



**INSTYTUT BADAŃ SYSTEMOWYCH  
POLSKIEJ AKADEMII NAUK**

**ANALIZA SYSTEMOWA W FINANSACH  
I ZARZĄDZANIU**

Wybrane problemy  
Tom 11

Pod redakcją  
Jerzego HOŁUBCA

Warszawa 2009

Wykaz opiniodawców artykułów zamieszczonych  
w niniejszym tomie:

prof. dr hab. inż. Jerzy HOŁUBIEC

dr inż. Lech KRUŚ

doc. dr hab. inż. Wiesław KRAJEWSKI

doc. dr hab. Jacek MALINOWSKI

dr inż. Edward MICHALEWSKI

prof. dr Adam SKOREK

dr hab. Ryszard SMARZEWSKI

prof. dr hab. inż. Andrzej STRASZAK

dr Dominik ŚLĘZAK

prof. dr hab. inż. Stanisław WALUKIEWICZ

doc. dr hab. Sławomir ZADROŻNY

© Instytut Badań Systemowych PAN  
Warszawa 2009

**ISBN 9788389475220**

Druk: Zakład Poligraficzny Jerzy Kosiński, Warszawa

# ZASTOSOWANIE INTUICJONISTYCZNYCH ZBIORÓW ROZMYTYCH W KONCEPCJI MODELU SYSTEMU DOBORU MATERIAŁÓW DYDAKTYCZNYCH PRZY KSZTAŁCENIU USTAWICZNYM

**Marcin Werdoni**

*Studia Doktoranckie IBS PAN*

*Systemy nauczania na odległość coraz częściej zastępują tradycyjne formy kształcenia. W artykule poruszono problem dopasowania takiego systemu do użytkownika oraz problem doboru materiału dydaktycznego do potrzeb klienta. Informacje dostarczone przez użytkownika do systemu wyszukującego są często nieprecyzyjne, w takim wypadku przydatne stają się zbiory rozmyte. Opisano koncepcję modelu systemu doboru materiałów dydaktycznych. Model oparty jest na intuicjonistycznych zbiorach rozmytych.*

**Słowa kluczowe:** *e-learning, nauczanie na odległość, intuicjonistyczne zbiory rozmyte*

## Wstęp

Adaptacja szkoleń e-learningowych do indywidualnych potrzeb i preferencji użytkowników jest aktualnie istotnym obszarem zainteresowań zarówno teoretyków, jak i praktyków, co znajduje odzwierciedlenie w rosnącej liczbie publikacji oraz pomysłów wdrożeń w tym zakresie.

Z praktyki wynika, że informacje dostarczone do systemu doboru są często nieprecyzyjne i niepewne, zbiory rozmyte<sup>1</sup> i ich uogólnienie są szczególnie przydatne przy reprezentacji wiedzy podawanej przez użytkownika. Autor proponuje zastosowanie intuicjonistycznych zbiorów rozmytych<sup>2</sup> (A-IFS - *Atanassov Intuitionistic fuzzy sets*).

Zastosowanie elementów teorii intuicjonistycznych zbiorów rozmytych przy doborze materiałów dydaktycznych, ze względu na wprowadzenie dodatkowego stopnia swobody przy opisie elementów zbioru daje możliwość lepszego opisu nieprecyzyjnie określonych przypadków oraz ujęcia ludzkiego sposobu wyrażania opinii (uwzględniając charakterystyczne dla człowieka niezdecydowanie i niepełną

---

<sup>1</sup> L.A. Zadeh *Fuzzy sets*, Information and Control, Nr 8, 1965.

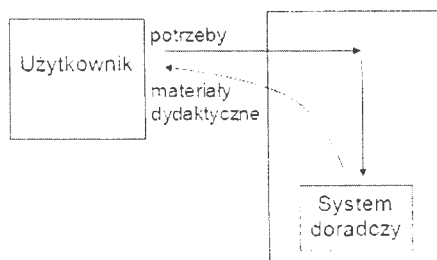
<sup>2</sup> K.T. Atanassov *Intuitionistic fuzzy sets*, Springer-Verlag, 1999.

wiedzę). W rezultacie powinno się otrzymać bardziej dopasowane do użytkownika wyniki wyszukiwania.

Zasadniczy problem badawczy dotyczy opracowania modelu systemu doboru materiałów dydaktycznych kształcenia ustawicznego z wykorzystaniem intuicjonistycznych zbiorów rozmytych.

### 1. Proponowany model

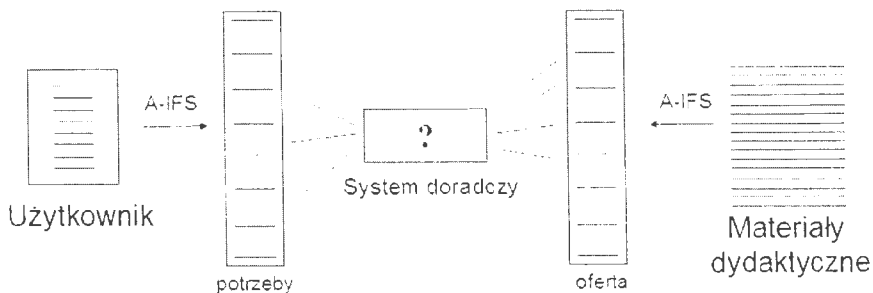
Autor proponuje stworzenie systemu doboru materiałów dydaktycznych (Rys. 1), który w założeniu będzie odwoływał się do obszernej bazy dostępnych szkoleń. Na podstawie preferencji użytkownika (w zakresie np. czasu, miejsca, obszaru tematycznego) wyszukiwane będzie szkolenie zbliżone do tych preferencji. Dopasowanie treści kursu do potrzeb użytkownika nastąpi między innymi poprzez uwzględnienie jego celów edukacyjnych, posiadanej wiedzy oraz stopnia szczegółowości w jakim chce on poznać określone zagadnienia.



Rys. 1. System doboru materiałów dydaktycznych

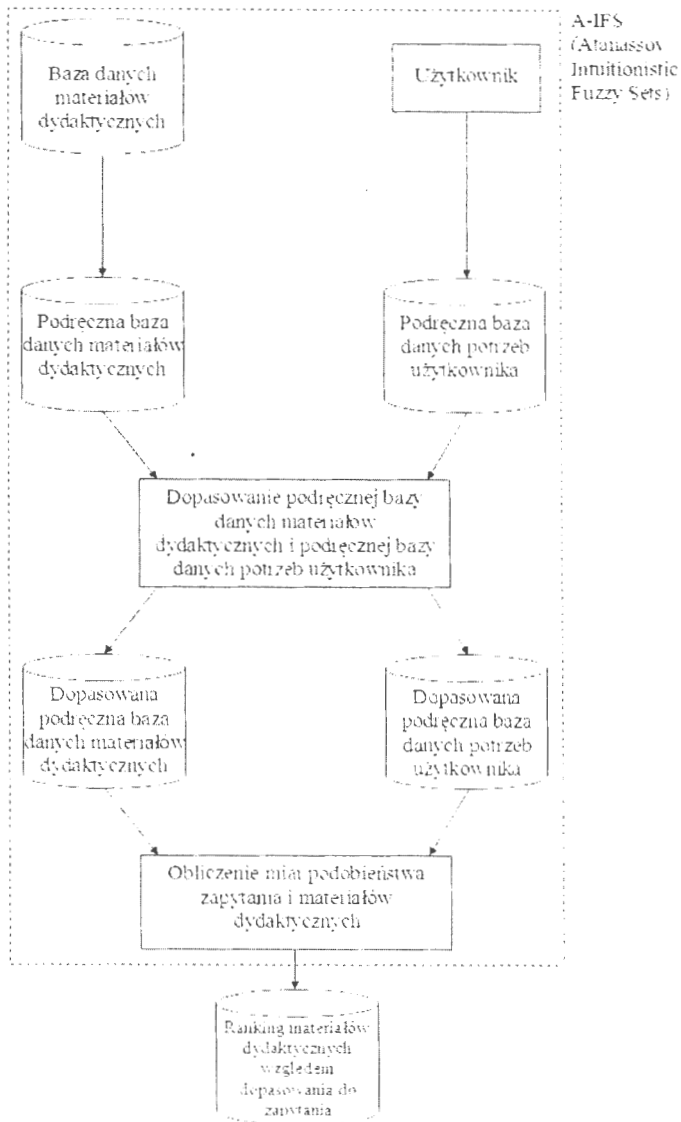
Źródło: opracowanie własne

Intuicjonistyczne zbiory rozmyte zostały zastosowane w transformatach języka opisu materiałów dydaktycznych oraz potrzeb użytkowników (Rys. 2).



Rys. 2. Transformata języka opisu materiałów dydaktycznych oraz potrzeb użytkowników. Źródło: opracowanie własne

Na rysunku 3 przedstawiono główne etapy pracy systemu doboru materiałów dydaktycznych.



Rys. 3. Schemat proponowanego algorytmu pracy systemu doboru  
Źródło: opracowanie własne

### 1.1.Reprezentacja materiałów dydaktycznych w proponowanym modelu

W proponowanym w modelu materiały dydaktyczne reprezentowane są przez słowa kluczowe. Każde słowo kluczowe modelowane jest przez intuicjonistyczne zbiory rozmyte. Określony jest w ten sposób stopień ważności danego słowa dla materiału dydaktycznego.

Waga przypisana słowu kluczowemu określa w sposób nieprecyzyjny na ile słowo to jest ważne dla reprezentacji treści materiału dydaktycznego. Określony jest każdemu słowu stopień przynależności ( $\mu$ ), stopień nieprzynależności ( $\nu$ ) słowa do danego materiału dydaktycznego oraz margines wahania ( $\pi$ ).

W celu dokonania badań opartych na proponowanym modelu należało zbudować bazę danych materiałów dydaktycznych. Aby to zrobić opracowano kwestionariusz ankietowy skierowany do pracowników Katedry Informatyki Gospodarczej i Logistyki Wydziału Zarządzania Politechniki Białostockiej.

Na podstawie kwestionariusza ankietowego, merytorycznej dyskusji i konsultacji z nauczycielami opracowano listę materiałów dydaktycznych wraz z reprezentującymi je słowami kluczowymi. Dla każdego słowa określono stopień przynależności, nieprzynależności oraz dopełnienie. Autor badaniom ankietowym ma zamiar poddać także studentów Politechniki Białostockiej, pracowników firmy KAN z Białegostoku oraz użytkowników IEEE Continuing Education.

Dane dotyczące materiałów dydaktycznych przechowywane są w pliku csv. Poszczególne informacje oddzielone są znakiem średnika. W pierwszej kolumnie znajduje się nazwa materiału dydaktycznego, w kolejnej słowo kluczowe, następne trzy kolumny to stopień dopasowania słowa kluczowego do danego materiału dydaktycznego ( $\mu$ ), stopień niedopasowania ( $\nu$ ) oraz niepewność ( $\pi$ ). Zgodnie z definicją zbiorów intuicjonistycznych suma tych trzech parametrów ma być równa 1.

W tabeli 1 przedstawiono fragment bazy danych materiałów dydaktycznych. W przyszłości autor zamierza zaimplementować automatyczny algorytm generowania słów kluczowych na podstawie streszczeń materiałów dydaktycznych. Automatyczne indeksowanie polega na obliczeniu wag poszczególnych słów kluczowych w dokumencie zgodnie z przyjętym schematem ważenia. Po znormalizowaniu wyliczonych wag (przez podzielenie przez maksymalną wagę słowa w danym dokumencie), stosuje się je jako wartości w szablonach terminów lingwistycznych<sup>3</sup>.

---

<sup>3</sup> G. Salton, C. Buckley *Term weighting approaches In automatic text retrieval*, Information Processing and Management, vol. 24, 1988.

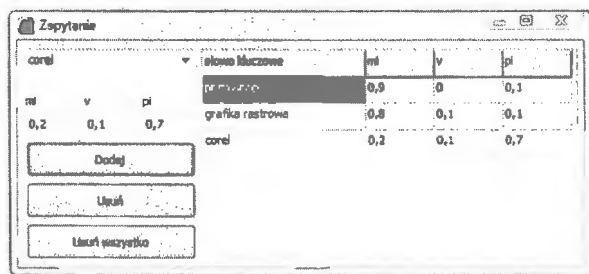
**M. Werdoni – Zastosowanie intuicjonistycznych zbiorów rozmytych...**

Nazwa materiału dydaktycznego	Słowo kluczowe	Stopień dopasowania ( $\mu$ )	Stopień niedopasowania ( $\nu$ )	Brak wiedzy ( $\pi$ )
Ekonometria	ekonometria	0,9	0	0,1
Ekonometria	ekonomia	0,675	0,175	0,15
Ekonometria	estymacja	0,9	0	0,1
Ekonometria	matematyka	0,9	0	0,1
Ekonometria	model	0,9	0	0,1
Ekonometria	prawdopodobieństwo	0,9	0	0,1
Ekonometria	statystyka	0,9	0	0,1
Ekonometria	test	0,9	0	0,1
Ekonometria	testy graniczne	0,585	0,285	0,13
Ekonomia	ekonomia	0,9	0	0,1
Ekonomia	gospodarka	0,9	0	0,1
Ekonomia	matematyka	0,54	0,34	0,12
Ekonomia	model	0,855	0	0,145
Ekonomia	nauki społeczne	0,855	0	0,145
Ekonomia	racjonalność	0,855	0	0,145
Ekonomia	system	0,9	0	0,1
Grafika	bitmapa	0,675	0,175	0,15

Tabela 1. Fragment bazy danych materiałów dydaktycznych

## 1.2. Reprezentacja zapytań

W proponowanym modelu zastosowano dwa rodzaje zapytań, proste oraz zaawansowane. Zapytanie składa się ze słów kluczowych podanych przez użytkownika. Każde słowo kluczowe modelowane jest przez intuicjonistyczne zbiory rozmyte. Określony w ten sposób jest stopień ważności danego słowa kluczowego. Waga przypisana słowu kluczowemu określa w sposób nieprecyzyjny na ile słowo to jest ważne dla reprezentacji zapytania. Każdemu słowu określony jest stopień przynależności ( $\mu$ ), stopień nieprzynależności ( $\nu$ ) słowa oraz margines wahania ( $\pi$ ). Okno stworzonego programu odpowiedzialne za tworzenie zapytań przedstawione jest na rysunku 4.



Rys. 4. Okno tworzenia zapytań. Źródło: *opracowanie własne*

a) zapytania proste

W tym wariancie zakłada się, że użytkownik podaje w zapytaniu wyłącznie słowa kluczowe, które uważa za ważne dla reprezentacji jego potrzeb informacyjnych. Tak więc w tym przypadku każde słowo kluczowe zostaje dodane do zapytania ze stopniem przynależności równym 1, stopień nieprzynależności oraz margines wahania są równe 0.

b) zapytania zaawansowane

W tym wariancie zakłada się, że użytkownik podaje w zapytaniu zarówno słowa kluczowe, które uważa za ważne dla reprezentacji jego potrzeb informacyjnych, słowa kluczowe, które uważa za nieważne oraz słowa co do których nie ma pewności. Tak więc w tym przypadku każde słowo kluczowe zostaje dodane do zapytania z pewnym stopniem przynależności, pewnym stopniem nieprzynależności oraz pewnym marginesem wahania. Należy pamiętać o tym, że:

$$\mu_A(x) + v_A(x) + \pi_A(x) = 1 \quad (1)$$

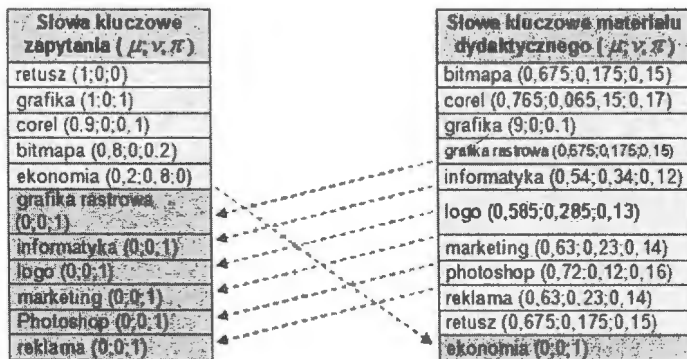
### 1.3. Ocena dopasowania

Przed przystąpieniem do oceny relewantności zapytania względem materiałów dydaktycznych umieszczonych w bazie, należy przyjrzeć się każdej parze *zapytanie – materiał dydaktyczny* oraz dokonać dopasowania danych (Rys. 5).

Ilość słów kluczowych określających zapytanie oraz ilość słów kluczowych określających materiał dydaktyczny musi być taka sama. Muszą to być takie same słowa kluczowe. Dopasowanie danych polega więc na wyszukaniu na liście słów kluczowych określających materiał dydaktyczny słów, które nie znajdują się na liście słów kluczowych określających zapytanie oraz dodanie tych słów do listy słów określających zapytanie. Podobnie w drugą stronę. Na liście słów kluczowych określających zapytanie zostają znalezione słowa, których brakuje na liście okre-



ślącej materiał dydaktyczny i zostają tam dodane. Każde słowo kluczowe opisane jest pewnym stopniem przynależności, pewnym stopniem nieprzynależności oraz pewnym marginesem wahania. Dodając brakujące słowa zostaje przypisany im stopień przynależności równy 0, stopień nieprzynależności równy 0, oraz margines wahania określający brak wiedzy równy 1. Powyżej opisany algorytm został zobrazowany na schemacie blokowym (Rys. 6).



Rys. 5. Dopasowanie list słów kluczowych opisujących zapytanie oraz materiał dydaktyczny. Źródło: opracowanie własne

Po dopasowaniu danych, dla każdej pary *zapytanie – materiał dydaktyczny* wyliczane jest w następujący sposób podobieństwo tych dwóch elementów<sup>4</sup> :

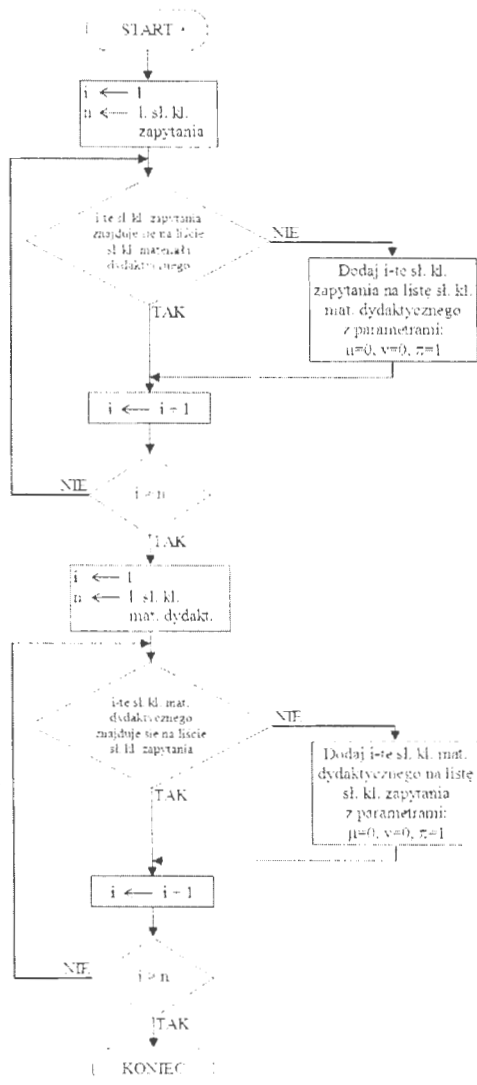
$$Sim(X, F) = \frac{l_{IFS}(X, F)}{l_{IFS}(X, F^C)} \tag{2}$$

gdzie:  $l_{IFS}(X, F)$  - odległość punktu  $X(\mu, \nu, \pi)$  od  $F(\mu, \nu, \pi)$ ,

$l_{IFS}(X, F^C)$  - odległość punktu  $X(\mu, \nu, \pi)$  do punktu  $F^C(\nu, \mu, \pi)$  - dopełnienia punktu F

Odległości  $l_{IFS}(X, F)$  i  $l_{IFS}(X, F^C)$  obliczane są przy użyciu trzech metod: unormowanej odległości Euklidesa:

<sup>4</sup> E. Szmidt *Intuicjonistyczne zbiory rozmyte i ich zastosowania w podejmowaniu decyzji*, Badania Systemowe – Teoria i Zastosowania, Warszawa, 2004.



Rys. 6. Schemat blokowy algorytmu odpowiedzialnego za dopasowanie list słów kluczowych opisujących zapytanie oraz materiał dydaktyczny.

Źródło: opracowanie własne

$$l_{IFS}(A, B) = \sqrt{\frac{1}{2n} \sum_{i=1}^n [(\mu_A(x_i) - \mu_B(x_i))^2 + (v_A(x_i) - v_B(x_i))^2 + (\pi_A(x_i) - \pi_B(x_i))^2]} \quad (3)$$

unormowanej odległości Hamminga:

$$l_{IFS}(A, B) = \frac{1}{2n} \sum_{i=1}^n (|\mu_A(x_i) - \mu_B(x_i)| + |v_A(x_i) - v_B(x_i)| + |\pi_A(x_i) - \pi_B(x_i)|) \quad (4)$$

unormowanej odległości Hausdorffa:

$$l_{IFS}(A, B) = \frac{1}{n} \sum_{i=1}^n \max\{|\mu_A(x_i) - \mu_B(x_i)|, |v_A(x_i) - v_B(x_i)|, |\pi_A(x_i) - \pi_B(x_i)|\} \quad (5)$$

Wyliczone według wzoru (2) podobieństwo zapytania do materiału dydaktycznego może zawierać się w przedziale od 0 do nieskończoności. Im wartość dopasowania jest bliższa zero, tym dopasowanie jest lepsze. Jeśli podobieństwo jest równe 1 wtedy materiał dydaktyczny jest neutralny w stosunku do zapytania. Im wartość miary jest większa od 1 tym materiał dydaktyczny jest bardziej nie pasujący w stosunku do zapytania.

W celach porównawczych dokonano również obliczeń z wykorzystaniem klasycznych zbiorów rozmytych. W tym przypadku skorzystano z powyższych wzorów, zmodyfikowanych w odpowiedni sposób aby były prawidłowe zgodnie z klasyczną logiką rozmytą.

Trzy użyte miary: Euklidesa, Hamminga oraz Hausdorffa dały bardzo zbliżone wyniki. Zawsze w ten sam sposób szeregowały materiały dydaktyczne. Czas obliczeń dla tych miar jest również bardzo przybliżony.

Uzyskane wyniki wstępnych badań są w pełni satysfakcjonujący i dające nadzieje na przyszłą rozwojową pracę. Rysunek 7 ukazuje przykładowe okno stworzonego programu z wynikami obliczeń.

W pierwszej kolumnie wyników znajdują się nazwy wszystkich materiałów dydaktycznych umieszczonych w bazie danych. Trzy kolejne kolumny to wyniki obliczeń z użyciem intuicjonistycznych zbiorów rozmytych, następne trzy to wyniki obliczeń z użyciem klasycznych zbiorów rozmytych. Obliczenia dokonywano korzystając z odległości Euklidesa (ieuk, reuk), Hamminga (iham, rham) oraz Hausdorffa (ihau, rhau). Nad poszczególnymi kolumnami znajduje się średni czas obliczeń. W podanym przykładzie według obliczeń opartych na intuicjonistycznych zbiorach rozmytych najbardziej dopasowany do zapytania jest materiał dydaktyczny nazwany „grafika”, obliczenia oparte na klasycznych zbiorach rozmytych „mówią”, że najbardziej dopasowany jest materiał dydaktyczny nazwany „test”. Wyniki te są bardzo satysfakcjonujące. W ten sposób stworzone zapytanie

jest najbardziej relewantne do materiału dydaktycznego „grafika”.

Wyniki		0,0391 ms	0,037 ms	0,0405 ms	0,0385 ms	0,0371 ms	0,037 ms
kurs	lruk	ham	rhau	lrak	rhau	rhau	rhau
grafika		0,8526	0,8526	1,5361	1,5751	1,5751	
ekonometria	1	1	1	2,769	3,7244	3,7244	
ekonomia	1	1	1	2,5375	3,3573	3,3573	
logistyka	1	1	1	1,7722	1,9661	1,9661	
visual basic dla aplikacj	1	1	1	2,3828	2,8664	2,8664	
biuro elektroniczne	1	1	1	1,6729	1,9661	1,9661	
informatyka	1	1	1	1,6898	2	2	
test	1	1	1	1,1305	1,1858	1,1858	

Rys. 7. Wyniki obliczeń. Źródło: opracowanie własne

Obliczenia były wykonywane dla następującego zapytania:

Słowa kluczowe zapytania ( $\mu, \nu, \pi$ )
photoshop (0,9;0,0,1)
grafika rastrowa (0,8;0,1;0,1)
corel (0,2;0,1;0,7)

## Podsumowanie

Autor w dalszej pracy zamierza opracować rekomendację dotyczących możliwości zastosowania opracowanej metody do doboru materiału dydaktycznego. Sformułowane zostaną także wnioski wskazujące możliwe sposoby poprawy działania algorytmów klasyfikujących materiały dydaktyczne. Opracowane zostaną miary oceny jakości modelowanych rozwiązań, wyznaczone zostaną szczegółowe zmienne decyzyjne wspomagające podejmowanie decyzji w zakresie doboru materiałów dydaktycznych przy kształceniu ustawicznym. Przeprowadzone zostaną

także badania i symulacje autorskiej metody doboru materiałów dydaktycznych przy kształceniu ustawicznym.

W odczuciu autora przedstawiony temat ma charakter innowacyjny, zaś jego opracowanie może przyczynić się do zwiększenia efektywności systemów dobierających materiał dydaktyczny przy kształceniu na odległość. Mając świadomość złożoności obliczeniowej i czasowej wykonywanych w pracy analiz oraz uwzględniając jej praktyczne aspekty, opracowany model wykazuje predyspozycje do implementacji powstałego na jego podstawie algorytmu, z wykorzystaniem metod optymalizacyjnych, przy wsparciu nowoczesnych technik informatycznych. Zaproponowany temat, jest zagadnieniem o dużym znaczeniu zarówno naukowym jak i praktycznym. Ma przed sobą szerokie perspektywy rozwojowe i poznawcze. Z całą pewnością tematyka pracy ma zastosowanie w aspektach wdrożeniowych.

## Literatura

- [1]. Arrigo M., Gentile M., Taibi D., Di Giuseppe O. (2005): *Specialized Search Engines for E-learning*. Recent Research Developments in Learning Technologies.
- [2]. Atanassov K.T. (1999): *Intuitionistic fuzzy sets*. Springer-Verlag.
- [3]. Salton G., Buckley C. (1988): Term weighting approaches In automatic text retrieval. *Information Processing and Management*, vol. 24.
- [4]. Szmidt E., Baldwin J. (2003): *New similarity measures for intuitionistic fuzzy set theory and mass assignment theory*. *Notes on Intuitionistic Fuzzy Sets*, Vol. 9, Nr 3.
- [5]. Szmidt E., Kacprzyk J. (2004): *A concept of similarity for intuitionistic fuzzy sets and its use in group decision making*.
- [6]. Szmidt E., Kacprzyk J. (2003): *A measure of similarity for intuitionistic fuzzy sets*. Proc. EUSFLAT.
- [7]. Szmidt E., Kacprzyk J. (2004): *A similarity measure for intuitionistic fuzzy sets and its application in supporting medical diagnostic reasoning*.
- [8]. Szmidt E., Kacprzyk J. (2000): Distances between intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, Vol. 114, No.3.
- [9]. Szmidt E., Kacprzyk J. (2006): *Entropy and similarity for intuitionistic fuzzy sets*. 11th Int. Conf. IPMU, Paris, 2375-2382.
- [10]. Szmidt E., Kacprzyk J. (2007): *Rozumowanie oparte na przypadkach z użyciem intuicjonistycznych zbiorów rozmytych*.
- [11]. Szmidt E., Kacprzyk J. (2008): *Using intuitionistic fuzzy sets in text categorization*. In: L. Rutkowski, R. Tadeusiewicz, L. Zadeh, J. Zurada (Eds.): *Artificial Intelligence*

- and Soft Computing - ICAISC 2008. Springer, Germany, 351-362. Lecture Notes in Artificial Intelligence.
- [12]. Szmidt E. (2004): *Intuicjonistyczne zbiory rozmyte i ich zastosowania w podejmowaniu decyzji*. Badania Systemowe – Teoria i Zastosowania, Warszawa.
- [13]. Werdoni M. (2007): *Nauczanie na odległość*. w: J. Hołubiec (red.), Analiza systemowa w finansach i zarządzaniu, Instytut Badań Systemowych PAN, Warszawa.
- [14]. Werdoni M., Demianiuk M. (2008): *Koncepcja systemu elearningowego podnoszącego kwalifikacje pracowników*. w: L. Kiełtyka (red.), Technologie i systemy komunikacji oraz zarządzania informacją i wiedzą, "Difin", Warszawa, 123-133.
- [15]. Werdoni M. (2008): *Oprogramowanie typu OpenSource w nauczaniu na odległość*. w: Międzynarodowa Konferencja Naukowa z okazji 15-lecia Wydziału Zarządzania Politechniki Białostockiej. Gospodarka i społeczeństwo: zarządzanie - wiedza - rozwój: abstrakty, EkoPress, Białystok, s. 212.
- [16]. Werdoni M. (2008): *Intuicjonistyczne zbiory rozmyte w doborze materiałów dydaktycznych w kształceniu na odległość*. w: J. Hołubiec (red.), Analiza systemowa w finansach i zarządzaniu, Instytut Badań Systemowych PAN, Warszawa.
- [17]. Zadeh L.A. (1965): Fuzzy sets. *Information and Control*, Nr 8.
- [18]. Zeshui X., Jian C., Junjie W. (2008): Clustering algorithm for intuitionistic fuzzy sets. *Information Sciences*, Elsevier.



ISBN 9788389475220