

Současné trendy v selekčních jazycích

Zpracoval Josef Schwarz (schwarzjv@seznam.cz)

Komentovaná základní literatura

Literatura je rozdělena do následujících hlavních oblastí:

- A. Věcné pořádkání informací, selekční jazyky – všeobecně
- B. Klasifikace, systematické selekční jazyky
- C. Předmětové selekční jazyky, deskriptorové selekční jazyky, tezaury
- D. Obsahová analýza, sémantická redukce dokumentů, indexace, klasifikace (proces), abstrahování
- E. Automatizované zpracování přirozeného jazyka
- F. Vyhledávání informací

V každé sekci jsou uvedeny vybrané položky s komentářem, obsahem a dalšími informacemi. Pořadí položek v jednotlivých oblastech je voleno podle důležitosti a charakteru jednotlivých prací.

A. VĚCNÉ POŘÁDÁNÍ INFORMACÍ, SELEKČNÍ JAZYKY – VŠEOBECNĚ

[AND03] ANDERSON, J.D. Organization of knowledge. In FEATHER, J., STURGES, P. (eds.). *International Encyclopedia of Information and Library Science*. 2nd ed. London : Routledge, 2003, s. 471-490.

Komentář: Rozsáhlé encyklopedické heslo systematicky identifikující základní procesy a prvky věcného pořádkání informací.

Vybavení publikace: rejstřík, literatura (cca 50 položek), doporučená četba

[FOS96] FOSKETT, A. C. *The subject approach to information*. 5th ed. London : Library Association Pub., 1996. 456 s.

Komentář: Monografie systematicky pokrývající celou oblast věcného zpracování informací. Popsány jsou vybrané SJ (DDC, MDT, Bliss, Ranganathan, třídění LC, předmětová hesla LC, Sears, vybrané oborové tezaury).

Vybavení publikace: rejstřík, literatura u jednotlivých kapitol

Obsah: 1. Theory of information retrieval systems; 2. Pre-coordinate indexing languages; 3. Post-coordinate systems; 4. Post-coordinate indexing languages; 5. Future prospects.

B. KLASIFIKACE, SYSTEMATICKÉ SELEKČNÍ JAZYKY

[BUC79] BUCHANAN, B. *Theory of library classification*. London : Clive Bingley, 1979. 141 s.

Komentář: Teoreticky fundovaný úvod do významu, teorie, funkcí a tvorby klasifikačních systémů. Autor se mj. systematicky věnuje tvorbě fazetové klasifikace a tvorbě notací.

Vybavení publikace: rejstřík, literatura u jednotlivých kapitol

Obsah (dle německého vydání): 1. Klassifikation: Definition und Einsatzmöglichkeiten; 2. Die verschiedenen Arten von Klassenbeziehungen; 3. Präkombinierte Klassifikationssysteme und Facettenklassifikationen; 4. Methodische Überlegungen zur Lösung von Ordnungsproblemen; 5.-6. Konstruktion einer Facettenklassifikation; 7.-9. Gestaltung des Notationssystems; 10. Das alphabetische Sachregister; 11. Universalklassifikationen; 12. Kritische Anmerkungen zur systematischen Ordnung; 13. Automatische Indexierung.

Německé vydání: BUCHANAN, B. *Bibliothekarische Klassifikationstheorie*. München : Saur, 1989. 151 s.

[DAL74] DAHLBERG, I. *Grundlagen universaler Wissensordnung : Probleme und Möglichkeiten eines universalen Klassifikationssystems des Wissens*. Pullach bei München : Verlag Dokumentation, 1974. 366 s.

Komentář: Komplexní pojednání o klasifikačních systémech zejména z hlediska univerzálních pořádacích soustav, zohledněny jsou terminologické, teoretické a historické aspekty problematiky

Další informace: 1., 2. a 7. kap. jsou dostupné v pracovním překladu

C. PŘEDMĚTOVÉ SELEKČNÍ JAZYKY, DESKRIPTOROVÉ SELEKČNÍ JAZYKY, TEZAURY

[BAK84] BAKO, M. *Informačné selekčné jazyky. III*. Bratislava : SPN, 1984. 245 s.

Komentář: Starší, ale teoreticky fundovaná učebnice zaměřená na deskriptorové SJ (tezaury) a zpracování přirozeného jazyka. Poskytuje rozsáhlý úvod do lingvistiky, která je základem pro teorii SJ; dále se věnuje významu, struktuře a tvorbě tezaurů, propojování SJ, automatické indexaci, klasifikaci a abstrahování.

Vybavení publikace: rejstřík typu KWIC, literatura u jednotlivých kapitol

Obsah: 1. Informácia a informačný systém; 2. Informácia a jej reprezentácia; 3. Základy jazykovedy; 4. Lingvistika a informatika; 5. Tezaurus a jeho výstavba; 6. Prepojitelnosť informačných jazykov; 7. Deskriptorové informačné jazyky a ich vzťah k iným informačným jazykom; 8. Automatické indexovanie a referovanie; 9. Automatická klasifikácia.

[SOE74] SOERGEL, D. *Indexing languages and thesauri : construction and maintenance*. Los Angeles : Melville Pub., 1974. 632 s.

Komentář: Starší, ale vyčerpávající teoretická monografie o významu, funkcích, teorii, typologii, struktuře, formátu, tvorbě a správě tezaurů.

Vybavení publikace: slovníček, rejstřík, literatura (cca 110 položek)

Obsah (podle orig. vydání): A. Überblick über Aufgaben und Aufbau eines Thesaurus. Aufwand für Herstellung eines Thesaurus; B. Begriffe und Begriffsbenennungen. Klassifikationssystem und Thesaurus und ihre Funktionen in einem Dokumentationssystem; C. Thesaurus-Struktur; D. Thesaurus-Format; E. Arbeitsablauf der Thesaurus-Herstellung; F. Regeln für Begriffsbenennungen, Rechtschreibung u. ä.; G. Weiterentwicklung von Thesauri; H. Thesauri als Grundlage für kooperative Dokumentation.

Originální vydání: SOERGEL, D. *Klassifikationssysteme und Thesauri*. Frankfurt am Main : DGD-Dokumentationsstelle, 1969. 224 s.

D. OBSAHOVÁ ANALÝZA, SÉMANTICKÁ REDUKCE DOKUMENTŮ, INDEXACE, KLASIFIKACE (PROCES), ABSTRAHOVÁNÍ

[FUG93] FUGMANN, R. *Subject analysis and indexing : theoretical foundation and practical advice*. Frankfurt am Main : Indeks Verlag, 1993. 250 s.

Komentář: Teoreticky fundovaná a formálně zpracovaná učebnice věnovaná obsahové analýze a indexaci. Podrobně rozebírá teoretické základy indexace (teorie pojmu, teorie pěti indexačních axiomů ad.), rozlišuje typy indexace (extrakce, přiřazování), rozebírá konzistenci indexace, metody abstrahování, tvorbu knižních rejstříků, tvorbu řízených slovníků, syntax SJ, hodnocení kvality

indexace. Neopomíjí ani praktické aspekty obsahové analýzy a indexace. Text je doprovázen řadou názorných schémat, každá kapitola končí několika otázkami, kterými si student může ověřit nabyté znalosti.

Vybavení publikace: systematický a abecední rejstřík, literatura (cca 50 položek), doporučená literatura
Obsah: 1. Introduction; 2. Information and information system; 3. Information system survival power; 4. Theoretical considerations on information storage and retrieval; 5. Indexing; 6. Some technology of information supply.

Další informace: Slovníček z publikace je dostupný v učebních materiálech

[LAN03] LANCASTER, F. W. *Indexing and abstracting in theory and practice*. 3rd ed. London : Facet, 2003. 464 s.

Komentář: Přehledová monografie zaměřená na věcné zpracování informací. Kromě indexace a klasifikace systematicky popisuje i tvorbu abstraktů, automatické procedury, hodnocení kvality a konzistence indexace, indexaci multimédií. Výklad je doprovázen četnými praktickými příklady.

Vybavení publikace: rejstřík, literatura (cca 600 položek)

Obsah: 1. Introduction; 2. Indexing principles; 3. Indexing practice; 4. Precoordinate indexes; 5. Consistency of indexing; 6. Quality of indexing; 7. Abstracts: types and functions; 8. Writing the abstracts; 9. Evaluation aspects; 10. Approaches used in indexing and abstracting services; 11. Enhancing the indexing; 12. On the indexing and abstracting of imaginative works; 13. Indexing multimedia sources; 14. text searching; 15. Automatic indexing, automatic abstracting and related procedures; 16. Indexing and the internet; 17. The future of indexing and abstracting; 18. Indexing exercises; 19. Abstracting exercises

E. AUTOMATIZOVANÉ ZPRACOVÁNÍ PŘÍROZENÉHO JAZYKA

[STR00] STROSSA, P. *Zpracování informačních fondů. Seš. č. 2, Algoritmizace a automatizace zpracování textových informací. 2.*, přeprac. a rozš. vyd. Praha : Vysoká škola ekonomická, Fakulta informatiky a statistiky, 2000. 175 s.

Komentář: Přehledný úvod do problematiky automatizovaného zpracování přirozeného jazyka a textu. Práce se zabývá otázkami automatické indexace (mj. popis systémů MOZAIKA a SÉMAN), metodami měření podobnosti dokumentů, automatickou tvorbou tezauru a automatickým abstrahováním.

Vybavení publikace: rejstřík, literatura (83 položek)

Obsah: 1. Automatické indexování; 2. Metody porovnávání dokumentů a dotazů založené na měření podobnosti; 3. Možnosti automatizace tvorby tezauru; 4. Některé další algoritmizovatelné operace.

[SCH02] SCHWARZ, J. *Současný stav a trendy automatické indexace dokumentů : přehledová studie*. Praha, říjen 2002-leden 2003. Dostupné na <http://full.nkp.cz/nkdb/docs/studie/MAIobsah.html>

Komentář: Současný stav a trendy vývoje, výzkumu a implementace automatické indexace dokumentů. Stručně jsou popsány základní koncepce a metody automatické indexace, mezi něž patří automatická extrakce a automatické přiřazování, a jsou uvedeny jejich vybrané techniky. Automatická indexace je zhodnocena i z hlediska možností její praktické implementace v současných informačních systémech a jsou uvedeny vybrané příklady úspěšných aplikací automatické indexace v ČR i zahraničí. Podrobněji je také popsána historie a současný stav automatické indexace v České republice.

Vybavení publikace: literatura (21 položek)

F. VYHLEDÁVÁNÍ INFORMACÍ

[BAE99] BAEZA-YATES, R., RIBEIRO-NETO, B. *Modern information retrieval*. New York : Addison-Wesley, 1999. 513 s.

Komentář: Moderně koncipovaná monografie, která se věnuje všem aspektům vyhledávání informací. Pro popis většiny vztahů používá matematický aparát, text obsahuje řadu názorných grafů a schémat. 2. kap. obsahuje vyčerpávající typologii a systematický popis vyhledávacích systémů, v 3. kap. je uveden popis TREC databáze – standardní databáze pro testování vyhledávacích systémů.

Vybavení publikace: slovníček, rejstřík, literatura (852 položek)

Obsah: 1. Introduction; 2. Modeling; 3. Retrieval evaluation; 4. Query languages; 5. Query operations; 6. Text and multimedia languages and properties; 7. Text operations; 8. Indexing and searching; 9. Parallel and distributed models and languages; 10. User interface and visualization; 11. Multimedia IR: models and languages; 12. Multimedia IR: Indexing and searching; 13. Searching the web; 14. Libraries and bibliographical systems; 15. Digital libraries.

Další informace: Úplný obsah, úvod a úplný rejstřík je dostupný na http://www.amazon.com/exec/obidos/tg/detail/-/020139829X/ref=lib_rd_ss/102-0801022-4978507?v=glance&s=books&vi=slide-show#reader-link