

Rešeršní činnost

ÚVODNÍ HODINA

22. 2. 2008

přednášející: Mgr. Petr Šmejkal

43262@phil.muni.cz

ÚK FF MU, Kabinet informačních studií a knihovnictví FF MU

Přednášky a cvičení

➤ témata:

- zaměření na uživatele, informační požadavek, dialog s uživatelem
 - formulace rešeršního dotazu, rešeršní strategie a taktiky v databázích
 - věcné vyhledávání
 - citační software
 - rešeršní strategie v prostředí internetu
 - hodnocení informací a informačních zdrojů
- **vystavení přednášek v učebních materiálech** na <https://is.muni.cz/auth/el/1421/jaro2008/VIKMA06/um/>

Osnova prezentace

- Požadavky na kolokvium
- Výklad termínu „vyhledávání informací“
- Složky procesu vyhledávání informací
- Rešeršní služby
- Zadání rešerše
- Úkol na příští hodinu
- Druhy rešerší
- Povinná literatura
- Doporučená literatura

Nalezení výkladu

☞ vyhledávání informací OR information retrieval

? otázka volby vhodného zdroje pro vyhledávání

? jak se k tomuto zdroji dostat


Různé možnosti nalezení výkladu

Příklady

- oborové slovníky, encyklopedie
 - dostupné z EIZ MU (<http://library.muni.cz/ezdroje/>)
 - Encyclopedia of Library and Information Science (ELIS)
 - Encyclopedia of Communication and Information (v balíčku GALE e-books)
 - ODLIS : Online Dictionary for Library and Information Science
 - Modern Information Retrieval Glossary
 - Terminologická databáze knihovnictví a informační vědy (TDKIV)
- vyhledávací stroje/vyhledávače/search engine
 - např. operátor define:information retrieval (<http://www.google.com>)
 - <http://www.ask.com>
 - <http://www.exalead.com>
- předmětové katalogy
 - Internet library for librarians <http://www.itcompany.com/inforetriever/>
 - Open directory project <http://www.dmoz.org/>
- monografická publikace, přehledy, učebnice (Introduction to Information Retrieval, Modern Information Retrieval)

... a různé výklady IR

IR jako disciplína - řeší problémy spjaté s IR na teoretické a praktické úrovni. Výzkum zahrnuje např. modelování, klasifikaci a kategorizaci dokumentů, systémovou architekturu, uživatelské rozhraní, vizualizaci dat, filtrování, dotazovací jazyky atd.

 **různé pohledy na problém:**

computer-centered - zpracování informačních dotazů co nejvýkonněji, algoritmy a techniky rešeršních systémů apod. (oblast počítačové vědy)

human-centered - chování uživatele, porozumění jeho potřebám, určování, jak toto porozumění působí na organizaci a operace rešeršního systému (pole informační vědy)

(více viz Baeza-Yates, Ricardo. *Modern Information Retrieval* [online]. Addison Wesley Longman Publishing. [cit. 2007-02-21].

Chapter 1, Introduction. Dostupné z

WWW: <http://sunsite.dcc.uchile.cl/irbook/1/node1.html>

IR jako proces

Výklad *vyhledávání informací* v TDKIV :

Činnost, jejímž cílem je identifikace relevantních dokumentů nebo informací v informačních zdrojích (např. plnotextové nebo bibliograf. databáze). Vyhledávání informací probíhá na základě konkrétního požadavku uživatele za pomoci dotazovacích a selekčních jazyků.

IR jako proces

Výklad podle J. Steinerové (1996:25, 47)

informačný priezkumu : komplex procesů směřující k vyhledání relevantních informací (tj. odpovídající informačnímu požadavku) z informačních zdrojů, přičemž výstupem jsou specifickým způsobem interpretované, strukturované a prezentované informace.

Výklad z ODLIS

Information retrieval (IR)

The process, methods, and procedures used to selectively recall recorded information from a file of data. In libraries and archives, searches are typically for a known item or for information on a specific subject, and the file is usually a human-readable catalog or index, or a computer-based information storage and retrieval system, such as an online catalog or bibliographic database. In designing such systems, balance must be attained between speed, accuracy, cost, convenience, and effectiveness.

Search

A systematic effort on the part of a library user or librarian to locate desired information by manual or electronic means, whether successful or not, as opposed to browsing a library collection casually with no clear intention in mind.

Složky vyhledávání informací

Užší pojetí dle Chu (2007:14):

- *database, search mechanism, language, interface*
- + people, information, information retrieval systems IRS – řešeršní systém (online systems, CD-ROM s., OPAC, Internet Retrieval Systems)

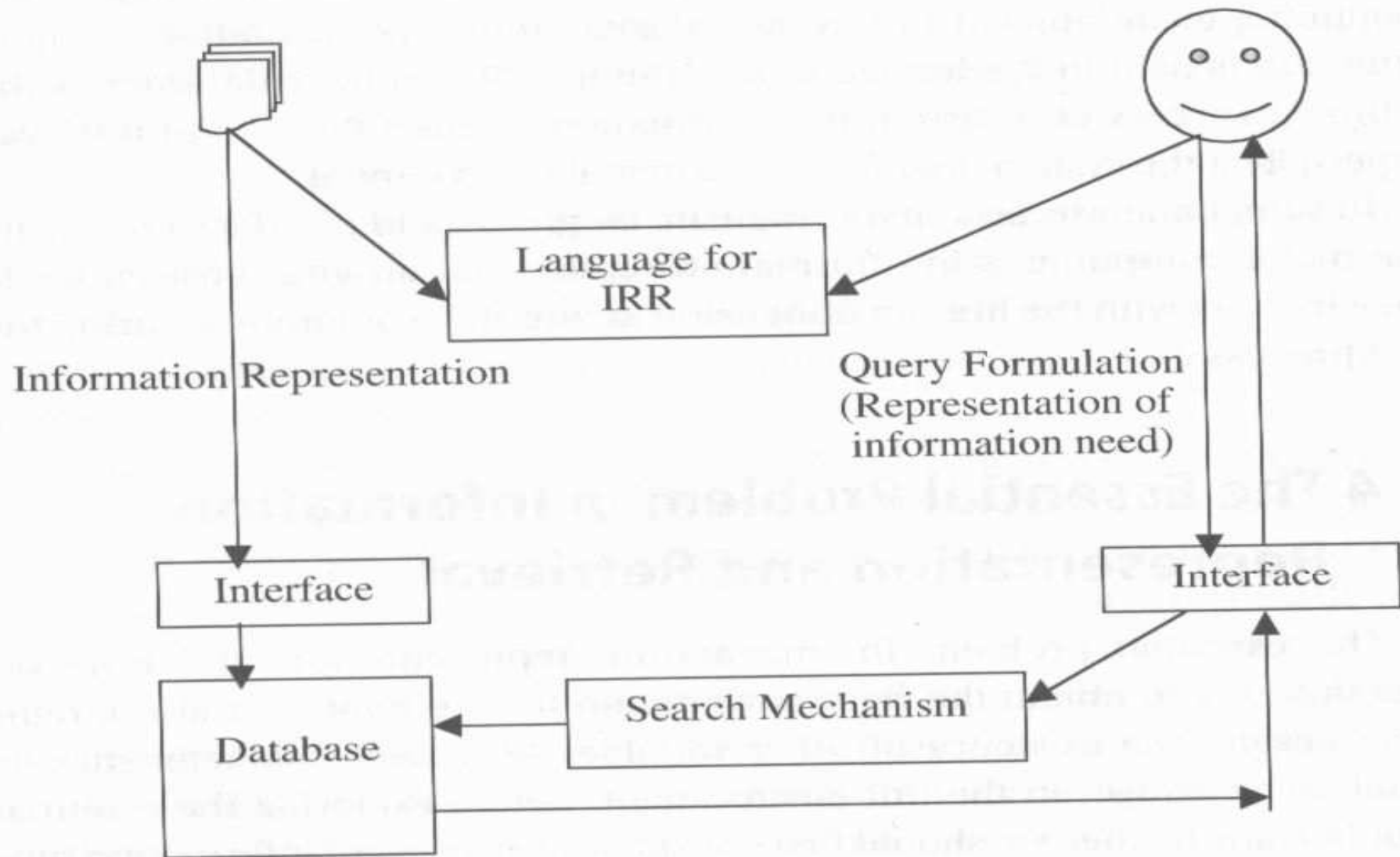


Figure 1.1 Process of Information Representation and Retrieval

Databáze

Sbírka nějakým způsobem reprezentovaných a uspořádaných informací/množina jednotně strukturovaných dat uložených v paměti počítače nebo na záznamovém médiu, jež jsou navzájem v určitém vztahu a tvoří určitý celek z hlediska obsažených informací (datový obsah databáze, TDKIV)

Dvojí pohled

- databáze = datový obsah + systém řízení báze dat (SŘBD, program pro práci s daty), např. v TDKIV
- zvlášť databáze a SŘBD = databázový systém
- ☞ řešeršní systém ≠ databáze, databáze je součástí řešeršního systému

Databáze

- db jako sbírka informací je organizována do **záznamů** (records) a ty se dále skládají z **polí** (fields)

Db je obvykle rozdělena na dvě části:

- vlastní data - **sekvenční soubory** (struktura pole-záznam-databáze)
- **indexy** – seznamy hodnot (slovo, fráze) z určitého pole nebo více polí s identifikací, z jakého záznamu hodnota pochází, umožňují přístup k záznamům v db dle položených dotazů - zrychlují proces vyhledávání
- více viz Bell, 2006:5-17

Rešeršní mechanismus

- informace z db mohou být vyhledány jen tehdy, když je poskytnut odpovídající rešeršní mechanismus
- od kapacity (definované rešeršní algoritmy) rešeršního mechanismu se odvíjí, jaké vyhledávací techniky budou dostupné a jak mohou být informace vyhledány
 - základní a pokročilé techniky vyhledávání viz Chu, 2007

Jazyk

Pro vyhledávání se používá

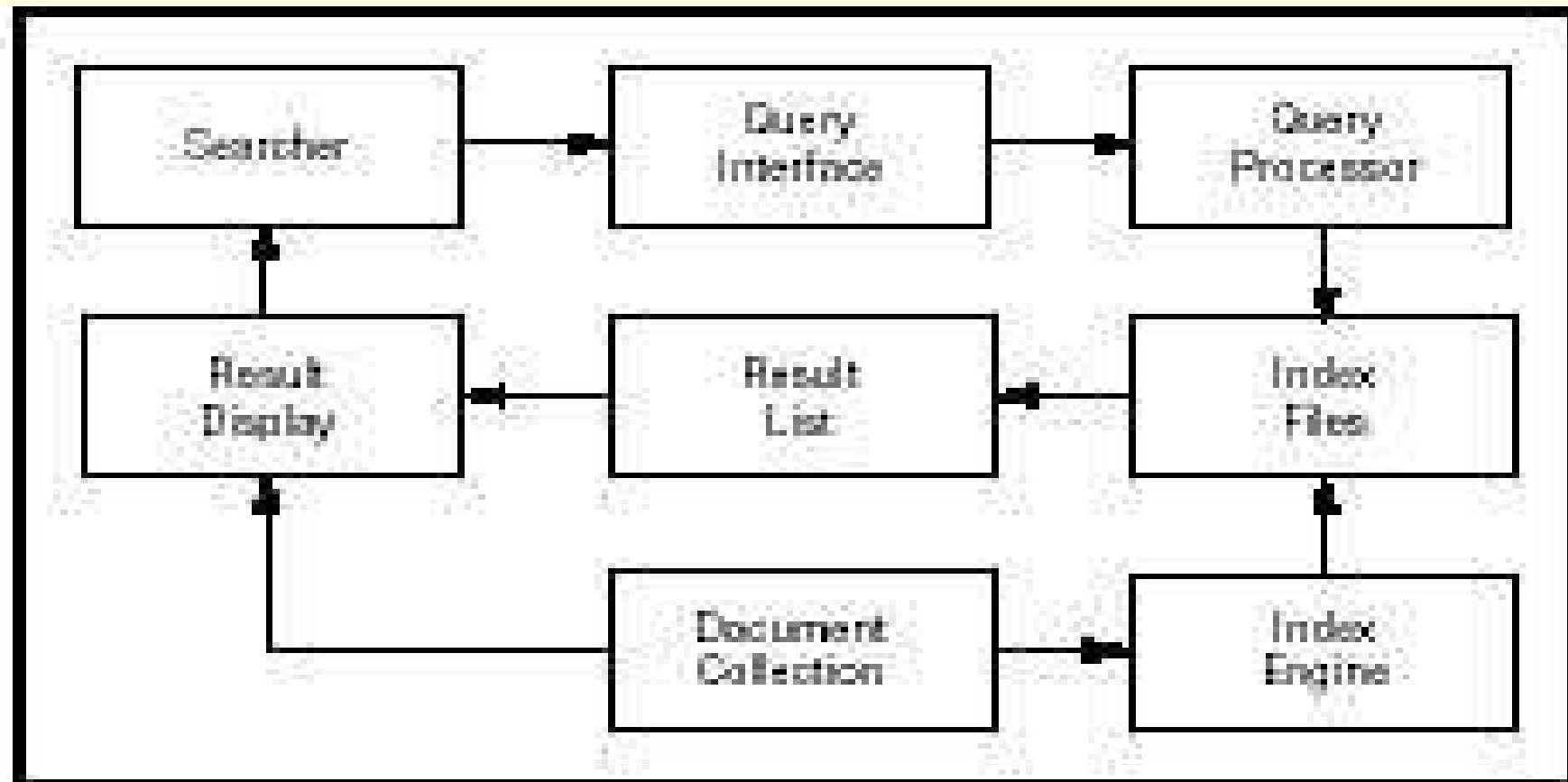
- přirozený jazyk
- umělý jazyk – věcný selekční jazyk

Uživatelské rozhraní

Uživatelské rozhraní - rozhraní mezi uživatelem a počítačovým programem... (def. z TDKIV)

- **slouží pro realizaci interakce člověk-počítač**
- **při vyhledávání informací odkazuje k interakci mezi uživatelem a vztahujícími se aktivitami**
- **koncový uživatel** – středobodem procesů týkajících se problematiky vyhledávání informací (human dimension in IR) – od toho se odvíjí **řešení designu rozhraní IRS** (např. standardy přístupnosti a použitelnosti)
- IRS – přívětivé uživatelské rozhraní na principu grafického uživatelského rozhraní x dotazy pomocí příkazového jazyka (zkušení řešeršéri)
- více viz Papík, 2001.

Složky rešeršního systému (IRS)



převzato z HEIDORN, 2002:889

Důležité termíny

- **relevantní informace** - informace formálně nebo věcně shodné s dotazem zadáným uživatelem v přirozeném či selekčním jazyku. Formální relevance je shoda formulace rešeršního dotazu se selekčními údaji dokumentu, věcná relevance je shoda obsahu informačního dotazu s obsahem vyhledaného dokumentu (výklad z TDKIV).
- **relevantní dokument** - dokument, který odpovídá kritériu relevance, tj. jeho selekční obraz je formálně nebo věcně shodný se selekčním obrazem informačního dotazu (výklad z TDKIV).
- **pertinentní dokument** – dokument, který uzná uživatel za užitečný (HEIDORN, 2002), subjektivní kategorie
- **míra přesnosti rešerše** - koeficient pro hodnocení relevance rešerše definovaný jako poměr počtu nalezených relevantních záznamů k celkovému počtu záznamů v rešerši. Obvykle se vyjadřuje v procentech (výklad z TDKIV).
 - jak velká část nalezených dok. je relevantní
- **míra úplnosti rešerše** - koeficient pro hodnocení relevance rešerši definovaný jako poměr počtu nalezených relevantních záznamů k počtu všech relevantních záznamů v databázi. Obvykle se vyjadřuje v procentech (výklad z TDKIV).
 - jak velká část relevantních dok. byla vyhledána

Rešeršní služby

V knihovnách a jiných informačních institucích se **služby**, které se zaměřují na **vyhledávání informací na požádání uživatelů**, označují jako **REŠERŠNÍ**.



REŠERŠNÍ SLUŽBY

- nadstavbové služby
- sestavování a zpřístupňování výstupů rešeršní činnosti (rešerší různých druhů) knihovnami a informačními institucemi
- jednotlivé instituce poskytují svým uživatelům různé druhy rešerší – z vlastních fondů, prostřednictvím externích informačních zdrojů
- jde o placené služby - ceníky

REŠERŠE

- **proces** vyhledávání informací o určité problematice na základě zadaného rešeršního požadavku
- výsledek tohoto vyhledávání informací (**produkt**) – množina informací odpovídající rešeršnímu požadavku, rešerše fixovaná na nosič se považuje za sekundární dokument - soupis záznamů dokumentů, popř. soupis jejich částí.
 - Dříve formální úprava norma ČSN 01 0198 : Formální úprava rešerší. Praha: Český normalizační institut, 1972 (neplatná norma)

Výklad rešerše v TDKIV

Výsledek (popř. proces) vyhledávání informací ve formě dokumentografických (bibliografických) nebo faktografických záznamů, popř. plných textů dokumentů.

Rešerše se zpracovává na základě rešeršního požadavku uživatele, který je zformulován pomocí dotazovacího jazyka do rešeršního dotazu.

☞ **informačním požadavkem** se rozumí požadované informace, které jsou považovány za nezbytné pro splnění úkolů a cílů (Steinerová, 1996)

☞ **rešeršní požadavek** - **konkrétně formulovaná žádost uživatele o zpracování rešerše**. Je obvykle formulován v přirozeném jazyce, prezentován písemně, ústně nebo jiným způsobem (výklad TDKIV).

- informační požadavek, který se stane předmětem procesu vyhledávání informací

Důležitý znak řešerše:

vyhotovuje se na základě *řešeršního požadavku* (zpracovává se jako odpověď na požadavek uživatele) – adresný charakter (uživatelská orientace)

- ☞ *konkrétně formulovaná žádost uživatele o zpracování řešerše*
- ☞ co by mělo být vymezeno: téma, časové vymezení, jazyk, geografické omezení, druhy dokumentů
- ☞ příklady: objednávka, požadavkový list

Ještě jednou...

Záznamy musí být vybrány podle **věcných a formálních hledisek** odpovídajících řešeršnému požadavku uživatele, např. tematika, časové vymezení, jazyk, druh dokumentů.

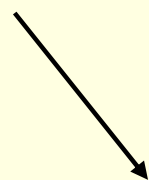
☞ informační/rešeršní požadavek – dotazovací jazyk – informační/rešeršní dotaz

Vyjádřený rešeršní požadavek nelze většinou použít přímo k získání požadované informace, např. na internetu prostřednictvím vyhledávacích nástrojů. Místo toho, uživatel musí nejprve **převést svůj požadavek do rešeršního dotazu**, který bude v souladu s **dotazovacím jazykem** daného rešeršního systému.

LISA: Library and Information Science Abstracts

reš. požadavek: uplatnění MDT v univerzitních knihovnách

reš. dotaz: de=(UDC) AND de=(university librar*)



formulace řešeršních dotazů –
dodržení principů dotazovacího jazyka
systému ILLUMINA

- užívání operátorů dotazovacího jazyka
- selekční pole (kódy a syntaxe, např. de=)

LLIS Library Literature & Information Science Fulltext

de=(UDC) AND de=(university librar*)

X

(Classification/Systems/Universal decimal) in SU
AND (College and university libraries) in SU

rešeršní dotaz v LLIS Library Literature &
Information Science Fulltext (systém WilsonWeb
LLIS)

Rešeršní dotaz/query

Vyjádření rešeršního požadavku uživatele v příslušném dotazovacím jazyce. Může jej formulovat informační pracovník nebo uživatel (výklad TDKIV).

- ☞ formulace rešeršního dotazu musí být v souladu s dotazovacím jazykem, který systém využívá

Dotazovací jazyk/query language

Jazyk určený ke komunikaci uživatele s vyhledávacím programem, umožňující vyhledávání a případnou úpravu požadovaných dat (výklad TDKIV).

Úkol na příští hodinu

- přečtení citované kapitoly
- DRÁBKOVÁ, Kateřina. **Informační potřeby, vnímání knihoven a informačních zdrojů ve společnosti** (se zaměřením na akademické prostředí). Brno: Masarykova univerzita, Filozofická fakulta, Ústav české literatury a knihovnictví, Kabinet knihovnictví, 2007. 110 s. + VII s. příl. Vedoucí diplomové práce Věra Jurmanová Volemanová. Dostupné též na WWW:
https://is.muni.cz/auth/th/74369/ff_m/kdrabkova_mgr_prace.pdf?fakulta=1421;obdobi=3704
- Kap. 2 INFORMAČNÍCH POTŘEBY A PROCES HLEDÁNÍ INFORMACÍ, s. 26-38

DRUHY REŠERŠÍ (Matoušová, 1988)

základní druhy rešerší (časové hledisko)

- **jednorázová** /retrospektivní/ r. - shrnuje materiál k danému tématu za stanovené období k určitému datu
 - výchozí datum – v závislosti na tématu rešerše, limitující faktor – retrospektiva dostupných sekundárních IZ, přihlédnutí k životnosti informací v jednotlivých oborech
 - konečné datum většinou současnost – snaha o podchycení nejnovější literatury

časové hledisko – pokr.

- **průběžná rešerše** – zjišťuje literaturu k danému tématu od data zadání směrem do budoucnosti
 - většinou se zpracovávají na základě sledování přírůstku jednoho sekundárního IZ
 - jednotlivé části rešerše mají předem dohodnutou periodicitu

služby poskytující tuto rešerši – *ARI adresní rozšiřování informací (SDI Selective Dissemination of Information)*

Klasické dělení podle úplnosti zahrnutého materiálu

- **úplná r.** – zahrnuje všechny zjistitelné informace nebo záznamy dokumentů (druhově komplexní, zachycení všech možných území)
- **výběrová r.** – obsahuje výběr materiálů zjištěných podle stanovených věcných a formálních hledisek
- **orientační r.** – poskytuje informace, které mají dát představu o předem málo známé problematice





Hledisko použitých pramenů

- **dokumentograf./bibliograf. r.** - zahrnuje dokumentografické (bibliografické) záznamy dokumentů
- **faktografická r.** – zahrnuje faktografické informace, tj. konkrétní fakta, data, parametry a další informace o věcech, jevech i osobách
 - statistická data, tabulky (chemické, fyzikální, technické), seznamy firem, výrobců

Podle druhu dokumentů

- **druhově komplexní r.** – zahrnuje záznamy všech druhů dokumentů, od nichž byly relevantní dokumenty zjištěny
- **jednodruhová r.** - záznamy dokumentů pouze jediného druhu dokumentů, např. patentová řešerše, normalizační řešerše
- **vícedruhová r.** - záznamy dokumentů dvou nebo více druhů dokumentů

Povinná literatura

-  Bell, S. S. *Librarian's guide to online searching*. Westport : Libraries Unlimited, 2006. 268 s. ISBN 1-59158-326-8.
dostupné v ÚK FF MU (registrační pult)
-  Heidorn, P. B.; Downie, J. S. *Retrieval of Information*. In *Encyclopedia of Communication and Information*. New York : Macmillan Reference USA, 2002, s. 888-896. Dostupné z WWW: <http://go.galegroup.com/ps/?authCount=1>
-  Chu, H. *Information representation and retrieval in the digital age*. Medford : Information Today, 2007. 248 s.
dostupné v ÚK FF MU (volný výběr)
-  Papík, R. *Vyhledávání informací II.: Uživatelské rozhraní a vlivy oboru "Human-computer interaction,,*. Národní knihovna, 2001, roč. 12, č. 2, s. 81-90. Dostupné též z WWW: <http://full.nkp.cz/nkkr/NKKR0102/0102081.html>

Doplňující literatura

- 👉 Baeza-Yates, R.; Ribeiro-Neto, B. *Modern information retrieval*. New York : ACM Press, 1999. 513 s. ISBN 0-201-39829-X. dostupné v ÚK FF MU
- 👉 Matoušová, M. *Rešerše : úvod do teorie a organizace rešeršních služeb*. Praha : ÚVTEI : Ústředí vědeckých, technických a ekonomických informací, 1988. 227 s.
dostupné v ÚK FF MU (registrační pult)
- 👉 Steinerová, J. *Teória informačného prieskumu*. Vyd. 1. Bratislava : Slovenská technická knižnica - Centrum VTI SR, 1996. 262 s. ISBN 80-85165-58-9.
dostupné v ÚK FF MU