



**POLSKA AKADEMIA NAUK**  
**Instytut Badań Systemowych**

---

**ROZMYTOŚĆ I BIPOLARNOŚĆ  
W INTELIGENTNYM WYSZUKIWANIU  
INFORMACJI**

**Sławomir Zadrozny**

**Warszawa 2013**



iBS PAN

**POLSKA AKADEMIA NAUK  
INSTYTUT BADAŃ SYSTEMOWYCH**

**Seria: BADANIA SYSTEMOWE  
Tom 73**

---

---

**Redaktor naukowy:  
Prof. dr hab. inż. Jakub Gutenbaum**

**Warszawa 2013**

## Rada redakcyjna serii: BADANIA SYSTEMOWE

Prof. Olgierd Hryniewicz - przewodniczący

Prof. Jakub Gutenbaum – redaktor naczelny

Prof. Janusz Kacprzyk

Prof. Tadeusz Kaczorek

Prof. Roman Kulikowski

Prof. Marek Libura

Prof. Krzysztof Malinowski

Prof. Zbigniew Nahorski

Prof. Marek Niezgódka

Prof. Roman Słowiński

Prof. Jan Studziński

Prof. Stanisław Walukiewicz

Prof. Andrzej Weryński

Prof. Antoni Żochowski

iBS PAN

**POLSKA AKADEMIA NAUK  
INSTYTUT BADAŃ SYSTEMOWYCH**

---

---

**Sławomir Zadrozny**

**ROZMYTOŚĆ I BIPOLARNOŚĆ  
W INTELIGENTNYM WYSZUKIWANIU  
INFORMACJI**

**Warszawa 2013**

**Copyright © by Instytut Badań Systemowych PAN  
Warszawa 2013**

**Autorzy:**

**Dr hab. Sławomir Zadrozny**

Instytut Badań Systemowych Polskiej Akademii Nauk

ul. Newelska 6, 01-447 Warszawa

*Slawomir.Zadrozny@ibspan.waw.pl*

**Recenzenci:**

**dr hab. inż. Maciej Krawczak**

**dr Marek Reformat**

**Skład:** Aneta M. Pielak

**Wydawca:**

**Instytut Badań Systemowych**

**Polskiej Akademii Nauk**

Newelska 6, 01-447 Warszawa

[www.ibspan.waw.pl](http://www.ibspan.waw.pl)

**ISSN 0208-8029**

**ISBN 83-894-7551-0**

# Rozdział 1

## Wstęp

Truizmem jest stwierdzenie, że żyjemy w erze informacji. Rozwój techniki w ostatnich dekadach umożliwił gromadzenie informacji na wielką skalę. Systemy informatyczne w supermarketach rejestrują dziennie miliony transakcji. Sieci telekomunikacyjne największych operatorów realizują dziennie setki milionów połączeń, których dane są zapisywane. Satelity obserwacyjne NASA przesyłają kilka gigabajtów danych na godzinę. Sekwencjonowanie DNA daje w wyniku terabajty danych w ciągu jednego dnia. Eksperymenty prowadzone w fizyce cząstek elementarnych są źródłem danych napływających w terabajtach na minutę.

Każdy z nas uczestniczy w procesach gromadzenia i przetwarzania danych, nawet jeśli nie jest tego świadom. Z drugiej strony coraz więcej osób w aktywny sposób poszukuje informacji i wchodzi w interakcje z systemami informatycznymi przy realizacji prozaicznych zajęć. Poszukiwanie informacji motywowane jest zwykle koniecznością podjęcia pewnej decyzji. Rosnący stopień skomplikowania świata powoduje, że podejmowanie decyzji staje się coraz większym wyzwaniem. Oczywiście postęp technologiczny zwalnia nas w dużym stopniu od podejmowania decyzji rutynowych - komputery robią to za nas lepiej i szybciej. Pomagają nam również w wyszukiwaniu informacji i w prostych przypadkach robią to bardzo dobrze. Często jednak nie do końca wiemy, jaka dokładnie informacja jest nam potrzebna. Wtedy komputery mogą nas wspomóc, ułatwiając nawigację wśród dostępnych zasobów danych i pomagając nam uświadomić sobie nasze potrzeby informacyjne.

Są jednak sytuacje, kiedy mamy względnie jasny obraz tego, co nas interesuje, ale nie jesteśmy w stanie "wytłumaczyć" tego komputerowi - przełożenie naszych potrzeb informacyjnych na standardowy język zapytań systemu informacyjnego jest bardzo trudne, o ile w ogóle możliwe.

Podobne zjawisko można zaobserwować po drugiej stronie kanału informacyjnego. Informacja, którą trzeba zapisać w bazie danych lub innym repozytorium informacji, jest często niekompletna, niepewna czy nieprecyzyjna. Nie zawsze można po prostu zignorować taką niedoskonałą informację. Poza tym mimo że informacja nie jest doskonała, to jednak przy zastosowaniu odpowiednich, inteligentnych technik można będzie być może podjąć na jej podstawie lepszą decyzję.

To właśnie obszar zainteresowania niniejszej książki. Interesować nas będą metody inteligentnego wyszukiwania i reprezentowania informacji pozwalające użytkownikowi lepiej i łatwiej wyartykułować jego preferencje, jak również skuteczniej reprezentować i przetwarzać informację niedoskonałą. Dotyczyć to będzie informacji ustrukturalizowanej, przechowywanej w bazach danych, jak i w być może nawet większym zakresie informacji tekstowej, której reprezentacja i wyszukiwanie stwarza istotnie większe trudności.

Szczególnie będzie nas interesowało zastosowanie metod szeroko rozumianej logiki rozmytej. Od chwili jej narodzin w roku 1965 logika rozmyta potwierdza swoją przydatność do modelowania preferencji człowieka, jego percepcji i przekonań. Ostatnio na jej gruncie rozwija się nowy nurt badań, który stawia sobie za cel między innymi opracowanie jeszcze lepszych metod reprezentacji i wyszukiwania danych. Punktem wyjścia jest tu obserwacja naturalnej skłonności człowieka do analizowania różnych aspektów podejmowanej decyzji w kategoriach pozytywne-negatywne. Dokonując wyboru spośród dostępnych opcji często zestawiać będziemy listy ich zalet i wad i na ich podstawie będziemy próbować podjąć decyzję. Taka bipolarność preferencji ma swoje uzasadnienie w fizjologii funkcjonowania mózgu.

Badania nad opracowaniem skutecznych metod wspomaganie użytkownika, biorących pod uwagę to zjawisko, są prowadzone na coraz szerszą skalę i z pewnością będą kontynuowane. W niniejszej książce odnotowujemy aktualny stan badań nad zastosowaniem rozmytości i bipolarności w zadaniach wyszukiwania i reprezentacji informacji.

Układ książki jest następujący. Rozdział 2 stanowi wprowadzenie do logiki rozmytej. Przedstawia się w nim pojęcie zbioru rozmytego i jego rozszerzeń. Omawia się teorię możliwości, która jest zasadniczym narzędziem do modelowania niepewności w ramach logiki rozmytej. Znajduje ona zastosowanie w większości spośród modeli inteligentnego wyszukiwania informacji omawianych w niniejszej książce. Rozdział zawiera również przegląd podstawowych pojęć z zakresu logiki klasycznej oraz bardziej szczegółowe omówienie wybranych zagadnień z zakresu logiki rozmytej

rozumianej jako logika wielowartościowa. Omawia się też logikę posybi-listyczną. Stanowi ona stosunkowo mniej znany formalizm, który jednak znajduje zastosowanie zarówno w omawianych modelach bipolarności, jak i w modelach wyszukiwania informacji tekstowej. Przedstawia się również ważne pojęcie zmiennej lingwistycznej, które odgrywa kluczową rolę w różnych podejściach modelowania rozmytego. Znaczna część rozdziału poświęcona jest operatorom agregacji, w tym kwantyfikatorom lingwistycznym. Ich zastosowanie ma istotne znaczenie dla uzyskania inteligentnego charakteru omawianych metod wyszukiwania informacji.

W rozdziale 3 omawia się podstawowe pojęcia z zakresu relacyjnych baz danych, niezbędne dla później przedstawianych zaawansowanych rozwiązań. Ważną część tego rozdziału stanowi opis wybranych metod modelowania informacji niekompletnej w bazach danych. Punktem wyjścia są klasyczne rozwiązania, stanowiące element relacyjnego modelu danych. Omawia się również wybrane modele rozmytych baz danych.

Rozdziały 4 i 5 zawierają główny materiał dotyczący inteligentnego wyszukiwania informacji w bazach danych, ze szczególnym uwzględnieniem bipolarności informacji i preferencji. Przedstawia się algebrę relacji rozmytych oraz nowsze języki wyszukiwania. Rozdział 5 omawia zjawisko bipolarności informacji i preferencji z punktu widzenia jego znaczenia w systemach przechowywania i wyszukiwania danych. Z wielu względów szczególne miejsce w tych rozważaniach zajmują zapytania bipolarne. Stanowią one naturalne wzbogacenie tradycyjnego aparatu bazodanowego i wychodzą naprzeciw potrzebom użytkowników realizujących złożone wyszukiwanie informacji. Znalazły one już duże zainteresowanie społeczności akademickiej udokumentowane licznymi publikacjami. Jednocześnie wiele zagadnień pozostaje otwartych i sama istota bipolarności z pewnością wymaga dalszych badań.

Rozdział 6 omawia zagadnienia związane z wyszukiwaniem informacji tekstowej. Dość szczegółowo przedstawia się różne aspekty i klasyczne modele wyszukiwania informacji tekstowej. Eksponuje się te zagadnienia, które związane są z możliwością zastosowania technik logiki rozmytej. Zwraca się uwagę na naturalną nieprecyzyjność i niepewność informacji przetwarzanej w procesie wyszukiwania.

W rozdziale 7 przedstawia się krótko historię badań nad zastosowaniami logiki rozmytej do modelowania i przetwarzania informacji tekstowej. W głównej części rozdziału omawia się nowy model wyszukiwania, pozwalający na jawne modelowanie nieprecyzyjności i niepewności dotyczącej reprezentacji dokumentów i preferencji użytkowników przy ich wyszukiwaniu.



**ISSN 0208-8029**  
**ISBN 83-894-7551-0**

---

**INSTYTUT BADAŃ SYSTEMOWYCH**  
**POLSKIEJ AKADEMII NAUK**

**tel.: (+48) 22 3810246 / 22 3810277 / 22 3810241 / 22 3810273**

**e-mail: [biblioteka@ibspan.waw.pl](mailto:biblioteka@ibspan.waw.pl)**

