zarekomendowanych w kroku i' ($i' < i$) (i zignorowanych) spełnia nierówność:

$$f(d_r, s_b^{(i)}) \leq \alpha^i \cdot 0,5 \cdot (\beta + \delta) + \beta \cdot \sum_{k=1}^i (\alpha^{i-k}) + \delta \cdot \sum_{k=1}^i (\alpha^{i-k}) - \gamma^{i-i-1} \quad (5.3.108)$$

Łatwo zauważyć, że im i' jest większe (im wyższy jest numer kroku, w którym dokument d_r był zarekomendowany) tym mniejsza jest wartość funkcji rekomendacji, gdyż $1 \geq \gamma > 0$. Pozwala to zapisać:

$$f(d_r, s_b^{(i)}) \leq \alpha^i \cdot 0,5 \cdot (\beta + \delta) + \beta \cdot \sum_{k=1}^i (\alpha^{i-k}) + \delta \cdot \sum_{k=1}^i (\alpha^{i-k}) - \gamma^{i-2} \quad (5.3.109)$$

$\sum_{k=1}^i (\alpha^{i-k}) = \sum_{k=1}^i (\alpha^{k-1})$ jest równe sumie i wyrazów ciągu geometrycznego o stałym ilorazie równym α i pierwszym wyrazie równym 1:

$$f(d_r, s_b^{(i)}) \leq \alpha^i \cdot 0,5 \cdot (\beta + \delta) + \beta \cdot \frac{1 - \alpha^i}{1 - \alpha} + \delta \cdot \frac{1 - \alpha^i}{1 - \alpha} - \gamma^{i-2} \quad (5.3.110)$$

Rozpatrzmy funkcję $f_{prec}(i) = \alpha^i \cdot 0,5 \cdot (\beta + \delta) + \beta \cdot \frac{1 - \alpha^i}{1 - \alpha} + \delta \cdot \frac{1 - \alpha^i}{1 - \alpha} - \gamma^{i-2}$. Do zbadania monotoniczności tej funkcji policzona zostanie jej pierwsza pochodna.

$$f'_{prec}(i) = 0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot \ln \alpha + \beta \cdot \frac{-\alpha^i \cdot \ln \alpha}{1 - \alpha} + \delta \cdot \frac{-\alpha^i \cdot \ln \alpha}{1 - \alpha} - \gamma^{i-2} \cdot \ln \gamma \quad (5.3.111)$$

Co pozwala otrzymać:

$$f'_{prec}(i) = 0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot \ln \alpha - (\beta + \delta) \frac{\alpha^i \cdot \ln \alpha}{1 - \alpha} - \gamma^{i-2} \cdot \ln \gamma \quad (5.3.112)$$

Wyłączając $(\beta + \delta) \cdot \alpha^i \cdot \ln \alpha$ z różnicy dwóch pierwszych wyrażeń otrzymujemy:

$$f'_{prec}(i) = 0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot \ln \alpha \cdot \left(1 - \frac{1}{1 - \alpha}\right) - \gamma^{i-2} \cdot \ln \gamma \quad (5.3.113)$$

i ostatecznie:

$$f'_{prec}(i) = 0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot \ln \alpha \cdot \left(\frac{-\alpha}{1 - \alpha}\right) - \gamma^{i-2} \cdot \ln \gamma \quad (5.3.114)$$

Biorąc pod uwagę, że dla $1 > \alpha > 0$: $\ln \alpha < 0$ oraz $\frac{-\alpha}{1 - \alpha} < 0$ można stwierdzić, że:

$$0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot \ln \alpha \cdot \left(\frac{-\alpha}{1 - \alpha}\right) > 0 \quad (5.3.115)$$

Podobnie, biorąc pod uwagę, że dla $1 > \gamma > 0$ jest spełnione $\ln \gamma < 0$ jest prawdą, że dla $i > 0$:

$$0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot \ln \alpha \cdot \left(\frac{-\alpha}{1 - \alpha}\right) - \gamma^{i-2} \cdot \ln \gamma > 0 \quad (5.3.116)$$

Funkcja $f_{prec}(i)$ jest, więc rosnąca w całej dziedzinie co oznacza, że jeśli $[m_r/m] \geq i > 0$ to największa wartość zostanie osiągnięta dla $i = [m_r/m]$. Co dodatkowo biorąc pod uwagę 5.3.110 i przechodność relacji większości pozwala zapisać:

$$f(d_r, s_b^{(i)}) \leq \alpha^{[m_r/m]} \cdot 0,5 \cdot (\beta + \delta) + \beta \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha} + \delta \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha} - \gamma^{[m_r/m]-2} \quad (5.3.117)$$

Z drugiej strony w twierdzeniu 5.3.6 pokazano, że dla $i > 1$ minimalna wartość funkcji rekomendacji dla relewantnego dokumentu nierekomendowanego w bieżącej sesji i spełniającego warunki zawarte w 5.3.104 jest równa:

$$f(d_m, s_b^{(i)}) \geq \beta \cdot (1 + \alpha) \cdot nv_{r \min} + \delta \cdot (1 + \alpha) \cdot sim_{r \min} \quad (5.3.118)$$

Podstawiając 5.3.117, i 5.3.118 do 5.3.106 otrzymujemy:

$$\beta \cdot (1 + \alpha) \cdot nv_{r_{\min}} + \delta \cdot (1 + \alpha) \cdot sim_{r_{\min}} > \quad (5.3.119)$$

$$\alpha^{[m_r/m]} \cdot 0,5 \cdot (\beta + \delta) + \beta \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha} + \delta \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha} - \gamma^{[m_r/m]-2}$$

Dzieląc obie strony przez zawsze dodatnie $1 + \alpha$ otrzymujemy:

$$\beta nv_{r_{\min}} + \delta \cdot sim_{r_{\min}} > \quad (5.3.120)$$

$$\frac{\alpha^{[m_r/m]} \cdot 0,5 \cdot (\beta + \delta) - \gamma^{[m_r/m]-2}}{1 + \alpha} + (\beta + \delta) \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2}$$

5.3.120 jest przekształceniem założeń twierdzenia zawartym w 5.3.102, co dowodzi prawdziwości twierdzenia. □

Interpretacja twierdzenie 5.3.8 jest analogiczna do interpretacji twierdzenia 5.3.6: jeśli wszystkie dokumenty relewantne mają dostatecznie wysokie współczynniki wzorców nawigacyjnych oraz odznaczają się odpowiednio dużym podobieństwem do dokumentów odwiedzonych w bieżącej sesji, wtedy już zarekomendowane i zignorowane dokumenty relewantne będą osiągały mniejsze wartości funkcji rekomendacji i w ten sposób nierekomendowane dokumenty relewantne będą rekomendowane w kolejnych krokach, a kompletność rekomendacji osiągnie wartość maksymalną.

Należy zauważyć, że parametry α , m , β , δ oraz γ a także liczba dokumentów relewantnych m_r determinują, jaka musi być różnica między $nv_{r_{\min}}$ i $nv_{r_{\max}}$ oraz między $sim_{r_{\min}}$ i $sim_{r_{\max}}$, aby system rekomendujący nie podpowiadał nierelewantnych dokumentów (trzecia i czwarta nierówność z twierdzenia). Dodatkowo nierówności pierwsza i druga określają na ile wysokie muszą być $nv_{r_{\min}}$ oraz $sim_{r_{\min}}$, aby raz zarekomendowany dokument cechował się niższą wartością funkcji rekomendacji od wartości funkcji rekomendacji dla relewantnego dokumentu, który nie był jeszcze rekomendowany.

Przykładowo, dla wartości parametrów zaproponowanych w tabeli 4.5.1 ($\alpha=0,25$, $\beta=1$, $\delta=1$, $\gamma=0,9$, $\rho_s=0,75$, $m=2$) i przy założeniu, że istnieje $m_r=30$ dokumentów relewantnych do potrzeb użytkownika oraz że $nv_{r_{\max}}=0,1$ oraz $sim_{r_{\max}}=0,1$ do osiągnięcia maksymalnej kompletności wystarczy, aby minimalne podobieństwo dokumentów relewantnych do dokumentów odwiedzonych w bieżącej sesji oraz minimalna wartość współczynnika wzorców nawigacyjnych nie były mniejsze niż 0,87.

Podobnie jak w przypadku twierdzeń 5.3.2, 5.3.6 drastyczne obniżenie prawych stron nierówności zawartych w twierdzeniu 5.3.8 nie zawsze jest korzystne, gdyż może spowodować zarekomendowanie dokumentów nierelewantnych. Analogicznie zbyt wysokie wartości wyżej wspomnianych współczynników spowoduje obniżenie się kompletności.

Wpływ zmiany wyrażeń podanych w twierdzeniu 5.3.8 na kompletność przedstawiony został w następującym wniosku (jest to wniosek analogiczny do wniosków 5.3.1, 5.3.5 oraz 5.3.11):

Wniosek 5.3.15

Jeśli w każdym kroku i ($i=2, \dots, n_{sb}$) bieżącej sesji s_b , użytkownik odwiedza dokument nierekomendowany w kroku poprzednim ($d^{(i)} \notin R^{(i-1)}$) i system hipertekstowy zawiera m_r ($m_r > m$) dokumentów relewantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{sb})}$, to spadek wartości następujących wyrażeń:

$$0,5 \cdot \alpha + \frac{nv_{n_{\max}}}{1-\alpha}, \quad 0,5 \cdot \alpha + \frac{sim_{n_{\max}}}{1-\alpha}, \quad \frac{\alpha^{[m_r/m]} \cdot 0,5}{1+\alpha} + \frac{1-\alpha^{[m_r/m]}}{1-\alpha^2} - \frac{0,5 \cdot \gamma^{[m_r/m]-2}}{\beta \cdot (1+\alpha)} \quad \text{oraz}$$

$$\frac{\alpha^{[m_r/m]} \cdot 0,5}{1+\alpha} + \frac{1-\alpha^{[m_r/m]}}{1-\alpha^2} - \frac{0,5 \cdot \gamma^{[m_r/m]-2}}{\delta \cdot (1+\alpha)}$$

będzie powodował wzrost kompletności metody rekomendacji *AdaptRank*.

Wniosek 5.3.15 pozwala przeanalizować jak będzie zmieniać się kompletność metody rekomendacji w zależności od doboru parametrów. Poniższe 3 wnioski wraz z twierdzeniem 5.3.9 przedstawiają wpływ parametrów α , γ , m oraz m_r na kompletność rekomendacji przy spełnionych założeniach warunku 5.3.15.

Parametr $0 < \alpha < \gamma < 1$ znajduje się w ujemnych licznikach wyrażeń z wniosku 5.3.15, co pozwala sformułować następujący wniosek:

Wniosek 5.3.16

Jeśli spełnione są warunki wniosku 5.3.15, to im większa jest wartość parametru γ tym większa jest kompletność metody rekomendacji *AdaptRank*.

Wniosek 5.3.16 jest zgodny z intuicją – gdyż parametr γ określa jak „mocno” fakt zignorowania danego dokumentu wpływa negatywnie na wartość jego funkcji rekomendacji w kolejnych krokach. Jeśli zignorowanie danego dokumentu wiąże się z drastycznym spadkiem wartości jego funkcji rekomendacji to łatwo dochodzi do sytuacji, w której wartość funkcji rekomendacji dla dokumentów relewantnych nierekomendowanych w bieżącej sesji przewyższy wartość funkcji rekomendacji dla dokumentów relewantnych zignorowanych w bieżącej sesji.

Parametry $1 \geq \beta > 0$ oraz $1 \geq \delta > 0$ znajdują się w ujemnych mianownikach wyrażeń z wniosku 5.3.12 co pozwala na sformułowanie następującego wniosku:

Wniosek 5.3.17

Jeśli spełnione są warunki wniosku 5.3.15, to im większa jest wartość parametrów β oraz δ tym mniejsza jest kompletność metody rekomendacji *AdaptRank*.

Duża wartość parametrów β oraz δ powoduje, wzrost wartości funkcji rekomendacji dla wszystkich dokumentów (patrz definicja 4.5.3) wskutek czego spadek wartości funkcji dla dokumentów zignorowanych nie będzie tak znaczny.

Twierdzenie 5.3.9

Jeśli spełnione są warunki wniosku 5.3.15, to im większa jest wartość parametru m tym większa jest kompletność metody rekomendacji *AdaptRank*.

Dowód:

W celu udowodnienia twierdzenia rozpatrzmy funkcję:

$$rec_m(x) = \frac{\alpha^x \cdot 0,5}{1 + \alpha} + \frac{1 - \alpha^x}{1 - \alpha^2} - \frac{0,5 \cdot \gamma^{x-2}}{\beta \cdot (1 + \alpha)} \quad (5.3.121)$$

Gdzie $x > 0$. W celu zbadania monotoniczności funkcji rec_m policzymy jej pierwszą pochodną.

$$rec_m'(x) = \frac{0,5 \cdot \alpha^x \cdot \ln \alpha^x}{1 + \alpha} - \frac{\alpha^x \ln \alpha^x}{1 - \alpha^2} - \frac{0,5 \cdot \gamma^{x-2} \cdot \ln \gamma}{\beta \cdot (1 + \alpha)} \quad (5.3.122)$$

Wyłączając z dwóch pierwszych wyrazów $0,5 \cdot \alpha^x \cdot \ln \alpha^x$, otrzymujemy:

$$rec_m'(x) = 0,5 \cdot \alpha^x \cdot \ln(\alpha^x) \cdot \left(\frac{1}{1 + \alpha} - \frac{1}{1 - \alpha^2} \right) - \frac{0,5 \cdot \gamma^{x-2} \cdot \ln \gamma}{\beta \cdot (1 + \alpha)} \quad (5.3.123)$$

I ostatecznie:

$$rec_m'(x) = 0,5 \cdot \alpha^x \cdot \ln(\alpha^x) \cdot \left(\frac{-\alpha}{1 - \alpha^2} \right) - \frac{0,5 \cdot \gamma^{x-2} \cdot \ln \gamma}{\beta \cdot (1 + \alpha)} \quad (5.3.124)$$

Biorąc pod uwagę, że dla $1 > \alpha > 0$ spełnione są nierówności: $\ln(\alpha^x) < 0$ oraz $\frac{-\alpha}{1 - \alpha^2} < 0$ można stwierdzić, że:

$$0,5 \cdot \alpha^x \cdot \ln(\alpha^x) \cdot \left(\frac{-\alpha}{1 - \alpha^2} \right) > 0 \quad (5.3.125)$$

Podobnie biorąc pod uwagę, że dla $1 > \gamma > 0$ prawdziwa jest nierówność: $\ln(\gamma) < 0$, można stwierdzić, że:

$$-\frac{0,5 \cdot \gamma^{x-2} \cdot \ln \gamma}{\beta \cdot (1 + \alpha)} > 0 \quad (5.3.126)$$

Co oznacza, że pierwsza pochodna funkcji rec_m jest dodatnia w całej dziedzinie, więc możemy stwierdzić, że funkcja rec_m jest rosnąca w całej dziedzinie, dzięki czemu możemy stwierdzić, że dla $1 > \alpha > 0$ wraz ze wzrostem m maleje wartość wyrażenia:

$$\frac{\alpha^{[m_r/m]} \cdot 0,5}{1 + \alpha} + \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} - \frac{0,5 \cdot \gamma^{[m_r/m]-2}}{\beta \cdot (1 + \alpha)} \quad (5.3.127)$$

Co biorąc pod uwagę wniosek 5.3.15 oznacza prawdziwość twierdzenia. □

Wzrost wartości parametru m powoduje wzrost kompletności, gdyż zarekomendowanie dużej liczby dokumentów w jednym kroku zmniejsza liczbę kroków potrzebnych do zarekomendowania wszystkich dokumentów relewantnych. Zmniejszenie liczby kroków jest korzystne, gdyż z każdym odwiedzionym krokiem, fakt zignorowania dokumentu w danym kroku wpływa coraz słabiej na wartość funkcji f dla dokumentów zignorowanych na początku bieżącej sesji. Powoduje to, że wartości funkcji rekomendacji dla dokumentów zarekomendowanych na początku bieżącej sesji będą coraz większe i podpowiedzenie nierekomendowanego dokumentu będzie coraz trudniejsze. Z drugiej jednak strony należy zauważyć, że duża liczba dokumentów podpowiadana w jednym kroku może powodować dezorientację użytkownika.

Biorąc pod uwagę dowód twierdzenia 5.3.9, a ściślej mówiąc rosnący charakter funkcji

$$rec_m(x) = \frac{\alpha^x \cdot 0,5}{1 + \alpha} + \frac{1 - \alpha^x}{1 - \alpha^2} - \frac{0,5 \cdot \gamma^{x-2}}{\beta \cdot (1 + \alpha)}$$

można sformułować następujący wniosek:

Wniosek 5.3.18

Jeśli spełnione są warunki wniosku 5.3.15, to im większa jest liczba dokumentów relewantnych m_r tym mniejsza jest kompletność metody rekomendacji *AdaptRank*.

Także ten wniosek jest zgodny z intuicją, gdyż kluczem do osiągnięcia maksymalnej kompletności jest czasowe „osłabienie” wartości odwiedzonych dokumentów relewantnych. Jeśli zbiór dokumentów relewantnych jest zbyt duży odwiedzone dokumenty przestają być „osłabiane”, a co za tym idzie nie jest możliwe zarekomendowanie wszystkich nierekomendowanych dokumentów relewantnych.

Twierdzenie 5.3.10

Jeśli spełnione są warunki wniosku 5.3.15 i $[m_r/m] > 1$, to im większa jest wartość parametru α tym mniejsza jest kompletność metody rekomendacji *AdaptRank*.

Dowód:

Do udowodnienia twierdzenia użyta zostanie funkcja:

$$rec_g(\alpha) = \frac{\alpha^{[m_r/m]} \cdot 0,5}{1 + \alpha} + \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} - \frac{0,5 \cdot \gamma^{[m_r/m]-2}}{\beta \cdot (1 + \alpha)} \quad (5.3.128)$$

gdzie $1 > \alpha > 0$. Do zbadania monotoniczności tej funkcji policzymy jej pierwszą pochodną:

$$\begin{aligned}
rec_g'(\alpha) = & \frac{0,5 \cdot ([m_r / m]) \cdot \alpha^{[m_r / m]-1} \cdot (1 + \alpha) - 0,5 \cdot \alpha^{[m_r / m]}}{(1 + \alpha)^2} - \\
& - \frac{\alpha^{[m_r / m]-1} (1 - \alpha^2) - 2\alpha \cdot (1 - \alpha^{[m_r / m]-1})}{(1 - \alpha^2)^2} + \frac{0,5 \cdot \gamma^{[m_r / m]-2}}{\beta \cdot (1 + \alpha)^2}
\end{aligned} \tag{5.3.129}$$

Po dokonaniu prostych przekształceń otrzymujemy:

$$\begin{aligned}
rec_g'(\alpha) = & \frac{\beta \cdot 0,5 \cdot \alpha^{[m_r / m]} ([m_r / m] (1 + \alpha) - \alpha) + 0,5 \cdot \gamma^{[m_r / m]-2}}{\beta (1 + \alpha)^2} - \\
& - \frac{\alpha^{[m_r / m]} - \alpha^2 + \alpha^{[m_r / m]-1} - 2\alpha}{(1 - \alpha^2)^2}
\end{aligned} \tag{5.3.130}$$

I ostatecznie:

$$\begin{aligned}
rec_g'(\alpha) = & \frac{\beta \cdot 0,5 \cdot \alpha^{[m_r / m]} ([m_r / m] (1 + \alpha) - \alpha) + 0,5 \cdot \gamma^{[m_r / m]-2}}{\beta (1 + \alpha)^2} - \\
& - \frac{\alpha^2 (\alpha^{[m_r / m]-2} - 1) + \alpha (\alpha^{[m_r / m]-2} - 2)}{(1 - \alpha^2)^2}
\end{aligned} \tag{5.3.131}$$

Gdy $1 > \alpha > 0$ następujące wyrażenie jest dodatnie

$$\frac{0,5 \cdot \alpha^{[m_r / m]} ([m_r / m] (1 + \alpha) - \alpha) + 0,5 \cdot \gamma^{[m_r / m]-2}}{(1 + \alpha)^2} > 0 \tag{5.3.132}$$

A dla $[m_r / m] > 1$ spełniona jest nierówność:

$$- \frac{\alpha^2 (\alpha^{[m_r / m]-2} - 1) + \alpha (\alpha^{[m_r / m]-2} - 2)}{(1 - \alpha^2)^2} \geq 0 \tag{5.3.133}$$

Oznacza to, że funkcja $rec_g(\alpha)$ jest rosnąca, gdy $[m_r / m] > 1$, co biorąc pod uwagę wniosek 5.3.15 oznacza prawdziwość twierdzenia.

□

5.3.3. Efektywność metody rekomendacji AdaptRank

W podpunktach 5.3.1 oraz 5.3.2 pokazano, przy jakich warunkach metoda rekomendacji *AdaptRank* osiąga odpowiednio maksymalną dokładność i kompletność. Niniejszy podpunkt poświęcony jest sformułowaniu warunków, których spełnienie gwarantuje maksymalną efektywność metody adaptacji *AdaptRank* niezależnie od długości bieżącej sesji. Jednocześnie

zostanie pokazane, że spełnienie tych warunków nie jest wystarczające do osiągnięcia maksymalnej efektywności przez metody składowe.

Pierwsze z twierdzeń przedstawionych w tym podpunkcie jest twierdzeniem pomocniczym wobec kolejnych twierdzeń i określa ogólne warunki, jakie musi spełniać **dowolna** funkcja rekomendacji, aby efektywność systemu rekomendującego była maksymalna (tj. równa 1).

Twierdzenie 5.3.11

Niech $D_{RN}^{(i)} \subseteq D_R^{(i)}$ oznacza zbiór dokumentów relewantnych w kroku i ($i=1, 2, \dots, n_{sb}$), które nie były zarekomendowane użytkownikowi w bieżącej sesji s_b . Dowolny system rekomendacji osiąga w bieżącej sesji s_b maksymalną efektywność, jeśli w każdym kroku tej sesji, w której zbiór $D_{RN}^{(i)}$ jest niepusty są spełnione założenia twierdzenia 5.3.5 i w każdym kroku, w którym zbiór $D_{RN}^{(i)}$ jest pusty spełnione są założenia twierdzenia 5.3.1.

Dowód:

Aby pokazać prawdziwość twierdzenia wystarczy udowodnić, że jeśli spełnione są warunki twierdzenia to zarówno dokładność jak i kompletność osiągają wartość maksymalną równą 1, gdyż:

$$E(SR, s_b) = \sqrt{1 \cdot 1} = 1 \quad (5.3.134)$$

Jeśli zbiór $D_{RN}^{(i)}$ jest niepusty to warunki twierdzenia 5.3.5 oznaczają, że system rekomendujący uzyskuje maksymalną kompletność, a dodatkowo co najmniej m dokumentów z tego zbioru ma wartości funkcji rekomendacji wyższe od wartości funkcji rekomendacji dla dokumentów nirelewantnych, co oznacza, że w krokach bieżącej sesji, w których zbiór $D_{RN}^{(i)}$ jest niepusty wszystkie rekomendowane dokumenty są relewantne:

$$\text{card}(R^{(i)} \cap D_R^{(i)}) = \text{card}(R^{(i)}) \quad (5.3.135)$$

Jest to warunek dostateczny do osiągnięcia maksymalnej dokładności. Jednoczesne osiągnięcie maksymalnej dokładności i kompletności jest równoznaczne z maksymalną efektywnością.

Jeśli zbiór $D_{RN}^{(i)}$ jest pusty oznacza to, że wszystkie dokumenty relewantne zostały już zarekomendowane użytkownikowi, co oznacza, że kompletność jest równa 1 niezależnie od tego czy system będzie rekomendował dokumenty relewantne czy też nie (oczywiście, pod warunkiem, że preferencje użytkownika nie ulegną zmianie). Dodatkowo z założeń udowodnianego twierdzenia wiadomo, że jeśli zbiór $D_{RN}^{(i)}$ jest pusty, to spełnione są założenia twierdzenia 5.3.1. Założenia twierdzenia 5.3.1 gwarantują maksymalną wartość dokładności rekomendacji. Możemy, więc stwierdzić, że także w tym przypadku założenia udowodnianego twierdzenia są wystarczające do osiągnięcia maksymalnej efektywności, a więc całe twierdzenie jest prawdziwe.

□

Interpretacja tego twierdzenia jest następująca: do osiągnięcia maksymalnej efektywności konieczne jest, aby we wszystkich krokach bieżącej sesji rekomendowane były tylko i wyłącznie dokumenty relewantne. Jednocześnie wielokrotne podpowiadanie tego samego doku-

mentu relewantnego może mieć miejsce tylko i wyłącznie wtedy, gdy nie istnieje żaden dokument relewantny, który nie był rekomendowany w bieżącej sesji.

Kolejne twierdzenie określa warunki, przy jakich metoda rekomendacji *AdaptRank* osiąga maksymalną efektywność przy założeniu, że użytkownik odwiedza jeden z dokumentów rekomendowanych w kroku poprzednim.

Twierdzenie 5.3.12

Jeśli w każdym kroku i ($i=2, \dots, n_{sb}$) bieżącej sesji s_b , użytkownik odwiedza jeden z dokumentów rekomendowanych w kroku poprzednim (tj. $d^{(i)} \in R^{(i-1)}$) i system hipertekstowy zawiera m_r ($m_r > m$) dokumentów relewantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{sb})}$ oraz w $[m_r/m]$ pierwszych krokach bieżącej sesji spełnione są następujące warunki:

$$\begin{aligned} nv_{r \min} &> 0,5 \cdot \alpha + \frac{nv_{n \max}}{1 - \alpha} \\ sim_{r \min} &> 0,5 \cdot \alpha + \frac{sim_{n \max}}{1 - \alpha} \\ nv_{r \min} &> (1 - \rho_s) \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} \\ sim_{r \min} &> (1 - \rho_s) \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} \end{aligned}$$

oraz począwszy od kroku $[m_r/m]+1$ spełnione są warunki:

$$nv_{r \min} > \frac{0,5 \cdot \alpha}{1 - \rho_s^{[m_r/m]}} + \frac{nv_{n \max}}{(1 - \rho_s^{[m_r/m]}) \cdot (1 - \alpha)} \quad \text{oraz} \quad sim_{r \min} > \frac{0,5 \cdot \alpha}{1 - \rho_s^{[m_r/m]}} + \frac{sim_{n \max}}{(1 - \rho_s^{[m_r/m]}) \cdot (1 - \alpha)},$$

to efektywność systemu rekomendującego w bieżącej sesji s_b będzie równa 1 niezależnie od liczby dokumentów odwiedzonych w tej sesji.

Dowód:

Zbiór dokumentów relewantnych jest taki sam we wszystkich krokach bieżącej sesji, więc w pierwszych $[m_r/m]$ krokach zbiór nierekomendowanych dokumentów relewantnych jest niepusty. Z założeń twierdzenia wiadomo, że w krokach tych spełnione są założenia twierdzenia 5.3.6, które to założenia są szczególnym przypadkiem założeń twierdzenia 5.3.5.

Począwszy od kroku $[m_r/m]+1$ zbiór nierekomendowanych dokumentów jest pusty, a twierdzenie spełnia założenia twierdzenia 5.3.2. Założenia twierdzenia 5.3.2 są szczególnym przypadkiem założeń twierdzenia 5.3.1. Z twierdzenia 5.3.11 wiadomo, że system rekomendacji *SR* w bieżącej sesji s_b osiąga maksymalną efektywność, jeśli w każdym kroku tej sesji, w której zbiór $D_{RN}^{(i)}$ jest niepusty są spełnione założenia twierdzenia 5.3.5 i w każdym kroku, w którym zbiór $D_{RN}^{(i)}$ jest pusty spełnione są założenia twierdzenia 5.3.1 co świadczy o prawdziwości udowodnianego twierdzenia.

□

Twierdzenie 5.3.12 jest uszczegółowieniem na potrzeby metody *AdaptRank* twierdzenia 5.3.11 podającego ogólne warunki na maksymalną efektywność. W sformułowaniu twierdzenia 5.3.12 wykorzystane zostały twierdzenia 5.3.6 i 5.3.2 określające warunki pozwalające metodzie rekomendacji *AdaptRank* osiągnąć maksymalną dokładność i kompletność. Należy zauważyć, że warunki zapewniające maksymalną efektywność są kombinacją warunków na maksymalną kompletność i dokładność (twierdzenia 5.3.2 i 5.3.6), wskutek czego ich interpretacja będzie identyczna do interpretacji powyższych twierdzeń. Oznacza to, że metoda *AdaptRank* będzie się odznaczać maksymalną efektywnością, gdy współczynniki wzorców nawigacyjnych dokumentów relewantnych będą odpowiednio wysokie i wyraźnie wyższe od współczynników nawigacyjnych dokumentów nierelewantnych. Innymi słowy, metoda *AdaptRank* osiągnie efektywność równą jedności, jeśli użytkownik uzna za relewantne, dokumenty, które były często odwiedzane w historycznych sesjach podobnych do bieżącej sesji. Analogicznie, podobieństwo dokumentów relewantnych do dokumentów odwiedzonych w bieżącej sesji powinno być odpowiednio duże i wyraźnie wyższe od podobieństwa pomiędzy dokumentami nierelewantnymi i dokumentami odwiedzonymi w bieżącej sesji. Innymi słowy, metoda *AdaptRank* osiągnie maksymalną efektywność, gdy użytkownik uzna za relewantne dokumenty, które posiadają dużo deskryptorów, występujących także w dokumentach odwiedzonych w bieżącej sesji.

Przykładowo, dla wartości parametrów zaproponowanych w tabeli 4.5.1 ($\alpha=0,25$, $\beta=1$, $\delta=1$, $\gamma=0,9$, $\rho_s=0,75$, $m=2$) i przy założeniu, że istnieje $m_r=30$ dokumentów relewantnych do potrzeb użytkownika oraz że $nv_{max}=0,1$ oraz $sim_{max}=0,1$ do osiągnięcia maksymalnej efektywności wystarczy, aby minimalne podobieństwo dokumentów relewantnych do dokumentów odwiedzonych w bieżącej sesji oraz minimalna wartość współczynnika wzorców nawigacyjnych nie były mniejsze niż 0,27 w pierwszych 15 krokach oraz były większe od 0,38 we wszystkich kolejnych krokach.

Warto jeszcze raz podkreślić, że powyższe warunki są zgodne z intuicją metod składowych i odzwierciedlają ich normalne działanie. Można więc stwierdzić, że warunki twierdzenia 5.3.12 będą często spełnione w praktyce. Jednocześnie należy podkreślić, że w twierdzeniach 5.3.13 oraz 5.3.14 zostanie pokazane, że warunki twierdzenia 5.3.12 nie są wystarczające do uzyskania maksymalnej efektywności przez metody składowe.

Bazując na twierdzeniach 5.3.11 i 5.3.12 można sformułować wniosek dotyczący warunków, przy jakich metoda rekomendacji *AdaptRank* osiąga maksymalną efektywność przy założeniu, że użytkownik odwiedza jeden z dokumentów nierekomendowanych w kroku poprzednim (dokumenty rekomendowane są ignorowane).

Wniosek 5.3.19

Jeśli w każdym kroku i ($i=2, \dots, n_{sb}$) bieżącej sesji s_b , użytkownik odwiedza jeden z dokumentów nierekomendowanych w kroku poprzednim ($d^{(i)} \notin R^{(i-1)}$) i system hipertekstowy zawiera m_r ($m_r > m$) dokumentów relewantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{sb})}$ oraz w $[m_r/m]$ pierwszych krokach bieżącej sesji spełnione są następujące warunki:

$$nv_{r \min} > \frac{\alpha^{[m_r/m]} \cdot 0,5}{1 + \alpha} + \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} - \frac{0,5 \cdot \gamma^{[m_r/m]-2}}{\beta \cdot (1 + \alpha)}$$

$$sim_{r \min} > \frac{\alpha^{[m_r/m]} \cdot 0,5}{1 + \alpha} + \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} - \frac{0,5 \cdot \gamma^{[m_r/m]-2}}{\delta \cdot (1 + \alpha)}$$

$$nv_{r \min} > 0,5 \cdot \alpha + \frac{nv_{n \max}}{1 - \alpha}$$

$$sim_{r \min} > 0,5 \cdot \alpha + \frac{sim_{n \max}}{1 - \alpha}$$

a począwszy od kroku $[m_r/m]+1$ spełnione są warunki:

$$nv_{r \min} > 0,5 \cdot \alpha + \frac{nv_{n \max}}{1 - \alpha} + \frac{0,5 \cdot \gamma^{[m_r/m]-1}}{\beta} \left(1 + \frac{\alpha}{1 - \alpha^{[m_r/m]}} \right) \text{ oraz}$$

$$sim_{r \min} > 0,5 \cdot \alpha + \frac{sim_{n \max}}{1 - \alpha} + \frac{0,5 \cdot \gamma^{[m_r/m]-1}}{\delta} \left(1 + \frac{\alpha}{1 - \alpha^{[m_r/m]}} \right)$$

to efektywność systemu rekomendującego w bieżącej sesji s_b będzie równa 1 niezależnie od liczby dokumentów odwiedzonych w tej sesji.

Wniosek 5.3.19 jest odpowiednikiem twierdzenia 5.3.12, z tą tylko różnicą, że dotyczy on bieżących sesji, w których użytkownik konsekwentnie ignoruje rekomendowane mu dokumenty. W tym przypadku dla wartości parametrów zaproponowanych w tabeli 4.5.1 ($\alpha=0,25$, $\beta=1$, $\delta=1$, $\gamma=0,9$, $\rho_s=0,75$, $m=2$) i przy założeniu, że istnieje $m_r=30$ dokumentów relewantnych do potrzeb użytkownika oraz że $nv_{r \max}=0,1$ oraz $sim_{n \max}=0,1$ do osiągnięcia maksymalnej efektywności wystarczy, aby minimalne podobieństwo dokumentów relewantnych do dokumentów odwiedzonych w bieżącej sesji oraz minimalna wartość współczynnika wzorców nawigacyjnych nie były mniejsze niż 0,86 w pierwszych 15 krokach oraz były większe od 0,45 we wszystkich kolejnych krokach. Jednocześnie należy podkreślić, że w twierdzeniach 5.3.13 i 5.3.14 zostanie pokazane, że warunki wniosku 5.3.19 nie są wystarczające do uzyskania maksymalnej efektywności przez metody składowe.

Twierdzenie 5.3.12 oraz wniosek 5.3.19 pokazują, że **niezależnie czy użytkownik odwiedza rekomendowane mu dokumenty, czy też je ignoruje**, system rekomendacji może osiągnąć maksymalną efektywność przy spełnieniu dwóch następujących założeń:

- Podobieństwo dokumentów relewantnych do dokumentów odwiedzanych w bieżącej sesji jest odpowiednio duże i wyraźnie wyższe od podobieństwa dokumentów nierelevantnych do dokumentów odwiedzanych w bieżącej sesji.
- Wartości współczynników wzorców nawigacyjnych dokumentów relewantnych są odpowiednio duże i wyraźnie wyższe od współczynników wzorców nawigacyjnych dla dokumentów nierelevantnych.

Twierdzenie 5.3.13 dotyczy efektywności rekomendacji metody składowej opartej na podobieństwie tekstowym, która została zdefiniowana w podpunkcie 4.2.1. W szczególności pokazuje ono, że maksymalna efektywność nie może być zagwarantowana, niezależnie jak wysokie jest podobieństwo dokumentów relewantnych do dokumentów odwiedzanych w bieżącej sesji i o ile jest ono wyższe od podobieństwa dokumentów nierelevantnych do dokumentów odwiedzanych w bieżącej sesji.

Twierdzenie 5.3.13

Jeśli system hipertekstowy, SR_t korzystający tylko z funkcji rekomendacji f_t , zawiera m_r ($m_r > m$) dokumentów relewantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{s_b})}$ i minimalne podobieństwo dokumentów relewantnych do dokumentów odwiedzanych w sesji s_b (sim_{rmin}) oraz maksymalne podobieństwo dokumentów nierelevantnych do dokumentów odwiedzanych w sesji s_b (sim_{nmax}) spełniają założenie:

$$\exists_{\substack{c \\ c > 0}} sim_{rmin} - sim_{nmax} \geq c$$

to niezależnie od wartości sim_{rmin} , sim_{nmax} , c i niezależnie czy w każdym kroku $i > 1$ bieżącej sesji s_b , użytkownik odwiedza jeden z dokumentów rekomendowanych w kroku poprzednim czy też nie, efektywność systemu rekomendującego SR_t , będzie zawsze spełniać nierówność:

$$E(SR_t, s_b) \geq \sqrt{\frac{m}{n_r}}$$

.gdzie n_r oznacza liczbę różnych dokumentów relewantnych, które mogły być zarekomendowane użytkownikowi w tej sesji:

$$n_r = m + \sum_{i=2}^{n_{s_b}} \min(m, \text{card} \left(D_R^{(i)} \setminus \bigcup_{j=1}^{i-1} R^{(j)} \right))$$

Dowód:

Z podrozdziału 2.2 wiadomo, że efektywność systemu rekomendującego SR w bieżącej sesji s_b jest liczona jako średnia geometryczna dokładności i kompletności:

$$E(SR_t, s_b) = \sqrt{\text{prec}(SR_t, s_b) \cdot \text{rec}(SR_t, s_b)} \quad (5.3.136)$$

Funkcja f_i dla dokumentu d_j w kroku i sesji s_b ma postać:

$$f_i(d_j, s_b^{(i)}) = \frac{1}{i} \cdot \sum_{k=1}^i \text{sim}(\mathbf{v}^{d_j}, \mathbf{v}^{d^{(k)}}) \quad (5.3.137)$$

Z założeń twierdzenia wiadomo, że wartość funkcji f_i dla dokumentu d_r będzie spełniać:

$$f_i(d_r, s_b^{(i)}) \geq \frac{1}{i} \cdot \sum_{k=1}^i \text{sim}_{r \min} \quad (5.3.138)$$

Łatwo zauważyć, że $\sum_{k=1}^i \text{sim}_{r \min} = i \cdot \text{sim}_{r \min}$ czyli ostatecznie:

$$f_i(d_r, s_b^{(i)}) \geq \text{sim}_{r \min} \quad (5.3.139)$$

Maksymalna wartość funkcji f_i dla dokumentu nierelevantnego d_n zgodnie z założeniami twierdzenia będzie spełniać nierówność:

$$f_i(d_n, s_b^{(i)}) \leq \frac{1}{i} \cdot \sum_{k=1}^i \text{sim}_{n \max} \quad (5.3.140)$$

Łatwo zauważyć, że $\sum_{k=1}^i \text{sim}_{n \max} = i \cdot \text{sim}_{n \max}$ czyli ostatecznie:

$$f_i(d_n, s_b^{(i)}) \leq \text{sim}_{n \max} \quad (5.3.141)$$

Korzystając z 5.3.139 oraz 5.3.141 można twierdzić, że przy spełnionych założeniach udowodnianego twierdzenia jest prawdziwe:

$$\forall_{n_{sb} \geq i > 0} \forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D \setminus D_R^{(i)}} (f(d_r, s_b^{(i)}) > f(d_n, s_b^{(i)})) \quad (5.3.142)$$

Co zgodnie z twierdzeniem 5.3.1 pozwala stwierdzić, że jeśli założenia udowodnianego twierdzenia są spełnione, to dokładność rekomendacji w bieżącej sesji jest równa 1.

Powyżej zostało pokazane, że przy spełnionych założeniach twierdzenia podpowiadane będą zawsze dokumenty relevantne. Niestety, postać funkcji f_i nie uwzględnia żadnych mechanizmów zmniejszających wartość funkcji dla dokumentów już odwiedzonych. W najgorszym przypadku w każdym kroku i bieżącej sesji dokładnie m tych samych dokumentów ze zbioru $D_R^{(i)}$ będzie odznaczać się najwyższym podobieństwem do dokumentów odwiedzonych, a w konsekwencji tylko te m dokumentów będzie stale podpowiadanych. Korzystając z definicji 4.1.6 dotyczącej kompletności systemu rekomendującego w bieżącej sesji otrzymujemy:

$$\text{rec}(SR_t, s_b) = \frac{\text{card}\left(\bigcup_{i=1}^{n_{sb}} (D_R^{(i)} \cap R^{(i)})\right)}{n_r} \geq \frac{m}{n_r} \quad (5.3.143)$$

Podstawiając 5.3.143 i $prec(SR, s_b)=1$ do 5.3.136 otrzymujemy:

$$E(SR_t, s_b) \geq \sqrt{1 \cdot \frac{m}{n_r}} \quad (5.3.144)$$

□

Należy zauważyć, że metoda składowa oparta na funkcji f_i nie gwarantuje osiągnięcia maksymalnej efektywności nawet, gdy $sim_{rmin}=1$ (dokumenty relewantne mają takie same deskryptory co dokumenty odwiedzone w bieżącej sesji) i $sim_{nmax}=0$ (dokumenty nirelewantne nie mają deskryptorów wspólnych z odwiedzionymi w bieżącej sesji).

W dowodzie twierdzenia 5.3.13 pokazano, że jego założenia są wystarczające do osiągnięcia maksymalnej dokładności, lecz nie są wystarczające do uzyskania maksymalnej kompletności, co w konsekwencji uniemożliwia uzyskanie maksymalnej efektywności. Kolejne twierdzenie określa warunki konieczne i wystarczające do jednoczesnego uzyskania maksymalnej dokładności i kompletności, a więc do osiągnięcia maksymalnej efektywności systemu rekomendującego działającego w oparciu o funkcję rekomendacji f_i .

Twierdzenie 5.3.14

Niech $D_{RN}^{(i)} \subseteq D_R^{(i)}$ oznacza zbiór dokumentów relewantnych w kroku i ($i=1, 2, \dots, n_{sb}$), które nie były zarekomendowane użytkownikowi w bieżącej sesji s_b . Efektywność systemu rekomendującego SR_t korzystającego z funkcji składowej f_i , będzie maksymalna w bieżącej sesji s_b , w której istnieje m_r ($m_r > m$) dokumentów relewantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{sb})}$, wtedy i tylko wtedy gdy minimalne podobieństwo dokumentów relewantnych do dokumentów odwiedzonych w sesji s_b sim_{rmin} i maksymalne podobieństwo dokumentów nirelewantnych do dokumentów odwiedzonych w sesji s_b sim_{nmax} spełniają założenie:

$$\exists_{l \geq c > 0} sim_{rmin} - sim_{nmax} \geq c$$

oraz w każdym kroku $i > 1$, w której zbiór $D_{RN}^{(i)}$ jest niepusty, co najmniej m_{aux} dokumentów należących do zbioru $D_{RN}^{(i)}$ spełnia:

$$\forall_{n_{sb} \geq i > 0} \exists_{D_{RN}^{(i)} \subseteq D_R^{(i)}} \forall_{d_r \in D_R^{(i)}} \forall_{d_m \in D_R^{(i)} \setminus D_{RN}^{(i)}} \left(sim(\mathbf{v}^{d_m}, \mathbf{v}^{d^{(i)}}) - sim(\mathbf{v}^{d_r}, \mathbf{v}^{d^{(i)}}) > (i-1) \cdot (f_i(d_m, s_b^{(i-1)}) - f_i(d_r, s_b^{(i-1)})) \wedge (card(D_{RN}^{(i)}) \geq m_{aux}) \right)$$

gdzie m_{aux} to mniejsza z dwóch następujących liczb: licznosc zbioru $D_{RN}^{(i)}$ oraz liczba dokumentów m podpowiadanych w jednym kroku.

Dowód:

Jak pokazano w dowodzie twierdzenia 5.3.13 warunek $\exists_{l \geq c > 0} sim_{r \min} - sim_{n \max} \geq c$ jest wystarczający do uzyskania maksymalnej kompletności. Do dowiedzenia prawdziwości twierdzenia wystarczy pokazać, że spełnienie jego założeń pozwoli uzyskać maksymalną kompletność.

Na początek udowodniony zostanie następujący lemat:

Lemat 1.

Wartość funkcji rekomendacji f_i dla dokumentu d w kroku i bieżącej sesji ($n_{sb} \geq i > 1$) może

$$\text{zostać zapisana jako: } f_i(d, s_b^{(i)}) = \frac{sim(\mathbf{v}^d, \mathbf{v}^{d^{(i)}})}{i} + \frac{(i-1) \cdot f_i(d, s_b^{(i-1)})}{i}$$

Dowód lematu:

Z definicji 4.2.1 wiadomo, że wartość funkcji rekomendacji f_i dla dokumentu d w kroku i ($n_{sb} \geq i > 1$) bieżącej sesji jest równa:

$$f_i(d, s_b^{(i)}) = \frac{1}{i} \cdot \sum_{k=1}^i sim(\mathbf{v}^{d_j}, \mathbf{v}^{d^{(k)}}) \quad (5.3.145)$$

Ze znaku sumy wyłączmy podobieństwo do dokumentu odwiedzonego w ostatnim kroku:

$$f_i(d, s_b^{(i)}) = \frac{sim(\mathbf{v}^d, \mathbf{v}^{d^{(i)}})}{i} + \frac{\sum_{k=1}^{i-1} sim(\mathbf{v}^d, \mathbf{v}^{d^{(k)}})}{i} \quad (5.3.146)$$

Po pomnożeniu licznika i mianownika drugiego ułamka przez $i-1$ otrzymujemy:

$$f_i(d, s_b^{(i)}) = \frac{sim(\mathbf{v}^d, \mathbf{v}^{d^{(i)}})}{i} + \frac{(i-1) \cdot \sum_{k=1}^{i-1} sim(\mathbf{v}^d, \mathbf{v}^{d^{(k)}})}{i \cdot (i-1)} \quad (5.3.147)$$

Łatwo zauważyć, że: $f_i(d, s_b^{(i-1)}) = \frac{\sum_{k=1}^{i-1} sim(\mathbf{v}^d, \mathbf{v}^{d^{(k)}})}{(i-1)}$ więc:

$$f_i(d, s_b^{(i)}) = \frac{sim(\mathbf{v}^d, \mathbf{v}^{d^{(i)}})}{i} + \frac{(i-1) \cdot f_i(d, s_b^{(i-1)})}{i} \quad (5.3.148)$$

Co dowodzi prawdziwości lematu.

Z twierdzenia 5.3.5 wiadomo, że do osiągnięcia maksymalnej kompletności w każdym kroku $i > 1$ wystarczy, aby wartość funkcji rekomendacji dla m dokumentów relewantnych nierekomendowanych jeszcze w tej sesji była większa od:

- wartości funkcji rekomendacji dla każdego z dokumentów nierekomendowanych;
- wartości funkcji rekomendacji dla każdego z dokumentów relewantnych już rekomendowanych w bieżącej sesji.

Założenie $\exists_{l \geq c > 0} sim_{r \min} - sim_{n \max} \geq c$ jest tożsamy z warunków na maksymalną kompletność. Do udowodnienia całego twierdzenia wystarczy, więc pokazać prawdziwość drugiego warunku. Drugi warunek będzie prawdziwy, jeśli dla każdego $i > 1$ wartość funkcji f_i dla nierekomendowanego dokumentu relewantnego $d_m \in D_{RN}^{(i)}$ będzie większa od wartości funkcji f_i dla rekomendowanego dokumentu relewantnego $d_r \in D_R^{(i)} \setminus D_{RN}^{(i)}$:

$$f_i(d_m, s_b^{(i)}) > f_i(d_r, s_b^{(i)}) \quad (5.3.149)$$

Korzystając z lematu 1 można napisać:

$$\frac{sim(\mathbf{v}^{d_m}, \mathbf{v}^{d^{(i)}})}{i} + \frac{(i-1) \cdot f_i(d_m, s_b^{(i-1)})}{i} > \frac{sim(\mathbf{v}^{d_r}, \mathbf{v}^{d^{(i)}})}{i} + \frac{(i-1) \cdot f_i(d_r, s_b^{(i-1)})}{i} \quad (5.3.150)$$

Mnożąc obie strony przez zawsze dodatnie i :

$$sim(\mathbf{v}^{d_m}, \mathbf{v}^{d^{(i)}}) + (i-1) \cdot f_i(d_m, s_b^{(i-1)}) > sim(\mathbf{v}^{d_r}, \mathbf{v}^{d^{(i)}}) + (i-1) \cdot f_i(d_r, s_b^{(i-1)}) \quad (5.3.151)$$

Dodając do obu stron nierówności: $-(i-1) \cdot f_i(d_m, s_b^{(i-1)}) - sim(\mathbf{v}^{d_r}, \mathbf{v}^{d^{(i)}})$, a następnie wyłączając $i-1$ przed nawias otrzymujemy:

$$sim(\mathbf{v}^{d_m}, \mathbf{v}^{d^{(i)}}) - sim(\mathbf{v}^{d_r}, \mathbf{v}^{d^{(i)}}) > (i-1) \cdot (f_i(d_r, s_b^{(i-1)}) - f_i(d_m, s_b^{(i-1)})) \quad (5.3.152)$$

□

Twierdzenie 5.3.14 pokazuje, że do zapewnienia maksymalnej efektywności systemu rekomendującego korzystającego z funkcji f_i jest konieczne, aby dokumenty relewantne już zarekomendowane w bieżącej sesji do pewnego kroku i cechowały się znacznie mniejszym podobieństwem do każdego dokumentu $d^{(i)}$ (gdzie $i > i'$) w porównaniu do podobieństwa między nierekomendowanymi dokumentami relewantnymi i każdym dokumentem $d^{(i)}$ ($i > i'$). Oznacza to, że po odwiedzeniu danego dokumentu jego podobieństwo do kolejnych odwiedzanych dokumentów musi znacznie się obniżyć. Warunek ten jest trudny do spełnienia gdyż:

- Decyzja o odwiedzanych dokumentach zależy od użytkownika, a nie od systemu. Jeśli dla danego dokumentu d_r podobieństwo do dokumentów odwiedzanych w kolejnych krokach sesji znacznie się obniża, oznacza to, że zmienił się zbiór deskryptorów występujących w odwiedzonych dokumentach. Jest to sprzeczne z normalnym zachowaniem użytkownika, który poszukując informacji na dany temat powinien odwiedzać dokumenty odznaczające się podobnym zbiorem deskryptorów.

- Drastyczne obniżenie podobieństwa do dokumentów odwiedzanych w kolejnych krokach sesji oznacza zmianę zainteresowań użytkownika przekładającą się na zmianę zbioru dokumentów relewantnych, co jest sprzeczne z założeniem twierdzenia.
- Nierówność: $sim(\mathbf{v}^{d_m}, \mathbf{v}^{d^{(i)}}) - sim(\mathbf{v}^{d_r}, \mathbf{v}^{d^{(i)}}) > (i-1) \cdot (f_t(d_r, s_b^{(i-1)}) - f_t(d_m, s_b^{(i-1)}))$ zawarta w twierdzeniu 5.3.12 pokazuje, że aby metoda osiągnęła maksymalną kompletność wraz ze wzrostem numeru odwiedzonych dokumentów, musi zwiększać się różnica między $sim(\mathbf{v}^{d_m}, \mathbf{v}^{d^{(i)}})$ i $sim(\mathbf{v}^{d_r}, \mathbf{v}^{d^{(i)}})$. Przykładowo, jeśli $f_t(d_r, s_b^{(10)})=0,9$ i $f_t(d_m, s_b^{(10)})=0,81$, to dla zapewnienia maksymalnej kompletności $sim(\mathbf{v}^{d_m}, \mathbf{v}^{d^{(11)}}) - sim(\mathbf{v}^{d_r}, \mathbf{v}^{d^{(11)}}) > 0,9$. Oznacza to, że dla długich sesji wymagana różnica może być większa od jedności (maksymalna efektywność w tych sesjach nigdy nie będzie osiągnięta).

Na mocy twierdzeń 5.3.12, 5.3.13 i 5.3.14 można sformułować następujący wniosek:

Wniosek 5.3.20

Jeśli spełnione są założenia twierdzenia 5.3.12, to efektywność metody rekomendacji *AdaptRank* będzie wyższa od efektywność metody składowej opartej na funkcji f_t we wszystkich bieżących sesjach, w których nie będzie spełniony jeden z następujących warunków:

$$\forall_{n_{sb} \geq i > 0} \exists_{D_{RN}^{(i)} \subseteq D_{RN}^{(i)}} \forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D_R^{(i)} \setminus D_{RN}^{(i)}} \left(sim(\mathbf{v}^{d_m}, \mathbf{v}^{d^{(i)}}) - sim(\mathbf{v}^{d_r}, \mathbf{v}^{d^{(i)}}) > (i-1) \cdot (f_t(d_r, s_b^{(i-1)}) - f_t(d_m, s_b^{(i-1)})) \wedge (card(D_{RN}^{(i)}) \geq m_{aux}) \right)$$

$$\text{lub } \exists_{c > 0} sim_{r \min} - sim_{n \max} \geq c$$

W sesjach, w których oba powyższe warunki są prawdziwe efektywności metody *AdaptRank* oraz metody składowej opartej na funkcji rekomendacji f_t będą takie same.

Na mocy twierdzeń 5.3.13, 5.3.14 i wniosku 5.3.19 można sformułować następujący wniosek:

Wniosek 5.3.21

Jeśli spełnione są założenia wniosku 5.3.19 to efektywność metody rekomendacji *AdaptRank* będzie wyższa od efektywności metody składowej opartej na funkcji rekomendacji f_i we wszystkich bieżących sesjach, w których nie będzie spełniony jeden z warunków:

$$\forall_{n_{sb} \geq i > 0} \exists_{D_{RN}^{(i)} \subseteq D_{RN}^{(i)}} \forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D_R^{(i)} \setminus D_{RN}^{(i)}} \left(\text{sim}(\mathbf{v}^{d_m}, \mathbf{v}^{d^{(i)}}) - \text{sim}(\mathbf{v}^{d_r}, \mathbf{v}^{d^{(i)}}) > (i-1) \cdot (f_i(d_r, s_b^{(i-1)}) - f_i(d_n, s_b^{(i-1)})) \wedge (\text{card}(D_{RN}^{(i)}) \geq m_{aux}) \right)$$

$$\text{lub } \exists_{l \geq c > 0} \text{sim}_{r \min} - \text{sim}_{n \max} \geq c$$

W sesjach, w których oba powyższe warunki są prawdziwe efektywności metody *AdaptRank* oraz metody składowej opartej na funkcji rekomendacji f_i będą takie same.

Wnioski 5.3.20 oraz 5.3.21 pokazują, że w dużej liczbie przypadków metoda *AdaptRank* cechuje się większą efektywnością od metody składowej opartej na funkcji f_i , gdyż tak jak to opisano powyżej, warunki, w których metoda oparta na funkcji składowej f_i osiąga maksymalną efektywność, są bardzo trudne do osiągnięcia w praktyce (patrz twierdzenie 5.3.14 oraz jego interpretacja), a jednocześnie w dużej liczbie przypadków metoda *AdaptRank* osiąga maksymalną efektywność (patrz twierdzenie 5.3.12 oraz wniosek 5.3.19).

Twierdzenia 5.3.15 i 5.3.16 są odpowiednikami twierdzeń 5.3.13 i 5.3.14 dla metody składowej opartej na zachowaniu użytkowników (patrz definicja 4.2.3).

Twierdzenie 5.3.15

Jeśli system hipertekstowy SR_c , korzystający tylko z funkcji rekomendacji f_c , zawiera m_r ($m_r > m$) dokumentów relewantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{s_b})}$ oraz f_{crmin} - minimalna wartość funkcji f_c dla dokumentów relewantnych, będzie większa od f_{cnmax} - maksymalnej wartości funkcji f_c dla dokumentów nirelewantnych:

$$\exists_{l \geq c > 0} f_{crmin} - f_{cnmax} \geq c$$

to niezależnie od wartości nv_{rmin} , nv_{nmax} , c i niezależnie czy w każdym kroku $i > 1$ bieżącej sesji s_b , użytkownik odwiedza jeden z dokumentów rekomendowanych w kroku poprzednim czy też nie, efektywność systemu rekomendującego SR_c będzie spełniać nierówność:

$$E(SR_c, s_b) \geq \sqrt{\frac{m}{n_r}}$$

gdzie n_r oznacza liczbę różnych dokumentów relewantnych, które mogły być zarekomendowane użytkownikowi w tej sesji: $n_r = m + \sum_{i=2}^{n_{s_b}} \min(m, \text{card}\left(D_R^{(i)} \setminus \bigcup_{j=1}^{i-1} R^{(j)}\right))$

Dowód:

Z podrozdziału 2.2 wiadomo, że efektywność systemu rekomendującego SR_c w bieżącej sesji s_b jest liczona jako średnia geometryczna dokładności i kompletności:

$$E(SR_c, s_b) = \sqrt{\text{prec}(SR_c, s_b) \cdot \text{rec}(SR_c, s_b)} \quad (5.3.153)$$

Z założenia $\exists_{l \geq c > 0} f_{crmin} - f_{cnmax} \geq c$ i z definicji 4.2.3 można stwierdzić, że:

$$\forall_{n_{s_b} \geq i > 0} \forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D \setminus D_R^{(i)}} (f_c(d_r, s_b^{(i)}) > f_c(d_n, s_b^{(i)})) \quad (5.3.154)$$

Oznacza to, że dokładność metody będzie równa jedności: $\text{rec}(SR_c, s_b) = 1$.

Powyżej zostało pokazane, że przy spełnionych założeniach twierdzenia podpowiadane będą zawsze dokumenty relewantne. Niestety, postać funkcji f_c nie uwzględnia żadnych mechanizmów zmniejszających wartość funkcji dla dokumentów relewantnych już odwiedzonych. W najgorszym przypadku w każdym kroku i bieżącej sesji m dokumentów ze zbioru $D_R^{(i)}$ będzie odznaczać się najwyższymi współczynnikami wzorców nawigacyjnych, a w konsekwencji tylko te m dokumentów będzie stale podpowiadanych. Korzystając z definicji 4.1.6 dotyczącej kompletności systemu rekomendującego w bieżącej sesji otrzymujemy:

$$rec(SR_c, s_b) = \frac{1}{n_r} \cdot card\left(\bigcup_{i=1}^{n_{sb}} (D_R^{(i)} \cap R^{(i)})\right) \geq \frac{m}{n_r} \quad (5.3.158)$$

Podstawiając 5.3.158 i $prec(SR_c, s_b)=1$ do 5.3.153 otrzymujemy:

$$E(SR_c, s_b) \geq \sqrt{1 \cdot \frac{m}{n_r}} \quad (5.3.159)$$

□

Podobnie jak w przypadku funkcji f_r , metoda składowa oparta na funkcji f_c nie gwarantuje osiągnięcia maksymalnej efektywności nawet, gdy $f_{cmin}=1$, czyli gdy dokumenty relewantne były zawsze odwiedzane w historycznych sesjach należących do wzorca nawigacyjnego najbardziej podobnego do wektora dokumentów odwiedzonych oraz $f_{cmax}=0$, czyli gdy dokumenty nirelewantne nigdy nie były odwiedzane w historycznych sesjach należących do wzorca nawigacyjnego najbardziej podobnego do wektora dokumentów odwiedzonych.

W dowodzie twierdzenia 5.3.15 pokazano, że jego założenia są wystarczające do osiągnięcia maksymalnej dokładności, lecz nie są wystarczające do uzyskania maksymalnej kompletności. Kolejne twierdzenie określa warunki wystarczające do jednoczesnego uzyskania maksymalnej dokładności i kompletności, a więc do osiągnięcia maksymalnej efektywności systemu rekomendującego działającego w oparciu o funkcję f_c .

Twierdzenie 5.3.16

Niech $D_{RN}^{(i)} \subseteq D_R^{(i)}$ oznacza zbiór dokumentów relewantnych w kroku i , które nie były zarekomendowane użytkownikowi w bieżącej sesji s_b . Efektywność systemu rekomendującego SR_c korzystającego z funkcji składowej f_c , będzie maksymalna w bieżącej sesji s_b , w której istnieje m_r ($m_r > m$) dokumentów relewantnych do potrzeb użytkownika, które są identyczne we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{sb})}$, wtedy gdy f_{cmin} - minimalna wartość funkcji f_c dla dokumentów relewantnych, będzie większa od f_{cmax} - maksymalnej wartości funkcji f_c dla dokumentów nirelewantnych:

$$\exists_{l \geq c > 0} f_{cr \min} - f_{cn \max} \geq c$$

oraz w każdym kroku $i > 1$, w której zbiór $D_{RN}^{(i)}$ jest niepusty, co najmniej m_{aux} dokumentów należących do zbioru $D_{RN}^{(i)}$ spełnia:

$$\forall_{n_{sb} \geq i > 0} \exists_{D_{RN}^{(i)} \subseteq D_{RN}^{(i)}} \forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D_R^{(i)} \setminus D_{RN}^{(i)}} \left(f_c(d_r, s_b^{(i)}) > f_c(d_n, s_b^{(i)}) \wedge (card(D_{RN}^{(i)}) \geq m_{aux}) \right)$$

gdzie m_{aux} to mniejsza z dwóch następujących liczb: licznosc zbioru $D_{RN}^{(i)}$ oraz liczba dokumentów m podpowiadanych w jednym kroku.

Dowód:

Jak pokazano w dowodzie twierdzenia 5.3.15 warunek $\exists_{1 \geq c > 0} f_{cr \min} - f_{cn \max} \geq c$ jest wystarczający do uzyskania maksymalnej dokładności. Dlatego do pokazania prawdziwości tego twierdzenia wystarczy pokazać, że spełnienie jego założeń jest równoważne z uzyskaniem maksymalnej kompletności.

Z twierdzenia 5.3.5 wiadomo, że do osiągnięcia maksymalnej kompletności wystarczy, aby w każdym kroku $i > 1$ wartość funkcji rekomendacji dla m dokumentów relewantnych nierekomendowanych jeszcze w tej sesji była większa od:

- wartości funkcji rekomendacji dla każdego z dokumentów nierekomendowanych
- wartości funkcji rekomendacji dla każdego z dokumentów relewantnych już rekomendowanych

Założenie $\exists_{1 \geq c > 0} f_{cr \min} - f_{cn \max} \geq c$ jest tożsamy z pierwszym z warunków na maksymalną kompletność. Drugi warunek podany jest w założeniach twierdzenia:

$$\forall_{n_{sb} \geq i > 0} \exists_{D_{RN}^{(i)} \subseteq D_{RN}^{(i)}} \forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D_R^{(i)} \setminus D_{RN}^{(i)}} (f_c(d_m, s_b^{(i)}) > f_c(d_r, s_b^{(i)}) \wedge (card(D_{RN}^{(i)}) \geq m_{aux})) \quad (5.3.160)$$

□

Twierdzenie 5.3.16 pokazuje, że do zapewnienia maksymalnej efektywności systemu rekomendującego korzystającego z funkcji f_c jest konieczne, aby po odwiedzeniu danego dokumentu jego współczynnik wzorców nawigacyjnych w kolejnych krokach znacznie się obniżył w stosunku do współczynnika wzorców nawigacyjnych dla nierekomendowanych dokumentów relewantnych. Warunek ten jest trudny do osiągnięcia w praktyce z następujących powodów:

- Wyżej wspomniane obniżenie musi wiązać się ze spadkiem podobieństwa wektora dokumentów odwiedzonych $z^{(i)}$ do wzorca nawigacyjnego c , który do tej pory był najbliższy. Aby nastąpił wyżej wspomniany spadek podobieństwa obecnie odwiedzany dokument musi nie być typowym przedstawicielem tego wzorca nawigacyjnego. Innymi słowy dokument $d^{(i)}$ nie powinien być odwiedzany w historycznych sesjach należących do tego wzorca. Oznacza to, że użytkownik nie powinien odwiedzać żadnego z dokumentów, które są odwiedzane przez użytkowników o podobnych zainteresowaniach. Jest to sprzeczne z intuicją, gdyż użytkownik, który poszukuje informacji na dany temat powinien zachowywać się podobnie jak inni użytkownicy o podobnych potrzebach informacyjnych.
- Zmiana najbliższego wzorca nawigacyjnego oznacza najczęściej zmianę zainteresowań użytkownika przekładającą się na zmianę zbioru dokumentów relewantnych, co jest sprzeczne z założeniem twierdzenia 5.3.16.

Na mocy twierdzeń 5.3.12, 5.3.15 i 5.3.16 można sformułować następujący wniosek:

Wniosek 5.3.22

Jeśli spełnione są założenia twierdzenia 5.3.12 to efektywność metody rekomendacji *AdaptRank* będzie wyższa od efektywność metody składowej opartej na funkcji rekomendacji f_c we wszystkich bieżących sesjach, w których nie będzie spełniony jeden z następujących warunków:

$$\exists_{l \geq c > 0} nv_{r \min} - nv_{n \max} \geq c \text{ lub } \forall_{n_{sb} \geq i > 0} \exists_{D_{RN}^{(i)} \subseteq D_{RN}^{(i)}} \forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D_R^{(i)} \setminus D_{RN}^{(i)}} \left(nv_{d_m}^{(i)} > nv_{d_r}^{(i)} \wedge \left(card(D_{RN}^{(i)}) \geq m_{aux} \right) \right)$$

W sesjach, w których oba powyższe warunki są prawdziwe efektywności metody *AdaptRank* oraz metody składowej opartej na funkcji rekomendacji f_c będą takie same.

Na mocy wniosku 5.3.19 i twierdzeń 5.3.15, 5.3.16 można sformułować następujący wniosek:

Wniosek 5.3.23

Jeśli spełnione są założenia wniosku 5.3.19 to efektywność metody rekomendacji *AdaptRank* będzie wyższa od efektywność metody składowej opartej na funkcji rekomendacji f_c we wszystkich bieżących sesjach, w których nie będzie spełniony jeden z następujących warunków:

$$\exists_{l \geq c > 0} nv_{r \min} - nv_{n \max} \geq c \text{ lub } \forall_{n_{sb} \geq i > 0} \exists_{D_{RN}^{(i)} \subseteq D_{RN}^{(i)}} \forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D_R^{(i)} \setminus D_{RN}^{(i)}} \left(nv_{d_m}^{(i)} > nv_{d_r}^{(i)} \wedge \left(card(D_{RN}^{(i)}) \geq m_{aux} \right) \right)$$

W sesjach, w których oba powyższe warunki są prawdziwe efektywności metody *AdaptRank* oraz metody składowej opartej na funkcji rekomendacji f_c będą takie same.

Wnioski 5.3.22 oraz 5.3.23 pokazują, że w dużej liczbie przypadków metoda *AdaptRank* cechuje się większą efektywnością od metody składowej opartej na funkcji f_c , gdyż tak jak to opisano powyżej, warunki, w których metoda oparta na funkcji składowej f_c osiąga maksymalną efektywność, są bardzo trudne do osiągnięcia w praktyce (patrz twierdzenie 5.3.16 oraz jego interpretacja), a jednocześnie w dużej liczbie przypadków metoda *AdaptRank* osiąga maksymalną efektywność (patrz twierdzenie 5.3.10 oraz wniosek 5.3.19).

Na zakończenie analizy efektywności metody *AdaptRank* podane zostaną wniosek i twierdzenie, które pokażą, że integracja dwóch metod składowych pozwala osiągnąć maksymalną efektywność nawet, gdy charakterystyka bieżącej sesji powoduje, że jedna ze składowych nie jest w stanie „odróżnić” dokumentów relewantnych od nirelewantnych.

Twierdzenie 5.3.17

Jeśli w każdym kroku i ($i=2, \dots, n_{sb}$) bieżącej sesji s_b , użytkownik odwiedza jeden z dokumentów rekomendowanych w kroku poprzednim (tj. $d^{(i)} \in R^{(i-1)}$) i system hipertekstowy zawiera m_r ($m_r > m$) dokumentów relewantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{sb})}$ i odwiedzane dokumenty nie posiadają deskryptorów, to jeśli w $[m_r/m]$ pierwszych krokach sesji spełnione są warunki:

$$nv_{r \min} > 0,5 \cdot \alpha + \frac{\delta}{\beta} \alpha + \frac{nv_{n \max}}{1 - \alpha}$$

$$nv_{r \min} > (1 - \rho_s) \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2}$$

oraz począwszy od kroku $[m_r/m]+1$ spełnione są warunki:

$$nv_{r \min} > \frac{0,5 \cdot \alpha}{1 - \rho^{[m_r/m]}} + \frac{0,5 \cdot \delta \cdot \alpha}{(1 - \rho^{[m_r/m]}) \cdot \beta} + \frac{nv_{n \max}}{(1 - \alpha) \cdot (1 - \rho^{[m_r/m]})},$$

to efektywność systemu rekomendującego w bieżącej sesji s_b będzie równa 1 niezależnie od liczby dokumentów odwiedzonych w tej sesji.

Dowód:

Warto zauważyć, że jeśli odwiedzane dokumenty nie posiadają deskryptorów to:

$$\forall_{n_{sb} \geq i > 0} \forall_{d \in D} \text{sim}(d, d^{(i)}) = 0.$$

W dowodzie wykorzystane zostaną dowody twierdzeń 5.3.2, 5.3.6 oraz 5.3.15.

Na początek udowodnimy, że dwie pierwsze nierówności z udowodnianego twierdzenia gwarantują maksymalną kompletność.

W dowodzie twierdzenia 5.3.6 pokazano, że metoda *AdaptRank* osiągnie maksymalną kompletność, jeśli dla każdego nieodwiedzonego dokumentu relewantnego $d_m \in D_{RN}^{(i)}$, dla każdego odwiedzonego dokumentu relewantnego $d_r \in D_R^{(i)} \setminus D_{RN}^{(i)}$ oraz dla każdego nirelewantnego dokumentu $d_n \in D \setminus D_R^{(i)}$ będzie spełnione:

$$f(d_m, s_b^{(i)}) > f(d_r, s_b^{(i)}) \wedge f(d_m, s_b^{(i)}) > f(d_n, s_b^{(i)}) \quad (5.3.161)$$

Także w twierdzeniu 5.3.6 pokazano, że $f(d_m, s_b^{(i)}) > f(d_n, s_b^{(i)})$ będzie prawdziwe, gdy:

$$\beta \cdot nv_{r \min} + \delta \cdot \text{sim}_{r \min} > 0,5 \cdot (\beta + \delta) \cdot \alpha + \left(\frac{1}{1 - \alpha}\right) \cdot (\beta \cdot nv_{n \max} + \delta \cdot \text{sim}_{n \max}) \quad (5.3.162)$$

Jak zauważono powyżej: $\forall_{n_{sb} \geq i > 0} \forall_{d \in D} \text{sim}(d, d^{(i)}) = 0$ co pozwala na następujące przekształcenie nierówności 5.3.132:

$$\beta \cdot nv_{r \min} > 0,5 \cdot (\beta + \delta) \cdot \alpha + \frac{nv_{n \max}}{1 - \alpha} \quad (5.3.163)$$

I ostatecznie dzieląc obie strony przez zawsze dodatnie β otrzymujemy:

$$nv_{r \min} > 0,5 \cdot \alpha + \frac{\delta}{\beta} \alpha + \frac{nv_{n \max}}{1 - \alpha} \quad (5.3.164)$$

Drugi z warunków, które muszą być spełnione do osiągnięcia maksymalnej kompletności to: $f(d_m, s_b^{(i)}) > f(d_r, s_b^{(i)})$. W dowodzie twierdzenia 5.3.6 pokazano, że wartość funkcji rekomendacji dla dokumentu relewantnego nieodwiedzonego w bieżącej sesji spełnia nierówność:

$$f(d_m, s_b^{(i)}) \geq \frac{1 - \alpha^i}{1 - \alpha} (\beta \cdot nv_{r \min} + \delta \cdot sim_{r \max}) \quad (5.3.165)$$

Jak zauważono powyżej: $\forall_{n_{sb} \geq i > 0} \forall_{d \in D} sim(d, d^{(i)}) = 0$, więc:

$$f(d_m, s_b^{(i)}) \geq \frac{1 - \alpha^i}{1 - \alpha} \cdot \beta \cdot nv_{r \min} \quad (5.3.166)$$

W dowodzie twierdzenia 5.3.6 pokazano także, że maksymalna wartość funkcji rekomendacji dla dokumentu odwiedzonego spełnia nierówność:

$$\begin{aligned} f(d_r, s_b^{(i)}) &\geq \beta \cdot \sum_{k=i+1-[m_r/m]}^i \left(\alpha^{i-k} \cdot nv_{d_r}^{(i)} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i-1+[m_r/m]}) \right) + \\ &+ \delta \cdot \sum_{k=i+1-[m_r/m]}^i \left(\alpha^{i-k} \cdot sim(d_r, d^{(i)}) \cdot \prod_{l=k}^i (1 - \rho_s^{l-i-1+[m_r/m]}) \right) \end{aligned} \quad (5.3.167)$$

Jak zauważono powyżej $\forall_{n_{sb} \geq i > 0} \forall_{d \in D} sim(d, d^{(i)}) = 0$, więc:

$$f(d_r, s_b^{(i)}) \geq \beta \cdot \sum_{k=i+1-[m_r/m]}^i \left(\alpha^{i-k} \cdot nv_{d_r}^{(i)} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i-1+[m_r/m]}) \right) \quad (5.3.168)$$

Po podstawieniu 5.3.165 i 5.3.168 do nierówności $f(d_m, s_b^{(i)}) > f(d_r, s_b^{(i)})$ otrzymujemy:

$$\frac{1 - \alpha^i}{1 - \alpha} \cdot \beta \cdot nv_{r \min} \geq \beta \cdot \sum_{k=i+1-[m_r/m]}^i \left(\alpha^{i-k} \cdot nv_{d_r}^{(i)} \cdot \prod_{l=k}^i (1 - \rho_s^{l-i-1+[m_r/m]}) \right) \quad (5.3.169)$$

Dokonanie przekształceń identycznych z przekształceniami z twierdzenia 5.3.6 pozwala otrzymać

$$nv_{r \min} > (1 - \rho_s) \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2} \quad (5.3.170)$$

5.3.164 i 5.3.170 są równe przekształceniom założeń twierdzenia co pozwala nam twierdzić, że dwie pierwsze nierówności z twierdzenia gwarantują maksymalną kompletność w $[m_r/m]$ pierwszych krokach bieżącej sesji. Jednocześnie w twierdzeniu 5.3.11 pokazano, że maksymalna kompletność w krokach, w których zbiór dokumentów relewantnych nierekomendowanych w bieżącej sesji jest niepusty, gwarantuje także maksymalną dokładność, a w konsekwencji maksymalną efektywność. Oznacza to, że twierdzenie jest prawdziwe dla $[m_r/m]$ pierwszych kroków bieżącej sesji.

Począwszy od kroku $[m_r/m]+1$ nie ma już dokumentów relewantnych, które nie były odwiedzone. Dlatego maksymalna efektywność będzie osiągnięta, gdy osiągnięta będzie maksymalna dokładność (ponownie korzystamy z dowodu twierdzenia 5.3.11). Na mocy twierdzenia 5.3.1 można stwierdzić, że wystarczy, więc, aby w każdym kroku bieżącej sesji począwszy od $[m_r/m]+1$, m dokumentów relewantnych posiadało wartości funkcji rekomendacji wyższe od wartości funkcji rekomendacji dla dokumentów nierelewantnych. W twierdzeniu 5.3.2 pokazano, że będzie to prawdą, gdy:

$$\begin{aligned} & (1 - \rho^{[m_r/m]}) \cdot \beta \cdot nv_{r \min} + (1 - \rho^{[m_r/m]}) \cdot \delta \cdot sim_{r \min} > \\ & 0,5 \cdot (\beta + \delta) \cdot \alpha + \left(\frac{1}{1 - \alpha} \right) \cdot (\beta \cdot nv_{n \max} + \delta \cdot sim_{n \max}) \end{aligned} \quad (5.3.171)$$

Korzystając ponownie z $\bigvee_{n_{sb} \geq t > 0} \bigvee_{d \in D} sim(d, d^{(i)}) = 0$ otrzymujemy:

$$(1 - \rho^{[m_r/m]}) \cdot \beta \cdot nv_{r \min} > 0,5 \cdot \beta \cdot \alpha + 0,5 \cdot \delta \cdot \alpha + \left(\frac{1}{1 - \alpha} \right) \cdot \beta \cdot nv_{n \max} \quad (5.3.172)$$

Dzieląc obie strony przez zawsze dodatnie $(1 - \rho^{[m_r/m]}) \cdot \beta$, można zapisać:

$$nv_{r \min} > \frac{0,5 \cdot \alpha}{1 - \rho^{[m_r/m]}} + \frac{0,5 \cdot \delta \cdot \alpha}{(1 - \rho^{[m_r/m]}) \cdot \beta} + \frac{nv_{n \max}}{(1 - \alpha) \cdot (1 - \rho^{[m_r/m]})} \quad (5.3.173)$$

5.3.173 jest równe ostatniej nierówności z twierdzenia. Całe twierdzenie jest więc prawdziwe. \square

Twierdzenie 5.3.17 pokazuje, że jeśli użytkownik odwiedza w bieżącej sesji dokumenty, które nie posiadają znaczącej informacji tekstowej, to przy odpowiednio wysokich wartościach współczynników wzorców nawigacyjnych dla dokumentów relewantnych metoda *AdaptRank* osiągnie maksymalną efektywność. Jednocześnie należy zauważyć, że jeśli użytkownik odwiedza dokumenty, które nie posiadają znaczącej informacji tekstowej to w każdym kroku bieżącej sesji funkcja f_i będzie równa 0 dla wszystkich dokumentów ze zbioru D . Co spowoduje, że składowa metoda rekomendacji korzystająca z funkcji f_i będzie we wszystkich kro-

kach rekomendowała dokumenty d_1, d_2, \dots, d_m niezależnie czy są one relewantne do potrzeb użytkownika. Pozwala to stwierdzić, że dla $n_{sb} > 1$ efektywność metody *AdaptRank* będzie wyższa od efektywności metody opartej na funkcji f_r . Jeśli, bowiem nawet, założymy, że m pierwszych dokumentów jest relewantnych, to ciągle rekomendowanie tych samych dokumentów spowoduje, że kompletność będzie wynosiła: $\sqrt{m/n_r}$ (patrz twierdzenie 5.3.13).

Kolejny wniosek jest podobny do twierdzenia 5.3.17, z tą tylko różnicą, że zerowe wartości podobieństwa zastąpiono zerowymi wartościami współczynników wzorców nawigacyjnych.

Wniosek 5.3.24

Jeśli w każdym kroku i ($i=1, 2, \dots, n_{sb}$) bieżącej sesji s_b , użytkownik odwiedza jeden z dokumentów rekomendowanych w kroku poprzednim (tj. $d^{(i)} \in R^{(i-1)}$) i system hipertekstowy zawiera m_r ($m_r > m$) dokumentów relewantnych do potrzeb użytkownika, które są takie same we wszystkich krokach bieżącej sesji, tj. $D_R^{(1)} = D_R^{(2)} = \dots = D_R^{(n_{sb})}$ oraz podobieństwo wektora dokumentów odwiedzonych do wszystkich wzorców nawigacyjnych jest równa 0:

$$\forall_{n_{sb} \geq i > 0} \forall_{n_c \geq j > 0} sim(\mathbf{z}^{(i)}, \mathbf{c}_j) = 0,$$

to jeśli w $[m_r/m]$ pierwszych krokach bieżącej sesji spełnione są następujące warunki:

$$sim_{r_{\min}} > 0,5 \cdot \alpha + \frac{\beta}{\delta} \cdot \alpha + \frac{sim_{n_{\max}}}{1 - \alpha} \quad i \quad sim_{r_{\min}} > (1 - \rho_s) \cdot \frac{1 - \alpha^{[m_r/m]}}{1 - \alpha^2}$$

oraz począwszy od kroku $[m_r/m]+1$ spełnione są warunki:

$$sim_{r_{\min}} > \frac{0,5 \cdot \alpha}{1 - \rho^{[m_r/m]}} + \frac{0,5 \cdot \beta \cdot \alpha}{(1 - \rho^{[m_r/m]}) \cdot \delta} + \frac{sim_{n_{\max}}}{(1 - \alpha) \cdot (1 - \rho^{[m_r/m]})},$$

to efektywność systemu rekomendującego w bieżącej sesji s_b będzie równa 1 niezależnie od liczby dokumentów odwiedzonych w tej sesji.

Wniosek 5.3.24 pokazuje, że jeśli użytkownik w bieżącej sesji odwiedza dokumenty, które nie były odwiedzane w żadnej historycznej sesji należącej do wyznaczonych wzorców nawigacyjnych, to przy odpowiednio wysokich wartościach podobieństwa między dokumentami odwiedzanymi i relewantnymi, metoda *AdaptRank* jest w stanie osiągnąć maksymalną efektywność w tej sesji. Jednocześnie należy zauważyć, że jeśli użytkownik odwiedza dokumenty, nieodwiedzane z innymi dokumentami to w każdym kroku bieżącej sesji funkcja f_c będzie równa 0 dla wszystkich dokumentów ze zbioru D . Spowoduje to, że składowa metoda rekomendacji korzystająca z funkcji f_c będzie we wszystkich krokach rekomendowała dokumenty d_1, d_2, \dots, d_m niezależnie czy są one relewantne do potrzeb użytkownika. Co pozwala stwierdzić, że dla $n_{sb} > 1$ efektywność metody *AdaptRank* będzie wyższa od efektywności metody opartej na funkcji f_c . Jeśli, bowiem, nawet założymy, że m pierwszych dokumentów jest relewantnych to ciągle rekomendowanie tych samych dokumentów spowoduje, że kompletność będzie wynosiła: $\sqrt{m/n_r}$ (patrz twierdzenie 5.3.15).

5.4. Własności metod składowych

Definicje dwóch składowych metody *AdaptRank* przedstawione w podrozdziale 4.2 były adaptacją rozwiązań znanych z literatury do formalnego modelu przedstawionego w niniejszej pracy. Jednocześnie w powyższym podrozdziale zaznaczono, że modyfikacje wprowadzone do składowych zostaną przedstawione po zdefiniowaniu metody *AdaptRank* oraz przedstawieniu jej własności. W niniejszym podrozdziale opisana zostanie postać metod składowych z uwzględnieniem wprowadzonych modyfikacji, a następnie zostanie pokazane, jakie korzyści zostały osiągnięte dzięki zaproponowanym zmianom. Dodatkowo wyodrębniona zostanie dodatkowa składowa rekomendująca tylko i wyłącznie w oparciu o przydatność dokumentu.

Pierwsza z definicji dotyczy składowej opartej na podobieństwie tekstowym. Funkcja przedstawiona w tej definicji jest modyfikacją funkcji f_t zdefiniowanej w podrozdziale 4.2.

Definicja 5.4.1

Wartość funkcji rekomendacji f'_t opartej na podobieństwie tekstowym dla dokumentu d_j ($j=1, 2, \dots, n$) w kroku i ($i=1, 2, \dots, n_{sb}$) bieżącej sesji s_b jest równa:

$$f'_t(d_j, s_b^{(i)}) = \sum_{k=1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right)$$

Modyfikacje wprowadzone względem funkcji $f_t(d_j, s_b^{(i)})$ przedstawionej w definicji 4.2.1 polegają na:

- Dodaniu mechanizmu osłabiającego wpływ dokumentów odwiedzonych w poprzednich krokach bieżącej sesji na wartość funkcji składowej f'_t .
- Dodaniu mechanizmów ograniczających możliwość ponownego podpowiedzenia już odwiedzonych dokumentu

Szczegółowa charakterystyka tych usprawnień zawarta jest w trzech kolejnych twierdzeniach. Twierdzenie 5.4.1 dotyczy malejącego wpływu dokumentów odwiedzonych w poprzednich krokach bieżącej sesji na wartość funkcji składowej f'_t .

Twierdzenie 5.4.1

Wpływ $\text{sim}(d_j, d^{(i)})$ - wartości podobieństwa dokumentu d_j nieodwiedzonego w bieżącej sesji s_b do dokumentu $d^{(i)}$ odwiedzonego w kroku i bieżącej sesji s_b (gdzie $0 < i' \leq n_{sb}$) na wartość funkcji składowej f'_t dla dokumentu d_j maleje w każdym kolejnym kroku i ($i > i'$) w tempie geometrycznym o stałym ilorazie równym α

Dowód:

Wartość funkcji rekomendacji f'_t w kroku i bieżącej sesji dla dokumentu d_j nieodwiedzonego w bieżącej sesji wynosi:

$$f'_t(d_j, s_b^{(i)}) = \sum_{k=1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \right) \quad (5.4.1)$$

Wartość funkcji rekomendacji z 5.4.1 może zostać zapisana w następujący sposób:

$$f'_t(d_j, s_b^{(i)}) = \sum_{k=1}^{i-1} \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \right) + \alpha^{i-i} \cdot \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j}) + \sum_{k=1}^{i+1} \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \right) \quad (5.4.2)$$

W ten sposób wyodrębniono wpływ $\text{sim}(d_j, d^{(i)})$ - wartości podobieństwa dokumentu d_j nieodwiedzonego w bieżącej sesji s_b do dokumentu odwiedzonego w kroku i' bieżącej sesji s_b . Oznaczmy go jako ciąg $x_i = \alpha^{i-i'} \cdot \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j})$. Policzmy iloraz dwóch kolejnych wyrazów ciągu x_i :

$$\frac{x_{i+1}}{x_i} = \frac{\alpha^{i+1-i'} \cdot \text{sim}(\mathbf{v}^{d^{(i+1)}}, \mathbf{v}^{d_j})}{\alpha^{i-i'} \cdot \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j})} = \alpha \quad (5.4.3)$$

□

Twierdzenie 5.4.2 pokazuje, że po odwiedzeniu danego dokumentu d_j w kroku i' podobieństwo do dokumentów odwiedzonych do kroku i' włącznie nie będzie miało żadnego wpływu na wartość funkcji składowej f'_t dla dokumentu d_j w kolejnych krokach począwszy od kroku $i'+1$.

Twierdzenie 5.4.2

Jeśli dokument d_j był odwiedzony w kroku i' w bieżącej sesji s_b , to w kolejnych krokach wartość funkcji składowej f'_t dla dokumentu d_j będzie zależeć tylko wyłącznie od podobieństwa dokumentu d_j do dokumentów odwiedzonych począwszy od kroku $i'+1$:

$$f'_t(d_j, s_b^{(i)}) = \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right)$$

Dowód:

Wyrażenie określające wartość funkcji składowej f'_t można przedstawić jako:

$$f'_t(d_j, s_b^{(i)}) = \sum_{k=1}^{i'} \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) + \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \quad (5.4.4)$$

Biorąc pod uwagę, że: $\prod_{l=k}^i (1 - z_j^{(l)}) = (1 - z_j^{(i)}) \cdot \prod_{l=1}^{i-1} (1 - z_j^{(l)}) \cdot \prod_{l=i}^i (1 - z_j^{(l)})$ otrzymujemy:

$$f'_i(d_j, s_b^{(i)}) = \sum_{k=1}^{i'} \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot (1 - z_j^{(i')}) \cdot \prod_{l=1}^{i'-1} (1 - z_j^{(l)}) \cdot \prod_{l=i'}^i (1 - z_j^{(l)}) \right) + \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \quad (5.4.5)$$

Z treści twierdzenia wiadomo, że dokument d_j został odwiedzony w kroku i' , czyli $z_j^{(i')}=1$, co pozwala zapisać:

$$f'_i(d_j, s_b^{(i)}) = \sum_{k=1}^{i'} \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot 0 \cdot \prod_{l=1}^{i'-1} (1 - z_j^{(l)}) \cdot \prod_{l=i'}^i (1 - z_j^{(l)}) \right) + \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \quad (5.4.6)$$

Wszystkie składniki pierwszej sumy zawierają mnożenie przez 0, więc ostatecznie:

$$f'_i(d_j, s_b^{(i)}) = \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) \quad (5.4.7)$$

□

Interpretacja tego twierdzenia jest następująca: po odwiedzeniu dokumentu d_j wartość funkcji rekomendacji dla dokumentu d_j jest zmniejszana do 0, a zachowanie użytkownika w poprzednich krokach przestaje mieć wpływ na ewentualną rekomendację tego dokumentu w przyszłości. Tak znaczne „osłabienie” wartości funkcji f dla dokumentu d_j ma na celu ograniczenie możliwości rekomendacji już odwiedzonego dokumentu w kolejnych krokach. Oczywiście, dokument d_j może zostać ponownie zarekomendowany, ale tylko wtedy gdy inne dokumenty będą odznaczały się bardzo niską wartością funkcji rekomendacji. Innymi słowy ten sam dokument będzie rekomendowany ponownie, gdy nie będzie innych dokumentów wyraźnie relewantnych do potrzeb użytkownika.

Kolejne twierdzenie także dotyczy osłabiania wartości funkcji dla odwiedzonego dokumentu i pokazuje, że po odwiedzeniu dokumentu d_j podobieństwo do kolejno odwiedzonych dokumentów będzie miało mniejszy wpływ na wartość funkcji składowej f'_i w porównaniu do wpływu, jakie miałyby to samo podobieństwo gdyby dokument d_j nie był odwiedzony.

Twierdzenie 5.4.3

Jeśli dokumenty d_j odwiedzony w kroku i oraz dokument d_m nieodwiedzony w bieżącej sesji cechują się takim samym podobieństwem do właśnie odwiedzanego dokumentu $d^{(i)}$ $sim(d_j, d^{(i)})=sim(d_m, d^{(i)})$ oraz $i>i'$ oraz ich wartości funkcji składowej f'_t w kroku poprzednim były równe $f'_t(d_j, s_b^{(i-1)})=f'_t(d_m, s_b^{(i-1)})$, to przyrost wartości funkcji składowej w kroku i dla dokumentu d_j będzie mniejszy od przyrostu wartości funkcji składowej w kroku i dla dokumentu d_m i wyrazi się zależnością:

$$\Delta f'_t(d_j, s_b^{(i)}) + \rho_s^{i-i'} \cdot sim(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j}) = \Delta f'_t(d_m, s_b^{(i)})$$

Dowód:

Przyrost funkcji składowej f'_t w kroku i dla dokumentu d_j wyraża się wzorem:

$$\Delta f'_t(d_j, s_b^{(i)}) = f'_t(d_j, s_b^{(i)}) - f'_t(d_j, s_b^{(i-1)}) \quad (5.4.8)$$

Przyrost funkcji składowej f'_t w kroku i dla dokumentu d_m wyraża się wzorem:

$$\Delta f'_t(d_m, s_b^{(i)}) = f'_t(d_m, s_b^{(i)}) - f'_t(d_m, s_b^{(i-1)}) \quad (5.4.9)$$

Z założeń twierdzenia wiadomo, że: $f'_t(d_j, s_b^{(i-1)})=f'_t(d_m, s_b^{(i-1)})$, co pozwala zapisać:

$$\Delta f'_t(d_m, s_b^{(i)}) = f'_t(d_m, s_b^{(i)}) - f'_t(d_j, s_b^{(i-1)}) \quad (5.4.10)$$

Odejmując od równania 5.4.8 równanie 5.4.10 otrzymujemy:

$$\begin{aligned} \Delta f'_t(d_j, s_b^{(i)}) - \Delta f'_t(d_m, s_b^{(i)}) &= f'_t(d_j, s_b^{(i)}) - f'_t(d_j, s_b^{(i-1)}) \\ &\quad - f'_t(d_m, s_b^{(i)}) + f'_t(d_j, s_b^{(i-1)}) \end{aligned} \quad (5.4.11)$$

Po wykonaniu odejmowania i podstawieniu wartości funkcji z definicji 5.4.1 otrzymujemy:

$$\begin{aligned} \Delta f'_t(d_j, s_b^{(i)}) - \Delta f'_t(d_m, s_b^{(i)}) &= \sum_{k=1}^i \left(\alpha^{i-k} \cdot sim(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) - \\ &\quad \sum_{k=1}^i \left(\alpha^{i-k} \cdot sim(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_m}) \cdot \prod_{l=k}^i (1 - z_m^{(l)}) \right) \end{aligned} \quad (5.4.12)$$

Wyłączając ostatni składnik z obu sum możemy zapisać:

$$\begin{aligned}
\Delta f'_t(d_j, s_b^{(i)}) - \Delta f'_t(d_m, s_b^{(i)}) &= \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j}) \cdot (1 - z_j^{(i)}) + \\
&+ \sum_{k=1}^{i-1} \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) - \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_m}) \cdot (1 - z_m^{(i)}) \\
&- \sum_{k=1}^i \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_m}) \cdot \prod_{l=k}^i (1 - z_m^{(l)}) \right)
\end{aligned} \tag{5.4.13}$$

Biorąc pod uwagę, że:

$$\begin{aligned}
&\sum_{k=1}^{i-1} \left(\alpha^{i-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) = \\
&= \alpha \cdot \sum_{k=1}^{i-1} \left(\alpha^{i-1-k} \cdot \text{sim}(\mathbf{v}^{d^{(k)}}, \mathbf{v}^{d_j}) \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right) = \alpha \cdot f'_t(d_j, s_b^{(i-1)})
\end{aligned} \tag{5.4.14}$$

można przekształcić 5.4.10 do następującej postaci:

$$\begin{aligned}
\Delta f'_t(d_j, s_b^{(i)}) - \Delta f'_t(d_m, s_b^{(i)}) &= \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j}) \cdot (1 - z_j^{(i)}) + \\
&+ \alpha \cdot f'_t(d_j, s_b^{(i-1)}) - \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_m}) \cdot (1 - z_m^{(i)}) - \alpha \cdot f'_t(d_m, s_b^{(i-1)})
\end{aligned} \tag{5.4.15}$$

Z założeń twierdzenia wiadomo, że: $f'_t(d_j, s_b^{(i-1)}) = f'_t(d_m, s_b^{(i-1)})$, co pozwala zapisać:

$$\begin{aligned}
\Delta f'_t(d_j, s_b^{(i)}) - \Delta f'_t(d_m, s_b^{(i)}) &= \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j}) \cdot (1 - z_j^{(i)}) - \\
&- \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_m}) \cdot (1 - z_m^{(i)})
\end{aligned} \tag{5.4.16}$$

Dokument d_j był odwiedzony w kroku i czyli $z_j^{(i)} = \rho_s^{i-i}$ dokument d_m nie był odwiedzony w bieżącej sesji ($z_j^{(i)} = 0$) więc:

$$\begin{aligned}
\Delta f'_t(d_j, s_b^{(i)}) - \Delta f'_t(d_m, s_b^{(i)}) &= \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j}) - \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j}) \cdot \rho_s^{i-i} - \\
&- \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_m})
\end{aligned} \tag{5.4.17}$$

Z założeń twierdzenia wiadomo, że: $\text{sim}(d_j, d^{(i)}) = \text{sim}(d_m, d^{(i)})$ czyli:

$$\Delta f'_t(d_j, s_b^{(i)}) - \Delta f'_t(d_m, s_b^{(i)}) = -\rho_s^{i-i} \cdot \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j}) \tag{5.4.18}$$

Po dodaniu do obu stron równania 5.4.18 $\Delta f'_t(d_m, s_b^{(i)}) + \rho_s^{i-i} \cdot \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j})$ otrzymujemy:

$$\Delta f'_t(d_j, s_b^{(i)}) + \rho_s^{i-i} \cdot \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j}) = \Delta f'_t(d_m, s_b^{(i)}) \tag{5.4.19}$$

□

Należy zauważyć, że różnice w przyrostach wartości funkcji dla dokumentu d_j i d_m będą coraz mniejsze ($\lim_{i \rightarrow \infty} \rho_s^{i-i'} \cdot \text{sim}(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j}) = 0$), tak więc można powiedzieć, że system będzie „zapominał”, że dokument d_j został odwiedzony.

Definicja 5.4.2 dotyczy funkcji składowej metody *AdaptRank*, która jest oparta na zachowaniu użytkowników:

Definicja 5.4.2

Wartość funkcji rekomendacji f_c dla dokumentu d_j ($j=1, 2, \dots, n$) w kroku i ($i=1, 2, \dots, n_{sb}$) bieżącej sesji s_b jest równa:

$$f_c(d_j, s_b^{(i)}) = \sum_{k=1}^i \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right)$$

Usprawnienia wprowadzone względem funkcji $f_c(d_j, s_b^{(i)})$ przedstawionej w definicji 4.2.3 polegają na:

- Dodaniu mechanizmu osłabiającego wpływ dokumentów odwiedzonych w poprzednich krokach bieżącej sesji na wartość funkcji składowej f_c
- Dodaniu mechanizmów ograniczających możliwość ponownego podpowiedzenia już odwiedzonych dokumentu
- Uwzględnieniu kilku najbliższych wzorców nawigacyjnych w obliczaniu wartości funkcji f_c , dzięki czemu można się spodziewać rekomendacji odznaczającej się większą dokładnością dla bieżących sesji, których wektor dokumentów odwiedzonych znajduje się na pograniczu kilku wzorców nawigacyjnych (patrz rysunek 5.4.2)

Szczegółowa charakterystyka tych usprawnień zawarta jest w kolejnych twierdzeniach i wniosku 5.4.1, który analogicznie do twierdzenia 5.4.1 dotyczy malejącego wpływu współczynników wzorców nawigacyjnych z poprzednich kroków na wartość funkcji składowej f_c .

Wniosek 5.4.1

Wpływ $nv_j^{(i)}$ - współczynnika wzorców nawigacyjnych dla dokumentu d_j w kroku i na wartość funkcji składowej f_c dla dokumentu d_j maleje w kolejnych krokach w tempie geometrycznym o stałym ilorazie równym α

Kolejne 2 twierdzenia są odpowiednikami twierdzeń 5.4.1 i 5.4.2 dlatego podane zostaną bez dowodu. Gwoli wyjaśnienia należy stwierdzić, że wartości funkcji składowej f_c dla dokumentów odwiedzonych są obniżane w ten sam sposób co wartości funkcji składowej f_t .

Twierdzenie 5.4.4

Jeśli dokument d_j był odwiedzony w kroku i' , to w kolejnych krokach wartość funkcji składowej f'_c dla dokumentu d_j będzie zależeć tylko wyłącznie od współczynników wzorców nawigacyjnych obliczonych od kroku $i'+1$:

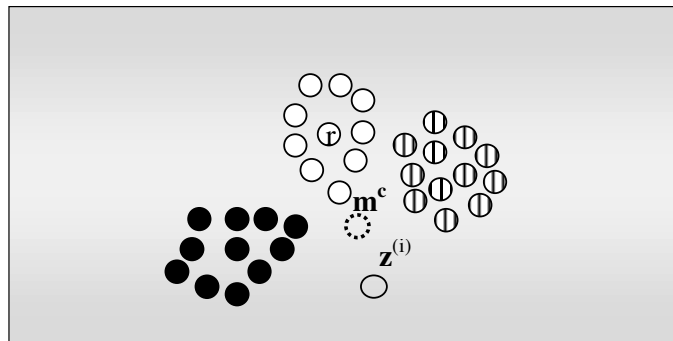
$$f'_c(d_j, s_b^{(i)}) = \sum_{k=i'+1}^i \left(\alpha^{i-k} \cdot nv_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right)$$

Twierdzenie 5.4.5

Jeśli dokument d_j odwiedzony w kroku i' oraz dokument d_m nieodwiedzony w bieżącej sesji cechują się takimi samymi wartościami współczynników wzorców nawigacyjnych w kroku i $nv_j^{(i)} = nv_m^{(i)}$ oraz $i > i'$ oraz ich wartości funkcji składowej f'_c w kroku poprzednim były równe $f'_c(d_j, s_b^{(i-1)}) = f'_c(d_m, s_b^{(i-1)})$, to przyrost wartości funkcji składowej w kroku i dla dokumentu d_j będzie mniejszy od przyrostu wartości funkcji składowej w kroku i dla dokumentu d_m i wyrazi się zależnością:

$$\Delta f'_c(d_j, s_b^{(i)}) + \rho_s^{i-i'} \cdot nv_j^{(i)} = \Delta f'_c(d_m, s_b^{(i)})$$

Twierdzenie 5.4.6 pokazuje jaką wartość przyjmuje współczynnik nawigacyjny, jeśli wektor dokumentów odwiedzonych do kroku i $\mathbf{z}^{(i)}$ jest tak samo podobny do najbliższych leżących centroidów i nie jest w ogóle podobny do innych centroidów. Sytuacja ta zilustrowana jest na rysunku 5.4.1. W tym przypadku wektor współczynników nawigacyjnych jest równy wektorowi o zwrocie i długości równej średniej centroidów reprezentujących najbliższe grupy dokumentów.



Rysunek 5.4.1. Podobieństwo między wektorem dokumentów odwiedzonych a wektorami historycznymi zostało przedstawione jako odległość między punktami w przestrzeni dwuwymiarowej.

Twierdzenie 5.4.6

Jeśli w kroku i ($i=1, 2, \dots, n_{sb}$) podobieństwo między wektorem dokumentów odwiedzonych $\mathbf{z}^{(i)}$ a m_c najbliższymi wzorcami nawigacyjnymi $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \dots, \mathbf{c}_{m_c}$ jest równe h i dodatkowo podobieństwo między wektorem aktualnej $\mathbf{z}^{(i)}$ i pozostałymi wzorcami nawigacyjnymi jest równe 0 to współczynnik nawigacyjny w kroku i dla dowolnego dokumentu d_j jest równy:

$$nv_j^{(i)} = h \cdot m_j^c$$

gdzie $\mathbf{m}^c = \frac{1}{m_c} \sum_{k=1}^{m_c} \mathbf{c}_k$ to średni wektor wzorców nawigacyjnych: $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \dots, \mathbf{c}_m$

Dowód:

Współrzędna j wektora \mathbf{m}^c jest równa:

$$m_j^c = \frac{1}{m_c} \sum_{k=1}^{m_c} c_{kj} \quad (5.4.20)$$

Z drugiej strony z definicji 4.5.2 wartość współczynnika nawigacyjnego w kroku i dla dokumentu d_j wynosi:

$$nv_j^{(i)} = \frac{1}{m_c} \sum_{k=1}^{m_c} sim(\mathbf{z}^{(i)}, \mathbf{c}_k) c_{kj} \quad (5.4.21)$$

Z treści twierdzenia wiadomo, że podobieństwo między wektorem dokumentów odwiedzonych $\mathbf{z}^{(i)}$, a m_c wzorcami nawigacyjnymi jest równe h , a podobieństwo do pozostałych wzorców jest równe 0, więc:

$$nv_j^{(i)} = \frac{1}{m_c} \sum_{k=1}^{m_c} h \cdot c_{kj} \quad (5.4.22)$$

Co po dokonaniu prostych przekształceń daje:

$$nv_j^{(i)} = \frac{h}{m_c} \sum_{k=1}^{m_c} c_{kj} \quad (5.4.23)$$

Biorąc pod uwagę 5.4.21 można napisać:

$$nv_j^{(i)} = h \cdot m_j^c \quad (5.4.24)$$

□

Tak jak już wspomniano powyżej jeśli wektor dokumentów odwiedzonych jest tak samo podobny do najbliższych wzorców nawigacyjnych współczynnik nawigacyjny jest wektorem, którego zwrot i kierunek są takie same jak wektora będącego średnią tych wzorców. Innymi słowy promowane są dokumenty będące najlepszymi reprezentantami hipotetycznego wzorca nawigacyjnego powstałego „na przecięciu” kilku wzorców nawigacyjnych. W praktyce, sesje, w których odwiedzane są dokumenty „przyporządkowane” do kilku wzorców nawigacyjnych nie są niczym niezwykłym.

Przykładowo, jeśli aktualna sesja użytkownika jest tak samo odległa od wzorca nawigacyjnego dotyczącego programowania w języku JAVA, dostępu do bazy danych oraz aplikacji urządzeń mobilnych to najprawdopodobniej podpowiadane będą dokumenty opisujące metody programowania w języku JAVA w środowisku mobilnym. Należy zwrócić uwagę, że strony te nie będą dobrymi przedstawicielami swoich macierzystych grup. Można np. domniemywać, że typowym przedstawicielem grupy dotyczącej programowania w języku JAVA będą dokumenty zawierające opis składni lub dostępnych funkcji i jest mało prawdopodobne, by strony zawierały informację na temat urządzeń mobilnych. Warto podkreślić, że faworyzowanie dokumentów „leżących na przecięciu” kilku wzorców nawigacyjnych wydaje się być trafniejsze niż rekomendowanie stron będących typowymi przedstawicielami najbliższego wzorca nawigacyjnego. Kolejny wniosek pokazuje co dzieje się z wartością współczynnika nawigacyjnego, gdy wektor dokumentów odwiedzonych podobny jest tylko do jednego wzorca nawigacyjnego.

Wniosek 5.4.2

Jeśli w kroku i podobieństwo między wektorem dokumentów odwiedzonych $\mathbf{z}^{(i)}$ a najbliższym wzorcem nawigacyjnym \mathbf{c}_{\max} jest równe h i dodatkowo podobieństwo między wektorem dokumentów odwiedzonych $\mathbf{z}^{(i)}$ i pozostałymi wzorcami nawigacyjnymi jest równe 0 to współczynnik nawigacyjny w kroku i dla dowolnego dokumentu d_j jest równy:

$$nv_j^{(i)} = z \cdot c_{\max j}$$

Wniosek 5.4.2 pokazuje, że w tym konkretnym przypadku wektor współczynnika nawigacyjnego zachowuje się jak to opisano w metodzie (Mobasher i inni, 2000b) tzn., ma taki sam zwrot jak najbliższy centroid pomnożony przez podobieństwo między wektorem dokumentów odwiedzonych i wyżej wspomnianym centroidem. Innymi słowy współczynnik wzorca nawigacyjnego może być potraktowany jako uogólnienie rekomendacji opartej na najbliższym centroidzie.

Kolejna definicja przedstawia nową oryginalną składową nieistniejącą w literaturze, która może być potraktowana jako metoda rekomendacji, która opiera się tylko i wyłącznie na przydatności dokumentu.

Definicja 5.4.3

Wartość funkcji rekomendacji f_q dla dokumentu $d_j \in D$ w i -tym kroku bieżącej sesji s_b jest równa:

$$f_q(d_j, s_b^{(i)}) = \alpha^i \cdot \left(q(d_j) \cdot \prod_{k=1}^i (1 - z_j^{(k)}) \right)$$

Kolejny wniosek pokazuje, że znaczenie jakości dokumentu wraz z wydłużaniem się sesji zmniejsza się. Interpretacja tego faktu jest następująca: na początku sesji preferencje użytkownika nie są znane, dlatego podpowiadane mu są dokumenty o wysokiej przydatności. W miarę jak użytkownik odwiedza coraz to nowe dokumenty i jego profil staje się coraz bardziej klarowny jakość dokumentów ustępuje miejsca podobieństwu tekstowemu oraz informacji o typowych wzorcach zachowań.

Wniosek 5.4.3

Dla dokumentów nieodwiedzonych w bieżącej sesji s_b funkcja jakości $f_q(d_j, s_b^{(i)})$ maleje w kolejnych krokach w tempie geometrycznym o przyroście $\alpha \in (0, 1)$

Trzy definicje przedstawione w tym podrozdziale oraz twierdzenie 5.1.1 pozwalają sformułować następujący wniosek:

Wniosek 5.4.4

Wartość funkcji rekomendacji f dla dokumentu $d_j \in D$ w i -tym kroku bieżącej sesji s_b może zostać zapisany jako kombinacja liniowa funkcji składowych pomniejszona o sumę odpowiednich współrzędnych dokumentów zignorowanych:

$$f(d_j, s_b^{(i)}) = 0,5 \cdot (\beta + \delta) \cdot f_q(d_j, s_b^{(i)}) + \beta \cdot f'_t(d_j, s_b^{(i)}) + f'_c(d_j, s_b^{(i)}) - \sum_{k=1}^i \left(\alpha^{i-k} \cdot g_j^{(k)} \cdot \prod_{l=k}^i (1 - z_j^{(l)}) \right)$$

5.5. Wpływ wartości parametrów na zachowanie się metody rekomendacji

Poprzednie podrozdziały pozwoliły przeanalizować jak zmieniać się będzie dokładność, kompletność oraz zdolności adaptacyjne metody w zależności od wartości parametrów. Wyniki przeprowadzonej analizy zostały zebrane w tabeli 5.5.1.

Zmienna	Dokładność		Kompletność		Zdolności adaptacyjne
	Rekomendacja nieignorowana	Rekomendacja ignorowana	Rekomendacja nieignorowana	Rekomendacja ignorowana	
α	Maleje	Maleje	Maleje	Maleje	Maleje
β	Obojętnie	Rośnie	Obojętnie	Maleje	Obojętnie
δ	Obojętnie	Rośnie	Obojętnie	Maleje	Obojętnie
γ	Nie dotyczy	Maleje	Nie dotyczy	Rośnie	Zależy (1)
ρ_s	Maleje	Nie dotyczy	Rośnie	Nie dotyczy	Zależy (2)
m_r	Rośnie	Rośnie	Maleje	Maleje	Zależy (3)
m	Maleje	Maleje	Rośnie	Rośnie	Zależy (4)

Tabela 5.5.1 Wpływ wzrostu wartości poszczególnych parametrów na dokładność, kompletność oraz zdolności adaptacyjne metody rekomendacji *AdaptRank*.

Poszczególne wiersze tabeli odpowiadają parametrom metody *AdaptRank*. Każda kolumna jest przypisana do danej własności metody rekomendacji. Poszczególne komórki w tabeli określają jak zmieniać się będzie wielkość znajdująca się w kolumnie przy wzroście wartości parametru znajdującego się w wierszu. Przykładowo, wraz ze wzrostem parametru m , zmaleje dokładność rekomendacji dla bieżących sesji, w których użytkownik w każdym kroku odwiedza jeden z rekomendowanych dokumentów. Jeśli komórka oznaczona jest etykietą „Zależy”, oznacza to, że wzrost wartości danego parametru może powodować wzrost lub spadek danej wielkości w zależności od wartości innych parametrów lub zachowania użytkownika.

Komórki tabeli 5.5.1 oznaczone kursywą wraz z odpowiadającym numerem odpowiadają zależnościom, niezbadanym analitycznie w poprzednich podrozdziałach. Zostaną one tutaj pokrótce omówione:

1. **Wpływ wzrostu parametru ρ_s na zdolności adaptacyjne.** Odwiedzenie rekomendowanego dokumentu d_r zawsze wpływa negatywnie na wartości funkcji rekomendacji dla dokumentu d_r w kolejnych krokach bieżącej sesji. To znaczy, że po odwiedzeniu danego dokumentu w kroku i , wartości funkcji rekomendacji w kroku i' będą pomnażane przez liczbę $(1-\rho_s^{i'-i})$. Jeśli więc przed zmianą zainteresowań użytkownika odwiedzane były dokumenty, które są relewantne do nowych potrzeb użytkownika to wzrost wartości parametru ρ_s spowoduje spadek zdolności adaptacyjnych. Jeżeli jednak dokumenty odwiedzane w bieżącej sesji są nierelwantne do nowych potrzeb użytkownika wzrost wartości parametru ρ_s spowoduje wzrost zdolności adaptacyjnych.
2. **Wpływ wzrostu parametru γ na zdolności adaptacyjne.** Podobnie jak w przypadku parametru ρ_s można stwierdzić, że zignorowanie rekomendowanego dokumentu zaw-

sze wpływa negatywnie na wartości funkcji rekomendacji dla tego dokumentu w kolejnych krokach bieżącej sesji. Jeśli, więc w przeszłości ignorowane były dokumenty, które są relewantne do nowych potrzeb użytkownika, to wzrost wartości parametru γ spowoduje spadek zdolności adaptacyjnych. Jeżeli jednak dokumenty zignorowane w bieżącej sesji są nierelatywne do nowych potrzeb użytkownika wzrost wartości parametru γ spowoduje wzrost zdolności adaptacyjnych.

3. **Wpływ wzrostu liczby dokumentów relewantnych na zdolności adaptacyjne metody.** Duża liczba dokumentów relewantnych powoduje, że duża liczba dokumentów będzie posiadała wysokie wartości funkcji *AdaptRank*. Jeśli po zmianie zainteresowań użytkownika wszystkie dokumenty uprzednio relewantne stałyby się nierelatywne, wzrost liczby dokumentów relewantnych przekładałby się bezpośrednio na spadek zdolności adaptacyjnych metody. Oczywistym jest, że nie zawsze zmiana zapotrzebowań użytkownika spowoduje, że wszystkie dokumenty dotychczas nierelatywne staną się relewantnymi, a relewantne nierelatywne, dlatego nie można podać tu jasnej zależności.
4. **Wpływ wzrostu liczby dokumentów rekomendowanych w jednym kroku na zdolności adaptacyjne metody.** Rozumowanie jest bardzo podobne do podanego w punkcie poprzednim. Jeśli po zmianie zainteresowań użytkownika wszystkie dokumenty uprzednio relewantne stałyby się nierelatywne, wzrost liczby dokumentów rekomendowanych w jednym kroku przekładałby się bezpośrednio na wzrost zdolności adaptacyjnych metody, gdyż po zarekomendowaniu danego dokumentu jego wartości funkcji rekomendacji w kolejnych krokach zmniejszą (z powodu parametru ρ_s jeśli dokument został odwiedzony lub z powodu parametru γ jeśli dokument został zignorowany).

Analiza tabeli 5.5.1 pozwala sformułować wiele ciekawych wniosków dotyczących doboru wartości parametrów:

- Parametry ρ_s i γ pozwalają określić czy priorytetem rekomendacji jest wysoka kompletność, czy wysoka dokładność. Innymi słowy służą one do podjęcia decyzji: czy podpowiadać różnorodną treść nie zawsze relewantną, czy też rekomendować mało zróżnicowaną treść odznaczającą się wysoką relewancją.
- Postać funkcji f powoduje, że parametry β i δ mają ograniczony wpływ na zachowanie się funkcji rekomendacji, dlatego przy doborze ich wartości należy kierować się, przede wszystkim, wpływem jakim podobieństwo tekstowe i wzorce nawigacyjne powinny mieć na wartości funkcji rekomendacji. Obniżenie wartości β może nastąpić w przypadku systemów hipertekstowych posiadających małą ilość danych na temat aktywności użytkowników (istnieje silne podejrzenie, że wzorce nawigacyjne zostały obliczone na niereprezentacyjnej liczbie historycznych sesji). Podobnie małe wartości parametru δ przeznaczone są dla systemów opartych na innych mediach, (np. obraz, dźwięk, film), w których tekst odgrywa rolę drugorzędną (przykładowo, służy tylko i wyłącznie do zapewnienia nawigacji między dokumentami).
- Należy zachować daleko idącą ostrożność w manipulacji parametrem m . Podanie wartości większej niż 4 może powodować kłopoty z prezentacją rekomendowanej treści, co w najgorszym przypadku może wywołać dezorientację użytkownika (zbyt duża liczba odsyłaczy hipertekstowych utrudnia zrozumienie treści i poruszanie się po strukturze hipertekstu).

- Oszacowanie liczby dokumentów relewantnych dla przeciętnej sesji umożliwia zapobieżeniu przy pomocy parametrów ρ_s i γ spadkowi kompletności lub dokładności rekomendowanej treści.

Przyglądając się zależnościom związanym z parametrem α można dojść do mylnego wniosku, że najlepszym rozwiązaniem jest nadanie mu wartości bliskiej 0. Rzeczywiście, w wielu przypadkach wzrost wartości parametru α prowadzi do spadku dokładności, kompletności oraz zdolności adaptacyjnych rekomendacji, ale tylko i wyłącznie, przy założeniu, że użytkownik porusza się po dokumentach relewantnych. Jednakże struktura wielu systemów hipertekstowych (przykładowo witryn WWW) zmusza użytkownika do okazjonalnego odwiedzenia dokumentów nirelewantnych, które służą tylko i wyłącznie do przemieszczania się z jednej strony relewantnej na drugą. Istnieje także spore prawdopodobieństwo, że użytkownik może znaleźć się na stronie nirelewantnej przez zupełny przypadek. Wydaje się niezbyt trafnym by użytkownik po odwiedzeniu 25 stron dotyczących piłki nożnej otrzymywał rekomendację na temat nieruchomości tylko i wyłącznie dlatego, że ostatnio odwiedzony dokument zawierał wywiad z prezesem firmy budowlanej, który przypadkowo posiada to samo nazwisko co znany napastnik jednego z czołowych zespołów ekstraklasy. Twierdzenia 5.5.1 i 5.5.2 pokazują jak parametr α wpływa na dokładność rekomendacji w przypadku odwiedzenia dokumentu nirelewantnego.

Twierdzenie 5.5.1

Jeśli w bieżącej sesji użytkownik w kroku $i > 1$ odwiedził dokument nirelewantny, taki, że:

$$\forall_{d_r \in D_R^{(i)}} (sim(d^{(i)}, d_r) = 0 \wedge nv_{d_r}^{(i)} = 0)$$

i system hipertekstowy zawiera m_r dokumentów relewantnych, z których co najmniej m nie było w tej sesji rekomendowanych oraz spełnione są następujące nierówności

$$nv_{r \min} > nv_{n \max} \cdot \frac{1 - \alpha^i}{\alpha - \alpha^i}$$

$$sim_{r \min} > sim_{n \max} \cdot \frac{1 - \alpha^i}{\alpha - \alpha^i}$$

$$\forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D_R^{(i)}} q(d_r) \geq q(d_n)$$

to żaden dokument nirelewantny nie będzie zarekomendowany w kroku i .

Dowód:

Na początek dokonamy przekształceń założeń twierdzenia. Pomnożymy obie strony pierwszej nierówności przez zawsze dodatni parametr β , a obie strony drugiej nierówności przez zawsze dodatni parametr δ

$$\begin{aligned} \beta \cdot nv_{r \min} &> \beta \cdot nv_{n \max} \cdot \frac{1 - \alpha^i}{\alpha - \alpha^i} \\ \delta \cdot sim_{r \min} &> \delta \cdot sim_{n \max} \cdot \frac{1 - \alpha^i}{\alpha - \alpha^i} \end{aligned} \quad (5.5.1)$$

Dodając stronami nierówności z 5.5.1 otrzymujemy:

$$\beta \cdot nv_{r \min} + \delta \cdot sim_{r \min} > \beta \cdot nv_{n \max} \cdot \frac{1 - \alpha^i}{\alpha - \alpha^i} + \delta \cdot sim_{n \max} \cdot \frac{1 - \alpha^i}{\alpha - \alpha^i} \quad (5.5.2)$$

Aby w kroku i żaden dokument nierelevantny nie został zarekomendowany, niezależnie od parametru m , należy udowodnić, że w kroku i wartości funkcji rekomendacji f dla dokumentów nierelevantnych jest mniejsza niż dla dokumentów relevantnych:

$$\forall_{d_n \in D \setminus D_R^{(i)}, d_r \in D_R^{(i)}} f(d_r, s_b^{(i)}) > f(d_n, s_b^{(i)}) \quad (5.5.3)$$

Pokażemy, że 5.5.3 jest prawdziwe nawet, gdy dokumenty ze zbioru $D^{(i)}_R$ osiągną minimalne wartości funkcji rekomendacji zgodne z założeniami twierdzenia, a dokumenty ze zbioru $D \setminus D^{(i)}_R$ odznaczają się wartościami maksymalnymi.

Maksymalna wartość funkcji f dla dokumentu nierelevantnego d_n spełniającego warunek twierdzenia będzie osiągnięta przy założeniach, że we wszystkich krokach podobieństwo do dokumentu odwiedzanego jest maksymalna tzn. $sim(d_n, d^{(i)}) = sim_{n \max}$, podobnie współczynnik wzorców nawigacyjnych w kroku i powinien być maksymalny tzn. $nv_{d_n}^{(i)} = nv_{n \max}$, a wartość funkcji przydatności $q(d_n) = q(d_r)$. Oznacza to, że wartość funkcji rekomendacji dla dokumentu nierelevantnego spełnia nierówność:

$$f(d_n, s_b^{(i)}) \geq 0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot q(d_n) + \beta \cdot nv_{n \max} \cdot \sum_{k=1}^i (\alpha^{i-k}) + \delta \cdot sim_{n \max} \cdot \sum_{k=1}^i (\alpha^{i-k}) \quad (5.5.4)$$

$\sum_{k=1}^i (\alpha^{i-k}) = \sum_{k=1}^i (\alpha^{k-1})$ to suma i wyrazów ciągu geometrycznego o pierwszym wyrazie równym 1 i stałym ilorazie równym α , więc:

$$f(d_n, s_b^{(i)}) \geq 0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot q(d_n) + \beta \cdot nv_{n \max} \cdot \frac{1 - \alpha^i}{1 - \alpha} + \delta \cdot sim_{n \max} \cdot \frac{1 - \alpha^i}{1 - \alpha} \quad (5.5.5)$$

Z drugiej strony przy spełnionych założeniach twierdzenia można stwierdzić, że w kroku i istnieje co najmniej m dokumentów relevantnych, które nie były w bieżącej sesji ani odwiedzone, ani zignorowane. W takich warunkach zgodnie z założeniami twierdzenia dokument osiągnie minimalną wartość funkcji rekomendacji, gdy: we wszystkich krokach oprócz ostatniego współczynnik wzorców nawigacyjnych oraz podobieństwo tekstowe do dokumentów odwiedzanych będą osiągać wartości minimalne równe odpowiednio $sim_{r \min}$ i $nv_{r \min}$. Z założeń twierdzenia wiadomo, że wartości podobieństwa do dokumentu odwiedzanego w kroku i

oraz współczynnik wzorców nawigacyjnych są równe 0, więc spełniona jest następująca nierówność:

$$f(d_r, s_b^{(i)}) \geq 0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot q(d_r) + \beta \cdot \sum_{k=1}^{i-1} (\alpha^{i-k} \cdot nv_{r_{\min}}) + \delta \cdot \sum_{k=1}^{i-1} (\alpha^{i-k} \cdot sim_{r_{\min}}) \quad (5.5.6)$$

$\sum_{k=1}^{i-1} (\alpha^{i-k})$ to suma $i-1$ wyrazów ciągu geometrycznego o pierwszym wyrazie równym α i stałym ilorazie równym α , więc:

$$f(d_r, s_b^{(i)}) = 0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot q(d_r) + \beta \cdot nv_{r_{\min}} \cdot \alpha \cdot \frac{1 - \alpha^{i-1}}{1 - \alpha} + \delta \cdot sim_{r_{\min}} \cdot \alpha \cdot \frac{1 - \alpha^{i-1}}{1 - \alpha} \quad (5.5.7)$$

Podstawiając 5.5.5 i 5.5.7 do 5.5.3 otrzymujemy:

$$\begin{aligned} & 0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot q(d_r) + \beta \cdot nv_{r_{\min}} \cdot \alpha \cdot \frac{1 - \alpha^{i-1}}{1 - \alpha} + \delta \cdot sim_{r_{\min}} \cdot \alpha \cdot \frac{1 - \alpha^{i-1}}{1 - \alpha} \\ & > 0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot q(d_n) + \beta \cdot nv_{n_{\max}} \cdot \frac{1 - \alpha^i}{1 - \alpha} + \delta \cdot sim_{n_{\max}} \cdot \frac{1 - \alpha^i}{1 - \alpha} \end{aligned} \quad (5.5.8)$$

Łatwo zauważyć, że jeśli $\forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D_R^{(i)}} q(d_r) \geq q(d_n)$, to:

$$\begin{aligned} & 0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot q(d_n) + \beta \cdot nv_{n_{\max}} \cdot \frac{1 - \alpha^i}{1 - \alpha} + \delta \cdot sim_{n_{\max}} \cdot \frac{1 - \alpha^i}{1 - \alpha} \\ & \geq 0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot q(d_n) + \beta \cdot nv_{n_{\max}} \cdot \frac{1 - \alpha^i}{1 - \alpha} + \delta \cdot sim_{n_{\max}} \cdot \frac{1 - \alpha^i}{1 - \alpha} \end{aligned} \quad (5.5.9)$$

Czyli jeśli prawdziwe jest 5.5.8 prawdziwe też będzie:

$$\begin{aligned} & 0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot q(d_r) + \beta \cdot nv_{r_{\min}} \cdot \alpha \cdot \frac{1 - \alpha^{i-1}}{1 - \alpha} + \delta \cdot sim_{r_{\min}} \cdot \alpha \cdot \frac{1 - \alpha^{i-1}}{1 - \alpha} \\ & > 0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot q(d_r) + \beta \cdot nv_{n_{\max}} \cdot \frac{1 - \alpha^i}{1 - \alpha} + \delta \cdot sim_{n_{\max}} \cdot \frac{1 - \alpha^i}{1 - \alpha} \end{aligned} \quad (5.5.10)$$

Odejmując od obu stron $0,5 \cdot (\beta + \delta) \cdot \alpha^i \cdot q(d_r)$, a następnie mnożąc obie strony przez $\frac{1 - \alpha}{\alpha - \alpha^i}$, które jest zawsze dodatnie dla $i > 1$, otrzymujemy:

$$\beta \cdot nv_{r_{\min}} + \delta \cdot sim_{r_{\min}} > \beta \cdot nv_{n_{\max}} \cdot \frac{1 - \alpha^i}{\alpha - \alpha^i} + \delta \cdot sim_{n_{\max}} \cdot \frac{1 - \alpha^i}{\alpha - \alpha^i} \quad (5.5.11)$$

5.5.9 jest tożsamy z przekształceniem założeń otrzymanych w 5.5.2 wskutek czego można wnioskować, że twierdzenie jest prawdziwe. \square

Twierdzenie 5.5.1 oznacza, że jeśli współczynniki dokumentów relewantnych w poprzednich krokach były odpowiednio wysokie, to żaden dokument nirelewantny nie będzie zarekomendowany nawet po odwiedzeniu dokumentu nirelewantnego, który nie posiada żadnego deskryptora wspólnego z żadnym dokumentem relewantnym ($sim(d^{(i)}, d_r) = 0$) ani nie jest odwiedzany przez użytkowników zachowujących się podobnie do właściciela bieżącej sesji s_b . Przykładowo, dla $\alpha=0,5$, $sim_{nmax}=0,1$, $nv_{nmax}=0,1$ oraz $i=30$, żaden dokument nirelewantny nie będzie zarekomendowany, gdy $sim_{rmin}>0,4$ oraz $nv_{rmin}>0,4$.

Do przeanalizowania wpływu parametru α na minimalne wartości współczynników nawigacyjnego i tekstowego gwarantujących niezarekomendowanie dokumentów nirelewantnych przy przypadkowym odwiedzeniu dokumentu nirelewantnego użyta zostanie specjalnie na tą okazję zdefiniowana zależność funkcyjna:

$$pv_{\min}(\alpha) = \frac{1-\alpha^i}{\alpha-\alpha^i} \text{ gdzie } 1>\alpha>0$$

Twierdzenie 5.5.2

Funkcja: $pv_{\min}(\alpha) = \frac{1-\alpha^i}{\alpha-\alpha^i}$ gdzie $0<\alpha<1$ jest malejąca dla $i>1$

Dowód:

Policzmy pierwszą pochodną funkcji $pv_{\min}(\alpha)$, korzystając z pochodnej ilorazu otrzymujemy:

$$pv_{\min}'(\alpha) = \frac{-i\alpha^{i-1} \cdot (\alpha - \alpha^i) - (1 - \alpha^i)(1 - i\alpha^{i-1})}{(\alpha - \alpha^i)^2} \quad (5.5.12)$$

Aby funkcja, była malejąca pierwsza pochodna musi być ujemna, co pozwala zapisać:

$$\frac{-i\alpha^{i-1} \cdot (\alpha - \alpha^i) - (1 - \alpha^i)(1 - i\alpha^{i-1})}{(\alpha - \alpha^i)^2} < 0 \quad (5.5.13)$$

Mianownik nierówności 5.5.13 jest zawsze dodatni dla $i>1$, co oznacza, że musi być spełniona następująca nierówność:

$$-i\alpha^{i-1} \cdot (\alpha - \alpha^i) - (1 - \alpha^i)(1 - i\alpha^{i-1}) < 0 \quad (5.5.14)$$

Po dokonaniu prostych przekształceń otrzymujemy:

$$-i\alpha^i + i \cdot \alpha^{2i-1} - (1 - i\alpha^{i-1} - \alpha^i + i\alpha^{2i-1}) < 0 \quad (5.5.15)$$

Co daje dalej:

$$-i\alpha^i + i \cdot \alpha^{2i-1} - 1 + i\alpha^{i-1} + \alpha^i - i\alpha^{2i-1} < 0 \quad (5.5.16)$$

I ostatecznie:

$$-i\alpha^i - 1 + i\alpha^{i-1} + \alpha^i < 0 \quad (5.5.17)$$

Na początek zostanie udowodnione, że dla $i=2$ nierówność 5.5.17 będzie spełniona niezależnie od parametru α :

$$-2\alpha^2 - 1 + 2\alpha + \alpha^2 < 0 \quad (5.5.18)$$

Po prostych przekształceniach otrzymujemy:

$$-(\alpha^2 - 2\alpha + 1) < 0 \quad (5.5.19)$$

I ostatecznie:

$$-(\alpha - 1)^2 < 0 \quad (5.5.20)$$

Biorąc pod uwagę, że $1 > \alpha > 0$, można stwierdzić, że dla $i=2$ nierówność 5.5.17 będzie spełniona niezależnie od parametru α . Do udowodnienia całego twierdzenia wystarczy pokazać, że ciąg:

$b_i = -i\alpha^i - 1 + i\alpha^{i-1} + \alpha^i$ jest malejący czyli, że:

$$-(i+1)\alpha^{i+1} - 1 + (i+1)\alpha^i + \alpha^{i+1} - (-i\alpha^i - 1 + i\alpha^{i-1} + \alpha^i) < 0 \quad (5.5.21)$$

Po dokonaniu prostych przekształceń otrzymujemy:

$$-i \cdot \alpha^{i+1} - \alpha^{i+1} - 1 + i \cdot \alpha^i + \alpha^i + \alpha^{i+1} + i\alpha^i + 1 - i\alpha^{i-1} - \alpha^i < 0 \quad (5.5.22)$$

i ostatecznie:

$$-i\alpha^{i+1} + 2i\alpha^i - i\alpha^{i-1} < 0 \quad (5.5.23)$$

Dzieląc obie strony przez zawsze dodatnie $i\alpha^{i-1}$ mamy:

$$-\alpha^2 + 2\alpha - 1 < 0 \quad (5.5.24)$$

I ostatecznie:

$$-(\alpha - 1)^2 < 0 \quad (5.5.25)$$

Biorąc pod uwagę, że $1 > \alpha > 0$, można stwierdzić, że ciąg b_i jest malejący, a całe twierdzenie jest prawdziwe.

□

Twierdzenie 5.5.2 pokazuje, że duża wartość parametru α może zapobiec zarekomendowaniu dokumentów nierelevantnych, gdy użytkownik przez przypadek odwiedza dokument, który nie ma nic wspólnego z jego obecnymi preferencjami.

Dobór parametrów przedstawionych w tym podrozdziale nie jest zadaniem łatwym i zależy od charakterystyk systemu hipertekstowego. Przykładowo, jeśli system hipertekstowy posiada bardzo małą liczbę historycznych sesji wskazane jest, aby $\delta > \beta$.

Wartości zaproponowane w tabeli 5.5.2 powinny być traktowane jako domyślne, stanowiące punkt wyjściowy do dostosowania parametrów do indywidualnych potrzeb danego systemu hipertekstowego.

Parametr	Proponowana wartość	Komentarz
α	0,25	Przy tak dobranej wartości, praktycznie tylko trzy ostatnio odwiedzone dokumenty mają wpływ na wartość funkcji rekomendacji $0,25^4 < 0,004$. W systemach hipertekstowych, w których dokumenty dotyczące jednej tematyki nie są połączone bezpośrednimi odsyłaczami należy zwiększyć wartość tego parametru (patrz twierdzenie 5.5.1 i 5.5.2).
β	1	Zakłada się, że w systemie hipertekstowym istnieje dostateczna liczba sesji historycznych, które gwarantują, że utworzone wzorce nawigacyjne odzwierciedlają rzeczywiste zainteresowania użytkowników systemu.
δ	1	Zakłada się, że w używanym systemie hipertekstowym informacja tekstowa jest głównym środkiem przekazu.
γ	0,9	Już po 22 krokach fakt zignorowania dokumentu ma ograniczony wpływ na wartość funkcji rekomendacji ($0,9^{22} < 0,1$), a po 43 krokach można mówić o zaniku tego wpływu $0,9^{43} < 0,01$.
ρ_s	0,75	Już po 9 krokach fakt odwiedzenia dokumentu ma ograniczony wpływ na wartość funkcji rekomendacji ($0,75^9 < 0,1$), a po 17 krokach można mówić o zaniku tego wpływu $0,75^{17} < 0,01$.
m	2	Duża liczba dokumentów rekomendowanych użytkownikowi w jednym kroku może utrudnić prezentację rekomendacji i powodować dezorientację użytkownika

Tabela 5.5.2 Proponowane wartości parametrów dla metody rekomendacji *AdaptRank*.

Tabela 5.5.3 przedstawia wartości sim_{rmin} oraz nv_{rmin} , które przy wartościach parametrów z tabeli 5.5.2 zagwarantują:

- Maksymalną dokładność i kompletność metody *AdaptRank* przy założeniu, że liczba dokumentów relewantnych $m_r=30$ oraz $sim_{rmax}=0,1$ i $nv_{rmax}=0,1$.
- Niezarekomendowanie dokumentów nierelevantnych w przypadku odwiedzenia dokumentu nierelevantnego po odwiedzeniu 30 dokumentów relewantnych
- Dodatkowo podano maksymalną liczbę kroków potrzebnych do zarekomendowania dokumentu $d_{max}^{(i)}$ maksymalnie relewantnego od kroku i (patrz definicja 5.2.3), gdy preferencje użytkownika zmieniły się po 30 krokach i dokument maksymalnie relewantny spełnia następujące zależności:

$$\forall_{i>i} \left(\forall_{d_j \in D \setminus \{d_{max}^{(i)}\}} \left(nv_{max}^{(i)} - nv_j^{(i)} > 0,2 \wedge sim(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_{max}^{(i)}}) - sim(\mathbf{v}^{d^{(i)}}, \mathbf{v}^{d_j}) > 0,2 \right) \right)$$

Wielkość	Wyrażenie
Maksymalna dokładność, gdy użytkownik odwiedza rekomendowane dokumenty	$sim_{rmin}>0,54$ i $nv_{rmin}>0,54$
Maksymalna dokładność, gdy użytkownik ignoruje rekomendowane dokumenty	$sim_{rmin}>0,45$ i $nv_{rmin}>0,45$
Maksymalna kompletność, gdy użytkownik odwiedza rekomendowane dokumenty	$sim_{rmin}>0,27$ i $nv_{rmin}>0,27$
Maksymalna kompletność, gdy użytkownik ignoruje rekomendowane dokumenty	$sim_{rmin}>0,87$ i $nv_{rmin}>0,87$
Liczba kroków, po których zostanie zarekomendowany dokument maksymalnie relewantny od kroku i	2
Wyrażenie zapewniające nie zarekomendowanie dokumentu nierelevantnego w przypadku odwiedzenia dokumentu nierelevantnego	$sim_{rmin}>0,4$ i $nv_{rmin}>0,4$

Tabela 5.5.3 Zachowanie się metody *AdaptRank* przy wartościach parametrów określonych w tabeli 5.5.2

Tabela 5.5.3 pozwala określić zachowanie się metody *AdaptRank* przy wartościach parametrów określonych w tabeli 5.5.2. Analizując jej zawartość można wysunąć następujące wnioski:

- Gdy użytkownik odwiedza rekomendowane dokumenty, priorytetem przy tak dobranych parametrach jest różnorodność rekomendacji (o wiele łatwiej osiągnąć maksymalną kompletność niż dokładność).
- Gdy użytkownik ignoruje rekomendowane dokumenty, priorytetem przy tak dobranych parametrach jest nierekomendowanie dokumentów nierelevantnych (o wiele łatwiej osiągnąć maksymalną dokładność niż maksymalną kompletność).
- Dobrane wartości parametrów zapewniają kompromis pomiędzy szybką adaptacją metody rekomendacji do nowych potrzeb użytkownika i nierekomendowaniem dokumentów nierelevantnych w przypadku odwiedzenia dokumentu nierelevantnego (patrz 2 ostatnie wiersze tabeli 5.5.3).

6. Eksperymentalna weryfikacja metody rekomendacji

Celem rozdziału jest przedstawienie wyników eksperymentów dotyczących efektywności metody *AdaptRank*. Rozdział składa się z dwóch części. W pierwszej z nich opisane są badania symulacyjne nad efektywnością metody *AdaptRank*. W drugiej, po krótkiej prezentacji autorskiego systemu zawierającego implementację metody, opisane są eksperymenty pozwalające określić efektywność rekomendacji w dwóch rzeczywistych systemach hipertekstowych.

6.1. Symulacyjne badania efektywności

W rozdziale 5 pokazano, że metoda *AdaptRank* osiąga wyższą efektywność od swoich składowych, gdy:

- Podobieństwo dokumentów relewantnych do dokumentów odwiedzanych w bieżącej sesji jest odpowiednio duże i wyraźnie wyższe od podobieństwa dokumentów nirelewantnych do dokumentów odwiedzanych w bieżącej sesji.
- Wartości współczynników wzorców nawigacyjnych dokumentów relewantnych są odpowiednio duże i wyraźnie wyższe od współczynników wzorców nawigacyjnych dla dokumentów nirelewantnych.

Celem badań symulacyjnych jest określenie czy metoda *AdaptRank* osiąga efektywność wyższą od swoich składowych, gdy:

- Dokumenty relewantne będą cechować się minimalnie wyższym podobieństwem do dokumentów odwiedzonych w bieżącej sesji w porównaniu do podobieństwa między dokumentami nirelewantnymi i dokumentami odwiedzonymi.
- Dokumenty relewantne będą odznaczać się minimalnie wyższymi współczynnikami wzorców nawigacyjnych w porównaniu ze współczynnikami wzorców nawigacyjnych dla dokumentów nirelewantnych.

Do przeprowadzenia badań symulacyjnych użyto specjalnie na tą okazję stworzony program o nazwie *AdaptRank Symulator*, którego głównym zadaniem jest symulowanie działania metody *AdaptRank* oraz jej składowych zdefiniowanych w podrozdziale 4.2. Tabela 6.1.1 przedstawia parametry wejściowe programu.

Nazwa parametru	Opis
n	Liczba dokumentów w symulowanym systemie hipertekstowym (liczność zbioru D)
$numberRelevant$	Liczba dokumentów relewantnych w zbiorze D .
$sessionLength$	Długość generowanych sesji (liczba dokumentów odwiedzonych w jednej sesji)
$sessionNumber$	Liczba sesji wygenerowanych w jednym kroku.
$rmin$	Minimalna dopuszczalna wartość dla $sim(d^{(i)}, d)$ oraz $nv_d^{(i)}$ gdzie $n_{sb} \geq i > 0$ oraz $d \in D_R^{(i)}$
$rmax$	Maksymalna dopuszczalna wartość dla $sim(d^{(i)}, d)$ oraz $nv_d^{(i)}$ gdzie $n_{sb} \geq i > 0$ oraz $d \in D_R^{(i)}$
$nmin$	Minimalna dopuszczalna wartość dla $sim(d^{(i)}, d)$ oraz $nv_d^{(i)}$ gdzie $n_{sb} \geq i > 0$ oraz $d \in D \setminus D_R^{(i)}$
$nmax$	Maksymalna dopuszczalna wartość dla $sim(d^{(i)}, d)$ oraz $nv_d^{(i)}$ gdzie $n_{sb} \geq i > 0$ oraz $d \in D \setminus D_R^{(i)}$
m	Liczba dokumentów rekomendowanych w jednym kroku
$paramAlfa$	Parametr α metody <i>AdaptRank</i>
$paramBeta$	Parametr β metody <i>AdaptRank</i>
$paramDelta$	Parametr δ metody <i>AdaptRank</i>
$paramGama$	Parametr γ metody <i>AdaptRank</i>
$paramRoS$	Parametr ρ metody <i>AdaptRank</i>

Tabela 6.1.1 Parametry wejściowe programu *AdaptRank Symulator*.

Po uruchomieniu programu tworzone jest repozytorium dokumentów o liczności określonej parametrem n . Pierwsze *numberRelevant* dokumentów oznaczane są jako relewantne, a następnie dla każdego dokumentu generowana jest losowo wartość funkcji przydatności ($q(d)$) – patrz podrozdział 4.5.1). Po wyznaczeniu repozytorium dokumentów, program rozpoczyna symulowanie bieżących sesji użytkownika zgodnie z wybranym zadaniem. W aktualnej wersji programu zdefiniowano następujące zadania:

- **Badanie efektywności w zależności od długości sesji** – użytkownik określa dwa dodatkowe parametry: min i max , które odpowiadają minimalnej i maksymalnej wartości parametru *sessionLength*. W kolejnych krokach działania programu parametrowi *sessionLength* przypisywane są kolejne liczby całkowite z przedziału $[min, max]$. Dla każdej wartości parametru generowane są bieżące sesje, których liczba jest równa wartości parametru *sessionNumber*.
- **Badanie efektywności w zależności od liczby dokumentów relewantnych** - użytkownik określa dwa dodatkowe parametry: min i max , które odpowiadają minimalnej i maksymalnej wartości parametru *numberRelevant*. W kolejnych krokach działania programu parametrowi *numberRelevant* przypisywane są kolejne liczby całkowite z przedziału $[min, max]$. Dla każdej wartości parametru generowane są bieżące sesje, których liczba jest równa wartości parametru *sessionNumber*.
- **Badanie efektywności w zależności od liczby dokumentów w repozytorium** - użytkownik określa dwa dodatkowe parametry: min i max , które odpowiadają minimalnej i maksymalnej wartości parametru n . W kolejnych krokach działania programu parametrowi n przypisywane są kolejne liczby całkowite z przedziału $[min, max]$. Dla każdej wartości parametru generowane są bieżące sesje, których liczba jest równa wartości parametru *sessionNumber*.
- **Badanie efektywności w zależności od maksymalnej wartości współczynników dla dokumentów nierelevantnych** - użytkownik określa trzy dodatkowe parametry: min , max oraz $step$, które odpowiadają minimalnej i maksymalnej wartości parametru $nmax$ oraz przyrostowi wartości parametru $nmax$. W pierwszym kroku program przypisuje parametrom $nmax$ i $rmin$ wartość min . W kolejnych krokach wartości parametrów $nmax$ i $rmin$ wzrastają o wartość określoną parametrem $step$, aż do osiągnięcia wartości max . W każdym kroku generowane są bieżące sesje, których liczba jest równa wartości parametru *sessionNumber*.

Każda wygenerowana sesja składa się z *sessionLength* kroków. W pierwszym kroku sesji odwiedzany jest losowo wybrany dokument. W następnych krokach odwiedzany jest jeden z rekomendowanych dokumentów relewantnych. Jeśli wszystkie dokumenty rekomendowane w poprzednim kroku były nierelevantne, odwiedzany jest losowy dokument. Jednocześnie w każdym kroku oraz dla każdego odwiedzonego dokumentu generowane są wartości współczynników wzorców nawigacyjnych oraz podobieństwa do obecnie odwiedzanego dokumentu. Wartości te zawierają się w przedziale $[nmin, nmax]$ dla dokumentów nierelevantnych oraz $[rmin, rmax]$ dla dokumentów relewantnych. Poniżej przedstawiono wyniki przeprowadzonych badań. Należy zauważyć, że wartości parametrów związanych bezpośrednio z metodą *AdaptRank* są równe wartościom proponowanym w podrozdziale 5.5 (patrz tabela 5.5.2), dla przypomnienia: $m=2$, $\alpha=0,25$, $\beta=1$, $\delta=1$, $\gamma=0,9$ oraz $\rho_s=0,75$.

6.1.1. Efektywność w zależności od długości sesji

W pierwszym z przeprowadzonych eksperymentów zbadano, jakie wartości będą osiągnięte przez dokładność, kompletność oraz efektywność metody *AdaptRank* i jej składowych, gdy długość sesji określona parametrem *sessionLength* będzie zmieniać się od 1 do 50. Pozostałe parametry

wejściowe przyjmują następujące wartości: $numberRelevant=20$, $n=100$, $nmin=0$, $nmax=0,5$, $rmin=0,5$ oraz $rmax=1$.

Warto zauważyć, że powyższe wartości parametrów dopuszczają sytuacje, w których współczynniki dokumentów nierelevantnych i relevantnych będą miały taką samą wartość równą 0,5. Jednocześnie należy podkreślić, że dobór parametrów zapewnia maksymalną dokładność metod składowych we wszystkich eksperymentach przedstawionych poniżej, warunek $rmin=nmax$ powoduje, że prawdziwe będzie następujące wyrażenie:

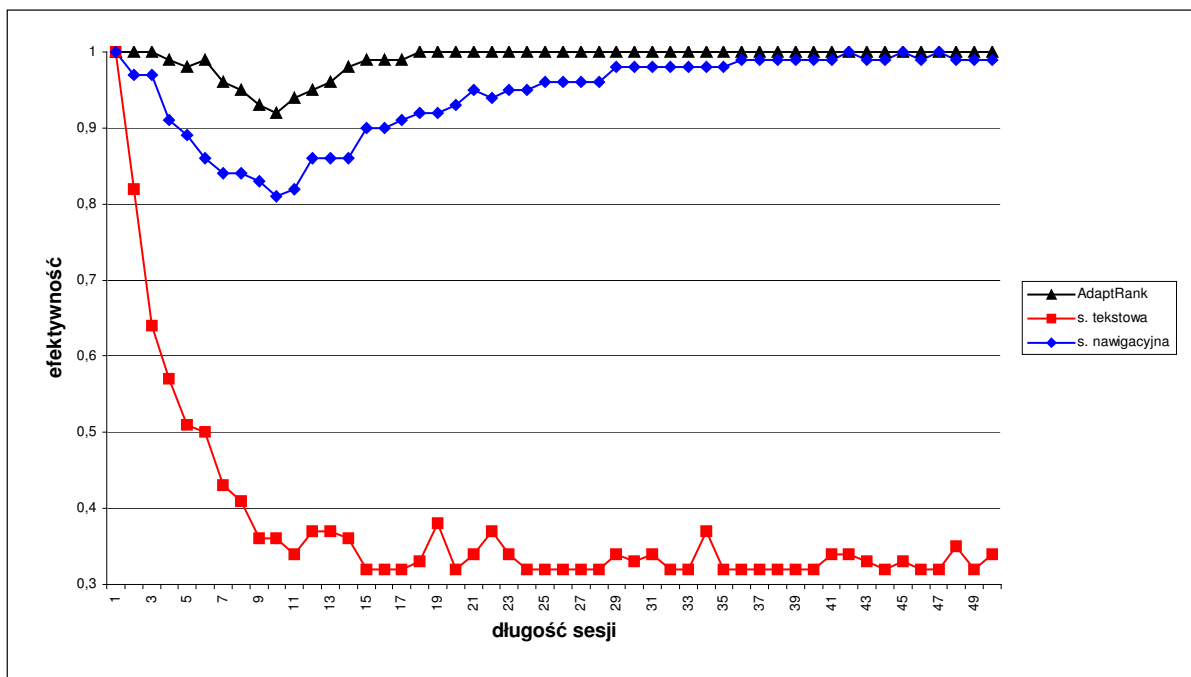
$$\forall_{n_{sb} \geq i > 0} \forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D \setminus D_R^{(i)}} (nv_{d_r}^{(i)} \geq nv_{d_n}^{(i)} \wedge sim(d_r, d^{(i)}) \geq sim(d_n, d^{(i)})) \quad (6.1.1)$$

Oznacza to, że metoda *AdaptRank*, aby wykazać się większą efektywnością będzie musiała odznaczać się znacząco wyższą kompletnością w porównaniu do metod składowych, przy jednoczesnym osiągnięciu dokładności zbliżonej do maksymalnej.

Rysunek 6.1.1 przedstawia wartości efektywności dla metody *AdaptRank* i jej składowych przy różnych długościach sesji. Wziąwszy pod uwagę, że trzy metody osiągnęły maksymalne wartości dokładności niezależnie od długości sesji, można stwierdzić, że na kształt krzywej z rysunku 6.1.1 wpływa tylko i wyłącznie osiągnięta kompletność (z tych właśnie powodów wykresy dokładności i kompletności zostały pominięte).

Niskie wartości efektywności metody składowej opartej na funkcji f_i (podobieństwo tekstowym) związane są z niskimi wartościami kompletności. Małe zróżnicowanie dokumentów rekomendowanych za pomocą tej metody jest spowodowane przypisywaniem wartościom funkcji f_i średniej arytmetycznej z podobieństwa danego dokumentu do wszystkich dokumentów odwiedzonych w tej sesji. Oznacza to, że dokument, który odznaczał się wysokimi podobieństwem do kilku odwiedzonych dokumentów będzie charakteryzował się względnie wysokimi wartościami funkcji f_i w kolejnych krokach, a w konsekwencji będzie często rekomendowany, przyczyniając się w znaczący sposób do obniżania kompletności.

Należy podkreślić, że niskie wartości kompletności dla tej składowej będą uwidaczniać się we wszystkich przeprowadzonych eksperymentach, co uzasadnia wprowadzenie modyfikacji opisanych w twierdzeniu 5.4.1 (modyfikacja ta pozwoliła stopniowo ograniczyć wpływ podobieństwa do dokumentu odwiedzanego w danym kroku na wartość funkcji rekomendacji w kolejnych krokach).

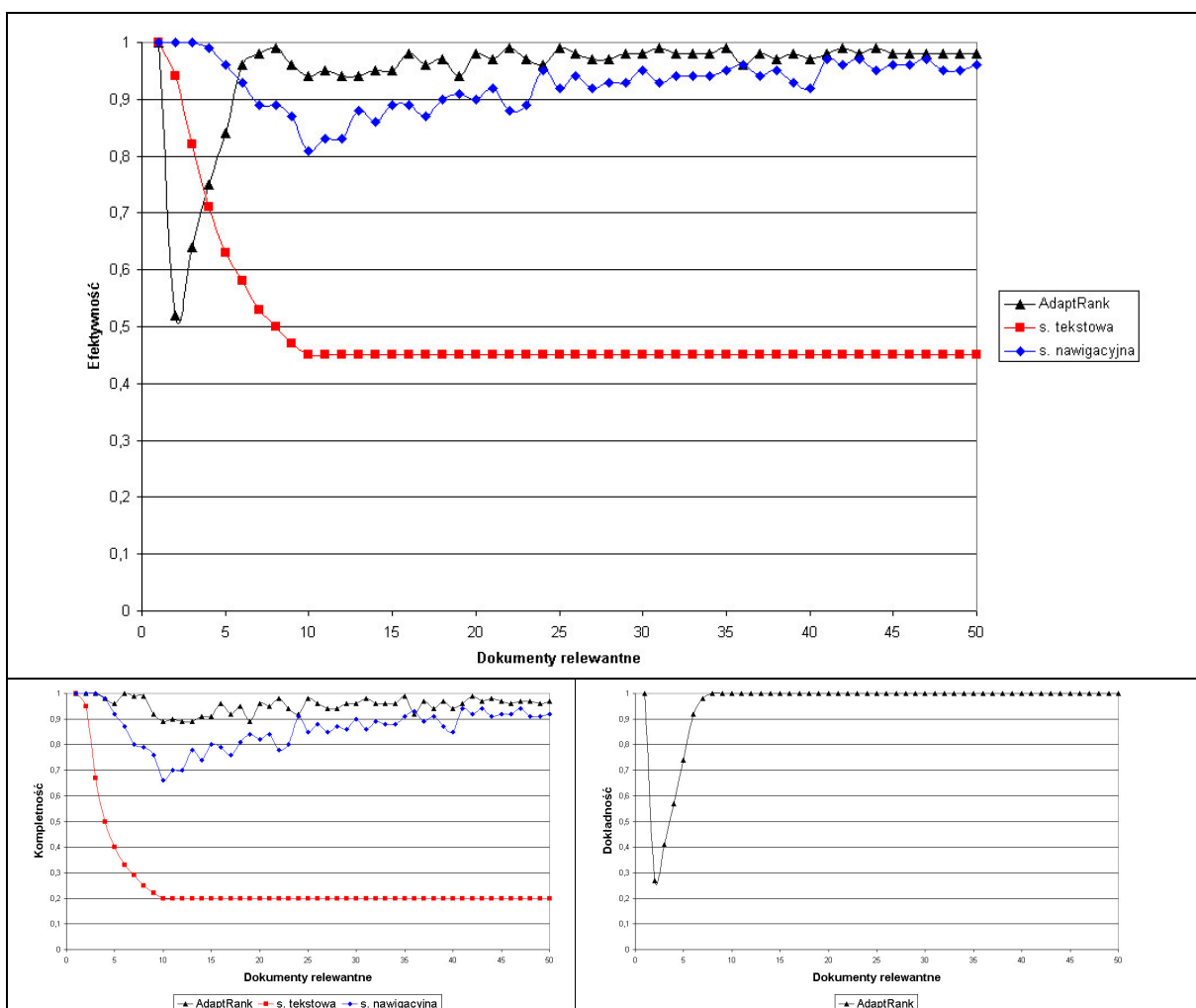


Rysunek 6.1.1 Zależność efektywności od długości sesji.

Kształt krzywych odpowiadających metodzie *AdaptRank* i jej składowej opartej na funkcji f_c (wzorcach nawigacyjnych) nie jest przypadkowy. Dla sesji bardzo krótkich (1-3 odwiedzanych dokumentów) obydwie metody bez trudu rekomendują różne dokumenty relewantne. Wraz ze wzrostem długości sesji kompletność się obniża, gdyż ponowne zarekomendowanie tego samego dokumentu jest coraz bardziej prawdopodobne. Dotyczy to szczególnie metod składowych, które nie posiadają żadnego mechanizmu zwiększającego różnorodność podpowiadanej treści. Oczywiście, gdy liczba dokumentów odwiedzanych jest wyraźnie wyższa od liczby dokumentów relewantnych, kompletność ponownie rośnie, gdyż jest coraz bardziej prawdopodobne, że wszystkie dokumenty relewantne zostaną zarekomendowane. Warto zauważyć, że metoda *AdaptRank* niezależnie od długości sesji odznacza się wyższą efektywnością w porównaniu do obydwu składowych.

6.1.2. Efektywność w zależności od numeru dokumentów relewantnych

W drugim z przeprowadzonych eksperymentów zbadano jak będzie zachowywać się dokładność, kompletność oraz efektywność metody *AdaptRank* i jej składowych, gdy liczba dokumentów relewantnych określona parametrem *numberRelevant* będzie zmieniać się od 2 do 50. Pozostałe parametry wejściowe posiadały następujące wartości: $n=100$, $nmin=0$, $nmax=0,5$, $rmin=0,5$ oraz $rmax=1$. W zależności od długości sesji rozpatrzono dwa przypadki: *sessionLength=5* oraz *sessionLength=40*. Na rysunku 6.1.2 przedstawiono wykresy odpowiadające wynikom eksperymentu dla sesji krótkich (długość sesji wynosiła 5 dokumentów). Na wykresie badającym wartości dokładności pominięto krzywe odpowiadającym metodom składowym, gdyż ich dokładność jest maksymalna niezależnie od numeru dokumentów relewantnych (wynika to z warunku $rmin=nmax$).



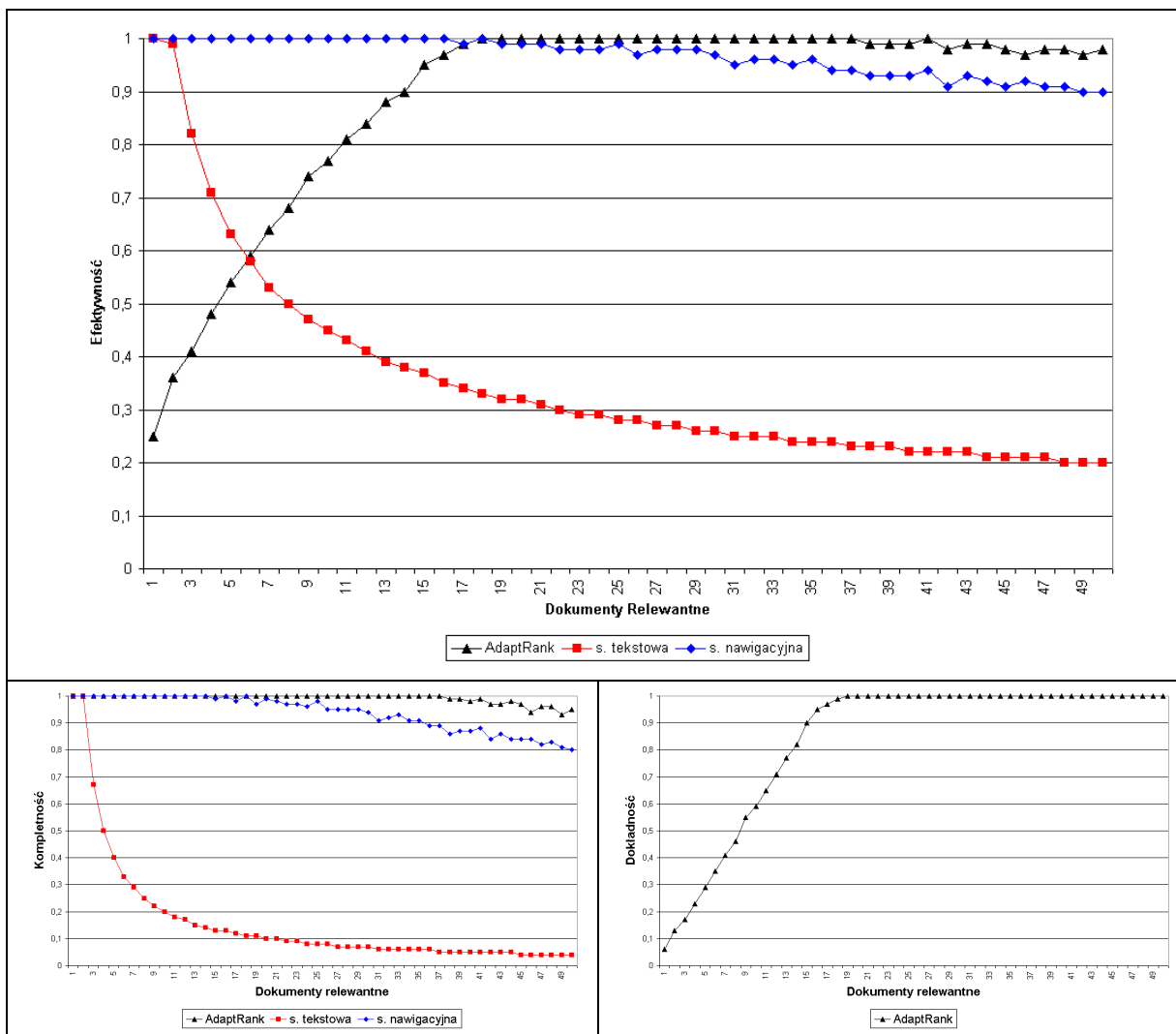
Rysunek 6.1.2 Zależność efektywności, kompletności i dokładności od liczby dokumentów relewantnych, dla krótkich sesji ($sessionLength=5$).

Analizując rysunek 6.1.2 łatwo dostrzec, że niezależnie od długości sesji metoda *AdaptRank* odznacza się wyższą kompletnością. Dodatkowo obserwuje się gwałtowne obniżenie dokładności metody *AdaptRank*, gdy zbiór dokumentów jest bardzo mały (od 2 do 4 elementów). Jest to związane z charakterystyką metody. Po zarekomendowaniu danego dokumentu relewantnego jego wartości funkcji rekomendacji w kolejnych kilku krokach ulegają obniżeniu. Mała liczba dokumentów relewantnych powoduje, że w pierwszych 2-3 krokach wszystkie dokumenty relewantne zostają zaproponowane użytkownikowi, co przekłada się na niskie wartości funkcji rekomendacji dla wszystkich dokumentów relewantnych w kolejnych krokach.

Gdy liczba dokumentów relewantnych jest dostatecznie duża (w tym przypadku większa niż 10) metoda *AdaptRank* odznacza się maksymalną dokładnością, dzięki wprowadzonemu mechanizmowi stopniowego zapominania. Powoduje to, że wartości funkcji rekomendacji dla dokumentów zarekomendowanych w pierwszych krokach sesji przestają być obniżane, a w konsekwencji w każdym kroku bieżącej sesji będą istniały co najmniej dwa dokumenty relewantne (2 to liczba dokumentów rekomendowanych w jednym kroku), które będą odznaczać się wyższymi wartościami funkcji f w porównaniu z wartościami tej samej funkcji dla dokumentów nirelewantnych.

Na rysunku 6.1.3 przedstawiono wykresy odpowiadające wynikom eksperymentu dla sesji długich (liczba odwiedzonych dokumentów wynosiła 40 dokumentów). Na wykresie badającym wartości dokładności pominięto krzywe odpowiadającym metodom składowym, ponieważ ich do-

kładność jest maksymalna niezależnie od numeru dokumentów relewanych (wynika to z warunku $rmin=nmax$).



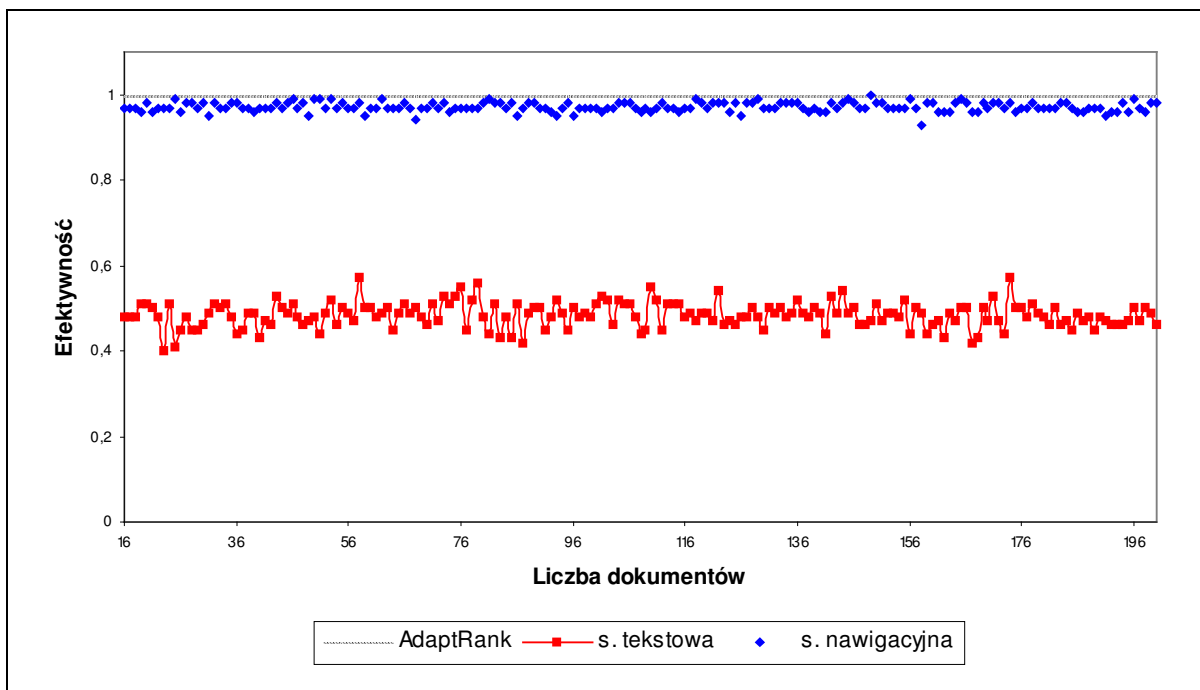
Rysunek 6.1.3 Zależność efektywności, kompletności i dokładności od liczby dokumentów relewanych, dla długich sesji ($sessionLength=40$).

Analizując rysunek 6.1.3 warto zauważyć, że w porównaniu do sesji krótkich, w sesjach długich potrzebna jest większa liczba dokumentów relewanych, aby metoda *AdaptRank* osiągnęła wysoką dokładność, a jednocześnie efektywność wyższą od efektywności składowych. Z drugiej jednak strony, do uzyskania maksymalnej dokładności, nie jest konieczne, aby liczba dokumentów relewanych była wyższa od liczby dokumentów odwiedzanych w danej sesji. Podobnie jak w przypadku krótkich sesji niska dokładność dla bieżących sesji odznaczających się małą liczbą dokumentów relewanych związana jest z czasowym osłabianiem wartości funkcji rekomendacji dla dokumentów podpowiedzianych w jednym z poprzednich kroków.

6.1.3. Efektywność w zależności od numeru dokumentów w systemie

W kolejnym eksperymencie przebadano jak będzie zmieniać się efektywność metody *AdaptRank* i jej dwóch składowych w zależności od liczebności zbioru dokumentów znajdujących się w systemie (parametr n przyjmował wartości od 16 do 200). Pozostałe parametry posiadały następujące wartości: $nmin=0$, $nmax=0,5$, $rmin=0,5$ oraz $rmax=1$, $sessionLength=20$, $numberRelevant=15$. Należy zauważyć, że minimalna wartość parametru $n=16$ zdeteminowana jest przez wartość parametru

numberRelevant. W wykonanym eksperymencie zarówno metoda *AdaptRank* jak i jej składowe osiągnęły maksymalną dokładność niezależnie od liczby dokumentów w systemie. Oznacza to, że kształty krzywych dotyczących kompletności i dokładności są bardzo podobne (gdy dokładność jest równa jedności efektywność jest równa pierwiastkowi kwadratowemu z kompletności). Dlatego też na rysunku 6.1.4 przedstawiono wyłącznie zależność efektywności od liczby dokumentów w zbiorze D .



Rysunek 6.1.4 Zależności efektywności metody *AdaptRank* i jej składowych od liczby dokumentów zaindeksowanych w systemie.

Analizując rysunek 6.1.4 można zauważyć, że nie ma wyraźnej zależności pomiędzy dokładnością, kompletnością i efektywnością metody, a liczbą dokumentów w systemie. Jest to zgodne z intuicją, gdyż nie liczba wszystkich dokumentów ma wpływ na wartości dokładności i kompletności, tylko liczba dokumentów relewantnych. Wyniki eksperymentu są także zbieżne z analizą formalną, gdyż warunki podane w podrozdziale 5.3 a określające kiedy metoda *AdaptRank* osiąga wyższą efektywność od efektywności jej składowych także nie zależały od liczby dokumentów znajdujących się w systemie.

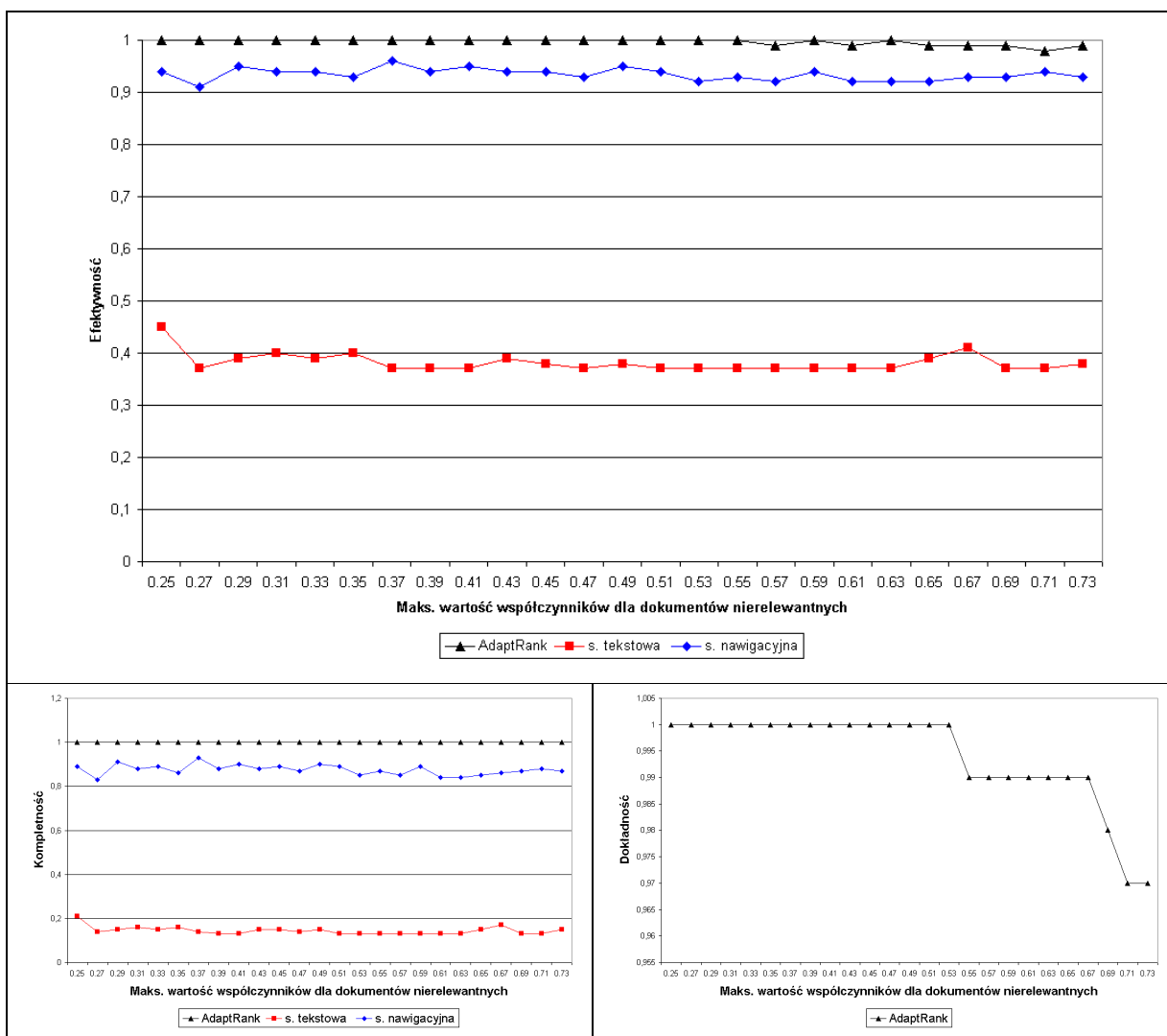
Innymi słowy: nie jest ważne czy dany system zawiera 30 czy 3000 dokumentów nerelewantnych, jeśli prawdziwy jest warunek: $\forall_{n_{sb} \geq i > 0} \forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D \setminus D_R^{(i)}} (nv_{d_r}^{(i)} \geq nv_{d_n}^{(i)} \wedge sim(d_r, d^{(i)}) \geq sim(d_n, d^{(i)}))$,

to dokumenty nerelewantne nie zostaną nigdy zarekomendowane wtedy i tylko wtedy, gdy istnieje dostateczna liczba dokumentów relewantnych. Warto też podkreślić, że w powyższym eksperymencie metoda *AdaptRank* niezależnie od liczby dokumentów zawsze osiągała maksymalną efektywność.

6.1.4. Efektywność w zależności od maksymalnej wartości współczynników dla dokumentów nerelewantnych

W kolejnym eksperymencie przebadano jak będzie zmieniać się efektywność metody *AdaptRank* i jej dwóch składowych w zależności od maksymalnej wartości współczynnika wzorców nawigacyjnych dla dokumentów nerelewantnych i maksymalnej wartości podobieństwa pomiędzy do-

kumentami odwiedzionymi, a dokumentami nierelevantnymi. Obydwie wielkości są zdeterminowane wartością parametru $nmax$. Pozostałe parametry wejściowe posiadały następujące wartości: $nmin=0$, $n=100$, oraz $rmax=1$, $sessionLength=20$, $numberRelevant=15$. Należy zauważyć, że warunek: $\forall_{n_{sp} \geq i > 0} \forall_{d_r \in D_R^{(i)}} \forall_{d_n \in D \setminus D_R^{(i)}} (nv_{d_r}^{(i)} \geq nv_{d_n}^{(i)} \wedge sim(d_r, d^{(i)}) \geq sim(d_n, d^{(i)}))$ obligował do jednoczesnej zmiany wartości parametru $rmin$. W eksperymencie tym przyjęto, że w każdym kroku $nmax=rmin$. Rysunek 6.1.5 przedstawia wyniki eksperymentu. Warto zauważyć, że maksymalna wartość współczynników dokumentów nierelevantnych nie ma większego wpływu na efektywność, dokładność i kompletność metod składowych. W przypadku metody *AdaptRank* dla bardzo wysokich wartości $nmax$ zanotowano minimalny spadek dokładności. Na wykresie badającym wartości dokładności pominięto krzywe odpowiadającym metodom składowym, gdyż ich dokładność jest maksymalna niezależnie od wartości parametru $nmax$ (wynika to z warunku $rmin=nmax$).



Rysunek 6.1.5 Zależności efektywności, dokładności i kompletności metody *AdaptRank* oraz jej składowych od maksymalnych wartości współczynników dokumentów nierelevantnych.

6.1.5. Podsumowanie wyników badań symulacyjnych

Powyższe eksperymenty pokazały, że gdy współczynniki dokumentów relevantnych (podobieństwo do dokumentów odwiedzionych w bieżącej sesji oraz współczynnik wzorców nawigacyjnych) są tylko minimalnie wyższe od analogicznych współczynników dla dokumentów nierele-

wantnych, w większości przypadków efektywność rekomendacji metody *AdaptRank* jest wyższa od efektywności jej składowych. Innymi słowy, gdy bieżąca sesja nie spełnia warunków opisanych w rozdziale 5, metoda *AdaptRank* także w wielu przypadkach odznacza się wyższą efektywnością w porównaniu do jej składowych.

Warto podkreślić, że założenia eksperymentu ($rmin=nmax$) przekładały się na maksymalną dokładność metod składowych niezależnie od zmian rozpatrywanych parametrów. W konsekwencji metoda *AdaptRank* do uzyskania wyższej efektywności, nie tylko musiała odznaczyć się wyższą kompletnością, ale także dokładnością zbliżoną do maksymalnej. Należy zauważyć, że we wszystkich eksperymentach metoda *AdaptRank* osiągnęła kompletność nie mniejszą (a częstokroć znacznie większą) niż jej metody składowe. Metoda *AdaptRank* nie uzyskała wyższej efektywności, tylko gdy liczba dokumentów relewantnych do potrzeb użytkownika bieżącej sesji była bardzo mała w stosunku do długości bieżącej sesji.

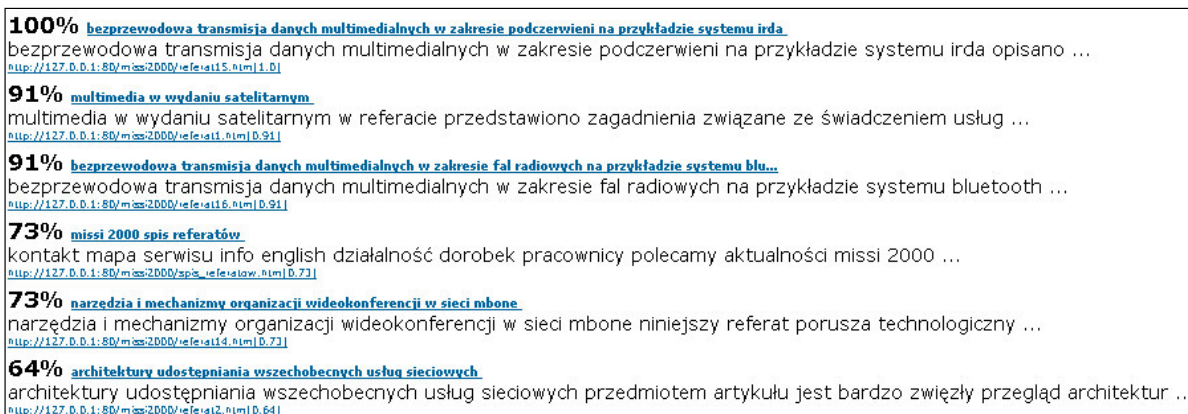
6.2. Zachowanie się metody w rzeczywistych systemach hipertekstowych

6.2.1. Implementacja metody *AdaptRank* w ramach systemu ROSA

Do zbadania zachowania się metody *AdaptRank* w rzeczywistych systemach hipertekstowych użyto autorskiego systemu *ROSA* (ang. *Remote Open Site Agent*), który zawiera pełną implementację metody *AdaptRank* oraz jej składowych. Główną ideą tej aplikacji jest optymalne wykorzystanie dwóch rodzajów danych związanych z konkretną witryną internetową. Pierwszy z nich to treść dokumentów, a drugi aktywność użytkowników. Aplikacja *ROSA* wykonuje się po stronie serwera i spełnia szereg pożytecznych funkcji. Do najważniejszych z nich należy zaliczyć:

- Rekomendację dokumentów relewantnych do potrzeb użytkownika według wielu zaimplementowanych metod. Jedną z nich jest *AdaptRank*.
- Rekomendację produktów z bazy danych. Szczegółowy algorytm, według którego rekomendowane są produkty został opisany w (Kazienko i Kiewra, 2004a).
- Wyszukiwanie informacji – system zawiera standardową wyszukiwarkę
- Statystyki obrazujące aktywność użytkowników (najczęściej odwiedzane strony, dokumenty, przez które użytkownicy wchodzą do systemu itp.)

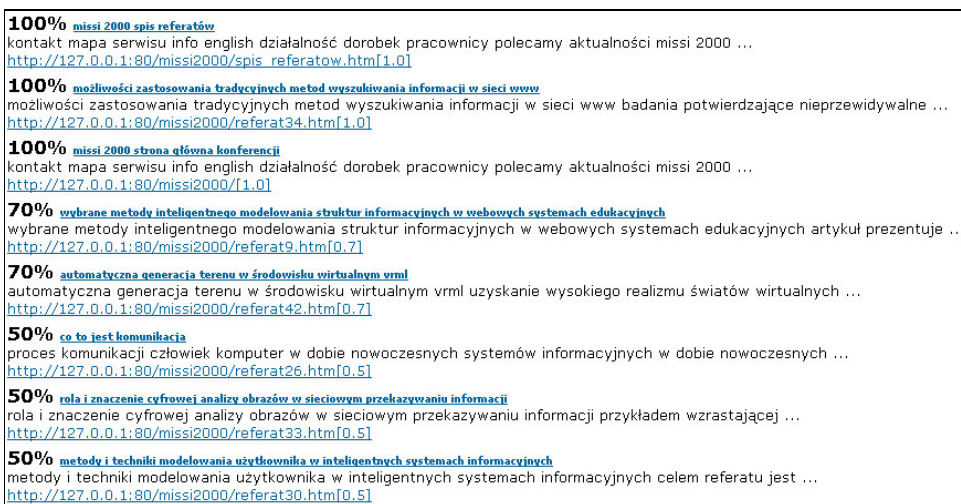
Opis jednej z pierwszych wersji aplikacji *ROSA* znajduje się w (Kazienko i Kiewra, 2003b). Na szczególną uwagę zasługuje eksplorator grup (wzorców nawigacyjnych), opisany po raz pierwszy w (Kiewra, 2005a) - patrz rysunek 6.2.1.



Rysunek 6.2.1 Eksplorator grup pozwala na szybka analize zawartosci poszczegolnych grup

Idea dzialania eksploratora grup jest bardzo prosta. Przedstawia on ranking dokumentow (w postaci tytulow wraz z krótkim streszczeniem) uporządkowanych malejaco ze względu na wartosc wspórzędnej rozpatrywanego wzorca nawigacyjnego, która odpowiada konkretnemu dokumentowi z prezentowanego rankingu (sposób wyznaczania wartosci wzorców nawigacyjnych podany zostal w definicji 4.5.2). Rysunek 6.2.1 przedstawia wzorec nawigacyjny, w którym uzytkownicy interesowali się bezprowodowa transmisja danych oraz uslugami sieciowymi.

Na podobnej zasadzie dziala eksplorator algorytmu *AdaptRank*, który w kazdym kroku biezacej sesji nie tylko pozwala „podejrzec” ranking dokumentow przeznaczonych do zarekomendowania (uporządkowany wedlug wartosci funkcji rekomendacji w ostatnim kroku biezacej sesji), ale takze umozliwia sledzenie zawartosci wektora dokumentow odwiedzonych oraz wektora dokumentow zignorowanych. Podobnie jak w przypadku eksploratora grup, dokumenty posortowane sa malejaco wedlug wartosci odpowiadajacych im wspórzędnych. Pierwsze *m* dokumentow przedstawionych na rysunku 6.2.2 to dokumenty, które zostaly zarekomendowane w aktualnym kroku biezacej sesji.



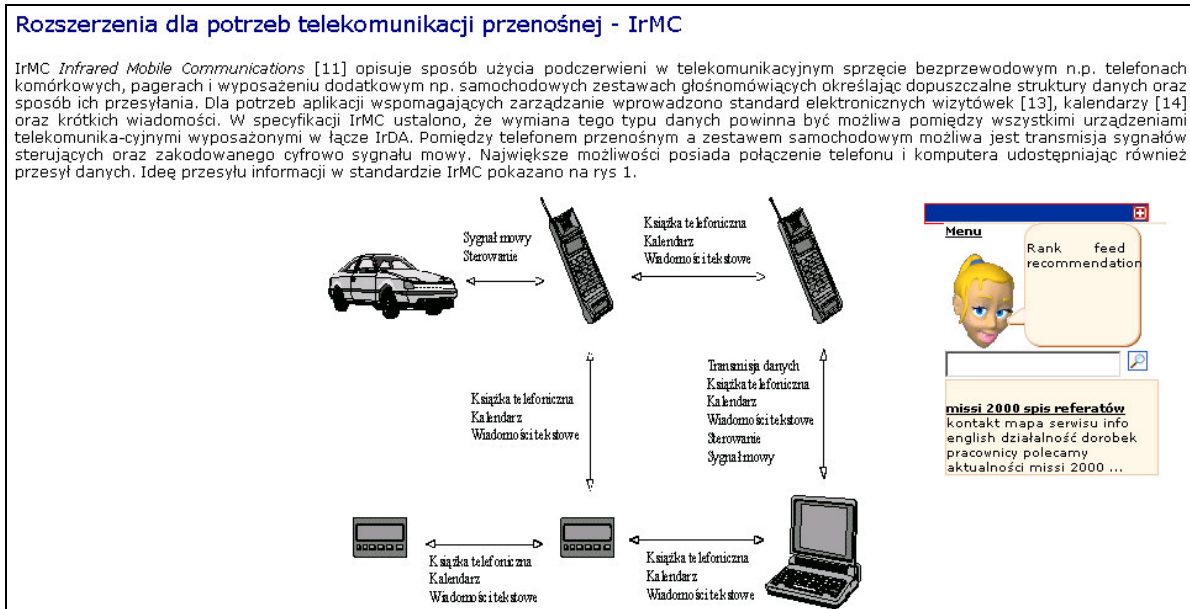
Rysunek 6.2.2 Za pomoca narzedzi w systemie ROSA w kazdym kroku mozna sprawdzic jak wyglada ranking dokumentow uporządkowany za pomoca wartosci funkcji rekomendacji.

System ROSA posiada architekture rozproszona, a kazdy z jego moduLOW traktowany jest jako inteligentny agent. Do implementacji metody *AdaptRank* uzyto nastepujacych agentow:

- *Szperacz internetowy (ang, crawler)* – odpowiedzialny jest za polaczenie z witryna za pomoca protokolu HTTP (WWW, 2005h) i stworzenie listy dokumentow skladajacych

się na zbiór D . Dodatkowo agent ten dla każdego „rozpoznanego dokumentu” tworzy zbiór terminów w nim występujących.

- *Zarządca treści* (ang. *content manager*) – na podstawie informacji uzyskanych przez Szperacza internetowego indeksuje dokumenty wybierając ze zbioru terminów podzbiór deskryptorów. Dodatkowo, dla każdego dokumentu d tworzy odpowiadający mu wektor \mathbf{v}^d .
- *Monitor sesji* (ang. *session monitor*) – jego głównym zadaniem jest przechwytywanie żądań pobrania dokumentów i łączenie ich w sesje w celu stworzenia wektora $\mathbf{z}^{(i)}$ (dokumentów odwiedzonych do kroku i) oraz wektora $\mathbf{g}^{(i)}$ (dokumentów zignorowanych do kroku i). Monitor sesji jest także odpowiedzialny za późniejsze przekształcenie sesji bieżących w sesje historyczne.
- *Eksplorator zachowań* (ang. *usage miner*) – jest odpowiedzialny za wyznaczenie wzorców nawigacyjnych na podstawie sesji historycznych utworzonych przez Monitora sesji.
- *Rekomendator odsyłaczy* (ang. *hyperlink recommender*) – zawiera implementację fazy interakcji metody *AdaptRank*.
- *Asystent użytkownika* (ang. *user assistant*) – jego zadaniem jest komunikacja z użytkownikiem, a ściślej mówiąc, przedstawianie rekomendacji dokumentów w postaci przewijalnej listy zawierającej tytuł i streszczenie podpowiadanego dokumentu (rysunek 6.2.3).



Rysunek 6.2.3 Integracja rekomendacji w obrębie przesłanego dokumentu ma postać przewijalnej listy, w której rekomendowane dokumenty zawierają tytuł i krótkie streszczenie

Warto dodać, że system zawiera też innych agentów, którzy nie uczestniczą w procesie rekomendacji *AdaptRank* (przykładowo *Rekomendator Produktów*, *Zarządca Statystykami* itp.). Technologie agencje umożliwiły wprowadzenie algorytmu aktualizacji danych, który został szczegółowo opisany w (Kazienko i Kiewra, 2003b). Algorytm ten zakładał, że poszczególne wielkości potrzebne do przeprowadzenia rekomendacji traktowane są jako swoista wiedza odpowiedzialnego za nie agenta. Przykładowo, pojawienie się nowych dokumentów w systemie hipertekstowym wykrywane przez szperacza internetowego wiąże się z aktualizacją:

- Wektorów deskryptorów posiadanych przez *Zarządcę treści*

- Wektorów historycznych sesji zarządzanych przez *Monitora sesji*
- Wzorców nawigacyjnych tworzonych przez *Eksploratora zachowań*

Należy zauważyć, że nawet w przypadku systemu hipertekstowego posiadającego kilkaset dokumentów aktualizacja wielkości posiadanych przez poszczególnych agentów może być operacją czasochłonną, dlatego zaistniała potrzeba zaprojektowania wyżej wspomnianego algorytmu, którego głównym zadaniem jest zapewnienie optymalnego użycia zasobów posiadanych przez każdego agenta (moc obliczeniowa, pamięć).

6.2.2. Opis przeprowadzonych eksperymentów

Do zbadania efektywności metody *AdaptRank* wykorzystano dwie witryny (serwisy internetowe) o następującej charakterystyce:

- Serwis internetowy Zakładu Systemów Informacyjnych Politechniki Wrocławskiej. W obrębie tego serwisu zaindeksowano 4784 dokumentów. Z pliku logów wyodrębniono 11.661 sesji historycznych. W witrynie tej zawartość tekstowa jest głównym źródłem informacji. Na uwagę zasługuje duża rozbieżność tematyczna. Witryna zawiera, między innymi, referaty z organizowanych konferencji, strony domowe studentów oraz materiały dydaktyczne. Zaindeksowane dokumenty napisane są w 3 językach: polskim (większość), angielskim oraz niemieckim.
- Serwis intranetowy hiszpańskiego oddziału Fujitsu, w obrębie którego zaindeksowano 6347 dokumentów. Z pliku logów wyodrębniono 19.036 sesji historycznych. Na uwagę zasługuje fakt, że witryna ta zawiera bardzo małą ilość informacji tekstowej. Dokumenty HTML stanowią bowiem tylko szkielet dający dostęp do zasobów zapisanych w innych formatach. Dokumenty napisane są w 2 językach: hiszpańskim (zdecydowana większość) oraz angielskim.

Dla obydwu witryn wybrano 10 przykładowych sesji bieżących, w których użytkownik usiłował znaleźć interesujące go informacje na z góry określony temat. Dla każdej sesji wyselekcjonowano ciąg adresów URL odpowiadających kolejno odwiedzanym dokumentom. Przy selekcji tej założono, że jeśli użytkownik nie mógł znaleźć relewantnego odsyłacza w aktualnie oglądanym dokumencie, to do wyboru następnego dokumentu używano wyszukiwarki dostępnej w systemie *ROSA*.

Po selekcji sesji obliczono wszystkie wielkości potrzebne do rekomendacji w czterech następujących metodach:

- Metoda *AdaptRank* – na podstawie analizy wpływu wartości parametrów na rezultaty metody przyjęto: $\alpha=0,25$, $\beta=1$, $\delta=1$, $\rho_s=0,75$, $\gamma=0,9$, $m=2$.
- Metoda składowa używająca funkcji f_t - opisana w podrozdziale 4.2.1.
- Metoda składowa oparta na wzorcach nawigacyjnych (używająca funkcji f_c) zaprezentowana w podrozdziale 4.2.2.
- Dokumenty odwiedzane razem – w tym przypadku rekomendowane były dokumenty najczęściej odwiedzane wraz z obecnie oglądanym dokumentem

6.2.3. Wyniki eksperymentów

Na skutek grupowania historycznych sesji utworzono 20 grup. Tabela 6.2.1 przedstawia krótką charakterystykę każdej z nich.

6.2.3.1. Serwis Zakładu Systemów Informacyjnych

Lp	Opis
1	Działalność naukowa i dydaktyczna pracowników ZSI
2	Projekt ROSA
3	Informacje na temat GPS
4	Wrocławskie mosty (serwis studencki)
5	Historia Wrocławia (serwis studencki)
6	Informacje ogólne o Wrocławiu (serwis studencki)
7	Kurs języka HTML
8	Zakupy we Wrocławiu (serwis studencki)
9	Informacje o Wrocławiu (w języku niemieckim)
10	Referaty z konferencji MISSI2000
11	Turystyka w okolicach Wrocławia (serwis studencki)
12	Referaty dotyczące nowoczesnych rozwiązań multimedialnych
13	Informacje dla studentów
14	Przegląd najważniejszych funkcji systemu ROSA (prezentacja)
15	Strony domowe pracowników Zakładu
16	Referaty z konferencji MISSI2004
17	Telekonsultacje
18	Informacje na temat organizowanych konferencji
19	Strony poświęcone konferencji MMIS 2004
20	Działalność dydaktyczna ZSI (w języku angielskim)

Tabela 6.2.1 Grupy odpowiadające typowym sesjom historycznym

Efektywność poszczególnych metod rekomendacji w każdej z 10 przykładowych sesji przedstawiona jest w tabeli 6.2.2

Opis sesji	Dług.	AdaptRank			Wzorce nawigacyjne			Podobieństwo tekst.			Wspólnie odwiedzane		
		D	K	E	D	K	E	D	K	E	D	K	E
Kurs HTML	10	0,70	0,60	0,65	0,10	0,10	0,10	1,00	0,50	0,71	0,00	0,00	0,00
Informacje dydaktyczne	8	1,0	0,75	0,87	0,00	0,00	0,00	1,00	0,50	0,71	0,88	0,38	0,57
ROSA tour	12	1,00	0,67	0,82	0,67	0,17	0,33	0,67	0,58	0,62	0,83	0,83	0,83
Informacje o Wrocławiu	8	0,88	0,88	0,88	0,00	0,00	0,00	0,75	0,75	0,75	0,75	0,63	0,68
Informacje o wykładowcach	8	0,75	0,75	0,75	0,63	0,38	0,48	1,00	0,63	0,79	0,38	0,13	0,22
Wstęp do języka JAVA	6	0,67	0,50	0,58	0,00	0,00	0,00	1,00	0,50	0,71	1,00	0,33	0,58
Komunikacja bezprzewodowa	12	0,92	0,92	0,92	1,00	0,50	0,71	0,92	0,75	0,83	0,92	0,50	0,68
Informacje o konferencji MMIS2004	4	1,00	1,00	1,00	1,00	0,50	0,71	0,50	0,50	0,50	1,00	0,75	0,87
Studencki klub naukowy	8	1,00	1,00	1,00	0,50	0,13	0,25	1,00	0,63	0,79	0,88	0,25	0,47
Działalność naukowa ZSI	8	1,00	0,88	0,94	1,00	0,25	0,50	1,00	0,50	0,71	0,50	0,13	0,25
Rezultaty końcowe		0,89	0,79	0,84	0,49	0,20	0,31	0,88	0,58	0,71	0,71	0,39	0,51

Tabela 6.2.2 Dokładność (kolumna oznaczona przez *D*), kompletność (kolumna oznaczona przez *K*), efektywność (kolumna oznaczona przez *E*) odpowiadająca 10 przykładowym sesjom z Zakładu Systemów Informacyjnych

Należy podkreślić, że grupy wyznaczone przez system ROSA (patrz tabela 6.2.1) zostały wygenerowane niezależnie od przykładowych sesji wyselekcjonowanych przez autora (pierwsza kolumna tabeli 6.2.2).

6.2.3.2. Serwis intranetowy

Na skutek grupowania historycznych sesji utworzono 27 grup. Tabela 6.2.3 przedstawia krótką charakterystykę każdej z nich.

Numer	Opis
1	Instrukcje wewnętrzne dla pracowników
2	Pomoc dla klientów i konserwacja produktów
3	Informacje ogólne o wytwarzaniu aplikacji
4	Dział kadr i oferty dla pracowników
5	Najnowsze wiadomości i komunikaty prasowe
6	Informacje na temat rozwiązań sieciowych dla klientów
7	Wielkie serwery Fujitsu
8	Informacje dla handlowców
9	Partnerzy strategiczni
10	Zarządzanie jakością i ochrona środowiska
11	Najnowsze informacje kwiecień 2004
12	Filozofia korporacji
13	Ankiety dla pracowników i klientów
14	Ceny sprzętu według kategorii (skanery, dyski twarde, drukarki itp.)
15	Imprezy i spotkania firmowe
16	Informacje dla klientów i prasy
17	Najnowsze wiadomości: czerwiec 2004
18	Najnowsze wiadomości lipiec 2004
19	Informacje na temat działu: Rozwiązania Sieciowe
20	Zasoby dotyczące przygotowywania ofert dla administracji publicznej
21	Informacje działu marketingu
22	Najnowsze informacje: październik 2004
23	Informacje ogólnoswiatowe w języku angielskim
24	Informacje ogólne działu usług
25	Informacje ogólne styczeń 2005
26	Kursy uzupełniające dla pracowników
27	Informacje na temat rozwiązań opartych na procesorach Intel

Tabela 6.2.3 Grupy odpowiadające typowym sesjom historycznym

Efektywność poszczególnych metod rekomendacji w każdej z 10 przykładowych sesji przedstawiona jest w tabeli 6.2.4

Opis sesji	Dług.	AdaptRank			Wzorce nawigacyjne			Podobieństwo tekst.			Wspólnie odwiedzane		
		D	K	E	D	K	E	D	K	E	D	K	E
Zarządzanie projektem	5	0,7	0,7	0,7	0	0	0	0,9	0,4	0,6	0,8	0,2	0,4
Aktualności	5	1	1	1	0	0	0	1	1	1	1	0,5	0,71
Sprzęt	5	1	0,7	0,84	1	0,4	0,63	0,8	0,3	0,49	0,8	0,5	0,63
Zdjęcia z imprez firmowych	5	0,2	0,2	0,2	0	0	0	0,1	0,1	0,1	0,1	0,1	0,1
Dział kadr	8	0,63	0,63	0,63	0,5	0,5	0,5	0,56	0,56	0,56	0,81	0,25	0,45
Centrala w Mardrycie	5	0,9	0,7	0,79	0,8	0,3	0,49	0,5	0,5	0,5	0,7	0,4	0,53
Współpraca z Microsoft	3	0,33	0,33	0,33	0,67	0,17	0,33	0,33	0,33	0,33	0,67	0,3	0,47
Zarządzanie jakością	3	1	1	1	0,67	0,17	0,33	1	0,67	0,82	0,67	0,33	0,47
Oferty dla pracowników	10	0,4	0,4	0,4	0,2	0,1	0,14	0,3	0,3	0,3	0,6	0,1	0,24
Dział integracji systemów	10	0,1	0,1	0,1	0	0	0	0,1	0,1	0,1	0,4	0,2	0,28
Rezultaty końcowe		0,63	0,58	0,6	0,38	0,16	0,24	0,56	0,43	0,48	0,65	0,29	0,43

Tabela 6.2.4 Dokładność (kolumna oznaczona przez *D*), kompletność (kolumna oznaczona przez *K*), efektywność (kolumna oznaczona przez *E*) odpowiadająca 10 przykładowym sesjom z Intranetu Fujitsu

Także w tym przypadku grupy wyznaczone przez system *ROSA* (patrz tabela 6.2.3) zostały wygenerowane niezależnie od przykładowych sesji wyselekcjonowanych przez autora (pierwsza kolumna tabeli 6.2.4).

6.2.3.3. Wyniki końcowe

Witryna	AdaptRank			Wzorce nawig.			Podobieństwo tekst			Odwiedzane razem		
	D	K	E	D	K	E	D	K	E	D	K	E
ZSI	0,8	0,79	0,84	0,49	0,2	0,31	0,88	0,58	0,71	0,71	0,39	0,51
Intranet	0,6	0,58	0,6	0,38	0,16	0,24	0,56	0,43	0,48	0,65	0,29	0,25

Tabela 6.2.5 Średnie: dokładność (kolumna oznaczona przez *D*), kompletność (kolumna oznaczona przez *K*), efektywność (kolumna oznaczona przez *E*) odpowiadająca witrynie ZSI oraz Intranetowi Fujitsu.

Metoda	ZSI			Intranet			Średnia		
	D	K	E	D	K	E	D	K	E
Najlepsza z pozostałych	0,88	0,58	0,71	0,65	0,43	0,48	0,72	0,51	0,6
AdaptRank	0,89	0,79	0,84	0,63	0,58	0,6	0,76	0,69	0,72
Procentowa poprawa	1,14	36,21	18,31	-3,08	34,88	25,00	5,56	35,29	20,00

Tabela 6.2.6 Metoda AdaptRank na tle najlepszych wyników „konkurencyjnych” metod (wyniki zaokrąglone są do dwóch miejsc po przecinku)

6.2.4. Ocena rezultatów i interpretacja wyników

Wyniki badań potwierdziły przypuszczenia. Metoda *AdaptRank* wykazała się najwyższą efektywnością w obydwu witrynach. Zgodnie z przewidywaniami na wzrost efektywności w głównej mierze miała wpływ wyższa kompletność (ponad 30%). Co ważne, mimo wyraźnego wzrostu kompletności, nie tylko nie zauważono obniżenia dokładności, lecz wręcz przeciwnie - odnotowano jej wzrost. Mała ilość informacji tekstowej w przypadku badanego serwisu intranetowego przyczyniła się do wyraźnie słabszej efektywności wszystkich metod rekomendacji (szczególnie jednak tych, które opierają się na podobieństwie tekstowym). Warto też podkreślić, że przeprowadzone eksperymenty pokazały, że metodę *AdaptRank* można z powodzeniem zaimplementować i efektywnie użytkować (zainstalowane metody rekomendacji nie spowodowały zauważalnych opóźnień przy pobieraniu dokumentów).

Największym zaskoczeniem są słabe wyniki metody opartej na wzorcach nawigacyjnych. Przyczyn tak słabego zachowania należy szukać w typowych dla tego podejścia wadach opisanych w rozdziale 3 (rozrzedzenie danych, problem nowych dokumentów, „nieprzystawalność” bieżącej sesji użytkownika do żadnego z wykrytych wzorców). Warto podkreślić, że w porównaniu do rekomendacji bazującej na wspólnie odwiedzonych dokumentach, rekomendacja oparta na wzorcach nawigacyjnych potrzebuje więcej czasu, aby „zauważyć” nowy dokument. Jest to w pełni wytłumaczalne mając na uwadze, że aby dany stał się „dobrym” reprezentantem danego wzorca nawigacyjnego dokument ten powinien zostać odwiedzony wiele razy i zawsze w towarzystwie, co najmniej kilku tych samych dokumentów.

Podczas eksperymentów uwidoczniła się także wyraźna przewaga rekomendacji opartej na podobieństwie tekstowym nad rekomendacją dokumentów odwiedzanych razem. Nie powinno to dziwić wzięwszy pod uwagę fakt, że w rekomendacji dokumentów odwiedzanych razem promowana jest strona główna serwisu oraz strony główne odpowiednich sekcji witryny (np. lista pracowników, zestawienie dostępnych kursów itp.). Przykładowo, w przypadku serwisu ZSI ponad 30% użytkowników rozpoczyna swoją wizytę w witrynie od strony głównej, a następnie przechodząc przez główną stronę danej sekcji dochodzi do interesującej ich informacji.

Podsumowując wyniki eksperymentu należałoby spojrzeć raz jeszcze na wzorce nawigacyjne utworzone w serwisie intranetowym, gdyż 5 z 27 grup utworzonych jest z historycznych sesji

użytkowników, którzy czytali najnowsze wiadomości z danego miesiąca (kwiecień 2004, czerwiec 2004, lipiec 2004, październik 2004, styczeń 2005). Grupy te związane są pośrednio ze specyficznym zachowaniem administratorów, którzy nie aktualizują stron z najnowszymi wiadomościami tylko zmieniają odsyłacze do najnowszych wiadomości. Przykładowo w maju 2005 strona główna zawierała odsyłacz do dokumentu *maj2005.html* jako do strony zawierającej najświeższe wiadomości. Po upływie miesiąca nie zmieniając treści dokumentu *maj2005.html*, dodano nową stronę o nazwie *czerwiec2005.html*, do której skierowano odsyłacz ze strony głównej. Oczywiście, tworzenie wzorców nawigacyjnych opisujących odwiedzanie mocno nieaktualnych wiadomości mija się z celem. Aby zapobiec tworzeniu się takich wzorców należałoby wprowadzić mechanizm stopniowego zapominania do procesu selekcji historycznych sesji. Innymi słowo historyczne sesje z ostatniego okresu czasu powinny mieć większy wpływ na tworzone wzorce nawigacyjne.

7. Podsumowanie

W niniejszej pracy przedstawiono hybrydową metodę rekomendacji o nazwie *AdaptRank*, która integruje dwie metody składowe. Pierwsza z nich oparta jest na analizie historycznych sesji, w których użytkownicy zachowywali się podobnie do właściciela bieżącej sesji. Druga zaś bazuje na podobieństwie tekstowym między dokumentami (proponowane są dokumenty posiadające dużą liczbę deskryptorów występujących także w ostatnio odwiedzonych dokumentach). Metoda *AdaptRank* ujęta została w formalny model rekomendacji zdefiniowany przez autora. Przedstawiony model zawiera usystematyzowanie i uporządkowanie pojęć związanych z rekomendacją pojawiających się w literaturze.

W rozdziale 5 zbadano formalne własności zaprezentowanej metody. W szczególności pokazano wiele przypadków, w których zaproponowana metoda daje wyższą efektywność niż metody składowe. Do najważniejszych z nich należy zaliczyć:

- Metoda hybrydowa zapewnia maksymalną efektywność rekomendacji, gdy podobieństwo dokumentów relewantnych do dokumentów odwiedzanych w bieżącej sesji jest odpowiednio duże i wyraźnie wyższe od podobieństwa dokumentów nierelewantnych do dokumentów odwiedzanych w bieżącej sesji a wartości współczynników wzorców nawigacyjnych dokumentów relewantnych są odpowiednio duże i wyraźnie wyższe od współczynników wzorców nawigacyjnych dla dokumentów nierelewantnych. Dodatkowo w pracy udowodniono, że metody składowe nie uzyskują maksymalnej efektywności przy takich samych założeniach.
- Metoda *AdaptRank* potrafi zapewnić maksymalną efektywność nie tylko wtedy, gdy użytkownik odwiedza rekomendowane dokumenty relewantne, ale także wtedy, gdy je ignoruje. Podobnie jak w poprzednim przypadku wymagane jest, by podobieństwo dokumentów relewantnych do dokumentów odwiedzanych w bieżącej sesji było odpowiednio duże i wyraźnie wyższe od podobieństwa dokumentów nierelewantnych do dokumentów odwiedzanych w bieżącej sesji a wartości współczynników wzorców nawigacyjnych dokumentów relewantnych były odpowiednio duże i wyraźnie wyższe od współczynników wzorców nawigacyjnych dla dokumentów nierelewantnych.. Analogicznie do poprzedniego przypadku, udowodniono, że metody składowe nie uzyskują maksymalnej efektywności przy takich samych założeniach.
- Uwzględnienie kilku najbliższych wzorców nawigacyjnych poprawia dokładność rekomendacji (więc pośrednio także efektywność), gdy aktualne zainteresowania użytkownika dotyczą kilku tematów.
- Integracja dwóch metod rekomendacji wykorzystujących różne źródła danych pozwala zapewnić wyższą efektywność od efektywności rekomendacji opartej tylko na podobieństwie dokumentów, gdy dokumenty należące do systemu hipertekstowego zawierają małą liczbę deskryptorów.
- Podobnie, metoda *AdaptRank* uzyskuje wyższą efektywność od metody rekomendacji opartej na wzorcach nawigacyjnych, gdy użytkownik odwiedza nowe dokumenty, które nie były jeszcze odwiedzane tak często, by uzyskać silną reprezentację w choć jednym wzorcu nawigacyjnym.

- Uwzględnienie wartości funkcji rekomendacji z poprzednich kroków bieżącej sesji pozwala zarekomendować dokument relewantny do potrzeb użytkownika nawet, gdy przez pomyłkę został odwiedzony dokument nierelevantny.

Do innych zadań zrealizowanych w ramach formalnych badań własności metody rekomendacji *AdaptRank* należy zaliczyć:

- Ocenę własności adaptacyjnych metody – w pracy pokazano, między innymi, że metoda *AdaptRank* jest w stanie w skończonej liczbie kroków dostosować się do nowych potrzeb informacyjnych użytkownika.
- Analizę własności metod składowych – w ramach tego zadania pokazano, jakie korzyści zostały osiągnięte dzięki zmianom wprowadzonym do metod składowych znanych z literatury.
- Przeanalizowanie wpływu wartości parametrów na zachowanie metody - w szczególności pokazano jak wzrost wartości odpowiednich parametrów wpływa na dokładność, kompletność, efektywność i zdolności adaptacyjne metody *AdaptRank*.

W praktycznej części pracy zrealizowano dwa rodzaje eksperymentów. Pierwszy polegał na symulacyjnym badaniu efektywności hybrydowej metody w porównaniu z jej metodami składowymi. Drugi zaś dotyczył badania efektywności metody w rzeczywistych witrynach.

Do przeprowadzenia badań symulacyjnych wykorzystano specjalnie na tą okazję napisany program o nazwie *AdaptRank Symulator*. Badania te zademonstrowały, że jeśli współczynniki dokumentów relewantnych (podobieństwo do dokumentów odwiedzonych w bieżącej sesji oraz współczynniki wzorców nawigacyjnych) są równe lub tylko minimalnie wyższe od analogicznych współczynników dla dokumentów nierelevantnych, to w większości przypadków efektywność rekomendacji metody *AdaptRank* jest wyższa od efektywności jej składowych. Innymi słowy, badania symulacyjne pokazały, że założenia, dla których formalnie udowodniono wyższą efektywność metody *AdaptRank* w porównaniu do jej składowych nie są jedyne, w których metoda *AdaptRank* odznacza się wyższą efektywnością. Należy zauważyć, że założenia badań symulacyjnych zapewniały maksymalną dokładność metod składowych we wszystkich przypadkach. W konsekwencji metoda *AdaptRank* w celu uzyskania wyższej efektywności, nie tylko musiała odznaczyć się wyższą kompletnością, ale także dokładnością zbliżoną do maksymalnej.

Biorąc pod uwagę analizę formalną i badania symulacyjne można stwierdzić, że metoda *AdaptRank* w większości przypadków uzyskuje wyższą efektywność od efektywności metod składowych w bieżących sesjach, w których metody składowe działają poprawnie tzn. gdy:

- Liczba deskryptorów wspólnych dla dokumentów relewantnych i odwiedzonych w bieżącej sesji jest większa od liczby deskryptorów wspólnych dla dokumentów nierelevantnych i dokumentów odwiedzonych w bieżącej sesji. Warunek ten przekłada się na wyższą wartość podobieństwa między dokumentami odwiedzonymi i relewantnymi od wartości podobieństwa dokumentów odwiedzonych do dokumentów nierelevantnych.
- Dokumenty relewantne w odczuciu użytkownika i odwiedzane w bieżącej sesji były wspólnie odwiedzane w historycznych sesjach należących do innych użytkowników (właściciel bieżącej sesji zachowuje się podobnie jak inni użytkownicy o podobnych preferencjach). Warunek ten oznacza, że współczynniki wzorców nawigacyjnych dla dokumentów relewantnych będą wyższe od tych samych współczynników dla dokumentów nierelevantnych.

Eksperymenty na rzeczywistych witrynach nie tylko potwierdziły wyraźnie lepszą efektywność metody *AdaptRank* od jej metod składowych, ale także pokazały, że implementacja metody nie nastęrcza większych problemów. Co ważne, wzrost kompletności był wyraźnie wyższy od wzrostu dokładności.

Na wysoką efektywność przedstawionej metody miały wpływ nowatorskie rozwiązania użyte w pracy, a mianowicie:

- Wektor dokumentów zignorowanych – jego współrzędne obniżają wartość funkcji rekomendacji dla dokumentów zarekomendowanych, lecz zignorowanych przez użytkownika, co pozwala zwiększyć kompletność rekomendacji. Jednoczesne zastosowanie mechanizmu „stopniowego zapominania” daje możliwość ponownego podpowiedzenia tego samego dokumentu, gdy wszystkie dokumenty relewantne zostaną zarekomendowane co najmniej raz w ramach bieżącej sesji.
- Współczynnik wzorców nawigacyjnych – jego wprowadzenie poprawia dokładność rekomendacji dla użytkowników, których zainteresowania znajdują się „na przecięciu” kilku wzorców nawigacyjnych.
- Użycie wektora dokumentów odwiedzonych do obniżenia wartości funkcji rekomendacji dla dokumentów pobranych w bieżącej sesji także przyczyniło się do poprawy kompletności. Podobnie jak w przypadku wektora dokumentów zignorowanych wprowadzono mechanizm „stopniowego zapominania”. Warto dodać, że koncepcja ta pojawiła się po raz pierwszy w artykule (Kazienko i Kiewra, 2003a), którego twórca niniejszej pracy był współautorem.

Uwagi wymaga także wprowadzenie początkowego rankingu przydatności, dzięki któremu podczas pierwszych interakcji z systemem, gdy profil użytkownika nie jest jeszcze znany dużą rolę w rekomendacji odgrywa jakość i popularność dokumentu (nie znając potrzeb użytkownika system rekomendujący proponuje dokumenty często odwiedzane i odznaczające się wysoką jakością).

Biorąc pod uwagę zunifikowane kryteria klasyfikacji systemów rekomendujących (patrz rozdział 2) należy stwierdzić, że zaproponowana metoda rekomendacji wykorzystuje najnowsze trendy rozwoju systemów rekomendujących dostępne w literaturze (anonimowość użytkownika, brak konieczności bezpośredniej oceny prezentowanej treści, zdolności adaptacyjne itp.)

Przedstawiona metoda, biorąc pod uwagę jej pełną implementację, może być punktem wyjścia do wielu nowych badań. Do najważniejszych należy zaliczyć:

- Adaptacja zastosowanego rozwiązania (szczególnie mechanizmów zwiększających kompletność) do innych typów rekomendacji (produkty w sklepie internetowym, najnowsze wiadomości itp.).
- Zastosowanie mechanizmu stopniowego zapominania do sesji historycznych, dzięki czemu sesje zarejestrowane dawniej miałyby mniejszy wpływ na dokonywaną rekomendację od sesji zarejestrowanych w ostatnim okresie czasu.
- Użycie zdefiniowanej metody w środowisku rozproszonym – umożliwiłaby to wymienianie informacji na temat potrzeb użytkowników w obrębie kilku lub nawet kilkunastu systemów hipertekstowych o podobnej tematyce.
- Wykorzystanie informacji znajdującej się w bieżącej sesji (wektor dokumentów odwiedzonych, wektor dokumentów zignorowanych, współczynnik wzorców nawigacyjnych) do poprawy porządku rezultatów wyszukiwania.

8. Literatura

- Adomavicius G., Tuzhilin A. (2003): *Recommendation technologies: Survey of current methods and possible extensions*. Working paper, Stern School of Business, New York University, New York
- Adomavicius G., Sankaranarayanan R., Sen S., Tuzhilin A. (2005): *In-corporating Contextual Information in Recommender Systems Using a Multidimensional Approach* ACM Transactions on Information Systems 23(1), s. 103-145
- Aggarwal C. C., Wolf J. L., Wu K. L., Yu P. S. (1999): *Horting Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering*. W: the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, s. 201–212
- Balabanovic M, Shoham Y (1997): *Fab: Content-based, collaborative recommendation*. Communications of the ACM, 40(3), s. 66-72
- Basu C., Hirsh H., Cohen W. (1998): *Recommendation as Classification: Using Social and Content-Based Information in Recommendation*. W: the 15th National Conference on Artificial Intelligence, Madison, WI, s. 714-720
- Billsus D, Pazzani M. J. (1999): *A Hybrid User Model for News Classification*. W: UM'99, Wien, New York: Springer-Verlag, s. 99–108
- Bodoff D. (2004): *Relevance models to help estimate document and query parameters*. ACM Transactions on Information Systems 22(3), ACM Press, s. 357-380
- Bookstein A. (1979): *Relevance*. Journal of the American Society for Information Science, 30(5), s. 269–273
- Breese J. S, Heckerman D, Kadie C. (1998): *Empirical analysis of predictive algorithms for collaborative filtering*. W: the Fourteenth Conference on Uncertainty in Artificial Intelligence, Madison, s. 43-52
- Brin S., Page L. (1998): *The anatomy of a large-scale hypertextual web search engine*, W: the 7th International World Wide Web Conference, Brisbane, Australia, 1998, s. 107-117
- Burke R. (2002): *Hybrid recommender systems: Survey and experiments*. W: User Modeling and User-Adapted Interaction 12(4), s. 331-370
- Bush V. (1945): *As we may think*. The Atlantic Monthly, 176(1), s. 101-108
- Carroll J. Rosson M. B. (1987): *The Paradox of the Active User*. In *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*, J. M. Carroll (Ed.), Cambridge, MA: MIT Press, s. 26–28.
- Chen L. Sycara K. (1998): *Webmate: A Personal Agent for Browsing and Searching*. In Proceedings of AGENTS '98, ACM, s. 132–139
- Claypool M., Gokhale A, Miranda T., Murnikov P., Netes D., Sartin M. (1999): *Combining Content-Based and Collaborative Filters in an Online Newspaper*. W: SIGIR '99 Workshop on Recommender Systems: Algorithms and Evaluation. Berkeley
- Cooley R., Tan P. N., Srivastava J. (1999): *WebSift: The Web Site Information Filter System*. W 1999 KDD Workshop on Web Mining. San Diego, CA: Springer-Verlag

- Froehlich T. J. (1994): *Relevance reconsidered—towards an agenda for the 21st century: introduction to special topic issue on relevance research*. Journal of the American Society for Information Science 45(3), s. 124-134
- Getoor L. Sahami M. (1999): *Using probabilistic relational models for collaborative filtering*. W: Workshop on Web Usage Analysis and User Profiling (WEBKDD'99), s. 103 - 145
- Goker M. H., Thompson C. A. (2000): *The Adaptive Place Advisor: A Conversational Recommendation System*. W: the 8th German Workshop on Case Based Reasoning, Lammerbuckel, Germany, s. 130-135
- Goldberg D, Nichols D, Oki B. M, Terry D. (1992): *Using Collaborative Filtering to Weave an Information Tapestry*. Communications of the ACM 35, s. 61–70
- Goldberg K., Roeder T., Gupta, D., Perkins C. (2001), *Eigentaste: A Constant Time Collaborative Filtering Algorithm*, Information Retrieval Journal, 4(2), s. 133-151
- Hoffman T. (2003): *Collaborative Filtering via Gaussian Probabilistic Latent Semantic Analysis*, W: Annual ACM Conference on Research and Development in Information Retrieval archive Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval table of contents, Toronto, Canada, s. 259 - 266
- Heine M.H. (1973): *Distance between sets as an objective measure of retrieval effectiveness*, Information Storage and Retrieval, 9, s. 181-198
- Kamba T, Bharat K. Albers M. C. (1995): *The Krakatoa Chronicle – an Interactive, Personalized, Newspaper on the Web*. W: the Fourth International World Wide Web Conference, s. 159–170
- Kazienko P., Kiewra M., (2003a): *Link Recommendation Method Based on Web Content and Usage Mining*, W: the International IIS: IIPWM'03 Conference, Advances in Soft Computing, Springer Verlag, s. 529-534
- Kazienko P., Kiewra M. (2003b): *ROSA - Multi-agent System for Web Services Personalization*. W: First Atlantic Web Intelligence Conference Proceedings, LNAI 2663, Springer Verlag, s. 297-306
- Kazienko P., Kiewra M. (2004a), *Integration of Relational Databases and Web Site Content for Product and Page Recommendation*. W 8th International Database Engineering & Applications Symposium, IEEE Computer Society, s. 111-116
- Kazienko P., Kiewra M. (2004b), *Personalized Recommendation of Web Pages*. Rozdział 10 w: Nguyen T. (ed.) Intelligent Technologies for Inconsistent Knowledge Processing. Advanced Knowledge International, Adelaide, South Australia, s. 163-183
- Kiewra M. (2005a): *Non-invasive Discovering of Web User Preferences for Data Warehouse Purposes*. Akceptowany rozdział do publikacji w: Ontologies and Soft Methods in Knowledge Management (Katarzyniak R.P., Ed.). Advanced Knowledge International, Adelaide, South Australia
- Kiewra M. (2005b): *RankFeed - Recommendation as Searching without Queries: New Hybrid Method of Recommendation*. Artykuł w: Journal of Universal Computer Science 11(2), Springer-Verlag , s. 229-249
- Kiewra M. (2005c): *Iterative Discovering of User's Preferences Using Web Mining*. Artykuł w: International Journal of Computer & Science Applications 2(2), s. 57-66

- Kiewra M., Nguyen N.T. (2005): *Non-textual Document Ranking Using Crawler Information and Web Usage Mining*. W KES 2005, Lecture Notes in Artificial Intelligence, s. 520-529
- Kleinberg J. (1999): *Authoritative sources in a hyperlinked environment*, Journal of the ACM, Vol. 46, No. 5, s. 604-632
- Kłopotek A. (2001): *Inteligente wyszukiwarki internetowe*, Akademicka Oficyna Wydawnicza EXIT
- Lalmas M. (1998): *Logical models in information retrieval: Introduction and overview*. Information Processing & Management, 34(1), s.19-33
- Krulwich, B., Burkey C. (1996): *Learning User Information Interests Through Extraction of Semantically Significant Phrases*. W: the AAAI Spring Symposium on Machine Learning in Information Access. Stanford, CA
- Krulwich B. (1997): *LifeStyle Finder: Intelligent User Profiling Using Large-Scale Demographic Data*, AI Magazine 18(2), s. 37-45
- Lewis D. D., Gale W. A. (1994): *A sequential algorithm for training text classifiers*. W: SIGIR94, 17th ACM International Conference on Research and Development in Information Retrieval, s. 3-12
- Li J., Zaiane O. R. Zaiane (2004): *Combining Usage Content and Structure Data to Improve Web Site Recommendation*. W: EC-Web 2004, Zaragoza, Spain. s. 305-315
- Lieberman H. (1995): *Letizia: An Agent that Assists Web Browsing*. W: the IJCAI'95, s. 924-929
- Meadow C. T. (1992): *Text Information retrieval Systems*, San Diego: Academic Press
- Mizzaro S. (1997): *Relevance: the whole history*. Journal of the American Society for Information Science 48(9), s. 810-832
- Mladenec D. (1996): *Personal WebWatcher: Implementation and Design*. Technical Report IJS-DP-7472, Department of Intelligent Systems. Slovenia: J. Stefan Institute
- Mobasher B, Cooley R, Srivastava J. (2000): *Automatic Personalization Based on Web Usage Mining*, Communications of the ACM 43(8), s. 142-151
- Mobasher B., Dai H, Luo T, Sun Y, Zhu J (2000): *Integrating Web Usage and Content Mining for More Effective Personalization*. LNCS 1875 Springer Verlag (2000), s. 156-176
- Mobasher B., Dai H, Luo T, Nakagawa M. (2001): *Effective Personalization Based on Association Rule Discovery from Web Usage Data*. W: the 3rd ACM Workshop on Web Information and Data Management (WIDM01), held in conjunction with the International Conference on Information and Knowledge Management (CIKM 2001), Atlanta, Georgia, s. 9-15
- Mobasher B, Jin X, Zhou Y. (2005) *A Maximum Entropy Web Recommendation System: Combining Collaborative and Content Features* In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'05), Chicago
- Montaner M, Lopez B, De La Rosa J. L. (2003): *A Taxonomy of Recommender Agents on the Internet*, Artificial Intelligence Review 19(4), s. 285-330
- Montaner M. (2003): *Collaborative Recommender Agents Based on Case-Based Reasoning and Trust*. Praca doktorska, Universitat de Girona

- Montaner M. Palau J., Lopez, de la Rosa B. (2004): *Collaboration Analysis in Recommender Systems using Social Networks*. W Cooperative Information Agents, Springer-Verlag, s. 137-151
- Mooney R. J., Roy L. (1999): *Content-Based Book Recommending Using Learning for Text Categorization*. SIGIR '99 Workshop on Recommender Systems: Algorithms and Evaluation. Berkeley, s. 195-204
- Moukas A. Amalthea (1997): *Information Filtering and Discovery using a Multiagent Evolving System*. Journal of Applied AI, 11(5),. Dublin, Ireland, Springer-Verlag, s. 437-457
- Nielsen J. (1990): *The Art of Navigating Through Hypertext*. Communications of the ACM 33(3), s. 296-310
- Pazzani M, Muramatsu J, Billsus D. (1996): *Syskill & Webert: Identifying Interesting Web Sites*. In Proceedings of the Thirteenth National Conference on Artificial Intelligence, s. 54-61
- Pazzani M. (1999): *A Framework for Collaborative, Content-Based and Demographic Filtering*. Artificial Intelligence Review, s. 393-408
- Pazzani M, Billsus D. (2000): *User modeling for adaptive news access*. *User Modeling and User-Adapted Interaction*, 10(2-3), s.147-180
- Rashid A. M., Albert I., Cosley D., Lam S. K., McNee S. M., Konstan J. A., Riedl J. (2002): *Getting to Know You: Learning New User Preferences in Recommender Systems*. W: the International Conference on Intelligent User Interfaces, San Francisco, California, s. 127-134
- Rasmussen E. (1992): *Clustering algorithms*. In: Frakes W. and Baeza-Yates R. (eds), *Information retrieval: data, structures & algorithms*. Chap. 16, Englewood Cliffs, NJ: Prentice Hall, s. 419-442
- Resnick P., Iacovou N., Sushak M., Bergstrom P., Riedl J. Grouplens (1994): *An open architecture for collaborative filtering of netnews* W the 1994 Computer Supported Collaborative Work Conference, Chapel Hill, North Carolina, s. 175-186
- Resnick P, Varian H. R. (1997): *Recommender Systems*. Communications of the ACM, 40 (3), s. 56-58
- Rijsbergen C.J. (1979): *Information Retrieval*. Glasgow, University
- Rocchio J. J. (1971): *Relevance feedback in information retrieval, The SMART retrieval system - experiments in automatic document processing* Prentice Hall, Englewood Cliffs, s. 313-323
- Salton G., McGill M. (1983): *Introduction to Modern Information Retrieval*, McGraw-Hill, New York
- Salton G. (1989): *Automatic Text Processing. The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley, Reading, MA
- Sarukkai R.R. (2000): *Link prediction and path analysis using Markov chains*. W: 9th international World Wide Web conference on Computer networks: the international journal of computer and telecommunications networking, s. 377-386
- Sarwar B, Karypis G, Konstan J, Riedl J. (2001): *Item-based Collaborative Filtering Recommendation Algorithms*. W the 10th International WWW Conference, s. 285-295
- Schafer J B, Konstan, J.A., and Riedl J. (2001): *E-Commerce Recommendation Applications, Data Mining and Knowledge Discovery*, Vol. 5 No. 1/2, s. 115-153.

- Schamber L., Eisenberg M., Nilan, M. (1990): *A re-examination of relevance: Toward a dynamic, situational definition*, Information Processing & Management, 26(6), s.755-776
- Shahabi C., Banaei-Kashani F., Chen Y. S., McLeod D. (2001): *Yoda: An accurate and scalable web-based recommendation system* W: Sixth International Conference on Cooperative Information Systems (CoopIS), s. 418-432
- Shah M. A. (1997): *ReferralWeb: A resource location system guided by personal relations*. Master's thesis, M.I.T.
- Shardanand U, Maes P. (1995): *Social information filtering: Algorithms for automating "word of mouth"*. W ACM CHI'95 Conference on Human Factors in Computing Systems, s. 210-217
- Sheth B. Maes P. (1993): *Evolving agents for personalized information filtering*. W the 9th IEEE Conference on Artificial Intelligence for Applications, s. 394-403
- Singhal A., Choi J., Hindle D., Lewis D., Pereira F. (1999): *AT&T at TREC-7*, The Seventh Text Retrieval Conference (TREC-7), NIST Special Publication, s. 239-252
- Smyth B, Cotter P. (2000): *A Personalized TV Listings Service for the Digital TV Age*. Knowledge-Based Systems 13, s. 53-59
- Sparck Jones K., Walker S., Robertson S.E. (2000): *A probabilistic model of information retrieval : Development and comparative experiments*, Parts 1 and 2. Information Processing and Management 36(6) s. 779-840
- Stefani A, Strappavara C. (1998): *Personalizing Access to Web Wites: The SiteIF Project*. W HYPERTEXT'98
- Svensson M, Laaksolahti J, Hook K, Waern A. (2000): *A Recipe Based On-line Food Store*, In Proceedings of the 2000 International Conference on Intelligent User Interfaces (IUI'2000), New Orleans, Louisiana, USA, s. 260 – 263
- Tran T, Cohen R. (2000): *Hybrid Recommender Systems for Electronic Commerce*. In Knowledge-Based Electronic Markets, Papers from the AAAI Workshop, AAAI Technical Report WS-00-04. Menlo Park, CA: AAAI Press, s. 78-83.
- Voiskunskii V.G. (1997): *Evaluation of search results: a new approach*. Journal of the American Society for Information Science, 48, s. 133-142
- Webb G., Kuzmycz M. (1996): *Feature Based Modelling: A Methodology for Producing Coherent, Consistent, Dynamically Changing Models of Agents* Competencies. In User Modelling and User-Adapted Interaction, volume 5, s. 117–150
- WWW (2005a) RFC 1738 - Uniform Resource Locators (URL), dostępne pod adresem: <http://www.faqs.org/rfcs/rfc1738.html>
- WWW (2005b) Wyszukiwarka DirectHit, dostępna pod adresem: <http://www.directhit.com>
- WWW (2005c) Elektroniczna biblioteka publikacji naukowych CiteSeer, dostępna pod adresem: <http://citeseer.ist.psu.edu>
- WWW (2005d) Sklep internetowy Amazon dostępny pod adresem: <http://www.amazon.com>
- WWW (2005e) Sklep internetowy CDNOW dostępny pod adresem: <http://www.cdnow.com>
- WWW (2005f) Narzędzie do zarządzania dokumentami i nieuporządkowaną informacją dostępne pod adresem: <http://www.autonomy.com>
- WWW (2005g) Oficjalna strona grupy badawczej GroupLens dostępna pod adresem: <http://www.grouplens.org>

WWW (2005h), RFC 2616, Hypertext Transfer Protocol -- HTTP/1.1. dostępne pod adresem: <http://www.faqs.org/rfcs/rfc2616.html>

WWW (2005i), CiteSeer Scientific Literature Digital Library, dostępne pod adresem: <http://citeseer.ist.psu.edu/>

WWW (2005j), Wyszukiwarka *DirectHit*, dostępna pod adresem: <http://www.directhit.com>

Yu K, Schwaighofer A, Tresp V, Xu X, Kriegel H. P. (2004): *Probabilistic Memory-based Collaborative Filtering*, W: IEEE Transactions on Knowledge and Data Engineering Volume 15, Number 1, s. 56-69

Zhang D. Dong Y. (2000): *An efficient algorithm to rank Web resources*. W: the International World Wide Web Conference, the international journal of computer and telecommunications networking, Amsterdam, Netherlands, s. 449-455

Zhu X., Gauch S. (2000): *Incorporating quality metrics in centralized/distributed information retrieval on the World Wide Web*, W Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval. ACM Press, s. 288-295

Zhu X. (2001): Using Markov Chains for Link Prediction in Adaptive Web Sites, W: User Modeling, s. 298-300