



POLSKA AKADEMIA NAUK
Instytut Badań Systemowych

**KOMPUTEROWE WSPOMAGANIE
ZARZĄDZANIA I PROCESÓW
DECYZYJNYCH W GOSPODARCE**

pod redakcją:
Jana Studzińskiego
Ludostawa Drelichowskiego
Olgierda Hryniewicza



**KOMPUTEROWE WSPOMAGANIE ZARZĄDZANIA
I PROCESÓW DECYZYJNYCH W GOSPODARCE**

Polska Akademia Nauk • Instytut Badań Systemowych

Seria: BADANIA SYSTEMOWE
tom 31

Redaktor naukowy:

Prof. dr hab. Jakub Gutenbaum

Warszawa 2002

KOMPUTEROWE WSPOMAGANIE ZARZĄDZANIA I PROCESÓW DECYZYJNYCH W GOSPODARCE

pod redakcją

Jana Studzińskiego, Ludosława Drelichowskiego
i Olgierda Hryniewicza

Książka zawiera wybór artykułów poświęconych omówieniu aktualnego stanu badań w kraju w zakresie rozwoju i zastosowań technologii, modeli i systemów informatycznych w gospodarce narodowej.

Recenzenci artykułów:

Prof. dr hab. inż. Olgierd Hryniewicz

Prof. dr hab. inż. Janusz Kacprzyk

Dr inż. Lech Kruś

Dr inż. Edward Michalewski

Prof. dr hab. inż. Andrzej Straszak

Dr inż. Jan Studzinski

Dr inż. Sławomir Zadrozny

© Instytut Badań Systemowych PAN, Warszawa 2002

Wydawca: Instytut Badań Systemowych PAN
ul. Newelska 6 01-447 Warszawa

Redakcja: Dział Informacji Naukowej i Wydawnictw IBS PAN
tel. 837-68-22
Barbara Kotuszewska

Druk: Zakład Poligraficzny Urzędu Statystycznego w Bydgoszczy
Nakład 200 egz. ark. wyd. 23,5 ark. druk. 20,0

ISBN 83-85847-73-1
ISSN 0208-8028

Rozdział 3

Metody i algorytmy obliczeniowe w systemach wspomagania decyzji

ELEMENTY LOGIKI ROMYTEJ W WYSZUKIWANIU INFORMACJI TEKSTOWEJ: ROZSZERZENIA I INTERPRETACJE JĘZYKA ZAPYTAŃ

Sławomir Zadrozny

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania
01-447 Warszawa, ul. Nowelska 6*

The use of the fuzzy logic related concepts for the information retrieval purposes is discussed. In particular, the interpretation of the Boolean weighted queries is analyzed. Some results presented in the literature on the fuzzy querying of databases are adapted to the information retrieval context.

Keywords: information retrieval, weighted queries, fuzzy logic

1. Wstęp

Szeroko rozumiane przetwarzanie informacji tekstowej jest jednym z tradycyjnych obszarów zastosowań informatyki. Obszar ten obejmuje wiele różnorodnych zagadnień wymagających rozwiązań opracowanych w ramach wielu dyscyplin naukowych. Spośród wielorakich problemów badawczych w niniejszej pracy zajmujemy się kwestią reprezentacji i wyszukiwania dokumentów tekstowych. Historycznie, obszar ten jest ściśle powiązany z systemami bibliotecznymi umożliwiającymi szybkie wyszukanie pozycji książkowej poszukiwanej przez czytelnika i określonej za pomocą kwerendy. Każda książka opisana jest zestawem atrybutów takich jak autor, wydawnictwo, itp. Ważnym elementem takiego opisu są słowa kluczowe opisujące tematykę książki. To właśnie ten element stanowi o specyfice wyszukiwania informacji. Wspomniane wcześniej atrybuty takie jak autor, tytuł itp. nakładają regularną strukturę na dane opisujące książkę. Dla ich przetwarzania właściwym narzędziem jest system zarządzania bazą danych ze swoimi metodami reprezentacji danych i językami zapytań. Operowanie na danych takie jak słowa kluczowe wymaga innych narzędzi. Znajdują one szczególne zastosowanie w wypadku Internetu oferującego ogromne zasoby informacji tekstowej.

Rozwój badań w tym zakresie stymulowany jest zarówno rozwojem poszczególnych „zainteresowanych” dyscyplin naukowych jak i burzliwym rozwojem Internetu, który stał się swoistym poligonem dla prac nad rozwojem metod analizy i przetwarzania dokumentów tekstowych. Przeważająca część

informacji dostępne w sieci ma charakter tekstowy, choć udział multimediiów staje się coraz bardziej znaczący. Poszukiwanie nowych, bardziej inteligentnych metod przetwarzania informacji tekstowej w sposób naturalny obejmuje również zastosowania logiki rozmytej, u której podłoża leży dążenie do modelowania świata rzeczywistego na sposób zgodny z percepcją człowieka. Metodologia oferowana na gruncie logiki rozmytej pozwala również zreinterpretować znane podejścia w nowy sposób.

Niniejsza praca ma następujący układ. W p. 2 omówione są krótko podstawy reprezentacji i wyszukiwania dokumentów tekstowych. Kolejny punkt omawia bliżej rozszerzenia jednego z modeli wyszukiwania informacji. Wreszcie p. 4 zawiera dyskusję możliwych interpretacji tzw. „ważonych zapytań” – wykorzystuje się przy tym elementy logiki rozmytej.

2. Klasyczne modele wyszukiwania informacji tekstowej

Zwykle przyjmuje się trzy podstawowe *modele systemów wyszukiwania informacji*: Boolowski, wektorowy i probabilistyczny. Wyróżnieniu tych modeli służą zasadniczo dwa elementy: sposób reprezentacji dokumentów oraz forma zapytań i ich interpretacja. Dla naszych dalszych rozważań istotne będą tylko dwa pierwsze modele. W ich krótkim opisie przyjmujemy następujące oznaczenia:

$D = \{d_i\}_{i=1,N}$ zbiór dokumentów tekstowych

$T = \{t_j\}_{j=1,M}$ słownik terminów indeksujących

Zbiór D może być dokładnie znany (jak np. w systemach bibliotecznych zbiór książek stanowiących księgozbiór biblioteki) lub nie do końca określony, poznawany w trakcie eksploracji (jak np. w wypadku niektórych scenariuszy wyszukiwania w Internecie). Podobnie słownik może zawierać z góry określone (automatycznie lub przez eksperta) terminy bądź może uwzględniać przede wszystkim terminy użyte w zapytaniu. Klasyczne modele wyszukiwania zakładają ustalony zbiór zarówno terminów jak i indeksowanych z ich pomocą dokumentów.

W modelu Boolowskim dokument reprezentowany jest jako zbiór terminów:

$$d_i = \{t_k\}_{k=1,K} \quad d_i \in D \quad (1)$$

W modelu wektorowym dokument reprezentowany jest jako wektor:

$$d_i = [w_1, \dots, w_M] \quad d_i \in A^M \quad A \subseteq R \quad (2)$$

gdzie w_j (*stopień ważności, waga*) określa na ile termin $t_j \in T$ istotny jest dla reprezentacji dokumentu. Często przyjmuje się $A = [0, 1]$.

Reprezentacja wektorowa dokumentu wymaga przypisania każdemu terminowi słownikowemu t_i odpowiedniej wartości w_i (wagi). Proces ten nazywamy *indeksowaniem*. Odpowiednie indeksowanie zbioru dokumentów stanowi jeden z

najważniejszych czynników decydujących o efektywności późniejszego wyszukiwania dokumentów z danego zbioru. Tradycyjnie, np. w systemach bibliotecznych, czynność ta była i jest wykonywana przez ekspertów. Jednak w wypadku medium takiego jak Internet jest to praktycznie niewykonalne ze względu na lawinowy wzrost dostępnej informacji tekstowej. W związku z tym stosuje się algorytmy automatycznie indeksujące dokumenty. Formalnie, zakłada się istnienie funkcji F takiej że:

$$F: D \times T \rightarrow [0, 1] \quad (3)$$

Funkcja ta powinna przypisywać terminowi t_j w dokumencie d_i wagę wprost proporcjonalną do tego jak często występuje on w tym dokumencie i odwrotnie proporcjonalną do tego jak często występuje on w całym zbiorze dokumentów. W literaturze zaproponowano wiele konkretnych postaci funkcji F . Jedną z popularniejszych form jest $tf \times idf$ zdefiniowana jak następuje:

$$F(d_i, t_j) = (f_{ij} / \arg \max_j f_{ij}) * (\log(N/n_j) / \arg \max_j \log(N/n_j))$$

gdzie f_{ij} jest częstością występowania terminu t_j w dokumencie d_i , N liczbą wszystkich indeksowanych dokumentów i n_j liczbą dokumentów w których term t_j występuje. Tak więc pierwszy człon jest znormalizowaną częstością występowania terminu t_j w dokumencie d_i , natomiast człon drugi jest odwróconą znormalizowaną częstością występowania w zbiorze D dokumentów zawierających choć jedno wystąpienie t_j .

Każdy z klasycznych modeli systemów wyszukiwania informacji poza reprezentacją dokumentów określa również formalizm stosowany do formułowania zapytań. W modelu Boolowskim zapytanie ma formę zdania rachunku zdań. Każdemu terminowi t_j odpowiada w tym języku zmienna zdaniowa, z_j . Zmienne zdaniowe mogą być połączone za pomocą dowolnych spójników logicznych, w szczególności koniunkcji i alternatywy. Takie zdanie-zapytanie interpretowane jest dla danego dokumentu poprzez przypisanie każdej zmiennej z_j wartości „prawda” jeśli $t_j \in d_i$ i wartości „fałsz” w przeciwnym wypadku (przypomnijmy, że zgodnie z (1) w modelu Boolowskim dokument traktowany jest jako zbiór zawartych w nim terminów słownikowych). Dany dokument zostaje uznany za pasujący do zapytania jeśli całe zdanie-zapytanie jest prawdziwe dla niego. Tak więc w klasycznym modelu Boolowskim dopasowanie dokumentu do zapytania ma charakter binarny: dopasowanie występuje lub nie – nie funkcjonuje tu pojęcie częściowego dopasowania.

W modelu wektorowym zapytanie ma postać wektora wag wybranych terminów słownikowych. Mamy więc tu do czynienia z jednolitą reprezentacją dokumentów i zapytań. Dopasowanie dokumentu i zapytania określa się jako miarę podobieństwa, $\text{sim}(d, q)$, obydwu wektorów. Najbardziej klasyczną miarą tego typu jest cosinus obu wektorów:

$$\text{sim}(d,q) = \frac{\sum_{i=1}^l w_i q_i}{\sqrt{\sum_{i=1}^l w_i^2} \sqrt{\sum_{i=1}^l q_i^2}}$$

gdzie $d = [w_1, \dots, w_l]$, $q = [q_1, \dots, q_l]$. Tak więc dopasowanie dokumentu i zapytania wyraża się liczbą z przedziału $[0,1]$.

W każdym z modeli możemy wyróżnić *zapytania elementarne*, to jest takie które odwołują się tylko do jednego terminu słownikowego. Formalnie wyrażone są one w postaci jednoelementowego zbioru terminów (model Boolowski) i wektora jednowymiarowego (model wektorowy).

3. Rozszerzenia modelu Boolowskiego

Model wektorowy został dość powszechnie zaakceptowany. W szczególności uznaje się, że użycie wag zarówno do reprezentacji dokumentów jak i w zapytaniach nadaje mu dużej elastyczności. Z drugiej strony, język zapytań w tym modelu jest dość ubogi. Praktycznie, użytkownik formułujący zapytanie może jedynie określić, że poszukiwane dokumenty mają zawierać wymienione przez niego terminy z przypisanymi im stopniami ważności możliwie bliskimi tym podanym w zapytaniu. Zapytanie takie odpowiada mniej więcej koniunkcji zapytań elementarnych. Kontrastuje to mocno z językiem zapytań modelu Boolowskiego gdzie użytkownik może swobodnie posługiwać się dowolnie skomplikowanymi kwerendami zbudowanymi z użyciem spójników logicznych I, LUB, NIE. Łatwo zauważyć, że większość wyszukiwarek internetowych udostępnia użytkownikowi możliwość formułowania zaawansowanych zapytań Boolowskich. Zainspirowało to liczne prace nad rozszerzeniem modelu Boolowskiego tak aby zachowując funkcjonalność swojego języka zapytań przejął on jednocześnie zalety modelu wektorowego. Rozszerzony model Boolowski stanowi również punkt wyjścia dla prac związanych z zastosowaniem elementów logiki rozmytej na użytek systemów wyszukiwania informacji. Wektor reprezentujący dokument może być w oczywisty sposób zinterpretowany jako *rozmyty zbiór terminów*.

Rozszerzenia modelu Boolowskiego mogą dotyczyć jedynie reprezentacji dokumentów lub zarówno reprezentacji dokumentów jak i zapytań. W tym pierwszym wypadku, zakłada się, że dokumenty reprezentowane są tak jak w modelu wektorowym natomiast język zapytań pozostaje bez zmian. Interpretacja klasycznych zapytań Boolowskich względem dokumentów reprezentowanych za pomocą wag przypisanych poszczególnym terminom może mieć różnoraki charakter.

Po pierwsze, można przyjąć, że zapytania są teraz zdaniem *rozmytej logiki zdań*. Tak więc mogą być one prawdziwe w pewnym stopniu z przedziału $[0,1]$. Jak

w klasycznym podejściu dopasowanie dokumentu do zapytania oblicza się jako stopień prawdziwości zdania-zapytania przy wartościowaniu określonym przez dany dokument. Poszczególnym zmiennym zdaniowym – odpowiadającym jak poprzednio poszczególnym terminom ze słownika T – przypisuje się wartości odpowiadające wagom tych terminów w rozważanym dokumencie. Następnie stosuje się tradycyjne operatory rozmyte min i max odpowiadające klasycznym spójnikom logicznym I i LUB użytym w zapytaniu. Możliwe są do uzyskania różne warianty tego podejścia poprzez zastosowanie różnych *norm i konorm trójkątnych* dla reprezentacji spójników logicznych.

Kolejne rozwinięcia modelu Boolowskiego uwzględniają możliwość przypisania wag również terminom występującym w zapytaniach. W klasycznym modelu wektorowym interpretacja tych wag jest prosta: poszukuje się dokumentów, w których poszczególne terminy mają przypisane możliwie podobne wagi do tych użytych w zapytaniu. W wypadku rozszerzonego modelu Boolowskiego rozważa się więcej możliwych interpretacji. W celu ich krótkiego omówienia przyjmijmy, że zapytanie przedstawione jest w normalnej postaci dyzjunktywnej:

$$q = ((t_{11}, w_{11}) \wedge \dots \wedge (t_{1u}, w_{1u})) \vee \dots \vee ((t_{d1}, w_{d1}) \wedge \dots \wedge (t_{dw}, w_{dw}))$$

gdzie t_{ij} jest terminem słownikowym a w_{ij} przypisana mu w zapytaniu waga. Przy założeniu tej postaci kanonicznej zapytania możemy skoncentrować się na określeniu stopnia dopasowania pojedynczej koniunkcji $(t_{11}, w_{11}) \wedge \dots \wedge (t_{1u}, w_{1u})$, zwanej *dyzjunktem*. Dopasowanie całego zapytania otrzymujemy poprzez agregację (np. z użyciem operatora max) wartości obliczonych dla poszczególnych dyzjunktów. W literaturze rozważa się zwykle trzy interpretacje wag w_{ij} (patrz np. Bordogna, Carrara i Pasi, 1995):

- jako względnej ważności,
- jako wymaganych progów ważności,
- jako punktu (wektora) idealnego

W pierwszej z tych interpretacji waga w_{ij} terminu t_{ij} w zapytaniu wskazuje na ile ważne jest z punktu widzenia spełnienia zapytania wystąpienie tego terminu w poszukiwanych dokumentach. Jeśli waga ta jest niska (bliska 0) to brak t_{ij} w dokumencie (czyli niska, być może równa 0, waga przypisana terminowi w danym dokumencie) nie dyskwalifikuje tego dokumentu. Jeśli waga ta jest wysoka (bliska 1) to termin musi pojawić się w dokumencie (czyli mieć przypisaną wysoką w nim wagę) aby został on uznany za pasujący do zapytania. Wydaje się, że interpretacja ta jest dość oczywista i intuicyjnie zrozumiała dla użytkownika systemu wyszukiwania informacji. Znajduje ona zastosowanie również w wypadku tradycyjnej Boolowskiej reprezentacji dokumentów jako nierozmytych zbiorów terminów.

Druga interpretacja określa, że waga poszczególnych terminów w poszukiwanych dokumentach musi przekroczyć wskazane w zapytaniu wartości progowe w_{ij} . W wypadku klasycznej Boolowskiej reprezentacji dokumentów wagi

takie nie wnoszą nic nowego w stosunku do klasycznych zapytań Boolowskich. Wydaje się, że taka interpretacja wag może być interesująca w wypadku formułowania bardzo specjalizowanych zapytań.

Trzecia interpretacja jest analogiczna do przyjętej w modelu wektorowym: poszukujemy dokumentów w których wagi poszczególnych terminów są podobne do wag użytych w zapytaniu. Stosują się tu uwagi analogiczne do wymienionych w odniesieniu do drugiej interpretacji.

4. Elementy logiki rozmytej w języku zapytań rozszerzonego modelu Boolowskiego

Konkluzją z rozważań poprzedniego rozdziału może być stwierdzenie, że najbardziej interesującym modelem systemu wyszukiwania informacji jest rozszerzony model Boolowski uwzględniający wagi dla terminów słownikowych zarówno w reprezentacji dokumentów jak i w zapytaniach. Trudność stanowi różnorodność możliwych interpretacji wag przypisywanych terminom występującym w zapytaniu. Pewną unifikację wnoszą tu prace wyrosłe na gruncie zastosowania elementów logiki rozmytej. Prace te oryginalnie często dotyczą zastosowań w kontekście języków zapytań dla systemów zarządzania bazą danych. Poniżej omawiamy je pokrótce adaptując je do kontekstu systemów wyszukiwania informacji. W dalszej części zakładamy, że zapytanie ma postać koniunkcji terminów opatrzonych wagami:

$$q = (t_1, w_1) \wedge \dots \wedge (t_k, w_k) \quad (4)$$

zaś poszukiwany stopień dopasowania dokumentu i zapytania oznaczać będziemy $\gamma(q, d)$. Zakładamy również, że stopień dopasowania zapytania elementarnego $q_i = t_i$ (z pominięciem wagi) dla dokumentu d równy jest wadze przypisanej w d terminowi t_i czyli $F(d, t_i)$. Jednocześnie stopień dopasowania dla całego zapytania (4) obliczamy jako:

$$\gamma(q, d) = \min_i \max(\gamma(q_i, d)) \quad (5)$$

Dubois i Prade (1997) analizowali kilka modeli stopni ważności przypisanych do elementarnych zapytań. Pierwszy z nich, oryginalnie zaproponowany przez Yagera (1987), wyraża się następującym wzorem dla zapytania elementarnego $q_i = (t_i, w_i)$:

$$\gamma(q_i, d) = \max(F(d, t_i), 1 - w_i), \quad (6)$$

Tak więc, w skrajnym wypadku, jeśli zapytanie elementarne jest całkowicie nieważne ($w_i = 0$) to nie nakłada ono żadnych ograniczeń na postać dokumentu, który ma je spełniać (stopień dopasowania jest zawsze równy 1). W przeciwnym wypadku, aby dokument spełniał zapytanie musi w nim występować termin t_i i to z

wysoką przypisaną mu wagą. W ten sposób model ten odpowiada interpretacji wi jako względnej ważności (patrz p. 3).

Dubois i Prade (1997) zaproponowali rozszerzenie tego modelu, które umożliwiałoby posługiwanie się zmienną wagą w_i zapytania elementarnego. Waga w_i zmieniałaby się automatycznie dla dokumentu zależnie od wartości $F(d, t_j)$. Dokładniej, zasadniczo waga ta byłaby taka jak podana w zapytaniu, chyba że wartość $F(d, t_j)$ byłaby bardzo niska. Wtedy waga w_j dla tego dokumentu byłaby zwiększona tak aby silniej oddać w całościowym stopniu dopasowania całego zapytania fakt braku występowania terminu t_j w danym dokumencie. Przykładowo, taka interpretacja stopnia ważności może być przydatna przy wyszukiwaniu dokumentów scharakteryzowanych następująco. Powinny one zawierać termin t_j , niekoniecznie jako pierwszoplanowy (czyli całkowicie, w stopniu 1, zadowolając nas dokumenty gdzie $F(d, t_j) < 1$), ale niedopuszczalne jest aby termin ten był w nich nieistotny (bardzo niska wartość $F(d, t_j)$ musi powodować, że stopień dopasowania tego dokumenty i naszego całego zapytania będzie bliski 0).

Kolejny model, również oryginalnie zaproponowany przez Yagera w roku 1984 (patrz: Dubois i Prade, 1997), odpowiada interpretacji wagi w_i jako wymaganego proggu ważności z p. 3. Wyraża się to następującym wzorem:

$$\gamma(q_i, d) = \begin{cases} 1 & \text{jeśli } F(d, t_i) \geq w_i \\ F(d, t_i) \text{ wpp} & \end{cases} \quad (7)$$

Aby usunąć nieciągłość tak zdefiniowanej funkcji dopasowania zaproponowano również inny wzór:

$$\gamma(q_i, d) = \begin{cases} 1 & \text{jeśli } F(d, t_i) \geq w_i \\ F(d, t_i) / w_i \text{ wpp} & \end{cases} \quad (8)$$

Podobne lub identyczne wzory zostały również zaproponowane specyficznie dla systemów wyszukiwania informacji – patrz Radecki(1979), Bookstein(1980), Bordogna i in. (2000).

Dubois i Prade (1997) zauważyli, że wszystkie powyższe modele są szczególnymi przypadkami ogólnego schematu postaci:

$$\gamma(q_i, d) = w_i \rightarrow F(d, t_i) \quad (9)$$

gdzie \rightarrow jest operatorem implikacji rozmytej. Na gruncie logiki rozmytej operator implikacji nie jest jednoznacznie określony. Do najczęściej stosowanych należą: Kleene-Dienes ($x \rightarrow y = \max(1 - x, y)$), Gödel ($x \rightarrow y = \text{jeśli } x \leq y \text{ to } 1 \text{ wpp. } y$), Goguen ($x \rightarrow y = \min(y/x, 1)$). Omawiane wyżej modele interpretacji wag w zapytaniach otrzymuje się przyjmując w schemacie (9) jako operator implikacji, odpowiednio

Kleene-Dienes dla (6), Gödel dla (7) oraz Goguen dla (8). Szczegółową analizę powyższego schematu zawiera praca Bordogna i in. (2000).

Interesującą propozycję różnicowania ważności poszczególnych elementów zapytania opracowali Lacroix i Lavency (1987). W ich podejściu zapytanie q ma dwie części: zasadniczą q_z , którą muszą spełniać wszystkie wyszukane dokumenty i pomocniczą, q_p , składającą się z jednego lub więcej dodatkowych zapytań, które służą uporządkowaniu dokumentów wybranych przez q_z . Możliwe jest zróżnicowanie ważności zapytań należących do q_p ; mogą być one wszystkie równie ważne lub ustalona jest pewna hierarchia ważności. W pierwszym wypadku dokumenty pasujące do q_z uporządkowane są według liczby spełnionych zapytań z q_p . W drugim wypadku stosowane jest uporządkowanie leksykograficzne względem spełnienia zapytań z q_p (najpierw uwzględniane są te zapytania, które są wyżej w hierarchii). Podobny operator został zaproponowany przez Bordogna i Pasi (1995).

Innym sposobem uelastycznienia języka zapytań jest zastosowanie tzw. miękkich operatorów agregacji. Klasyczne spójniki logiczne I i LUB (odpowiadające kwantyfikatorom ogólnemu i szczegółowemu) są często krytykowane jako operatory agregacji (w tym wypadku stopni spełnienia zapytań elementarnych). Uwzględniają one przy agregacji stopień spełnienia tylko jednego zapytania elementarnego: najslabiej i najlepiej spełnionego, odpowiednio dla I i LUB. Lukę te zapełniają kwantyfikatory lingwistyczne wprowadzone przez Zadeha (1983). Pozwalają one na wyrażenie zapytania żądającego spełnienia np. *większości* składowych zapytań elementarnych. Również w tym wypadku możliwe jest uwzględnienie różnej ważności zapytań elementarnych. Stosowny wzór przyjmuje następującą postać:

$$\gamma(q, d) = \mu_Q \left(\frac{\sum_{i=1}^l \min(F(d, t_i), w_i)}{\sum_{i=1}^l w_i} \right) \quad (10)$$

gdzie Q jest zbiorem rozmytym reprezentującym rozważany kwantyfikator lingwistyczny (patrz Zadeh(1983)). Idea wykorzystania kwantyfikatorów lingwistycznych w językach zapytań do baz danych została oryginalnie zaproponowana przez Kacprzyka i Ziółkowskiego (1986); patrz też Kacprzyk i in. (1989). Została ona przeniesiona na grunt systemów wyszukiwania informacji przez Bordogna i Pasi (1995).

6. Uwagi końcowe

Głównym tematem niniejszej pracy jest kwestia interpretacji wag w języku zapytań rozszerzonego modelu Boolowskiego. Oczywiście daleko nie pokrywa to całości prac prowadzonych na świecie w zakresie wykorzystania elementów logiki

rozmytej w systemach wyszukiwania informacji. Wśród obszarów studiowanych z użyciem aparatu logiki rozmytej wymienić można: próby uwzględnienia struktury wyszukiwanych dokumentów (Bordogna i Pasi, 1995), lingwistyczne określanie stopnia ważności (Bordogna i Pasi, 1995), operacja dzielenia algebry relacji rozmytych w zastosowaniu do wyszukiwania informacji tekstowej (Bordogna i in., 2000), rozmyte tezaury (Miyamoto, 1990). Liczne wyniki osiągnięto w dziedzinie rozmytych algorytmów analizy skupień – obszaru tradycyjnie leżącego w centrum zainteresowania wyszukiwania informacji.

Literatura

- Baeza-Yates, R. i B. Ribeiro-Neto (1999) *Modern Information Retrieval*. ACM Press, Addison-Wesley.
- Bookstein, A. (1980) Fuzzy requests: an approach to weighted Boolean searches. *Journal of the American Society for Information Sciences*, 31, 240-247.
- Bordogna, G. i G. Pasi (1995) Linguistic aggregation operators in fuzzy information retrieval. *International Journal on Intelligent Systems*, 10(2), 233-248.
- Bordogna, G., P. Carrara i G. Pasi (1995) Fuzzy approaches to extend Boolean information retrieval. W: P. Bosc i J. Kacprzyk (red.), *Fuzziness in Database Management Systems*, Physica Verlag, Heidelberg, 231-274.
- Bordogna, G., P. Bosc i G. Pasi (2000) Extended Boolean information retrieval in terms of fuzzy inclusion. W: O. Pons, M.A. Vila i J. Kacprzyk (red.), *Knowledge Management in Fuzzy Databases*, Physica Verlag, Heidelberg, New York, 234-246.
- Dubois, D. i H. Prade (1997) Using fuzzy sets in flexible querying: why and how? W: *Flexible Querying Answering Systems*, T. Andreasen, H. Christiansen, and H.L. Larsen (red.) Kluwer Academic Publishers.
- Kacprzyk, J., G. Pasi, P. Vojtáš i S. Zadrozny (2000) Fuzzy querying: issues and perspectives. *Kybernetika*, 6(36), 605-616.
- Kacprzyk, J., S. Zadrozny i A. Ziółkowski (1989). FQUERY III+: a "human-consistent" database querying system based on fuzzy logic with linguistic quantifiers. *Information Systems*, 14, 443-453.
- Kacprzyk, J. i A. Ziółkowski (1986). Database queries with fuzzy linguistic quantifiers. *IEEE Transactions on Systems, Man and Cybernetics*. SMC, 16, 474 - 479.
- Kraft, D.H. i D.A. Buell (1992) Fuzzy Sets and Generalized Boolean Retrieval Systems. W: D. Dubois, H. Prade i R.R. Yager (red.) *Readings in Fuzzy Sets for Intelligent Systems*, Morgan Kaufmann Publishers, San Mateo.
- Kraft, D.H., G. Bordogna i G. Pasi (1994) An Extended Fuzzy Linguistic Approach to Generalize Boolean Information Retrieval, *Journal of Information Sciences*, 2(3), 119-134.
- Kraft, D.H., G. Bordogna i G. Pasi (1999) Fuzzy Set Techniques in Information Retrieval. W: J.C. Bezdek, D. Dubois i H. Prade (red.) *Fuzzy Sets in Approximate Reasoning and Information Systems*, vol. 3, The Handbook of Fuzzy Sets Series, Kluwer Academic Publishers, Norwell.

- Lacroix, M. i P. Lavency (1987) Preferences: Putting more knowledge into queries. *Proc. of the 13 International Conference on Very Large Databases*, Brighton, UK, 217-225.
- Miyamoto, S. (1990) *Fuzzy Sets in Information Retrieval and Cluster Analysis*. Kluwer Academic Publishers, Dordrecht, Boston, London.
- Radecki, T. (1979) Fuzzy set theoretical approach to document retrieval. *Information Processing and management* 15, 247-260.
- Salton, G., E.A. Fox i H. Wu (1983) Extended Boolean information retrieval. *Communications of ACM*, 26(11), 1022-1036.
- Yager R.R. (1987) A note on weighted queries in information retrieval systems. *Journal of the American Society for Information Science* 38, 23-24.
- Zadeh, L.A. (1983). A Computational Approach to Fuzzy Quantifiers in Natural Languages. *Computational Mathematics Applications*, 9, 149-184.

ISSN 0208-8028
ISBN 83-85847-73-1

**W celu uzyskania bliższych informacji i zakupu dodatkowych egzemplarzy
prosimy o kontakt z Instytutem Badań Systemowych PAN
ul. Newelska 6, 01-447 Warszawa
tel. 837-35-78 w. 241 e-mail: bibliote@ibspan.waw.pl**