

Obsah

1	Základní pojmy informačních systémů	7
2	Klasifikace informačních systémů	8
3	Vyhledávací systémy	10
4	Vyhledávací algoritmy	11
4.1	Elementární algoritmus	12
4.2	Vyhledávací metody s předzpracováním vzorků	13
4.2.1	Vyhledávací metody s předzpracováním vzorků se sousměrným vyhledáváním	13
4.2.1.1	Vyhledávací stroj	13
4.2.1.2	Knuth-Morris-Prattův algoritmus	14
4.2.1.3	Algoritmus Aho-Corasickové	16
4.2.1.4	Vyhledávací konečné automaty	19
4.2.1.5	Vyhledávání nekonečné množiny vzorků	23
4.2.2	Vyhledávací metody s předzpracováním vzorků s protisměrným vyhledáváním	34
4.2.2.1	Boyer-Mooreův algoritmus	34
4.2.2.2	Algoritmus Commentz-Walterové	36
4.2.2.3	Protisměrné vyhledávání nekonečné množiny vzorků	39
4.2.2.4	Dvoucestný automat se skokem	41
4.2.3	Shrnutí	43
4.3	Vyhledávací metody s předzpracováním textu – indexové metody	44
4.3.1	Implementace indexových systémů	45
4.3.1.1	Použití invertovaného souboru	45
4.3.1.2	Použití seznamu dokumentů	45
4.3.1.3	Souřadnicový systém s ukazateli	45
4.3.2	Metody indexování	47
4.3.2.1	Analýza textu	47
4.3.2.2	Jednoduchá metoda automatického indexování	49
4.3.2.3	Řízené indexování	49
4.3.2.4	Konstrukce tezauru	50
4.3.2.5	Vyhledávání vzorků pomocí fragmentových indexů	51
4.4	Vyhledávací metody s předzpracováním textu a vzorků – signaturové metody	51
4.4.1	Řetězené a vrstvené kódování	52
4.4.2	Metody tvorby signatur	53
4.4.3	Vyhledávání vzorku pomocí signatury a indexu	54
4.5	Vyhledávání vzorků pomocí dvojitého slovníku	54

5	Jazyky pro vyhledávání	56
6	Kompresce dat	60
6.1	Základní pojmy komprese dat	60
6.1.1	Kódování a dekódování	60
6.1.2	Entropie a redundance	62
6.2	Predikce a modelování	62
6.3	Reprezentace celých čísel	64
6.3.1	Fibonacciho kódy	64
6.3.2	Eliasovy kódy	65
6.4	Statistické metody komprese dat	67
6.4.1	Shannon-Fanovo kódování	67
6.4.2	Huffmanovo kódování	68
6.4.2.1	Statické Huffmanovo kódování	69
6.4.2.2	Vlastnosti Huffmanových stromů	70
6.4.2.3	Adaptivní Huffmanovo kódování	72
6.4.2.4	Implementační poznámky	74
6.4.3	Aritmetické kódování	76
6.5	Slovníkové metody komprese dat	78
6.5.1	Statické slovníkové metody	78
6.5.2	Semiadaptivní slovníkové metody	79
6.5.3	Adaptivní slovníkové metody	80
6.5.3.1	Metoda posuvného okna	80
6.5.3.2	Metody s rostoucím slovníkem	82
6.5.3.3	Slovníkové metody s restrukturalizací slovníku	87
6.6	Syntaktické metody	89
6.6.1	Derivační kódování	89
6.6.2	Analytické kódování	90
6.7	Kontextové modelování	91
6.7.1	Konečné kontextové modely	91
6.7.2	Modely založené na konečných automatech	92
6.7.2.1	Automaty s konečným kontextem	92
6.7.2.2	Dynamické Markovovo modelování	93
7	Kontrola správnosti textu	95
7.1	Kontrola textu pomocí frekvenčního slovníku	96
7.2	Kontrola textu pomocí dvojitého slovníku	96
7.3	Interaktivní kontrola textu	97
7.4	Kontrola textu založená na pravidelnosti slov	98