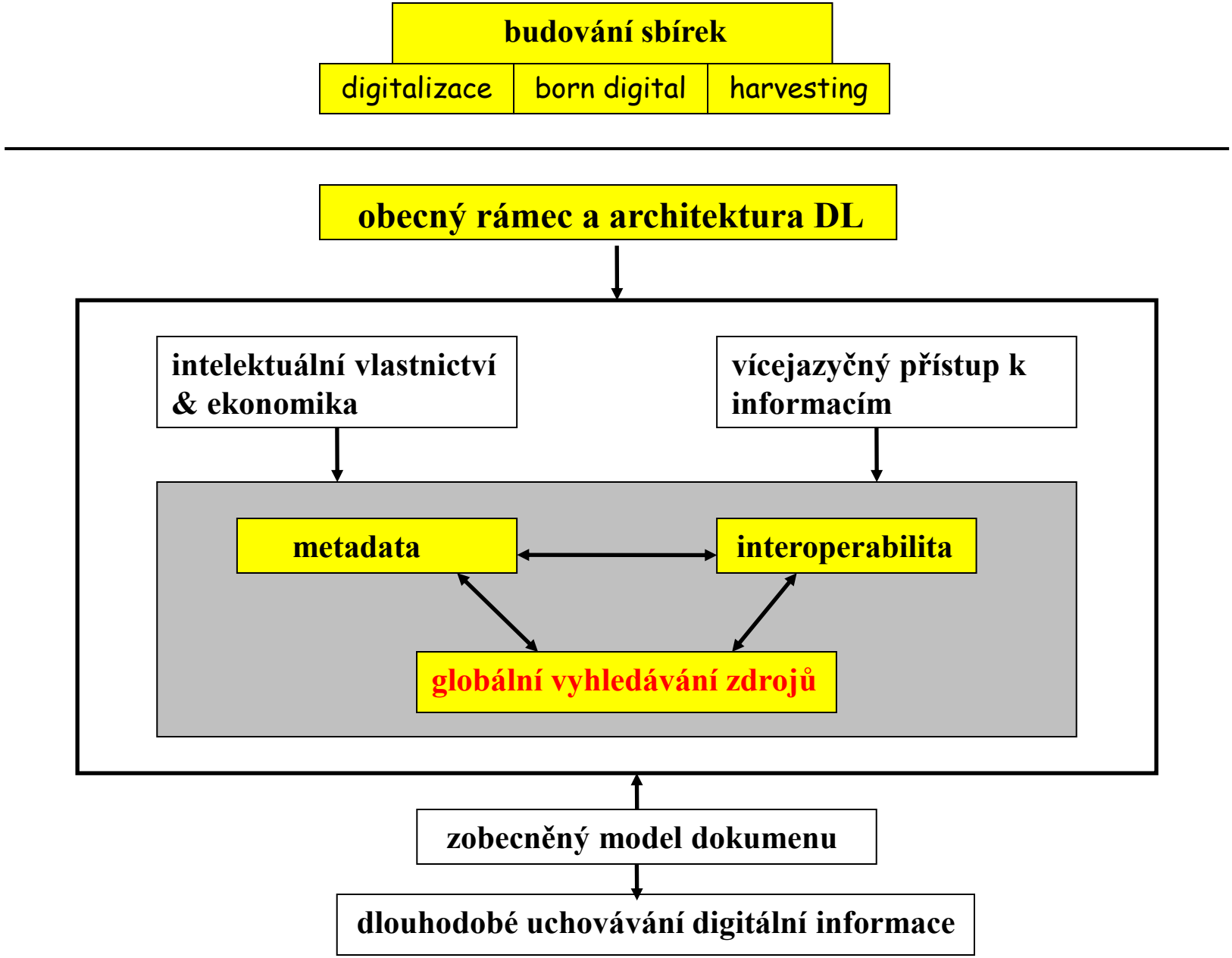
Digitální knihovny
Globální vyhledávání zdrojů

Miroslav Bartošek

Ústav výpočetní techniky MU

Knihovnicko-informační centrum MU


budování sbírek
digitalizace
born digital
harvesting
obecný rámec a architektura DL
intelektuální vlastnictví & ekonomika
vícejazyčný přístup k informacím
metadata
interoperabilita
globální vyhledávání zdrojů
zobecněný model dokumenu
dlouhodobé uchovávání digitální informace

Obsah přednášky

1. Úvod a přehled
2. Vyhledávání na webu
3. Federativní a metavyhledávání
4. DL a vyhledávače na webu
5. Unicode
6. Sémantický web

1. Úvod a přehled

1. Vyhledávání v globální DL

DL : globální systém, vysoce

- distribuovaný
- decentralizovaný
- dynamický

- Jak v DL efektivně vyhledávat?
- Vyhledávání v DL x vyhledávání na Internetu

1.1 Vyhledávání – oblasti výzkumu

- **organizace**
  při distribuovaném vyhledávání má každé řešení svůj organizační aspekt; vždy musí existovat určitá forma koordinace – má-li být vyhledávání efektivní
- **systémy**
  systémová infrastruktura podporující vyhledávání (routing dotazů, mezirepozitářové protokoly, bezpečnost, soukromí, autentifikace, placení)
- **digitální obsah**
  logický výběr inf.bází, dotazování netextových zdrojů, ratings, filtrace,
  přechod od vyhledávání explicitní informace k získávání implicitních poznatků
  (knowledge discovery, sémantický web)
- **rozhraní**
  HCI: konstrukce dotazů, prezentace/vizualizace výsledků, task understanding, proces exposure
- **metriky**
  taxonomie pro vyhodnocování různých řešení, testbeds

1.2 Pokroky ve vyhledávání

Nejlepší výsledky zatím přináší hrubá síla (brutte force):

- vyhledávání informací - webovské vyhledávače
- porozumění sémantice dokumentů - Deliver
- vyhodnocování výsledků - Google
- archivace digitálního dědictví - Internet Archive
- citační analýza - CiteSeer
- reference linking - OpenURL
- extrakce metadat z multimediálních zdrojů - Informedia
- automatický referenční knihovník – Univ Washington

1.3 Míry pro vyhledávání

Jak porovnávat/měřit kvalitu různých vyhledávacích systémů?

Relevance dokumentu = míra uspokojení informační potřeby

Míry efektivnosti vyhledávání

- **přesnost** (**precision**) **NalRel / Nal**
  jaká část nalezených dokumentů je relevantní
- **úplnost, výtežnost** (**recall**) **NalRel / RelAll**
  jak velká část všech exist. relevantních dok. byla nalezena

Další parametry vyhledávacího systému

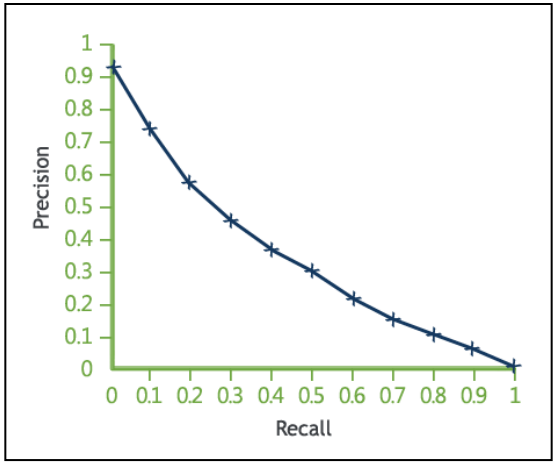
- **pokrytí** (**coverage**)
  jak velká část informačního prostoru je zachycena v DB vyhledávacího systému
- **odpad** (**false drop**)
  kolik bude vybráno nerelevantních dokumentů

1.3 Míry pro vyhledávání

	vyhledáno	nevyhledáno	CELKEM
relevantní	**A**	**B**	A + B
irelevantní	**C**	**D**	C + D
CELKEM	A + C	B + D	A + B + C + D

- **přesnost (precision)** **P = A / (A+C)**
- **úplnost (recall)** **R = A / (A+B)**

- **vztah precision-recall** = nepřímá úměra
  - týká-li se vše, co jste našli, přesně daného tématu, pravděpodobně jste přišli o nějaké informace
  - čím víc se blížíte úplnému zachycení tématu, tím více irelevantního materiálu vyhledáte



2. Vyhledávání na webu

2.1 Z historie vyhledávání na webu

- **1989/90 – návrh služby WWW – Tim Berners-Lee**
- 1991 první webová stránka
- 1993 Mosaic – první grafický prohlížeč
- 1994 WebCrawler, Lycos, Yahoo!
- 1995 Magellan, Excite, Infoseek, Inktomi, AltaVista, HotBot
- 1998 Google (Stanford)
- 2000 Google dominantní vyhledávač , Baidu (ČLR)
- 2005 MSN Search (Microsoft) – 2009 Bing
- Aktuálně nejpopulárnější (http://www.ebizmba.com/articles/search-engines)
  Google, Bing, Yahoo! Search, Ask (Ask Jeeves), Aol Search, MyWebSearch, WebCrawler, WoW, Infospace, Dogpile, DuckDuckGo, Info, Lycos, Excite
- Speciální: Yippy (deep-web), Mahalo, KartOO, …

http://searchenginewatch.com/

2.1 Z historie vyhledávání na webu

- 1999 800 mil veřej. www-stránek (15TB)
- 2000 2 miliardy www-stránek
- 2013 50 miliard www-stránek

✓ denně změněno 23% stránek
(studie Stanford Univ, 2000)

✓ poločas rozpadu 10 dnů
(1/2 URL neplatná)

Přístup k info: množství databází, archivů, vyhledávačů

Jak najít v „moři informací“ právě tu potřebnou ?

Vyhledávací stroje – historicky soupeření o pokrytí (velikost indexu 2001):

Google	*Fast*	*WebTop*	*Inktomi*	*AltaVista*
602 mil	500	500	500	400 mil www stránek

Dnes – inteligentní vyhledávání, nové formáty

Ale!

Pouze povrchový web (bez dynamických, zaheslovaných, netextových, … dokumentů)

nikoliv *hluboký* web (až 500 x větší, 7.500TB oproti 19TB na povrchu)

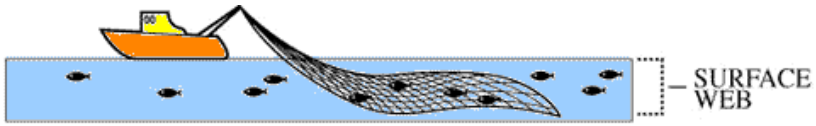
2.2 Hluboký web

Hluboký web (deep web) – skryté informační bohatství (studie 2001)

- 500 x větší (7.5 PB deep, 19 TB surface)
- 550 miliard www-stránek (oproti 1 mld na povrchu)
- 200.000 www sídel (oproti 5 miliónům na povrchu)
- 60 největších sídel = 750 TB (40x větší než celý povrch !!)
- podstatně vyšší kvalita informací
- víc jak polovina ve specializovaných předmětových DB
- až 95% info veřejně přístupná, bez poplatků
- M.K.Bergman. **The deep web: Surfacing hidden value**
  http://www.press.umich.edu/jep/07-01/bergman.html
- strmý nárůst oproti 2001 (3/555 mil domén v 2001/2013)

- volně nedostupné informace
- dynamické stránky
- speciální formáty dat

- BrightPlanet DeepWeb University Blog http://www.brightplanet.com/deep-web-university/



2.3 Big Data

- **Big Data** – objem dat, které jsou na/za hranici možnosti zpracování soudobými databázovými systémy a tradičními datovými aplikacemi
- Aktuální hranice – **Exabyte dat 10^{18}** (pohyblivý cíl)
  - dnešní denní produkce dat na světě
- **Problémy a nové výzvy**
  - uchovávání
  - zálohování
  - **vyhledávání**
  - přenosy, zpracování, vizualizace, **analýzy** …

- Výzkum – obrovské možnosti/potenciál ve využití BigData
- „future lies not in how much data we generate, but how we will work with them to obtain relevant information for solving problems in real time“

Tony Hey, Microsoft Research Connection

2.3 Big Data

• Kilobyte	kB	2^{10}	10^{3}		
• Megabyte	MB	2^{20}	10^{6}	milion	počet živočišných druhů
• Gigabyte	GB	2^{30}	10^{9}	miliarda	počet obyvatel Indie
• Terabyte	TB	2^{40}	10^{12}	bilion	počet všech ryb v oceánech
• Petabyte	PB	2^{50}	10^{15}	biliarda	počet mravenců na Zemi
• Exabyte	EB	2^{60}	10^{18}	trilion	inflace v Zimbabwe 2009
• Zettabyte	ZB	2^{70}	10^{21}	triliarda	počet zrnek písku na Zemi
• Yottabyte	YB	2^{80}	10^{24}	kvadrilion	počet hvězd ve Vesmíru

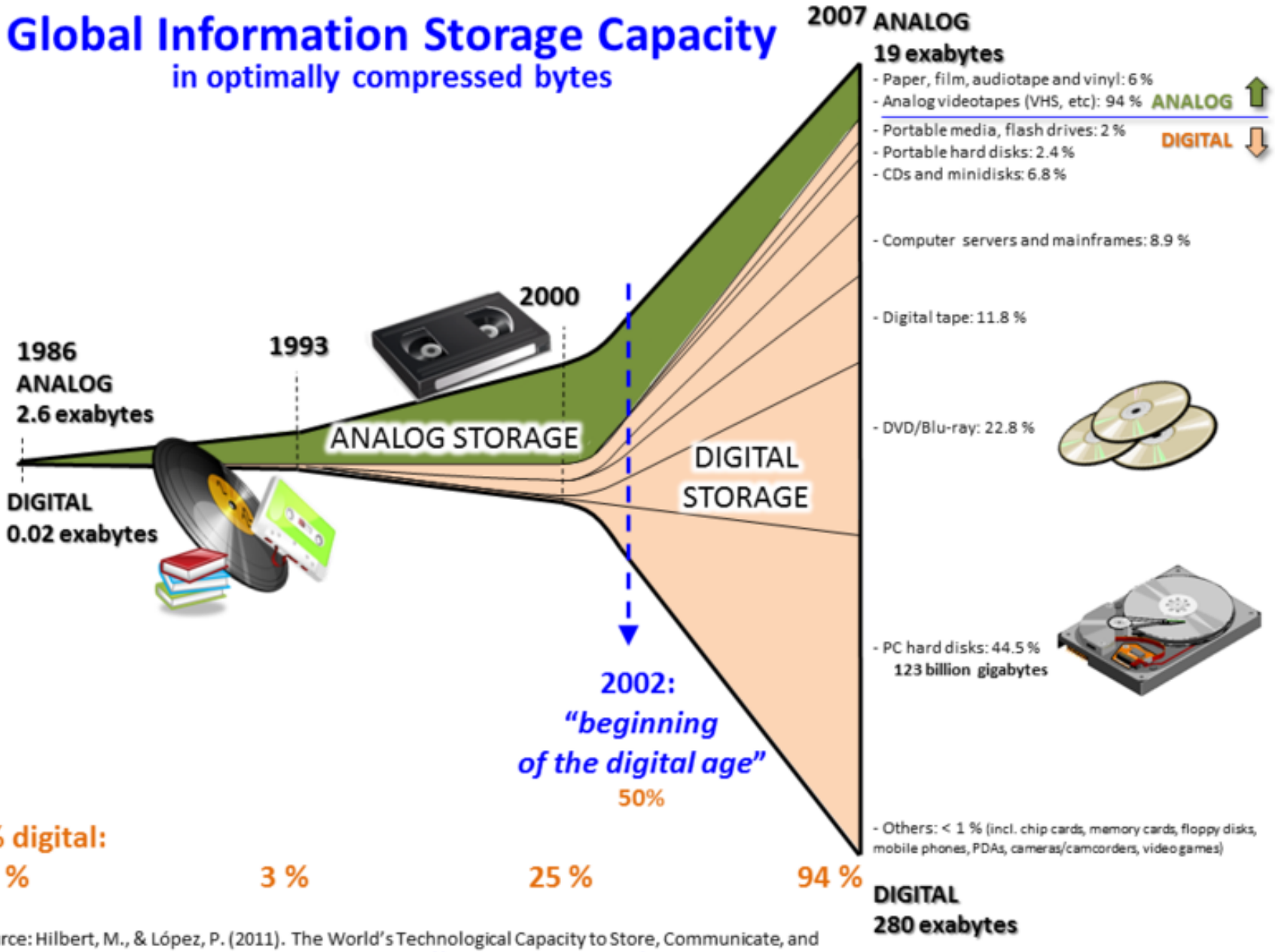
- Počet atomů na Zemi 10^{50} (10^{78} – 10^{82} ve Vesmíru)

2.3 Big Data

- **2,7 ZB** – globální objem všech dat na konci roku 2012 (o 48 % více než v r. 2011, odhad IDC)
- **33 ZB/2018** , 175/2025 – odhad EC
- **1 ZB** – objem datových přenosů v Internetu 2016 (odhad Cisco)
- **172 800 000** – denně zpracovaných platebních VISA transakcí
- **500 000 000** – denně odeslaných tweetů
- **1,15 miliardy** aktivních uživatelů Facebooku, denně generujících sociální data
- **5 miliard lidí** – generujících data denně přes mobily a internet

Global Information Storage Capacity

in optimally compressed bytes

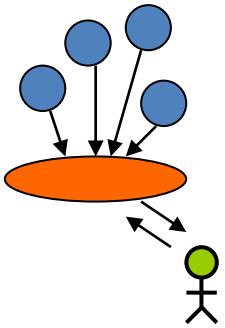


Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65. http://www.martinhilbert.net/WorldInfoCapacity.html


alphaspirit/Shutterstock

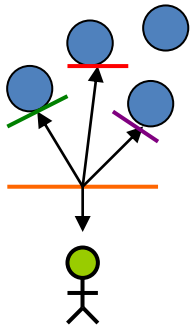
3. Federativní vyhledávání a metavyhledávání

3. Dva přístupy k vyhledávání



a) federativní vyhledávání (Google, OAI, discovery)

- předběžný sběr velkého množství dat do 1 hromady
- předzpracování nashromážděných dat ještě před dotazem uživatele
- po zadání dotazu se prohledává jen nasbíraná hromada
  - just-in-case processing (předzpracování dat ještě před dotazem)



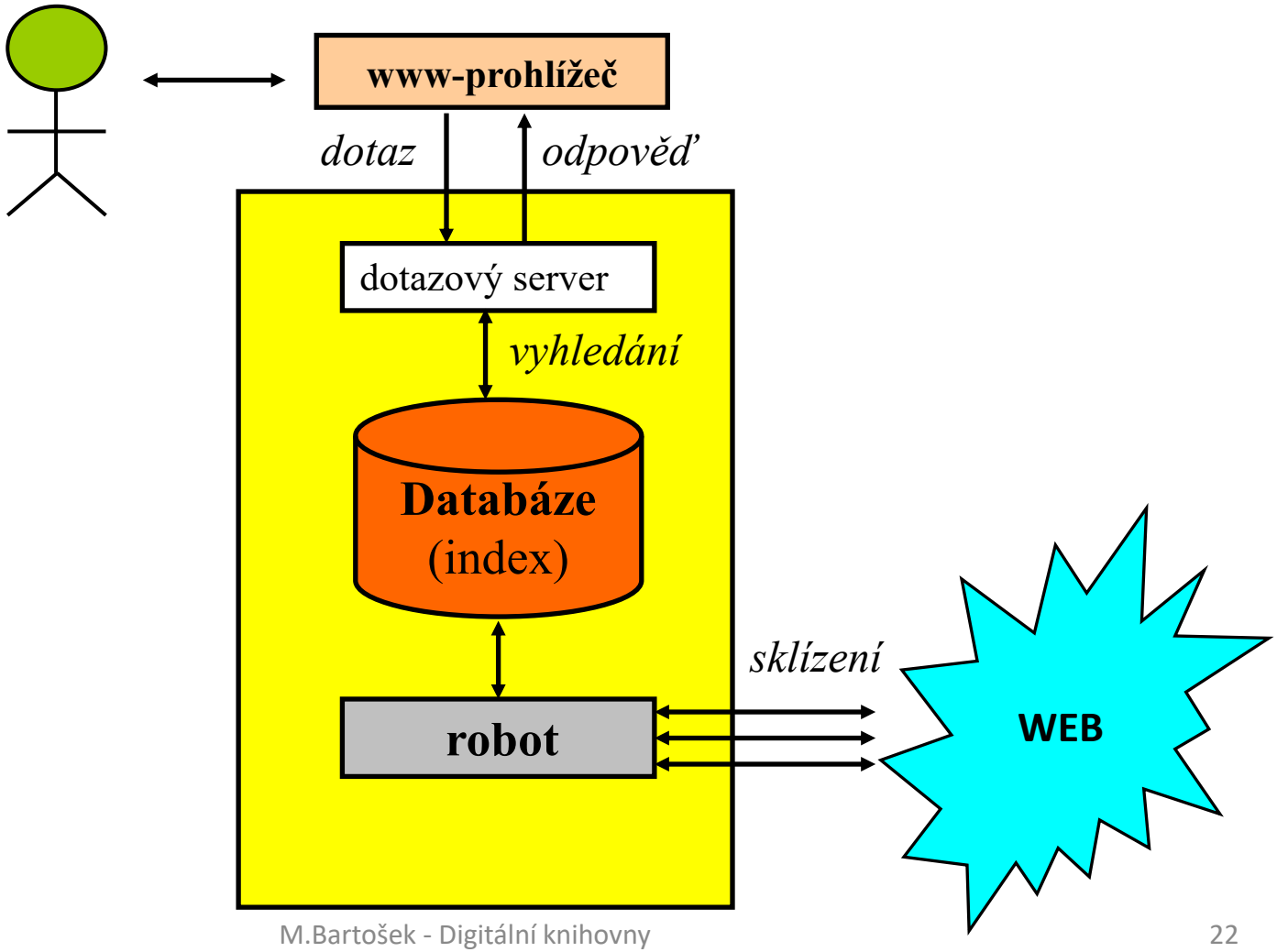
b) meta-vyhledávání (Z39.50, SRW/U, Metalib)

- integrated/parallel/simultaneous/cross-db searching
- dotaz rozeslán souběžně do všech (heterogenenních) zdrojů (každý zdroj provede vlastní vyhledávání)
- Integrace dílčích výsledků do výsledné odpovědi
  - just-in-time processing (veškeré zpracování probíhá až po dotazu)

V obou případech jediné vyhledávací rozhraní

Federativní vyhledávání

3.1 Federativní vyhledávání



3.1 Federativní vyhledávání

- aktuálně populárnější – googlomanie
  - rychlá, okamžitá odezva
  - obrovský rozsah prohledávaných zdrojů (miliardy)
  - jednoduchý přístupný vyhledávací mechanismus
  - relevance ranking
  - záplava nových služeb (maps, scholar, books, news, voice …)
- **Ale**
  - **aktuálnost dat?**
  - **nelze prohledávat dynamické webové stránky (jen statické zdroje)**
  - **prohledávání jen veřejně přístupných zdrojů (licencované DB?)**
  - **kvalita a autenticita zdrojů informací (při plošném sběru)?**
  - **jaké je pokrytí? A kontrola uživatele nad pokrytím?**

3.1.1 Discovery služby

Discovery services – nový šlágr v oblasti EIZ pro VaV

- **Primo Central** (ExLibris)
- **Ebsco Discovery Service** (EBSCO)
- **Summon** (Serial Solutions)
- **AquaBrowser Library**
- **VuFind** (open source) aj.

MU od 2013/10

- **velký centrální index** – předzpracovaný, nasbíraný z různých zdrojů
- **jednotné vyhledávací prostředí** (EIZ, knihovna, DL)
  - vyhledávaní informací
  - dodávání informací (napojení na linkovací služby)
  - objevování nového

3.1.2 discovery.muni

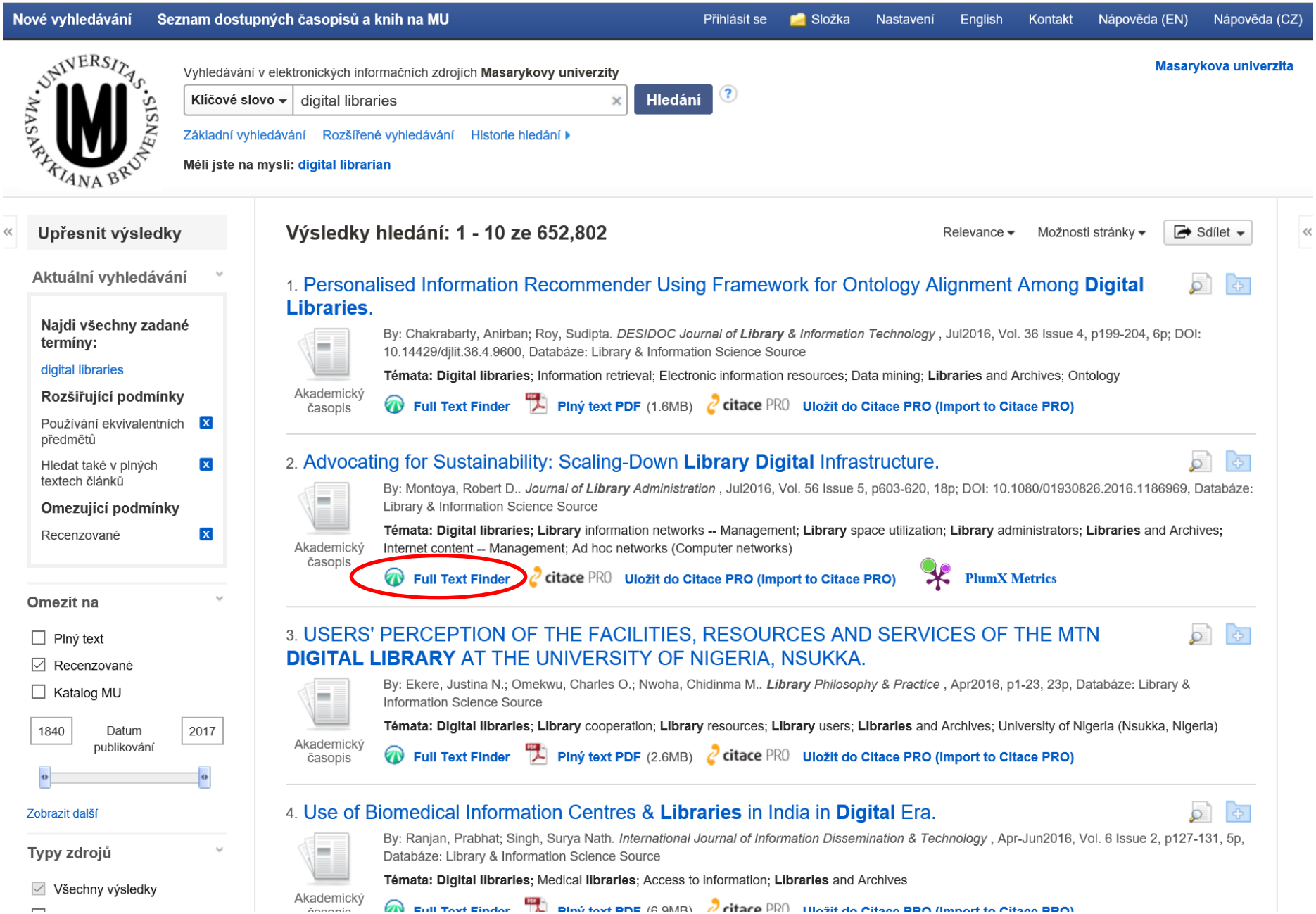
http://discovery.muni.cz

- **Centrální index** (obrovský, přes miliardu záznamů)
  - data od všech světových vydavatelů odborné literatury
  - licencované EIZ a vědecké databáze dostupné na MU
  - lokální informační zdroje MU (knihovní katalog, archiv VŠ diplomek, …)
- **Vyhledávání**
  - prohledávání všech odborných inf.zdrojů z jednoho místa
  - jednoduché vyhledávací rozhraní ala Google
  - zpřesňování výsledků pomocí filtrování
- **Linkovací služba**
  - **FullText Finder** – směrování na plný text vyhledaných výsledků v databázích dostupných pro uživatele MU
- **A-to-Z**
  - vyhledávání e-časopisu / e-knihy dostupné na MU

EDS – Ebsco Discovery Service

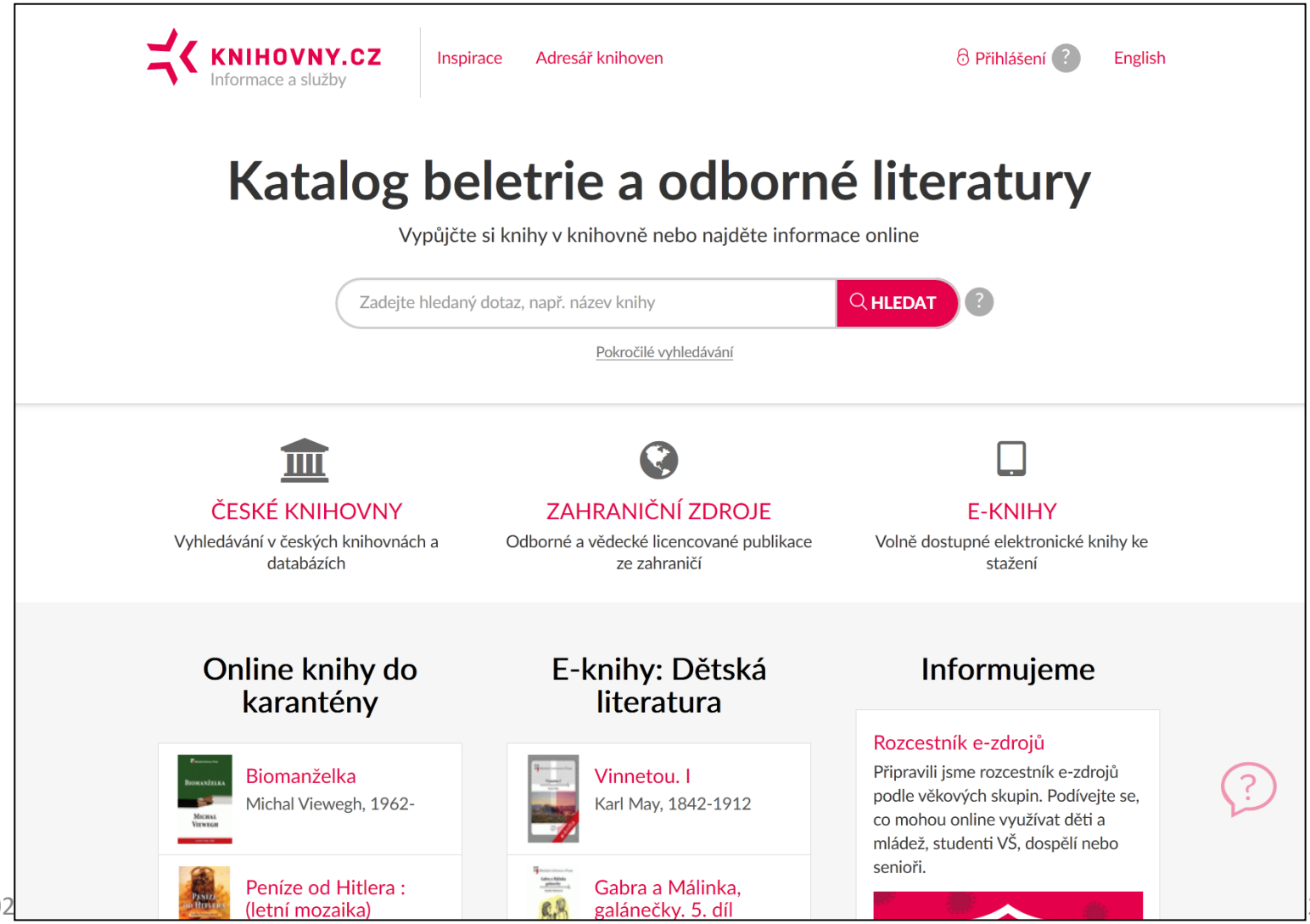
3.1.2 discovery.muni

http://discovery.muni.cz



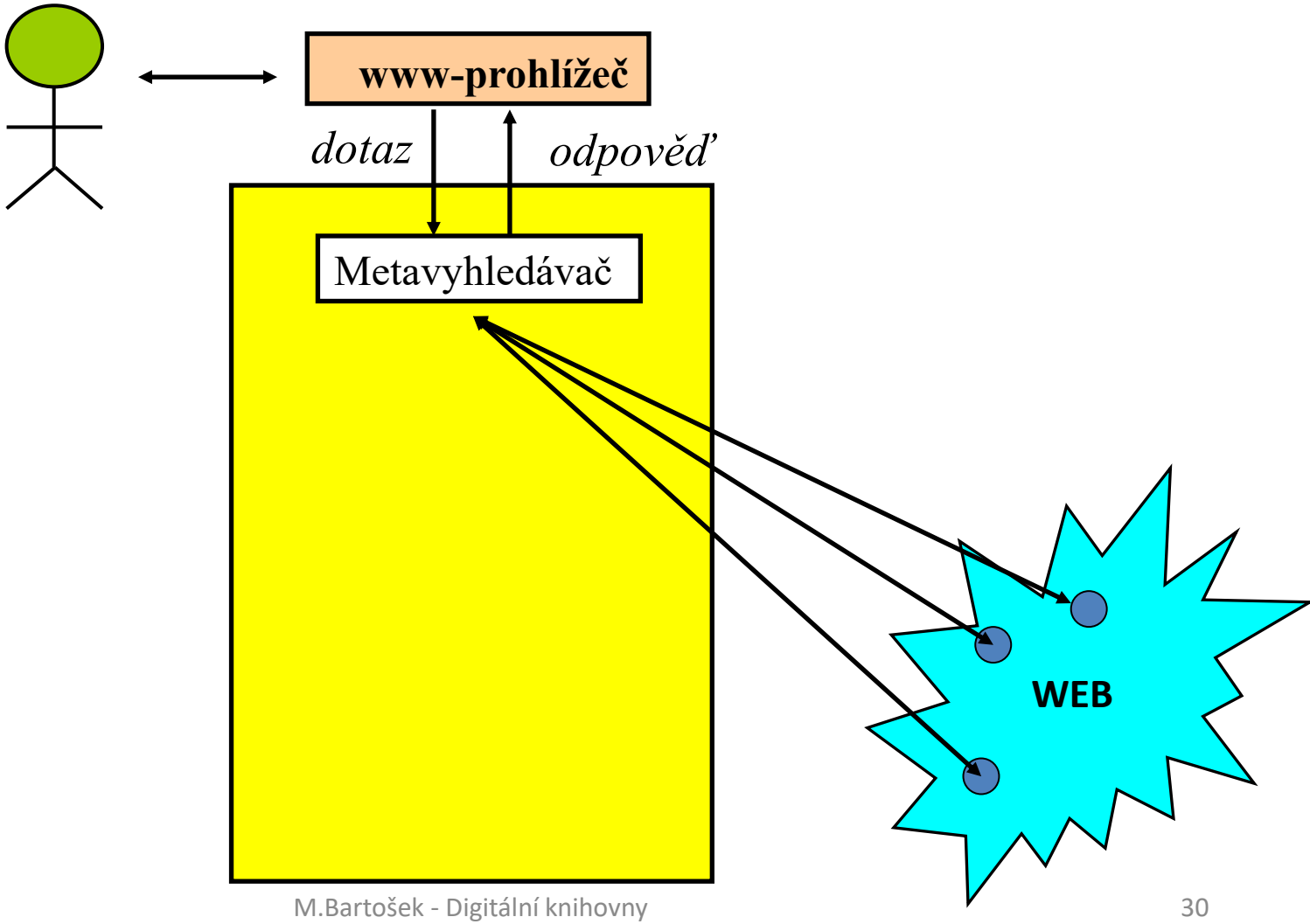
3.1.3 Portál knihoven ČR

www.knihovny.cz



Metavyhledávání

3.2 Metavyhledávání



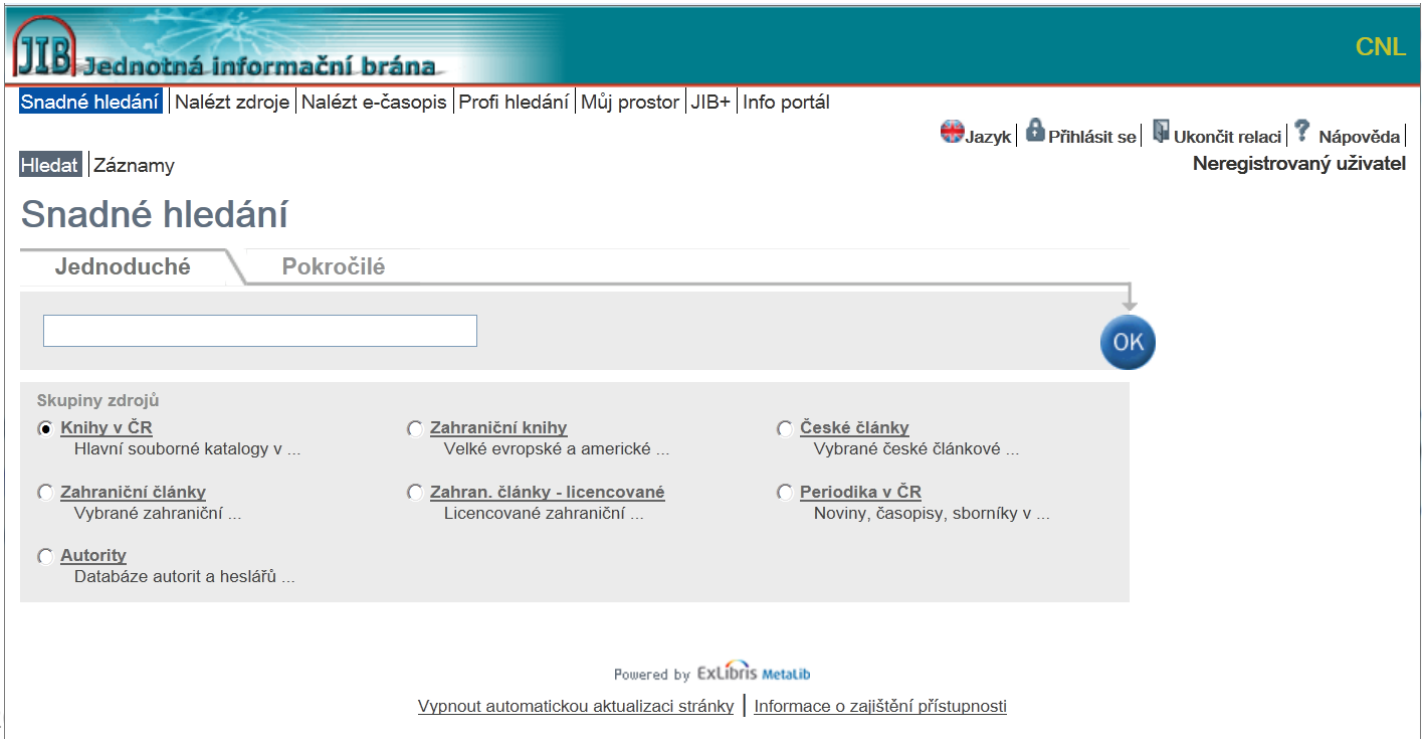
3.2 Metavyhledávání

- propojení heterogenních zdrojů
  (lokace, formát, technologie, typ materiálů)
- každý zdroj vlastní SE (search-engine)

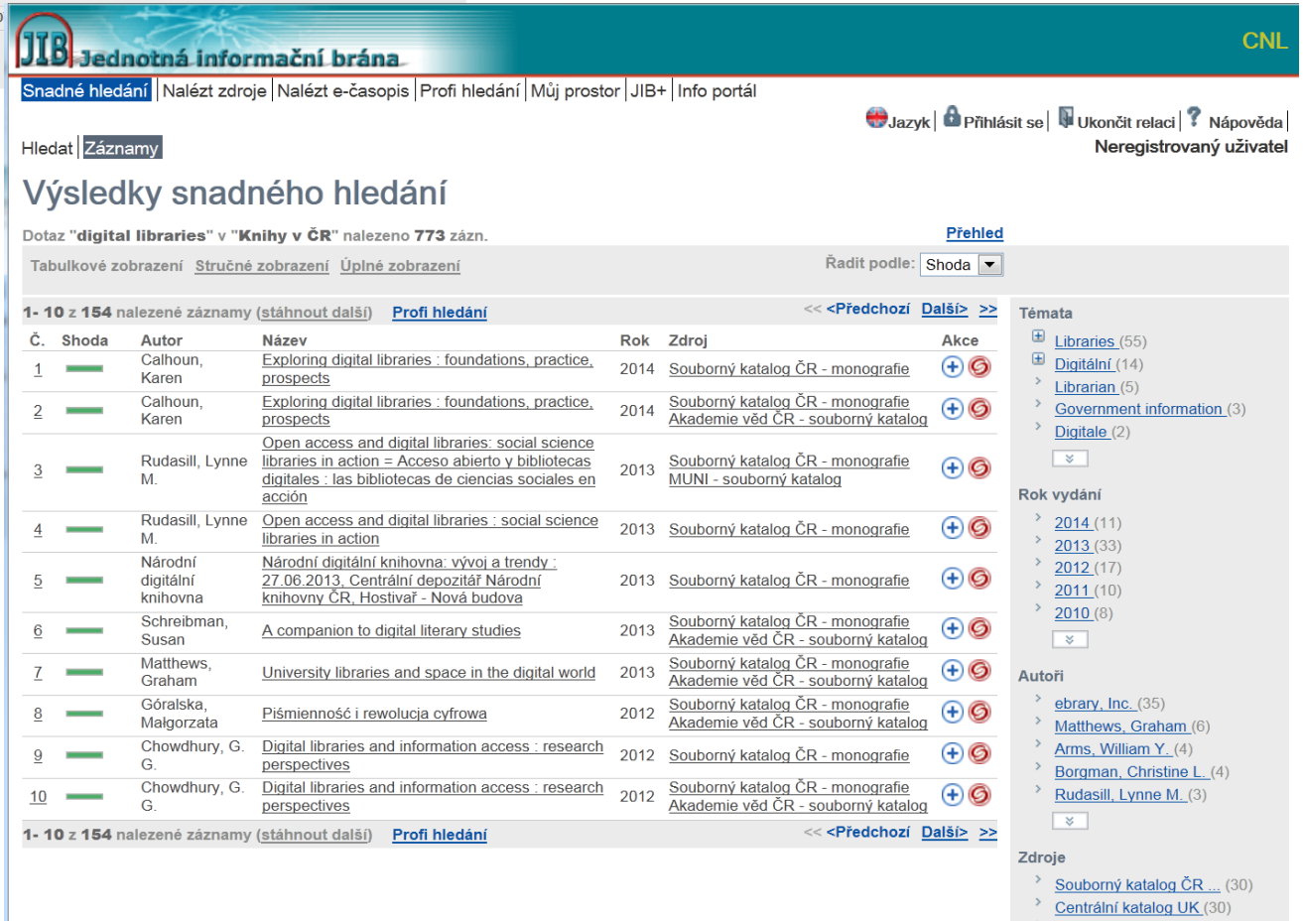
- musí se řešit
  - potřebné informace o jednotlivých zdrojích
  - přenos uživatelského dotazu k různým SE (adaptace dotazu)
  - konverze výsledků do jednotného formátu
  - zpracování unifikovaných výsledků
    - slučování
    - deduplikace
    - konzistentní prezentace

3.2.1 Metavyhledávání - JIB

- Příklad: **Jednotná informační brána NK ČR**
  - http://www.jib.cz/ (dnes již mimo provoz)
  - Od 31.12.2018 provoz ukončen, funkci JIB nahrazuje „Portál knihoven ČR“
- Technologie: MetaLib od ExLibris




Jednotná informační brána
Snadné hledání | Nalézt zdroje | Nalézt e-časopis | Profi hledání | Můj prostor | JIB+ | Info portál
Jazyk | Přihlásit se
Hledat | Záznamy
Snadné hledání
Dotaz "digital libraries" v "Knihy v ČR"
Probíhá vyhledáván
Zobrazit stažené záznamy Zrušit
Název zdroje Status Nalezeno Staženo
Souborný katalog ČR - monografie PROVEDENO 185 30
Centrální katalog UK PROVEDENO 275 30
MUNI - souborný katalog PROVEDENO 150 30
VUT Brno - souborný katalog PROVEDENO 5 5
VŠE - souborný katalog STAHUJÍ SE ZÁZNAMY 129
ČVUT - souborný katalog PROVEDENO 5 5
Akademie věd ČR - souborný katalog PROVEDENO 21 21
SKAT - souborný katalog PRO


Jednotná informační brána
CNL
Snadné hledání | Nalézt zdroje | Nalézt e-časopis | Profi hledání | Můj prostor | JIB+ | Info portál
Jazyk | Přihlásit se | Ukončit relaci | Nápověda
Neregistrovaný uživatel
Hledat | Záznamy
Výsledky snadného hledání
Dotaz "digital libraries" v "Knihy v ČR" nalezeno 773 zázn.
Přehled
Tabulkové zobrazení Stručné zobrazení Úplné zobrazení
Řadit podle: Shoda
1- 10 z 154 nalezené záznamy (stáhnout další) Profi hledání
<< <Předchozí Další> >>
Č. Shoda Autor Název Rok Zdroj Akce
1 Calhoun, Karen Exploring digital libraries : foundations, practice, prospects 2014 Souborný katalog ČR - monografie
2 Calhoun, Karen Exploring digital libraries : foundations, practice, prospects 2014 Souborný katalog ČR - monografie Akademie věd ČR - souborný katalog
3 Rudasill, Lynne M. Open access and digital libraries: social science libraries in action = Acceso abierto y bibliotecas digitales : las bibliotecas de ciencias sociales en acción 2013 Souborný katalog ČR - monografie MUNI - souborný katalog
4 Rudasill, Lynne M. Open access and digital libraries : social science libraries in action 2013 Souborný katalog ČR - monografie
5 Národní digitální knihovna Národní digitální knihovna: vývoj a trendy : 27.06.2013, Centrální depozitář Národní knihovny ČR, Hostivař - Nová budova 2013 Souborný katalog ČR - monografie
6 Schreibman, Susan A companion to digital literary studies 2013 Souborný katalog ČR - monografie Akademie věd ČR - souborný katalog
7 Matthews, Graham University libraries and space in the digital world 2013 Souborný katalog ČR - monografie Akademie věd ČR - souborný katalog
8 Góralska, Małgorzata Piśmienność i rewolucja cyfrowa 2012 Souborný katalog ČR - monografie Akademie věd ČR - souborný katalog
9 Chowdhury, G. G. Digital libraries and information access : research perspectives 2012 Souborný katalog ČR - monografie
10 Chowdhury, G. G. Digital libraries and information access : research perspectives 2012 Souborný katalog ČR - monografie Akademie věd ČR - souborný katalog
1- 10 z 154 nalezené záznamy (stáhnout další) Profi hledání
<< <Předchozí Další> >>
Témata
Libraries (55)
Digitální (14)
Librarian (5)
Government information (3)
Digitale (2)
Rok vydání
2014 (11)
2013 (33)
2012 (17)
2011 (10)
2010 (8)
Autoři
ebrary, Inc. (35)
Matthews, Graham (6)
Arms, William Y. (4)
Borgman, Christine L. (4)
Rudasill, Lynne M. (3)
Zdroje
Souborný katalog ČR ... (30)
Centrální katalog UK (30)

3.2.2 Metavyhledávání

- bližší knihovníkům a DL (přesnější, cílenější, pod kontrolou)
- bližší producentům dat (lepší ochrana IPR)
- vyhledávání i v „profesionálních“ zdrojích (DB)
- potřeba řady standardů (NISO Metasearch Initiative)
  - Access Management (autentifikace, autorizace)
  - Collection Description, Service Description (explain)
  - Search/Retrieve
- vazba na výzkum sémantického webu

infrastruktura

- oproti federativnímu vyhledávání:
  - složitější, křehčí
  - větší potenciální možnosti (NISO: „stojí za to to zkusit“)

4. DL a vyhledávače na webu

4. DL a webové vyhledávače

„Prakticky všechno co je nejlepší v digitálních knihovnách, je mizerné u webovských vyhledávačů – a naopak“

- webové-vyhledávače
  - *rychlá první informace*
  - \+ prakticky realizované, široce dostupné,
  - \+ užitečné, propojení na zdroje z otevřeného přístupu
  - \- vysoké pokrytí a úplnost, malá přesnost
  - \- jen povrchový web (500x větší hluboký-web nedostupný)
- DL
  - *kvalitní cílená informace*
  - \+ perspektivní, teoreticky dobře podložené
  - \+ kvalitnější vyhledávání, širší rozsah služeb
  - \- zatím ještě ne plně zvládnuté, globálně nerozvinuté

4.1 Slabá místa vyhledávačů

Webové vyhledávače – skvělý pomocník, ale mají své nedostatky

- Příliš mnoho výsledků
  - Nelze všechny systematicky projít
  - Vysoká míra redundance (chybí clusterování výsledků)
- Mechanické vyhledávání podle klíčových slov
  - Chybí porozumění dotazu
- Netransparentní řazení výsledků
  - Veřejné/skryté triky pro lepší viditelnost webové stránky
- Žádná garance „důvěryhodnosti“ výsledků
  - Akceptujeme, že výsledkům nemůžeme plně věřit (rozpornost, (ne)ověřitelnost, (ne)aktuálnost)

4.2 Wikipédie – nový typ DL?

- Specifická digitální knihovna (od 2001)
- Velký úspěch přístupu „Wisdom of the Crowd“
- Posuny ve vnímání kvality a přínosu
  - Potenciálně slabá místa
    - Chyby z neznalosti/neprofesionality tvůrců
    - Subjektivní či nevyvážený popis, občasné excesy, vandalismus
    - Různorodost jazykových verzí
    - (ne)Použitelnost pro studentské či odborné/vědecké práce?
  - Silné stránky
    - Rozsah a aktuálnost v porovnání s tradičními encyklopediemi
    - Větší různorodost pohledů (ne jediný vlastník „pravdy“)
    - Vyvíjející se samokorekční mechanismy
    - Široká všeobecná dostupnost

5. Unicode

5. Kódování znaků

- tisíce různých jazyků
- stovky abeced (latinka, azbuka, hebrejština, arabština, indické znakové systémy; [fonetický] znak reprezentuje zvuk)
- ideografické systémy (čínština, korejština, japonština; znak = pojem)
- různá kódování znaků v rámci jednoho jazyka
- čeština - kódování
  - 7-bitové ASCII
  - CP1250 (MS Windows Latin 2)
  - ISO 8895-2 (Unix, Latin 2)
  - CP 852 (PC Latin 2 – MS DOS)
  - kódování Kamenických (MJK)
  - KOI8-cs (T602)
  - APPLE CE
  - ….

8-bitů

漢 汉
字 字

5.1 Unicode

Globální DL –> potřeba **jednotné reprezentace všech znaků**

- 1987: Apple+Xerox – práce na nástupci ASCII (**Unicode**)
- 1991: mezinárodní Unicode Consortium
- 1993: ISO-10646
- nyní: všechny skripty všech používaných jazyků na světě (113.000 znaků)
- další: historické jazyky (egyptské hieroglyfy), hudební notace, …

- přímá podpora Unicode v moderních program. jazycích (Java,…), operačních systémech, browserech, …
- Unicode knihovny pro starší systémy (C, Perl, …)
- ***round-trip compatibility*** :
  - každá existující znaková sada může být mapována do Unicode
  - výsledný Unicode-soubor lze převést do původní znakové sady *bez ztráty jakékoliv informace*

5.1 Unicodový prostor

ISO Unicode

- Unicode = **masívní** standard (94.000 znaků) (32 resp. 21 bitů)
- celkem 32 *úrovní* (planes), každá 65.536 znaků (16 bitů)
  - **Basic Multilingual Plane** (živé jazyky)
  - **Supplementary Multilingual Plane** (historické skripty, matem.symb)
  - **Supplementary Ideographic Plane** (40.000 starověkých čínských)
  - … Unicode v 5.2-2009 – 245.000 znaků /symbolů z 90 jazyků/abeced
- 1.část: **Basic Multilingual Plane** (živé jazyky, 49K zn, 1000 stran)
- znaky nerozlišovány podle jazyků, ale dle skriptů (typu písma)
- Kódový prostor = 5 *zón skriptů* :

– alfabetické	0000-33FF	latinka,azbuka,hebrej,arab,ind,
– ideografické	3400-BFFF	CJK (Chinese-Japan-Korean)
– ostatní	A000-D7FF	Yi, Hangul (11.172 kódů)
– zástupné	D800-DFFF	
– rezervované	E000-FFFF	

5.2 Basic Multilingual Plane

zóna	oblast	kód	skript	#kódů
alfabet	obecné	0000	Basic Latin (US ASCII)	128
		0080	Latin-1 (ISO 8859-1)	128
		0100	Latin Extended	336
		0300	Combining Diacritical Marks	112
		0370	Greek	144
		0400	Cyrillic	256
		0530	Armenian	96
		0590	Hebrew	112
		0600	Arabic	256
		…		
	symboly	2000	General Punctuation	112
		2070	Superscripts and Subscripts	48
		20A0	Currency Symbols	48
		…		
ideogr		3400	CJK Unified Ideographs, Ext A	6656
		…		

5.3 Kódování znaků Unicode

Unicode 21 bitů (U+000000-U+10FFFF), ISO 32 bitů

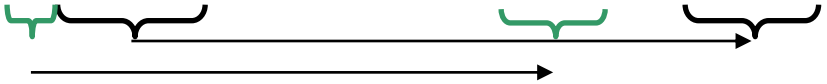
UTF – Unicode character set Transformation Format

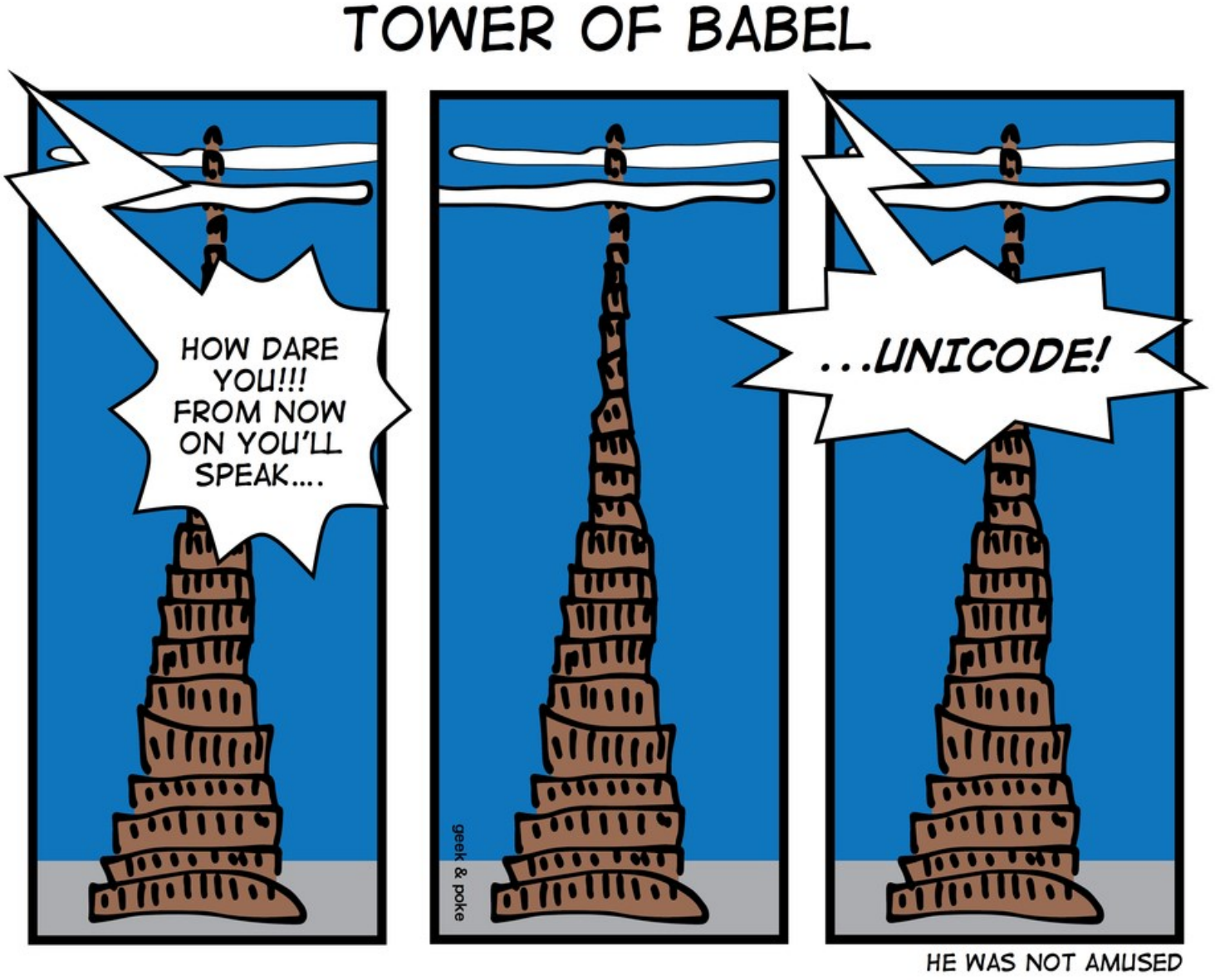
- **UTF-32** - 4 byty na 1 znak "G" = U+000047
- **UTF-16** - 2 byty na 1 znak (Basic M-Plane) "G" = U+0047
- **UTF-8** - 1-4 byty na 1 znak "G" = U+47

Unicode hodnota	21-bit binární kód	UTF-8
U+000000-U+00007F	00000000000000wwwwwwww	0wwwwwww
U+000080-U+0007FF	00000000000wwwwwxxxxxx	110wwwww 10xxxxxx
U+000800-U+00FFFF	00000wwwwxxxxxxyyyyyy	1110wwww 10xxxxxx 10yyyyyy
U+010000-U+1FFFFF	wwwxxxxxxyyyyyyzzzzzz	11110www 10xxxxxx 10yyyyyy 10zzzzzz

- 1 byte = 7-bitové ASCII (začíná vždy 0) **G = U+47 47**
- 2 byte = vše až po indické skripty (počet 1 = počet B) **ä = U+E4 C3A4**

ä: E4 =11100100 -> C3 A4 = 11000011 10100100





geek-and-poke.com

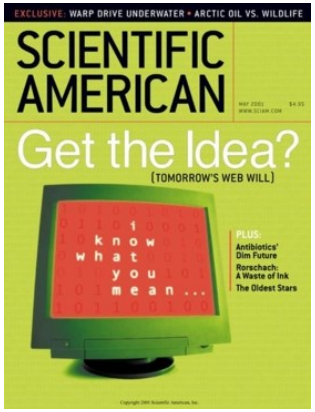
6. Sémantický web

6. Co je sémantický web

- Web dnes: repozitář dokumentů určených pro člověka
- Sem-Web: repozitář dat a info zpracovatelných počítačem

- **Tim Berners-Lee**
  - The semantic web is an **extension of the current Web** in which information is given well-defined meaning, enabling computers and people to work in better cooperation.
  - The Semantic Web is a vision: the idea of having data on the web defined and linked in a way **that it can be used by machines not just for display purposes**, but for automation, integration and reuse of data across various applications.

- [1] Tim Berners-Lee, James Hendler, Ora Lassila: *The Semantic Web*. Scientific American, May 2001
- **W3C – Semantic Web Working Group**
  http://www.w3.org/2001/sw/



6.1 Modelový scénář SemW - dle článku [1]

- **Lucie volá Petrovi**: „Jsem s matkou u obvodního lékaře. Doporučil ji vyšetření a léčbu u specialisty.“
- **Petr**: „Vyber a objednej na příští týden nějakého dobrého doktora, já vás k němu odvezu.“
- **Lucie**: Hned ještě u lékaře zaúkoluje přes mobil svého **agenta** (softwarového asistenta pro SW).
- **Agent**: - spojí se s obvoďákovým agentem a stáhne si od něj informace o matčině předepsané léčbě;
  - najde na webu několik seznamů odborných lékařů příslušné specializace a projde si je;
  - vybere specialisty, kteří mají smlouvu s matčinou zdravotní pojišťovnou, mají ordinaci ve vzdálenosti do 20km od matčina domu a jsou hodnoceni jako výborní až velmi dobří v přehledech od důvěryhodných hodnotitelských agentur.
- **Agent**: - porovná volné objednací termíny specialistů (poskytnuté jejich SW agenty) s nabitými diáři Petra a Lucie;
  - během pár minut pošle Petrovi a Lucii nejvýhodnější variantu.
- **Petr**: - nabídnutá varianta se mu nelíbí: do ordinace vybraného specialisty by musel matku vézt přes střed města a ještě k tomu se vracet v čase, kdy vrcholí dopravní špička;
  - zadá povel svému vlastnímu agentovi, aby provedl nový výběr – tentokrát se striktnějšími preferencemi ohledně doby a místa schůzky. -- Petrův agent se spojí s Lucčiným agentem.
- **Lucčin agent**: ověří si důvěryhodnost Petrova agenta v dané věci a předá mu veškeré dosud zjištěné údaje.
- **Petrův agent**: během chvilky představí novou variantu – přidá k ní ale dvě upozornění:
  1. Petr by si musel přeplánovat několik méně důležitých schůzek.
  2. Daný lékař nemá v databázi matčiny pojišťovny uvedenu potřebnou specializaci, ale z jiných důvěryhodných zdrojů agent prověřil, že lékař tuto specializaci opravdu má.
     Přeje si Petr k tomu bližší údaje?
- **Petr**: Zamumlá: „ušetři mně zbytečných detailů“ a variantu potvrdí. Téměř současně vydá potvrzení i Lucie – a tím je vše zařízeno. Agent matku objedná a poznačí schůzku v diářích jejich dětí.

6.2 Charakteristiky SemW

Hlavní: SemW již není určen jen pro lidi, ale i pro stroje (počítače)

- Sémantický web = Web s významem
- program (inteligentní agenti), který má zpracovávat data na webu, se může tento význam dozvědět a využít ho pro svou činnost
- se SemW pracují nejen lidé, ale také agenti (stroje), kteří sbírají různorodá data z různých zdrojů, automatizovaně je zpracovávají, odvozují z nich nové poznatky, vyměňují si informace mezi sebou navzájem, …
- v SW již nevyhledáváme stránky obsahující jen stejná slova, ale také podobné pojmy (sémantické vyhledávání)

6.3 Možnosti SemW

- **inteligentní pojmové vyhledávání**
  identifikace relevantních dokumentů a jejich řazení podle míry vhodnosti
- **zodpovídání jednoduchých otázek**
  Kdo je prezidentem České republiky?
- **zodpovídání složitých otázek**
  Jaká je současná situace v Egyptě?

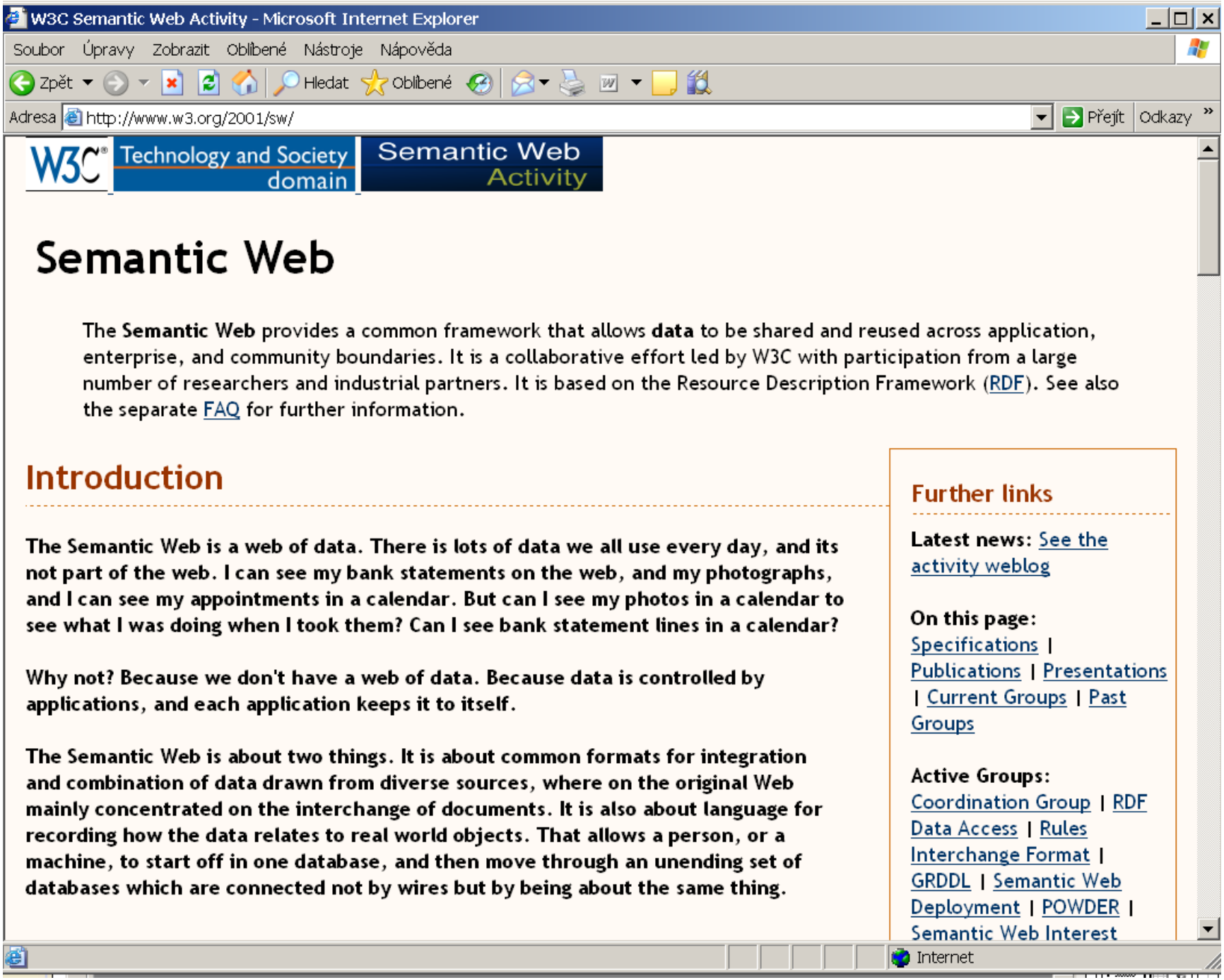
V. Sklenák. Sémantický web. INFORUM 2003.

6.4 Komponenty SemW

1. označkování webových stránek (struktura dat - **XML**)
2. vyznačit význam (sémantika - **RDF**)
3. vyhledávání pojmů napříč různými oblastmi (**ontologie**)
4. odvozování (logika, **odvozovací pravidla**)
5. vyhledávání znalostí a souvislostí (**agenti**)

akčnost SW

Metadata přidaná k datům na webu, která poskytují formální sémantiku obsahu webu


W3C Semantic Web Activity - Microsoft Internet Explorer
Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda
Zpět Hledat Oblíbené
Adresa http://www.w3.org/2001/sw/
Přejít Odkazy
W3C Technology and Society domain
Semantic Web Activity
Semantic Web
The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF). See also the separate FAQ for further information.
Introduction
The Semantic Web is a web of data. There is lots of data we all use every day, and its not part of the web. I can see my bank statements on the web, and my photographs, and I can see my appointments in a calendar. But can I see my photos in a calendar to see what I was doing when I took them? Can I see bank statement lines in a calendar?
Why not? Because we don't have a web of data. Because data is controlled by applications, and each application keeps it to itself.
The Semantic Web is about two things. It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing.
Further links
Latest news: See the activity weblog
On this page:
Specifications | Publications | Presentations | Current Groups | Past Groups
Active Groups:
Coordination Group | RDF Data Access | Rules Interchange Format | GRDDL | Semantic Web Deployment | POWDER | Semantic Web Interest
Internet

Literatura

Doplňková literatura

- Vyzkoušejte (a používejte) **discovery.muni.cz**
- **Can the Web turn into a digital library?**
  Herman Maurer, Heimo Mueller, Intl Journal on Digital Libraries 13/2, March 2013
  https://link.springer.com/article/10.1007/s00799-012-0097-9
- Časopis Ikaros, roč. 2011: Anna Matějková – trilogie o sémantickém webu:
  - **Sémantický web**
    http://www.ikaros.cz/semanticky-web
  - **Technologie sémantického webu**
    http://www.ikaros.cz/technologie-semantickeho-webu
  - **Současnost sémantického webu**
    http://www.ikaros.cz/soucasnost-semantickeho-webu
- Tim Berners-Lee, James Hendler, Ora Lassila: ***The Semantic Web***. Scientific American, May 2001
  https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf