

SPIS TREŚCI

PRZEDMOWA	9
1. SYSTEMY INFORMACYJNE.	13
1.1. Wprowadzenie	13
1.2. Definicja systemu informacyjnego	14
1.3. Własności systemów informacyjnych	19
1.4. Zależność atrybutów	26
1.5. Systemy zredukowane	33
1.6. Podsystemy informacyjne	36
1.7. Łączenie systemów informacyjnych	40
2. JĘZYKI SYSTEMÓW INFORMACYJNYCH	48
2.1. Wprowadzenie	48
2.2. Składnia języka	51
2.3. Semantyka języka	53
2.4. Reguły przekształcania termów	58
2.5. Postać normalna termów	60
2.6. Dokładność i efektywność języka	64
2.7. Instrukcje	67
2.8. Instrukcje liczbowe	68
2.9. Instrukcje numeryczne.	69
2.10. Instrukcje relacyjne	73
2.11. Instrukcje logiczne	75
2.12. Instrukcje porządkujące	77
3. SYSTEMY WIELOSTOPNIOWE I HIERARCHICZNE	79
3.1. Wprowadzenie	79
3.2. Systemy dwustopniowe	79
3.3. Systemy wielostopniowe	83
3.4. Systemy hierarchiczne	84
3.5. Uwagi końcowe	86
4. SYSTEMY ROZPROSZONE	87
4.1. Wprowadzenie	87
4.2. System rozproszony i jego języki	88
4.3. System z rozproszonymi atrybutami	89
4.4. System z rozproszonymi obiektami	93
4.5. Użytkownik centralny a użytkownik lokalny	94
4.6. Pamięć centralna i pamięć lokalna	96
4.7. Uwagi końcowe	98

5. SYSTEMY INFORMACYJNE WIELOWARTOŚCIOWE	100
5.1. Wprowadzenie	100
5.2. Przykład systemu wielowartościowego	100
5.3. Definicja systemu informacyjnego wielowartościowego	101
5.4. Składnia języka systemów wielowartościowych	105
5.5. Semantyka języka systemów wielowartościowych	105
5.6. Aksjomaty języka wielowartościowego	108
6. PRZYBLIŻONE SYSTEMY INFORMACYJNE.	111
6.1. Wprowadzenie	111
6.2. Definicje przybliżonego systemu informacyjnego	112
6.3. Przykład systemu przybliżonego	112
6.4. Składnia języka przybliżonego systemu informacyjnego	114
6.5. Semantyka języka przybliżonego systemu informacyjnego	115
6.6. Obliczanie wartości termów w języku przybliżonym	117
6.7. Obliczanie przybliżone wartości termów języka dokładnego	122
7. STOCHASTYCZNE SYSTEMY INFORMACYJNE	125
7.1. Wprowadzenie	125
7.2. Stochastyczny system informacyjny	126
7.3. Język stochastycznych systemów informacyjnych	128
7.4. Podsumowanie	132
8. PRZYBLIŻONA KLASYFIKACJA OBIEKTÓW	134
8.1. Wprowadzenie	134
8.2. Zbiory przybliżone	135
8.3. Własności przybliżeń	140
8.4. Przybliżony opis podzbiorów	145
8.5. Próbkę zbioru	148
8.6. Przykład zastosowania	149
8.7. Klasyfikacja wielowartościowa	151
8.8. Dokładność przybliżonej klasyfikacji	153
8.9. Uwagi końcowe	154
9. NIEPEŁNA KLASYFIKACJA OBIEKTÓW	155
9.1. Uwagi wstępne	155
9.2. Niepełna klasyfikacja dwuwartościowa	155
9.3. Przybliżona równość zbiorów	157
9.4. Przypadek wielu ekspertów	160
10. O PROCESIE UCZENIA SIĘ	163
10.1. Uwagi wstępne	163
10.2. Proces uczenia	163
10.3. Dokładność uczenia się — oszacowanie od dołu	165
10.4. Uwagi końcowe	167
ZAKOŃCZENIE	168

Dodatek A. POJĘCIA MATEMATYCZNE UŻYWANE W PRACY	170
A1. Notacja logiczna	170
A2. Zbiory	170
A3. Relacje	171
A4. Relacje równoważności	172
A5. Relacje porządku	172
A6. Funkcje	172
Dodatek B. UWAGI O REALIZACJI SYSTEMÓW INFORMACYJNYCH METODĄ SKŁADOWYCH ATOMOWYCH	174
B1. Uwagi wstępne	174
B2. Zbiory elementarne	174
B3. Wyszukiwanie zbiorów elementarnych.	176
B4. Postać normalna	177
B5. Wielodostęp	177
B6. Systemy rozproszone	177
B7. Konsekwencje sprzętowe.	178
LITERATURA	179
SKOROWIDZ	184

PRZEDMOWA

W niniejszej książce są rozważane dwa zasadnicze zagadnienia. Pierwsze z nich to, jak znaleźć obiekty, gdy opis ich własności jest dany w pewnym języku formalnym. Drugie jest w pewnym sensie zagadnieniem odwrotnym: dany jest zbiór obiektów, a należy znaleźć opis charakterystyczny tego zbioru w zadanym języku formalnym. Oba te problemy mają zasadnicze znaczenie we współczesnych zastosowaniach maszyn liczących.

Pierwszy z nich dotyczy systemów wyszukiwania informacji, drugi zaś jest związany z ważnym działem zastosowań maszyn liczących, a mianowicie z tzw. sztuczną inteligencją i wiąże się przede wszystkim z problemami automatyzacji rozumowania indukcyjnego.

Punktem wyjścia do obu tych problemów jest pojęcie systemu informacyjnego oraz jego języka, dlatego wyszukiwanie informacji oraz rozumowanie indukcyjne, zagadnienia na pierwszy rzut oka bardzo odległe, znalazły się w tym tekście obok siebie.

Celem niniejszego opracowania nie jest przedstawienie stanu tych dwu dziedzin, lecz raczej sprecyzowanie niektórych podstawowych pojęć dotyczących wyszukiwania informacji oraz rozumowania indukcyjnego.

Istnieje bardzo bogata literatura dotycząca podstaw teoretycznych wyszukiwania informacji oraz rozumowania indukcyjnego, jednakże w zasadzie nie będziemy się do niej odwoływać, gdyż przedstawione tu podejście różni się dość istotnie od dotychczasowego traktowania tych zagadnień. Punktem wyjścia do nowych rozważań jest, jak to wspomniano poprzednio, pojęcie systemu informacyjnego oraz związanego z nim języka formalnego; pojęcia te były sformułowane przez autora [89] oraz zmodyfikowane w pracy Marka i Pawlaka [63], a następnie były rozwijane w kilku ośrodkach w kraju, stąd niemal cała literatura dotycząca tego tematu jest głównie pióra autorów polskich.

Podane tu pojęcie systemu informacyjnego jest dość bliskie modelom rozważanym przez Maedę [56], Saltona [103], Wonga i Chianga [115], Babada [3], Cherniawskiego i Schneidera [8], a przede wszystkim Codda [9], jednakże nie pokrywa się ono całkowicie ze wspomnianymi modelami. Różnic tych jednak nie będziemy tu dyskutowali. Zainteresowany Czytelnik z łatwością może wyrobić sobie pogląd — na podstawie cytowanej literatury — na temat podobieństw i różnic między proponowaną definicją systemu informacyjnego a innymi modelami systemów informacyjnych.

Wprowadzone pojęcie systemu informacyjnego zostało użyte nie tylko do zbadania pewnych własności systemów wyszukiwania informacji, lecz także do sprecyzowania niektórych problemów związanych z automatyzacją rozumowania indukcyjnego. Ta część pracy została zainspirowana wynikami Michalskiego [79], i wykazuje największe podobieństwo do jego podejścia do tego zagadnienia. Główna różnica między proponowanym podejściem do problematyki rozumowania indukcyjnego a podejściem Michalskiego i innych polega na przyjęciu jako punktu wyjścia zdefiniowanego pojęcia systemu informacyjnego. Pozwoliło to między innymi na proste określenie dokładności rozumowania (w naszej terminologii — dokładności uczenia), co, jak się wydaje, nie byłoby tak proste w innych modelach.

Książka ta jest przeznaczona przede wszystkim dla studentów i wykładowców, jako podręcznik pomocniczy do przedmiotów „systemy informacyjne” oraz „sztuczna inteligencja”. Pewne partie mogą również zainteresować projektantów systemów informacyjnych oraz systemów klasyfikacyjnych, rozpoznających czy też uczących się. Chociaż w tej książce nie mówi się bezpośrednio o praktycznych zastosowaniach podanych idei, wiele z nich może znaleźć, a niektóre już znalazły takie zastosowanie.

Z podanych tu własności systemu informacyjnego wynika nowa metoda wyszukiwania informacji. Jest ona dość zbliżona do metody zaproponowanej przez Wonga i Chianga [115], chociaż zachodzą między nimi również istotne różnice. Występują też w niektórych koncepcjach daleko idące podobieństwa zaproponowanego ujęcia rozwiązań w modelu relacyjnym Codda, podejścia te jednak w istocie różnią się zasadniczo.

Proponowana metoda wyszukiwania informacji została nazwana roboczo *metodą składowych atomowych*. Została ona zrealizowana i zbadana od strony użytkowej, najpierw przez Margańskiego [71], a następnie niezależnie przez Błaszczyka [5] i Telukową [107]. Zrealizowane na jej podstawie systemy pracują od dłuższego czasu użytkowo i wykazują wiele zalet eksploatacyjnych. Kilka dalszych realizacji eksperymentalnych jest w toku, między innymi w Instytucie Matematyki Politechniki Warszawskiej, pod kierunkiem mgra K. Grzelaka.

Metoda ta została również z powodzeniem zastosowana przez A. T. Francisza do statystycznej bazy danych — na podstawie koncepcji podanych w pracach [4] i [7]¹⁾.

Także z rozważań nad rozumowaniem indukcyjnym wynikają nowe metody realizacji systemów rozpoznających, klasyfikujących i innych. Jedna z takich metod została zrealizowana [11] w celach eksperymentalnych, nie została ona jednak do tej pory w pełni zbadana.

Od strony teoretycznej podane tu pojęcia zostały rozszerzone w różnych kierunkach (patrz np. prace [22], [23], [48], [49], [108], [85], [86], [82], [111], [3], [6], jednakże jak dotąd nie zbadano możliwości ich praktycznego wykorzystania.

¹⁾ Patrz A. T. Francis: Application of the atomic constituents method to statistical data base, praca doktorska, Instytut Podstaw Informatyki PAN, 1982.

Lektura książki nie wymaga w zasadzie żadnego przygotowania specjalistycznego, jednakże osoby, które mają wiedzę praktyczną z zakresu budowy maszyn liczących i ich zastosowań, a w szczególności ci, którzy zrealizowali samodzielnie jakiś system, na pewno znajdą tu znacznie więcej materiału do przemyśleń niż Czytelnicy bez takiego doświadczenia, którzy będą w stanie zrozumieć jedynie warstwę „zewnętrzną” podanych rozważań.

Do zrozumienia książki jest potrzebna znajomość matematyki nie przekraczająca zasobu wiedzy absolwenta szkoły średniej. Czytelnikowi nie znającemu używanego tu formalizmu polecam książki Rasiowej [95], Kuratowskiego [36] oraz Marka i Onyszkiewicza [59], w których można znaleźć szczegółowe wyjaśnienie użytych pojęć i notacji.

Książka składa się z dziesięciu rozdziałów i dwu dodatków. Rozdziały od 1 do 7 dotyczą systemów wyszukiwania informacji, rozdziały 8÷10 rozumowania indukcyjnego. Do zrozumienia rozdziałów 8÷10 jest potrzebna znajomość rozdziałów 1, 2, w których są wyjaśnione pojęcia systemu informacyjnego oraz jego języka — stanowiące punkt wyjścia do rozważań nad systemami wyszukiwania informacji oraz do rozumowań indukcyjnych.

W dodatku A podano w skrócie wyjaśnienia najważniejszych pojęć matematycznych użytych w pracy, dodatek B zawiera pewne ogólne informacje związane z realizacją podanych tu koncepcji.

Pragnę wyrazić podziękowanie za wiele cennych uwag pierwszym Czytelnikom maszynopisu: Małgorzacie Pawlak, dr Alicji Wakulicz-Deja, prof. Mirosławowi Dąbrowskiemu oraz prof. Konradowi Fiałkowskiemu. Szczególną wdzięczność jestem winien doc. Ewie Orłowskiej oraz prof. Wiktrowi Markowi, których uwagi pozwoliły mi uniknąć wielu istotnych niedociągnięć i błędów.

Zdzisław Pawlak

Warszawa, październik 1981