

**Andrzej Małachowski, Maciej Idzikowski, Maciej Wasiak**

## **WERYFIKACJA ISTOTNOŚCI KRYTERIÓW OCENY PROJEKTOWANIA WITRYN INTERNETOWYCH ZA POMOCĄ ALGORYTMÓW GENETYCZNYCH**

### **1. Wstęp**

Witryna internetowa przedsiębiorstwa jest bardzo istotnym elementem aktywów nowoczesnego przedsiębiorstwa. Bardzo trudno wyobrazić sobie właściwie funkcjonującą organizację pomijającą tę formę komunikacji z otoczeniem. Podczas prac projektowych i kodowania łatwo jednak popełnić błąd – zespół specjalistów projektujący serwis internetowy nie jest w stanie w pełni *a priori* stwierdzić, co się podoba, a co się nie podoba internautom, jakie elementy zostały w serwisie pominięte, a jakie wyeksponowano za bardzo, co przyciąga ludzi do przeglądania stron internetowych, a co skłania ich do rezygnacji z surfowania po danej stronie internetowej. By usunąć te wątpliwości, za pomocą stron www prowadzi się badania ankietowe wśród użytkowników sieci mające na celu:

- określenie istotności poszczególnych kryteriów projektowych witryn internetowych,
- zbadanie wpływu witryn internetowych wybranych przedsiębiorstw na ich wizerunek wśród internautów,
- ocenę wybranych witryn internetowych przedsiębiorstw.

### **2. Ocena istotności kryteriów projektowych witryn internetowych**

Podstawowym celem prowadzonych badań było uzyskanie informacji (ocen) o jakości udostępnianych w Internecie witryn wybranych grup przedsiębiorstw. Dobrze zaprojektowany i skonstruowany serwis internetowy jest jednym z najważniejszych kanałów komunikacyjnych przedsiębiorstwa.

Jako obiekty badania zostały wybrane witryny przedsiębiorstw w sieci www. Dobór przedsiębiorstw odbył się według następującego klucza:

- 30 losowo wybranych przedsiębiorstw spośród 200 największych przedsiębiorstw branży IT,
- 30 losowo wybranych przedsiębiorstw notowanych na warszawskiej GPW, nie związanych z branżą IT.

Jako metodę zbierania informacji wybrano badania ankietowe wśród polskich użytkowników sieci www. Użytkownicy z kolei zostali wybrani w sposób celowo-losowy – do pierwszych 50 osób został wysłany list elektroniczny z prośbą o wypełnienie ankiety i przesłanie informacji dalej. Następnie list z prośbą o wypełnienie ankiety był wysyłany na zasadach marketingu wirusowego.

Zbierane w badaniu ankietowym informacje zostały podzielone na trzy grupy:

- 1) dane podstawowe – opisujące internautę (województwo, płeć, wiek, wykształcenie, szybkość dostępnego łącza internetowego),
- 2) podstawowe kryteria oceny witryn internetowych przedsiębiorstw – na podstawie literatury przedmiotu i własnych doświadczeń autorów zaproponowano dziewięć podstawowych kryteriów (ocenianych przez ankietowanych w skali od 0 do 10), które, jak się wydaje, dobrze charakteryzują każdą witrynę internetową zarówno pod względem wizualnym, informacyjnym, jak i użytkowym; były to:
  - szybkość ładowania witryny,
  - pierwsze wrażenia internauty,
  - łatwość nawigacji,
  - zawartość multimedialna,
  - ekspozycja wizerunku przedsiębiorstwa,
  - zawartość treści,
  - aktualność witryny,
  - dostępność alternatywnych możliwości eksploracji witryny,
  - dostępność dodatkowych informacji.

W ankiecie wzięły udział 1842 osoby, z czego 64% stanowili mężczyźni, a 36% – kobiety. Wszyscy ankietowani mieszkają w Polsce.

Najważniejsze etapy badań miały za zadanie analizę zgromadzonego materiału dotyczącego kryteriów i subkryteriów oceny witryn internetowych. Dla każdego badanego kryterium została obliczona średnia liczba punktów, a w celu uporządkowania i zobiektywizowania wyników otrzymane wartości poddano normalizacji. Przypisane przez internautów punkty pozwalają na rangowanie kryteriów według ich ważności.

Kolejnym efektem przeprowadzonego badania było określenie ważności kryteriów oceny witryny internetowej. Na podstawie wypełnionych przez internautów ankiet sporządzono uporządkowaną listę najważniejszych kryteriów oceny witryn internetowych. Wyniki przedstawia tab. 1.

Tabela 1. Ważność kryteriów oceny witryny internetowej

| Pozycja | Kryterium                                        | Punktacja |
|---------|--------------------------------------------------|-----------|
| 1       | aktualność witryny                               | 8,67      |
| 2       | zawartość treści                                 | 8,54      |
| 3       | szybkość ładowania witryny                       | 8,19      |
| 4       | łatwość nawigacji                                | 7,78      |
| 5       | pierwsze wrażenia internauty                     | 6,55      |
| 6       | ekspozycja wizerunku przedsiębiorstwa            | 6,12      |
| 7       | dostępność dodatkowych informacji                | 5,47      |
| 8       | zawartość multimedialna                          | 4,87      |
| 9       | dostępność alternatywnych możliwości eksploracji | 4,68      |
|         | Średnia ważność kryterium                        | 6,76      |

Źródło: [5].

Dla internautów najważniejszymi kryteriami oceny jakości projektowania i użytkowania witryny jest jej aktualność, zawartość treści, szybkość ładowania witryny oraz łatwość nawigacji. Wszystkie te elementy zostały ocenione przez ankietowanych znacznie powyżej średniej ważności kryteriów (6,76).

Nawet najlepiej zaprojektowana witryna, korzystająca z najnowszych osiągnięć technologii sieciowej, nie zdobędzie popularności wśród internautów, jeśli nie będzie przedstawiała aktualnych informacji dotyczących przedsiębiorstwa, oferowanych przez nie produktów i usług, a prezentowane informacje nie będą bardziej kompleksowe niż w tradycyjnych środkach przekazu. Typowym przykładem są strony biur podróży, których oferta zmienia się codziennie. Bez zastosowania mechanizmów bieżącej aktualizacji danych i prezentacji pełnej informacji o oferowanych wyjazdach witryna, a zarazem firma, szybko straci wiarygodność w oczach klienta.

Zawartość treści to kolejny niezwykle ważny dla internautów element poprawnie skonstruowanej witryny internetowej. Warstwa informacyjna serwisów internetowych musi być łatwo przyswajalna i dobrze zorganizowana. Bardzo istotna jest również wiarygodność publikowanych informacji – wiadomości niesprawdzone, nieprawdziwe, opublikowane bez podania autora, mogą zmniejszyć zaufanie internautów do przedsiębiorstwa. Istotnym elementem publikowania informacji w sieci jest również odpowiednie wykorzystanie technologii hiperlinków. Jak powinno wyglądać właściwe ich użycie, można zobaczyć na stronie Wielkiej Internetowej Encyklopedii Multimedialnej (<http://wiem.onet.pl>) – w momencie pojawienia się w opisie trudniejszych słów lub zwrotów użytkownik ma możliwość automatycznego przejścia na strony objaśniające ich znaczenie.

Szybkość ładowania witryny jest wciąż problemem polskich internautów. Brak ogólnie dostępnych łączy szerokopasmowych uniemożliwia kreowanie bardziej interaktywnego przekazu, gdyż zbyt długi czas ładowania witryny zwykle zniechęca użytkowników do odwiedzania takich stron przedsiębiorstwa. Należy więc pamię-

tać o odpowiedniej pod tym względem optymalizacji elementów strony, przystosowując ją do publikacji w Internecie.

Łatwość nawigacji to kolejna cecha witryny internetowej, do której wagę przywiązują badani. Łatwość poruszania się po witrynie firmowej przedsiębiorstwa, uproszczenie dotarcia do poszukiwanych informacji, przejrzysta mapa serwisu są więc istotnymi cechami, na które należy zwrócić uwagę podczas projektowania witryny.

Kolejne miejsca w rankingu zajmują elementy ściśle związane z zasadami projektowania witryn internetowych – są nimi pierwsze wrażenia oraz ekspozycja wizerunku przedsiębiorstwa. Należy zauważyć, że w prowadzonych badaniach, według oceny internautów, ważność tych kryteriów przekracza granicę 60%, co powoduje, że należy je uznać za bardzo ważne czynniki budowy strony internetowej.

Najmniej doceniane przez internautów są zaawansowane rozwiązania multimedialne oraz dostępność alternatywnych możliwości eksploracji. Na niewielką ważność rozwiązań multimedialnych (wyszukana grafika, animacje, przekaz wideo) z całą pewnością wpływa jakość łącz internetowych posiadanych przez badanych – zajmujące niejednokrotnie po kilkaset kilobajtów animacje czy filmy ładują się w oknie przeglądarki internauty korzystającego z modemu niezwykle długo, czym skutecznie zniechęcają go do wizyt na stronach www przedsiębiorstwa publikującego tego typu zawartość.

Niedoceniając przez internautów dostępności alternatywnych możliwości eksploracji witryny wiąże się z coraz większą funkcjonalnością przeglądarek stron internetowych, które, podobnie do edytorów tekstu, mają obecnie możliwość dowolnego wpływu na wygląd strony (np. możliwość wyłączenia grafiki podczas ładowania strony HTML). Nowoczesne przeglądarki oferują również możliwość wydrukowania dowolnego fragmentu wyświetlanej strony oraz precyzyjnego przeniesienia treści strony do dowolnej aplikacji (np. edytora tekstu). Tworzenie dodatkowych stron budowanych i przeznaczonych tylko do drukowania staje się powoli anachronizmem.

Warto również zwrócić uwagę na odchylenia standardowe dotyczące poszczególnych kryteriów. Wartość odchylenia standardowego dla wszystkich kryteriów wahała się w granicach 1,52-2,71 (tab. 2). Ankietowani internauci byli najbardziej zgodni co do ważności kryterium „aktualność witryny”, najmniejsza zgodność panowała natomiast w stosunku do kryterium „zawartość multimedialna”. Wartość odchylenia dla trzech najważniejszych, według internautów, kryteriów projektowych wahała się w granicach 1,52-1,76. Ankietowani dość jednomyślnie ocenili więc ważność tych kryteriów. Zwraca uwagę wysoka wartość odchylenia standardowego (2,33) dla kryterium łatwość nawigacji – przy dość dużej ważności tego kryterium (7,78 pkt) zaobserwować można jednak znaczną rozpiętość ocen. Można więc wnioskować, że istnieje podział pomiędzy użytkownikami Internetu odnośnie do ważności tego kryterium.

Stosunkowo wysokie wartości odchylenia standardowego (2,35-2,71) mają również mniej ważne kryteria oceny witryny internetowej: dostępność dodatkowych informacji, zawartość multimedialna oraz dostępność dodatkowych możliwości eksploracji, co można zinterpretować jako potwierdzenie istnienia wśród internautów znacznych rozbieżności w ocenie ważności (istotności) tych kryteriów projektowania witryn.

Tabela 2. Odchylenia standardowe dla poszczególnych kryteriów oceny witryny internetowej

| Pozycja | Kryterium                                        | Odchylenie standardowe |
|---------|--------------------------------------------------|------------------------|
| 1       | aktualność witryny                               | 1,52                   |
| 2       | zawartość treści                                 | 1,76                   |
| 3       | szybkość ładowania witryny                       | 1,68                   |
| 4       | łatwość nawigacji                                | 2,33                   |
| 5       | pierwsze wrażenia internauty                     | 1,98                   |
| 6       | ekspozycja wizerunku przedsiębiorstwa            | 1,95                   |
| 7       | dostępność dodatkowych informacji                | 2,35                   |
| 8       | zawartość multimedialna                          | 2,71                   |
| 9       | dostępność alternatywnych możliwości eksploracji | 2,62                   |

Źródło: [5].

### 3. Logika rozmyta i algorytm genetyczny

W celu zweryfikowania niektórych wyników przedstawionych w poprzednim punkcie postanowiono wykorzystać techniki sztucznej inteligencji, takie jak algorytm genetyczny wraz z logiką rozmytą.

Zastosowano narzędzie wykorzystane już wcześniej do oceny kondycji finansowej przedsiębiorstwa, opisane w [12]. Jest to program budujący rozmyty system wnioskujący za pomocą algorytmu genetycznego (AG). AG w trakcie kolejnych generacji poszukuje takich parametrów systemu, jak: optymalne kształty zbiorów rozmytych dla zmiennych wejściowych, właściwe konkluzje dla reguł. Zastosowana metoda pozwala na odwzorowanie wysoce nieliniowych zależności. Przede wszystkim jednak AG selekcjonuje zmienne, tak aby końcowy model zawierał ich ograniczoną liczbę. Liczbę tę trzeba określić przed rozpoczęciem uczenia systemu – zazwyczaj wynosi ona od 1 do 5 zmiennych. Ta funkcja narzędzia miała w opisanych badaniach pierwszoplanowe znaczenie.

Do procesu optymalizacji użyto algorytmu kodującego liczby rzeczywiste (RCGA – *real coded genetic algorithm*). Jest to algorytm wykorzystujący w pełni ideę konwencjonalnego algorytmu genetycznego, z tą różnicą, że w chromosomach zamiast bitów umieszczone są bezpośrednio wartości parametrów.

Za użyciem takiego algorytmu przemawia:

- rzeczywisty charakter większości szukanych parametrów,

- zwiększona efektywność algorytmu ze względu na brak operacji konwersji kodu binarnego do liczb rzeczywistych,
- zwiększona precyzja w odnajdywaniu szukanych wartości (brak granulacji spowodowanej ograniczoną długością ciągu bitowego),
- brak zakłóceń, które w ciągach binarnych powodowała operacja krzyżowania,
- bardziej zrozumiałe działanie współczynnika mutacji i łatwiejsze jego określanie.

Przy stosowaniu RCGA należy nieco bardziej szczegółowo zdefiniować współczynnik mutacji. Oprócz cechy odpowiadającej za prawdopodobieństwo mutacji należy określić również jej skalę. Najczęściej skala ta jest odchyleniem standardowym możliwej zmiany według rozkładu normalnego lub zakresem możliwej zmiany według rozkładu prostokątnego. W badaniu użyto drugiej z przedstawionych możliwości.

Podstawowe parametry zastosowanego AG to: liczebność populacji – 100, całkowicie losowe generowanie początkowej populacji, maksymalnie 200 generacji. Przeprowadzono 20 przebiegów algorytmu, każdy z nieco innym prawdopodobieństwem (od 0,05 do 0,2) mutacji. Prawdopodobieństwo krzyżowania wynosiło zawsze 0,8.

Do wnioskowania rozmytego użyto modelu Mamdaniego, agregacji według zasady MAX-MIN i zastosowano taki sam typ inferencji. Do opisanie zbiorów rozmytych zastosowano funkcje sigmoidalne, dopełniające się w ramach jednej zmiennej do jedności. Baza reguł była zawsze kompletna i niesprzeczna, tzn. dla każdej kombinacji przesłanek istniała jedna i tylko jedna konkluzja. Sposób wyostrzania przeprowadzano metodą środka ciężkości (COG, COA).

W algorytmie genetycznym zastosowano kodowanie rzeczywiste, selekcję metodą ruletki z użyciem *steady-state*. Rodzaj użytego systemu rozmytego jest dosyć typowy, natomiast sam algorytm genetyczny był poddany wielu modyfikacjom i usprawnieniom w celu uzyskiwania jak najlepszych wyników. Szczegóły dotyczące budowy i funkcjonowania tego narzędzia można odnaleźć w [12].

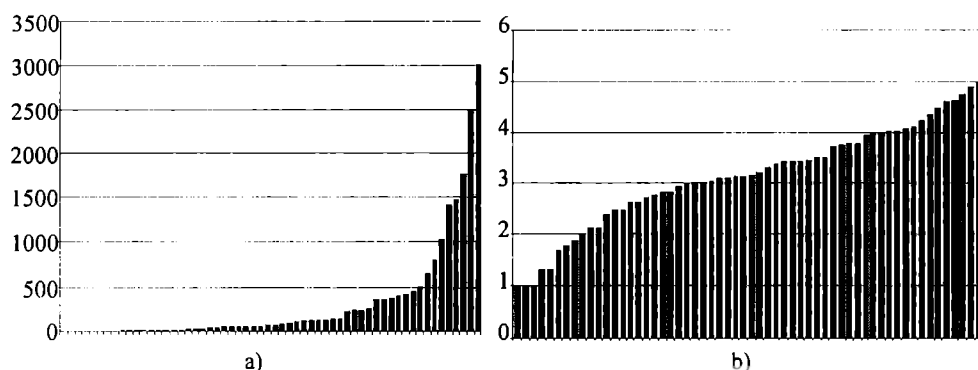
## 4. Przygotowanie danych

W trakcie badania ankietowego respondenci odpowiadali na pytania o istotność każdego z 9 kryteriów oceny witryn oraz odrębnie oceniali 60 witryn internetowych w skali od 0 do 10. Niezwykle interesujące byłoby zweryfikowanie podanych przez respondentów preferencji odnośnie do istotności poszczególnych cech stron, zawartych w tab. 1. Aby to uczynić, należałoby porównać podane przez respondentów oceny punktowe cech konkretnych witryn z obiektywnie dobraną jakością witryny. Takim kryterium może być *link popularity*, czyli liczba linków prowadzących do witryny.

Ze zbioru danych usunięto „Gazetę Wyborczą”, do której liczba odnośników była 9-krotnie większa niż do drugiej w kolejności witryny. Ze względu na charakter rozkładu tej cechy (zob. rys. 1) wielkości te zlogarytmowano, poddając je dodatkowo drobnemu przekształceniu liniowemu, dzięki czemu zostały sprowadzone do przedziału od 1 do 5.

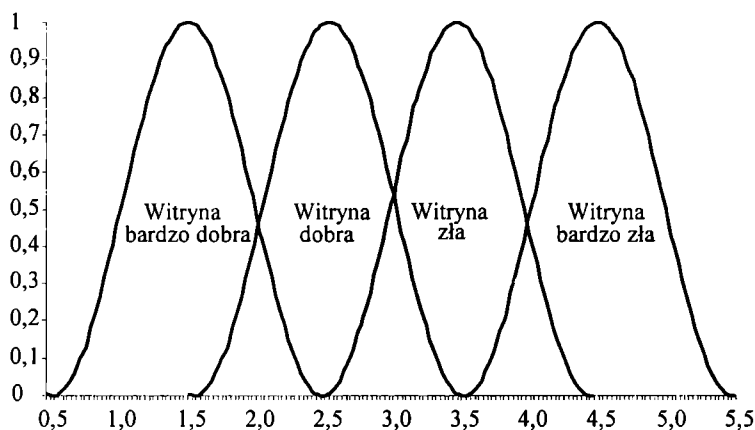
Ta informacja stanowiła ciąg uczący dla algorytmu genetycznego, który na jej podstawie miał za zadanie wyselekcjonować kluczowe cechy, mogące mieć związek z jakością strony.

Przyjęto, podobnie jak we wcześniejszym badaniu, że witryny zostaną podzielone na 4 klasy: bardzo złe (wartości od 1 do 2), złe (od 2 do 3), dobre (od 3 do 4) i bardzo dobre (od 4 do 5).



Rys. 1. a) *Link popularity* badanych witryn; b) *Link popularity* po zlogarytmowaniu

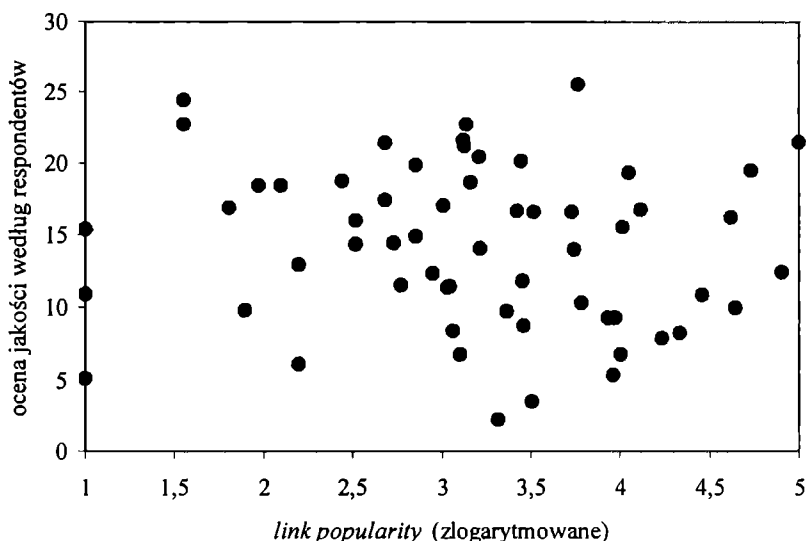
Źródło: opracowanie własne.



Rys. 2. Zbiory rozmyte opisujące jakość witryny

Źródło: opracowanie własne.

Na rysunku 3 przedstawiono położenie każdej witryny względem obu klasyfikacji.



Rys. 3. Zależność między *link popularity* witryn a ich oceną przez respondentów

Źródło: opracowanie własne.

Od razu można zauważyć, że ocena jakości witryn według kryterium *link popularity* jest zupełnie inna niż we wcześniejszym badaniu – można by oczekiwać silnego związku, ale nie ma go w ogóle. Mimo to badanie kontynuowano.

Postanowiono użyć algorytmu genetycznego do zbudowania dwóch rodzajów modeli:

- a) modelu rozmytego wykorzystującego dwie zmienne,
- b) modelu rozmytego wykorzystującego trzy zmienne.

W obu przypadkach zmienne wejściowe były opisywane za pomocą trzech zbiorów rozmytych (stany: niski, średni, wysoki).

Zazwyczaj celem budowy systemu opartego na sztucznej inteligencji jest późniejsze użycie go do klasyfikacji lub wnioskowania. W tym przypadku cel był inny – szukając maksymalnego dopasowania (trafności) do danych uczących, AG miał wskazać te cechy, które mają największy wpływ na popularność strony. W punkcie 5 przedstawiono wyniki badania.

## 5. Uzyskane wyniki

Przeprowadzono 10 testów mających na celu wyselekcjonowanie 2 kluczowych cech oraz 10 testów selekcyjnych 3 kluczowe cechy. Najlepsze rezultaty przedstawiono w tab. 3. Kolejność zmiennych nie ma tu znaczenia.

Tabela 3. Najważniejsze cechy witryn wybrane przez AG

| System    | Trafność | Cechy                                                             |
|-----------|----------|-------------------------------------------------------------------|
| 2 zmienne | 83,7%    | – zawartość treści<br>– aktualność witryny                        |
| 3 zmienne | 85,5%    | – zawartość treści<br>– aktualność witryny<br>– łatwość nawigacji |

Źródło: opracowanie własne.

respondenci subiektywnie wskazywali, w jakim stopniu istotne są dla nich poszczególne cechy witryn. W badaniu z wykorzystaniem technik sztucznej inteligencji jako informację uczącą wybrano kryterium *link popularity*, które w pierwszym badaniu nie było w ogóle brane pod uwagę. Co więcej, w zasadzie nie istnieje widoczny związek, między jakością witryn według ocen respondentów a *link popularity* (por. rys. 3). Jest jednak zaskakujące to, że wyniki opisanego badania są w dużym stopniu zbieżne z subiektywnymi ocenami respondentów.

W kolejnym etapie badań autorzy planują zwiększyć liczbę analizowanych witryn oraz porównać wyniki z innymi obiektywnymi kryteriami oceny jakości witryny, takim jak np. *page-rank*.

Przy wyborze dwóch najważniejszych cech wyniki są całkowicie zbieżne ze wskazaniami respondentów – są to: aktualność witryny oraz zawartość treści.

Przy wyborze trzech najważniejszych zmiennych AG uwzględnił dodatkowo łatwość nawigacji, którą respondenci umieścili na czwartym miejscu pod względem istotności.

Oba opisanie badania były przeprowadzone na podstawie odrębnych danych. W trakcie pierwotnego badania

## Literatura

- [1] Duch W., Adamczak R., Grąbczewski K., *A new Methodology of Extraction, Optimization and Application of Crisp and Fuzzy Logical Rules*, IEEE Transactions on Neural Networks, 2001.
- [2] Goldberg D.E., Richardson J., *Genetic Algorithms with Sharing for Multimodal Function Optimization*, Genetic Algorithms and Their Applications: Proceedings of the SICoGA, 1987, s. 41-49,
- [3] Gupta M.M., Yamakawa T., *Fuzzy Logic in Knowledge-Based Systems, Decision and Control*, North-Holland/Elsevier Science Publishers B.V., Amsterdam 1988.
- [4] Herrera F., Lozano M., Verdegay J.L., *Generating Fuzzy Rules from Examples Using Genetic Algorithm*, Proceedings Fifth International Conference of Information Processing and Management of Uncertainty in Knowledge-Based Systems, Paris 1994.
- [5] Idzikowski M., *Wybrane aspekty kreowania tożsamości i wizerunku przedsiębiorstwa w przestrzeni Internetu*, rozprawa doktorska, AE, Wrocław 2004.

- 
- [6] Jin Y., Member, Seelen W., Sendhoff B., *On Generating FC3 Fuzzy Rule Systems from Data Using Evolution Strategies*, IEEE Transactions On Systems, Man, And Cybernetics – Part B: Cybernetics, vol. 29, nr 6, 1999.
  - [7] Michalewicz Z., *Algorytmy genetyczne + struktury danych = programy ewolucyjne*, Wyd. Naukowo-Techniczne, Warszawa 1998.
  - [8] Nauck D., Nauck U., Kruse R., *Generating Classification Rules with Neuro-Fuzzy System NEFCLASS*, NAFIPS'96, Berkeley 1996.
  - [9] Piegat A., *Modelowanie i sterowanie rozmyte*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 1999.
  - [10] *Proceedings 3<sup>rd</sup> International Conference on Genetic Algorithms*, CA, Morgan Kaufmann, San Mateo 1989.
  - [11] Schaffer J.D., *A Study of Control Parameters Affecting On-Line Performance of Genetic Algorithms for Function Optimization*, Proceedings 3<sup>rd</sup> International Conference on Genetic Algorithms, CA, Morgan Kaufmann, San Mateo 1989.
  - [12] Wang H.F., Tsaur R.C., *Resolution of Fuzzy Regression Model*, „European Journal of Operational Research” 2000, vol. 126, nr 1.
  - [13] Wasiak M., *Rozmyty system wnioskujący, oceniający kondycję finansową przedsiębiorstw*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 991, *Pozyskiwanie wiedzy z baz danych*, Wrocław 2002.

## THE USE OF GENETIC ALGORITHM IN CHOOSING MOST IMPORTANT CRITERIA OF WEB PAGES DESIGN

### Summary

This article presents the results of the research which covers the design of corporate websites. It also verifies data acquired with the traditional method with the genetic algorithms method.

---

**Dr hab., prof. AE Andrzej Malachowski i dr Maciej Idzikowski** są pracownikami Katedry Komunikacji Gospodarczej, **dr Maciej Wasiak** jest pracownikiem Katedry Systemów Sztucznej Inteligencji w Akademii Ekonomicznej we Wrocławiu.