Vyhledávání informací

Informační objekty

Formulace rešeršního dotazu

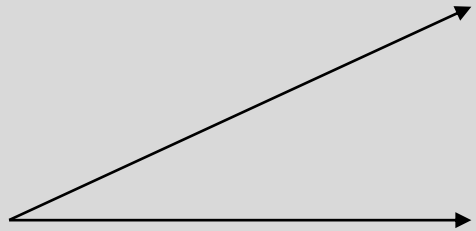
8. 10. 2021

Rozehřívací vyhledávačka

Potřebuji vědět, ve kterém roce byl objeven neobydlený ostrov, který je vzdálený nějak 1350, 1400 nebo možná až 1450 mil od nejbližšího člověka. Nemůžu si vzpomenout, kolik přesně to bylo mil, ale někde jsem o tom četl...

Opáčko: proces vyhledávání

- Informační zájem
- Informační potřeba
- Informační požadavek



- Osobní
- Profesní

- Fyzická
- Vědomá
- Formalizovaná
- Kompromisní

Referenční rozhovor
Rešeršní dotaz

Proces vyhledávání

- Referenční rozhovor
- Volba strategie a volba zdrojů
    1. pojmová analýza
    2. hledání synonym a souvisejících pojmů
    3. převedení na výrazy řízeného slovníku
    4. aplikace operátorů
    5. aplikace dalších vyhledávacích technik

Formulace rešeršního dotazu

Aktéři vyhledávacího procesu

- uživatelé
- informační profesionálové
- informační objekty, dokumenty, zdroje
- metody, typy vyhledávání
- vyhledávací nástroje

Informační objekty

- informace (DIKW)
- dokumenty
- informační zdroje
- *elektronické informační zdroje*

Dokumenty

Informační pramen tvořený nosičem informací a množinou informací na něm fixovaných a sloužící k přenosu dat v čase a prostoru. Dokumenty se dělí podle řady kritérií, např. podle způsobu záznamu dat, podle odvozenosti obsahu, podle kontinuity (periodické a neperiodické), podle stupně zveřejnění (zveřejněné, nezveřejněné, interní).

Dokumenty

- podle kontinuity
- podle míry zveřejněné
- podle odvozenosti obsahu
  - primární (monografie, článek…)
  - sekundární (odkazuje na primární)
  - terciární (odkazuje na sekundární)

Dokumenty

- podle druhu způsobu záznamu dat
  - textové / netextové
  - obrazové, zvukové, audiovizuální, strojem čitelné
  - smíšené

Informační zdroj

Informační objekt (informace nebo skupina informací tvořících jednotný celek bez ohledu na typ nebo formát), který obsahuje dostupné informace odpovídající informačním potřebám uživatele. Informační zdroj může být tištěný, zvukový, obrazový nebo elektronický (včetně zdrojů dostupných online).

Informační zdroje

- podle typu
  - bibliografické
  - plnotextové
  - faktografické

- podle obsahu
  - polytematické / monotematické

- podle druhu dokumentů

EBSCO
PSYCINFO

Informační zdroje

- producent, poskytovatel (- *informační instituce*)
- dostupné veřejně / neveřejné
- komerční / nekomerční

- *je třeba hodnotit kvalitu IZ*
- *využívat metazdroje*

Vyhledávačka

Najděte terciární informační zdroj, ve kterém budete moci identifikovat relevantní sekundární či primární informační zdroje pro následující vyhledávání v oblasti medicíny.

Identifikace IZ

- zjistím, že nějaký zdroj existuje
- pomocí sekundárních a terciárních zdrojů
- zjistím jeho vlastnosti, včetně možností vyhledávání
- zvolím vhodný zdroj i s ohledem na obsahové aspekty
- zjistím (a zajistím) dostupnost IZ

Metodika vyhledávání

- typy vyhledávání
- nástroje vyhledávání

Typy vyhledávání

- Identifikační vyhledávání
- Věcné vyhledávání
- Faktografické vyhledávání

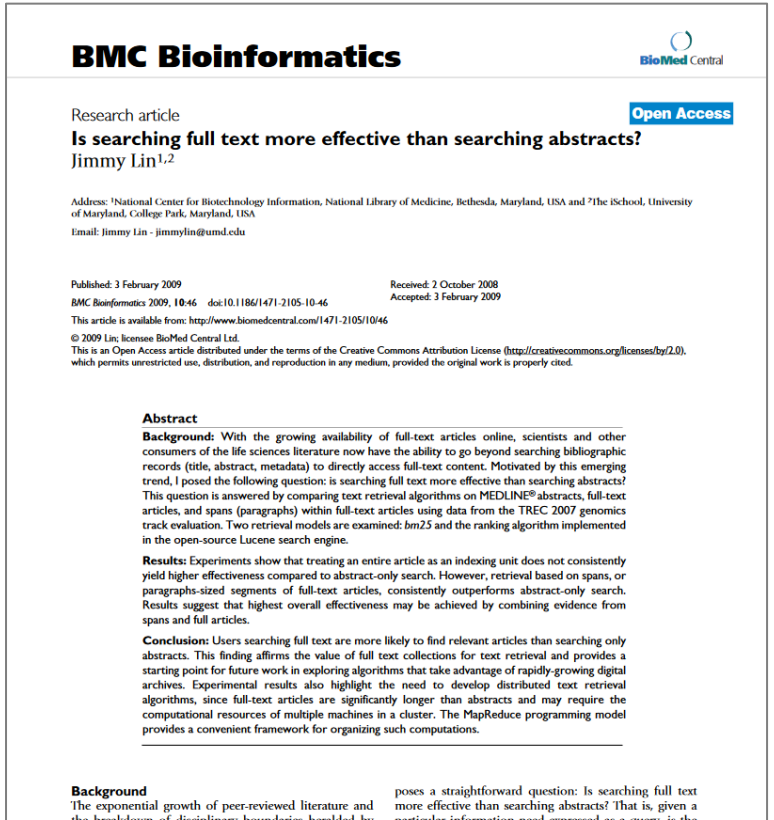
Typy vyhledávání

- **searching (vyhledávání)** – query modely
- **browsing (prohlížení)** – vyhledávání jako interaktivní aktivita měnící se podle momentální potřeby, vyskytuje se bez ohledu na strukturu vyhl. systému; heuristické a oportunistické
- **filtering (filtrování)** – odstraňování dat z příchozího toku informací – filtry, fasety, osobní profily
- **data mining** – procházení skrze velké množství dat s účelem odhalit vzory a vztahy pro řešení otázek

Typy vyhledávání

- Nestrukturované – freetextové (jednoduché)
- Strukturované – metadata (selekční obraz dokumentu)
  pokročilé vyhledávání; řízené slovníky
- Plnotextové

fulltextové vyhledávání



BMC Bioinformatics

BioMed Central

Research article

Open Access

Is searching full text more effective than searching abstracts?

Jimmy Lin[1,2]

Address: [1]National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland, USA and [2]The iSchool, University of Maryland, College Park, Maryland, USA

Email: Jimmy Lin - jimmylin@umd.edu

Published: 3 February 2009

BMC Bioinformatics 2009, **10**:46 doi:10.1186/1471-2105-10-46

Received: 2 October 2008

Accepted: 3 February 2009

This article is available from: http://www.biomedcentral.com/1471-2105/10/46

Abstract

Background: With the growing availability of full-text articles online, scientists and other consumers of the life sciences literature now have the ability to go beyond searching bibliographic records (title, abstract, metadata) to directly access full-text content. Motivated by this emerging trend, I posed the following question: is searching full text more effective than searching abstracts? This question is answered by comparing text retrieval algorithms on MEDLINE® abstracts, full-text articles, and spans (paragraphs) within full-text articles using data from the TREC 2007 genomics track evaluation. Two retrieval models are examined: *bm25* and the ranking algorithm implemented in the open-source Lucene search engine.

Results: Experiments show that treating an entire article as an indexing unit does not consistently yield higher effectiveness compared to abstract-only search. However, retrieval based on spans, or paragraphs-sized segments of full-text articles, consistently outperforms abstract-only search. Results suggest that highest overall effectiveness may be achieved by combining evidence from spans and full articles.

Conclusion: Users searching full text are more likely to find relevant articles than searching only abstracts. This finding affirms the value of full text collections for text retrieval and provides a starting point for future work in exploring algorithms that take advantage of rapidly-growing digital archives. Experimental results also highlight the need to develop distributed text retrieval algorithms, since full-text articles are significantly longer than abstracts and may require the computational resources of multiple machines in a cluster. The MapReduce programming model provides a convenient framework for organizing such computations.

Background

The exponential growth of peer-reviewed literature and

poses a straightforward question: Is searching full text more effective than searching abstracts? That is, given a

Nástroje vyhledávání

- Operátory – logické, proximitní
- Další vyhledávací nástroje a techniky
- Selekční jazyky (identifikační, věcné, MDT, DDT,...)
- Uživatelské rozhraní (grafické rozhraní, filtry, atp.)
- Jednoduché / pokročilé vyhledávání
- Vyhledávací jazyky (např. Common Command Language)

GEOBIBLINE

- česká geografická databáze
- obory: teoretická, fyzická a sociální geografie, kartografie, demografie…
- CCL search *(wyr=2008 and wau=novák)*
- *https://geobibline.cz/*
- ISO 8777

- Náhradní řešení: https://aleph.cvut.cz/

Vyhledávačka

Nalezněte v našem Alephu, zda se i v něm dá vyhledávat pomocí dotazovacího jazyka CCL a pokud ano, tak jak. Následně zjistěte skrze CCL dotaz, kolik je v databázi knih od autorů Čapka, Nerudy nebo Máchy, které mají přímo v názvu titulu varianty slova literatura a nevydalo je nakladatelství Odeon.

Vyhledávačka

Mělo by jich být 8.

((wau=čapek or neruda or mácha) and (tit=liter?)) not (pub=odeon)

Proces vyhledávání

- Referenční rozhovor
- Volba strategie a volba zdrojů
    1. pojmová analýza
    2. hledání synonym a souvisejících pojmů
    3. převedení na výrazy řízeného slovníku
    4. aplikace operátorů
    5. aplikace dalších vyhledávacích technik

Formulace rešeršního dotazu

1. Pojmová analýza

- identifikace klíčových pojmů
- reprezentace pojmů

„Potřeboval bych něco o nemocech srdce.“

1. Pojmová analýza

- identifikace klíčových pojmů
- reprezentace pojmů

„Potřeboval bych něco o nemocech srdce.“

„V předmětu řešíme různé typy problémů s tlakem u starších lidí. Hledám odborné zdroje, které se tomuto tématu věnují. Nejlépe maximálně deset let staré, v češtině i angličtině.“

2. Hledání synonym

- identifikace synonym a příbuzných výrazů
- hledání výrazů v dalších jazycích

- Proč?
- pro výběr vhodného vyhledávacího výrazu
- pro převod na výraz věcného selekčního jazyka
- pro rozšiřování a zužování tématu

3. Výrazy řízeného slovníku

- výraz = výraz slovníku (př. USE – preferovaný)
- výraz RT k výrazu slovníku (related)
- existuje pouze BT slovníku (broader)
- existuje pouze NT slovníku (narrower)

Slovník lexikálních jednotek selekčního jazyka
uspořádaný specifickým způsobem (např. zahrnuje
vztahy ekvivalence, hierarchie a asociace), který slouží
pro indexaci a vyhledávání dokumentů.

Příště...

- Dokončení tématu
- Boolovské operátory
- Modely vyhledávání
- Co to je, k čemu to je
- Proč to řešíme
- Rešerše

Mgr. Tomáš Marek
marek@kisk.cz