

# Strukturování a označování informací

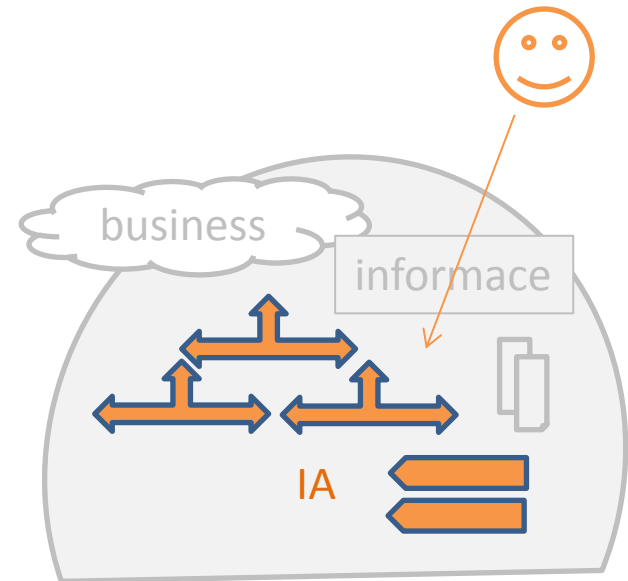
ZIM 2015

Přemek Brada

KIV ZČU Plzeň

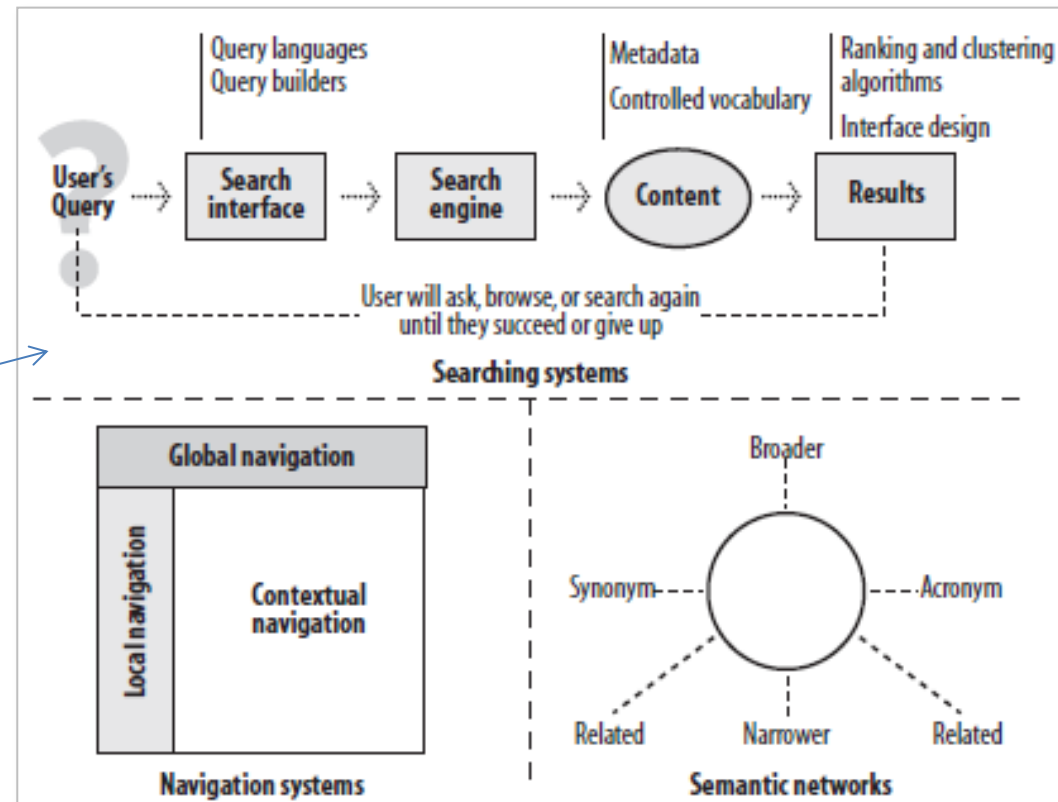
# Výchozí problém

- Rostoucí objem informací  
=> potřeba orientace
- Příklady
  - kavenu.eu ~1.2tis položek
  - alza.cz ~70tis položek
  - předměty KIV ~100
  - Courseware ZČU ~2000



# Organizování, označování informací: jádro IA

- Organizování
  - seskupování do celků, kategorií
- Označování
  - pojmenování celků



- K tomu IA používá

# Systemy organizace informací

- Seskupování do celků
  - Informací, prvků obsahu – stránek, textů, ...
- Odráží způsoby, jakými jde organizovat
- Složky
  - **Schémata** = společné charakteristiky prvků obsahu -> skupiny
  - **Struktury** = typy vztahů mezi prvky obsahu a skupinami

# Organizace: schémata

- Jednoznačná
  - [Abecedně](#)
  - [Chronologicky](#)
  - Geograficky
- Nejednoznačná
  - Tématicky
  - [Podle činností](#)
  - [Podle publika](#)
  - Podle metafory
  - (Hybridní)

News and Media > Regions
<a href="#">Directory</a> > <a href="#">News and Media</a> > <a href="#">By Region</a> > <a href="#">Regions</a>
CATEGORIES <a href="#">(What's This?)</a>
<ul style="list-style-type: none"><li>• <a href="#">Africa</a> (32)</li><li>• <a href="#">Antarctica</a> (10)</li><li>• <a href="#">Asia</a> (102)</li><li>• <a href="#">Caribbean</a> (11)</li><li>• <a href="#">Europe</a> (90)</li></ul>

Office
<ul style="list-style-type: none"><li>• Vše o Microsoft Office</li><li>• Zkušební verze Office</li><li>• Jak nakoupit Office</li><li>• Návod a postupy</li><li>• Za 1690 Kč pro studenty a domácnosti</li><li>• Jak funguje Superkancelář?</li><li>• Jak otevřít Office 2007 ve verzi 2003?</li></ul>

# Organizace: struktury informací

- Uspořádání skupin a prvků obsahu -> dohledatelnost
- Různá míra volnosti
  - Hierarchie (taxonomie)
  - Fazetová klasifikace
  - Datový model
  - Štítkování (folksonomie)
  - Hypertext
- Týká se také metadat

## Marcus Tullius Cicero

*Tento článek pojednává o římském politikovi. O typografické jednotce pojednává článek [cicero](#) (typografie).*

**Marcus Tullius Cicero** (3. ledna 106 př. n. l. Arpinum – 7. prosince 43 př. n. l.) byl římský řečník, politik a spisovatel

Pocházel z nepříliš bohaté rodiny, první vzdělání získal v Arpinu, později se vzdělával ještě v Římě. Studoval především rétoriku, ale i právo a filosofii. V době studií překládal z řečtiny.

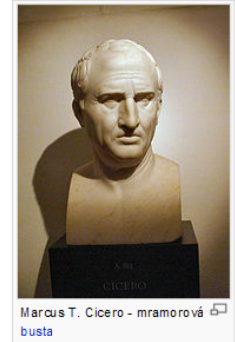
Roku 79 př. n. l. začal vystupovat veřejně, především u soudu. V tomto roce byl nucen ze zdravotních důvodů opustit Řím, odjel do Řecka, později cestovali i po Asii. Na těchto cestách se dále vzdělával a především cvičil v rétorice. O dva roky později 77 př. n. l. se vrátil do Říma, zde se oženil s Terentií a začal se věnovat veřejným záležitostem, především **obhajobám** u soudu. Svou činností získal značnou popularitu a roku 76 př. n. l. byl zvolen **quaestorem** na Sicílii, po ukončení **quaestury** se stal **senátorem**.

Roku 70 př. n. l. se stal **aedilem** kurulským a v roce 67 př. n. l. se stal **městským praetorem**. Roku 64 př. n. l. se stal **konzulem**, v této funkci odvrátil **pozemkový zákon** předložený **Publiem Serviliem Rullem** a **Catilinovo** spiknutí, obě tyto akce usilovaly o **potlačení demokracie** v Římě.

Roku 58 př. n. l. musel odejít do **vyhnání**, ale již roku 57 př. n. l. mu byl povolen návrat do Říma. Zde se dále snažil vystupovat na obranu republiky, ale vzhledem k tomu, že veškerá moc již byla v rukou triumvirů, odešel z veřejného života a věnoval se výuce především **gramatiky** a **rétoriky**, veřejně vystupoval pouze při soudních obhajobách.

Před občanskou válkou se přiklonil na stranu **Pompeiovu**, po jeho porážce odmítl vedení republikánského loďstva a raději odprosil **Caesara**, se kterým se vrátil do Říma jako přítel. Od roku 47 př. n. l. nevykonával žádnou veřejnou funkci a věnoval se pouze **filosofii**. Na veřejnosti vystupoval pouze v případech, kdy některý z jeho přátel potřeboval pomoc (*Oratio pro Marcello*, *Oratio pro Ligario* a *Oratio pro rege Deiotario*).

Po Caesarově smrti věřil v obrát k republice, ale byl zklamán, proto opustil Řím, cestou ho zastihla nepravdivá zpráva, že se Antonius vzdal moci. Po návratu zjistil, že se tato zpráva nezakládala na pravdě a začal proti **Antoniovi** vystupovat, přednesl **14 řečí proti Antoniovi**. Po těchto projevech senát vyhlásil Antoniovi válku, jejíž vedení svěřil **Octavianovi**, ten se ale spojil s Antoniem a tím se moc senátu definitivně rozpadla. Cicero se pokusil o útek z Říma, avšak byl dostižen a popraven. Jeho hlava a pravá ruka byly vystaveny na řečišti v Římě.



# Pojmenování: Kontrolované slovníky

- Druhy

- *synonymické kruhy*

- *authority* file (preferované pojmy, +alternativy)

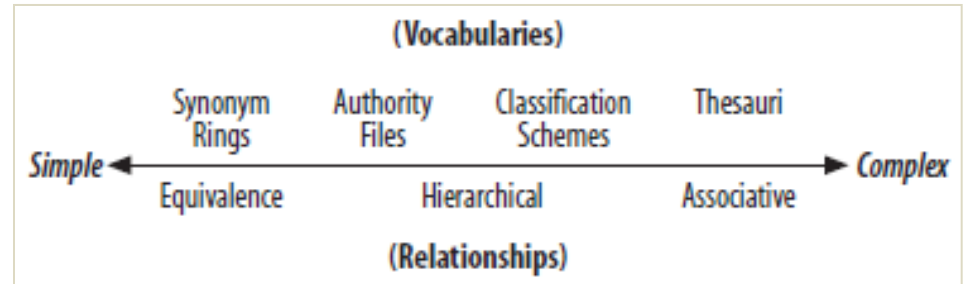
- *tezaury* (vztahy mezi pojmy): nutno řešit slovní formu, pád, rod, číslo, hláskování, ..., psaní zkratek

- Využití

- zvýšení úplnosti hledání (x přesnost)

- poučení uživatelů (terminologie v oblasti)

- navigace (konzistentně preferované pojmy)



# Označování informací

- Návěští (label) = reprezentace informace
  - výrazně významnější než v interaktivní komunikaci
  - přiroz. jazyk, laická práce => možné problémy

- **Typy návěští**

- kontextové odkazy
- nadpisy
- volby v navigaci
- rejstříkové pojmy
- ikony





# Jak získat systém návěští

- Systém: granularita, úplnost, způsob prezentace
- Ze **stávajících** informací
  - analýza webu, dokumentů – inventář pojmů
  - podobné a konkurenční weby – patterns of labels
  - slovníky, tezaury
- Pro **nový** (budoucí) obsah
  - analýza obsahu – inventář pojmů
  - konzultace s experty – definovat oblast, získat pojmy
  - skupinová technika „třídění karet“
  - analýza logů – zejména hledané výrazy
  - vytěžování folksonomií

Výsledky nutno  
vyčistit, revidovat,  
doplnit

# Problémy a výzvy

- Nejednoznačnost jazyka
  - homonyma, klasifikace termínů
- Heterogenita obsahu
  - typy
  - granularita
- Odlišnost pohledů
  - tvorba taxonomií subjektivní
- Lokální politika
  - ovlivnění dat osobními zájmy

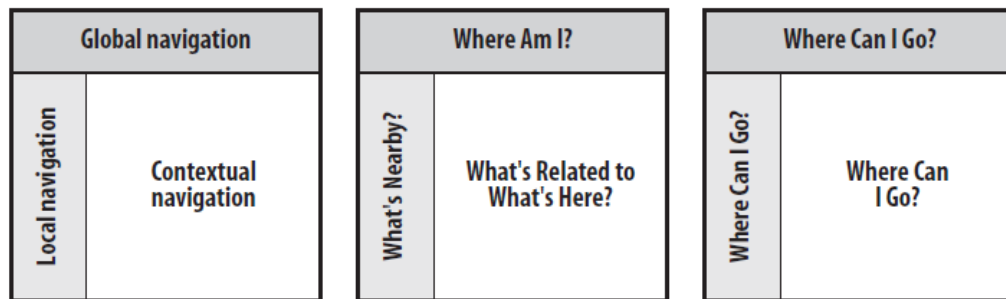
# NAVIGACE

# Navigační systémy

- Výsledek organizace a pojmenování obsahu
- Systém navigace po website = kolekce prvků
  - vestavěné ve stránkách
  - doplňkové formy
  - vlastnosti prohlížeče
- Důležité vlastnosti
  - jednotnost
  - srozumitelnost – slovník, architektura webu
  - [neutopit obsah](#)

# „Vestavěná“ navigace

- V každé stránce => porozumění IA webu



– kontext stránky (test: „tisk“ nebo „z vyhledávače“)

- Druhy

– *globální*: stejná na [všech subsite](#); autor=IA

– *lokální*: možnosti v daném kontextu, stejný fmt

– *kontextová*: odkazy/rámeček, [cross-sell](#); autoři=editoři

# Doplňková navigace

- Alternativa k hlavní, možnost přímo na obsah
- Druhy
  - *vyhledávání* (viz dále)
  - *site map*: pár úrovní velkého hierarchického webu
  - *rejstřík*: abecední schema bez hierarchie;  
granularita? ručně? slovník? terminologie nestálá
  - návody a *průvodci*: lineární, stručnost,  
přehlednost

# Souvisí s navigací

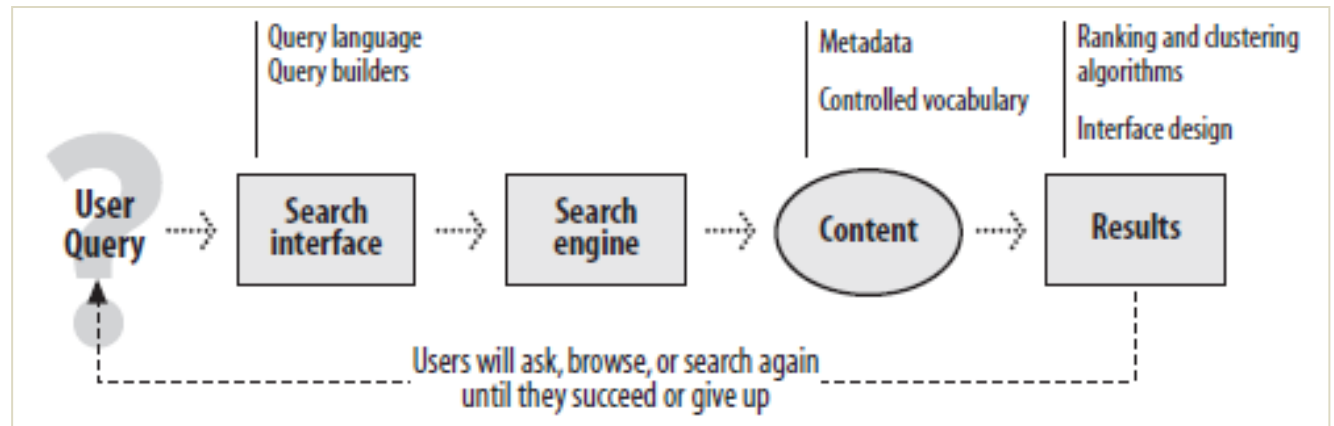
- Další způsoby, jak pomoci uživatelům dohledat
  - velká hype, skromná realita
- *Personalizace* = předpřipravené stránky podle publika
  - vhodné pro intranety
- Přizpůsobitelnost = uživatel může sám ovlivnit
- Vizualizace
- *Sociální* navigace = dle feedbacku uživatelů
  - Web 2.0 pomáhá u všech tří

# VYHLEDÁVÁNÍ



# Vyhledávání

- Velká oblast, velké očekávání, velké peníze
- Základní anatomie



- Rozhraní
  - „The Box“ (+ „Advanced“)
  - SERP = Search Engine Results Page

# Proč a co hledat

- Vyhledávač není IT ale IA => požadavky
- Co vyhledávat
  - *typy obsahu*
  - *oblasti webu* (search zones) – indexované separátně, zvýšení relevance
  - *koncové (obsahové) stránky*
  - *části prvků obsahu* – titulek, hlavní text, odkazy, metadata (klíčová slova), img alt texty

# Search Box

- Aspekty
  - expertíza a informační potřeby uživatelů
  - typ a množství dat (informací)
- The Box – předpoklady
  - celý web, jakékoli slovo, stačí Enter
  - synonyma? fráze? gramatika? tlačítko?
- Advanced – má smysl? usability x effort x use
  - gramatika výrazu, prohledávané oblasti, jazyk, ...
  - Pro normální web je lepší dát možnost pracovat s výsledky hledání.

# SERP

- Co zobrazit
  - dotaz, [nastavení](#) (z advanced), počet výsledků
  - části obsahových prvků (title? text? vybrané věty?)
  - související (rubriky, kategorie, stránky)
- Kolik zobrazit
  - počet na stránce
  - stránkování (rozlišení obrazovky, kognitivní limity)
- Jak zobrazit
  - abecedně, chronologicky
  - relevance (ranking), [filtrování](#) a [kategorie](#) výsledků

# Parametry vyhledávače

- Kvalitativní
  - přesnost, úplnost, F-míra (P=0.45, R=0.61, F=0.50)
- Algoritmy
  - stemming
  - stop-slova
  - spelling, fonetika („ofis suit“)
  - kontrolovaný slovník
  - sémantika dotazu („[Kdy se narodil Galileo?](#)“)
- Relevance výsledků (ranking)
  - ručně x automaticky x sociální x placené
  - shoda s výrazem, výskyt v dokumentu, důležitost stránky ...
  - Google PageRank (U.S. Patent 6,285,999) – „předávání“

# SOUHRN

# Z čeho se skládá IA daného webu

- Základní složky
  - Organizace a značkování informací
  - Navigační systémy
  - Vyhledávací systémy
- Další kategorie
  - Procházení vs hledání
  - Viditelné a neviditelné
- Každá stránka

