

Spis treści

Wstęp	9
Rozdział 1	
JĘZYK I JEGO BADANIE	
1.1. Język	13
1.1.1. Uwagi ogólne	13
1.1.2. Geneza języka i uniwersalia językowe	13
1.1.3. Pojęcie i funkcje języka	15
1.1.4. Pismo	18
1.2. Językoznawstwo	19
1.2.1. Definicja i struktura językoznawstwa	19
1.2.2. Główne koncepcje w językoznawstwie współczesnym	20
1.3. Znaczenie koncepcji modelu w badaniach nad językiem	29
Rozdział 2	
STATYSTYCZNE MODELE OPISU ZAWARTOŚCI DOKUMENTÓW I ICH TAKSONOMIA	
2.1. Model zawartości dokumentu tekstowego	31
2.2. Wymiary klasyfikacji	32
2.3. Poziom opisu lingwistycznego	34
2.3.1. Poziom morfologiczny	34
2.3.2. Poziom frazeologiczny	34
2.3.3. Poziom syntaktyczny	35
2.3.4. Poziom semantyczny	36
2.3.5. Podsumowanie	36
2.4. Przeznaczenie modelu	37
2.4.1. Uwagi ogólne	37
2.4.2. Szacowanie podobieństwa	38
2.4.3. Wyszukiwanie informacji	39
2.4.4. Streszczanie dokumentów	39
2.4.5. Analiza wydźwięku	40
2.4.6. Tłumaczenie maszynowe	40
2.4.7. Podsumowanie	41
2.5. Podstawy teoretyczne modelu	41
2.6. Zakres wykorzystywanej wiedzy	43
Rozdział 3	
ANALIZA LINGWISTYCZNA TEKSTU	
3.1. Znakowanie dokumentu	45
3.2. Segmentacja tekstu	46
3.3. Lematyzacja	46

3.4. Analiza morfologiczna	47
3.4.1. Uwagi ogólne	47
3.4.2. Regułowe metody ujednoznaczniania morfologicznego	48
3.4.3. Model Markowa	49
3.4.4. Ukryty model Markowa	50
3.4.5. Wykorzystanie ukrytych łańcuchów Markowa do identyfikacji części mowy	61
3.4.6. Odległość pomiędzy łańcuchami znaków	69
3.5. Anotacja składniowa	73
3.5.1. Gramatyki formalne i ich klasyfikacja	73
3.5.2. Gramatyki formalne jako narzędzie opisu języków naturalnych	76
3.5.3. Parsowanie języków bazkontekstowych	77
3.6. Anotacja semantyczna i identyfikacja nazw własnych	87

Rozdział 4

MODELE ALGEBRAICZNE

4.1. Model przestrzeni wektorowej	89
4.2. Modelowanie ukrytej struktury semantycznej	92
4.2.1. Uwagi ogólne	92
4.2.2. Analiza głównych składowych	93
4.2.3. Analiza ukrytych wymiarów semantycznych	98
4.3. Odległość pomiędzy dokumentami w modelach algebraicznych	102
4.4. Wykorzystanie miar podobieństwa w procesie klasyfikacji bezwzorcowej dokumentów	105
4.5. Analiza streszczeń artykułów publikowanych w czasopiśmie „Przegląd Organizacji”	111

Rozdział 5

MODELE PROBABILISTYCZNE

5.1. Podejście probabilistyczne w eksploracyjnej analizie dokumentów	121
5.2. Probabilistyczna analiza ukrytych wymiarów semantycznych	124
5.3. Wnioskowanie bayesowskie	126
5.3.1. Ogólna charakterystyka wnioskowania bayesowskiego	126
5.3.2. Metoda maksymalnego prawdopodobieństwa a posteriori	129
5.3.3. Metoda oparta na aproksymacji Laplace'a	130
5.3.4. Metody symulacyjne oparte na zastosowaniu łańcuchów Markowa	132
5.4. Rozkład Dirichleta w modelowaniu zawartości dokumentów tekstowych	136
5.4.1. Rozkład beta i rozkład Dirichleta	136
5.4.2. Probabilistyczna analiza dokumentów monotematycznych	138
5.5. Ukryta alokacja Dirichleta	141
5.5.1. Postać modelu	141
5.5.2. Rozkład <i>a priori</i>	143
5.5.3. Funkcja wiarygodności	144
5.5.4. Wnioskowanie	145
5.5.5. Algorytm LDA	152
5.6. Analiza streszczeń prac doktorskich	155

Rozdział 6

METODY WYKORZYSTUJĄCE WIEDZĘ DZIEDZINOWĄ

6.1. Wstępna charakterystyka ontologii	170
6.2. Podobieństwo klas i instancji ontologii	174
6.2.1. Uwagi ogólne	174
6.2.2. Miary podobieństwa pomiędzy konceptami	174
6.2.3. Uwzględnienie kontekstu w wyznaczaniu podobieństwa pomiędzy klasami	176
6.3. Podobieństwo pomiędzy zbiorami klas	184
6.4. Analiza i pomiar podobieństwa dokumentów oparty na modelu ontologicznym	187
6.5. Analiza treści kształcenia wybranych przedmiotów akademickich	189

Rozdział 7

OPIS ZAWARTOŚCI DOKUMENTÓW ZA POMOCĄ SŁÓW I FRAZ KLUCZOWYCH

7.1. Klasyfikacja metod identyfikacji słów i fraz kluczowych	197
7.2. Metody oparte na częstości występowania terminów indeksujących	199
7.3. Metody oparte na teorii informacji	201
7.4. Metody oparte na współwystępowaniu wyrazów	201
7.4.1. Algorytm Matsuo-Ishizuki	201
7.4.2. Algorytm RAKE	204
7.5. Metody oparte na wynikach analizy taksonomicznej	206
7.6. Algorytm KEA	208

Rozdział 8

ANALIZA OPINII KONSUMENCKICH

8.1. Przedmiot badań	210
8.2. Rodzaje opinii i typy ich analizy	212
8.3. Rodzaje analiz opinii konsumenckich	212
8.4. Metody badawcze stosowane w analizie opinii	214
8.5. Analiza opinii nabywców telefonów komórkowych	214

Zakończenie	219
--------------------	------------

Literatura	221
-------------------	------------