Debating-Oriented Information Retrieval
(Towards the WHY Search Engine)

Norbert Fuhr
University of Duisburg-Essen, Germany
Retrieval, Seeking & Work Task

Socio-organizational & cultural context

Work task context

Seeking context

Docs
Repr
DB

Request
Repr
Query

Match
Result

Seeking Task
Seeking Process
Seeking Result

Work Task
Work Process
Task Result

Kekälainen & Järvelin 2002
Purposes of a WHY engine

• Decision Support
  – in the workplace
    (knowledge workers spend 40% of their time with info access)
  – buying products
  – for or against medical treatments
  – for whom should I vote?

• Understanding/Learning
  – ...

Current Web Search Technology

- Mainly word based
- Biased results
Schloss Dagstuhl: Über Dagstuhl
https://www.dagstuhl.de/
All past and future conferences in Dagstuhl are listed on the web site.

Schloss Dagstuhl: Dagstuhl-Seminare
https://www.dagstuhl.de/programm/dagstuhl-seminare/
Immer wieder wird bestätigt, dass Dagstuhl-Seminare die produktivsten wissenschaftlichen Veranstaltungen sind, die Wissenschaftler erlebt haben. Die auf ...

Schloss Dagstuhl: About Dagstuhl
https://www.dagstuhl.de/en/about-dagstuhl/
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH (English: Schloss Dagstuhl – Leibniz Center for Informatics) is the world's premier venue for informatics ...

Dagstuhl – Wikipedia
Die im Jahre 1290 erstmals erwähnte Burg Dagstuhl war die Keimzelle der späteren Herrschaft Dagstuhl. Nach 1375 kam die Herrschaft an die Familien der ...
Geographie - Geschichte - Politik - Kultur und Sehenswürdigkeiten
Why queries in Web search


- Analysed 311 Bing queries from 2006
- 25.2% of the why-questions are answered at the search results page
- 74.4% of the why-questions get at least one relevant answer in the top 10
For what reasons should I choose C# over Java and C++?
programmers.stackexchange.com/.../for-what-reas... ▼ Diese Seite übersetzen
17.12.2011 - Yes, C# is a better language with more modern features than C++ and Java, but that is not the most important thing for choosing C#.

The art of programming: Java vs C# - Stack Overflow
stackoverflow.com/.../the-art-of-programming-java-... ▼ Diese Seite übersetzen
04.03.2009 - Throughout my university career I used Java to code projects until I ... For years C# was playing catchup with Java. That ended with C#. ... I'm a ...

Java vs. C# - Computer Science
www.cs.yale.edu/homes/hudak/CS112F06/java.html ▼ Diese Seite übersetzen
You might have heard of the programming language called Java, since it has gotten a lot of press in recent years, perhaps more than any language ever.

Is C# better than Java? - Quora
https://www.quora.com/Is-C-better-than-Java - Diese Seite übersetzen
To quote from an introductory CS class at Yale, "[C#] is better than Java in certain ways, but is so similar in most ways that once you have learned how to p..."
Web Search for Decision Making (2)

is dagstuhl better than shonan

[PDF] 4 - DROPS - Schloss Dagstuhl
drops.dagstuhl.de/opus/volltexte/2013/.../annual-report-2012-web.pdf ▼
von S Dagstuhl - 2013 - Ähnliche Artikel
for face-to-face meetings is even more important now than it tik-Teildisiplinen ...... “Nil Shonan Meeting Memorial Symposium – For Making. Als eingeladener ...

Schloss Dagstuhl : GI-Dagstuhl Seminars
https://www.dagstuhl.de/en//gi-dagstuhl-seminars/ ▼ Diese Seite übersetzen
The proposal should be informal and should not have more than 1-2 pages. The proposal should contain the names of the organizers as well as a description of ...
Es fehlt: shonan

Schloss Dagstuhl : Call for Proposals
10.09.2015 - More information about Dagstuhl Seminars and Dagstuhl ... Including more than four organizers in total, or more than two organizers with the ...
Es fehlt: shonan

A workshop revival
dl.acm.org/ft_gateway.cfm?id=2160726 - Diese Seite übersetzen
von P Hyman - 2012
Dagstuhl-like venues, such as Shonan. Village in the Tokyo ... attend more and more specialized conferences ... cept—and then find an attractive place for it, get ...
Bias in Web Search

R. White: *Beliefs and Biases in Web Search*, SIGIR 2013

Studied results for 680 yes/no questions on health issues

1) Engines more likely to rank Yes above No, and return more Yes

2) People much more likely to click on Yes than No, even when control for availability and rank position

3) Engine had wrong answer @ top rank for half of questions
Limitations of Current Web Search

- “Positivism”: Engines search for words only, do not understand semantic relationships
  Is Barack Obama a Muslim/Christian?
  Is vaccination harmful/safe?

- Clickthrough bias: search engines prefer pages clicked more often
  - “Yes” answers
  - Spectacular answers:
    Causes for headache → brain tumor
Towards Argument Retrieval

D. Swanson: *Two Medical Literatures that are Logically but not Bibliographically Connected.* JASIST, 1987

- If "A causes B" and "B causes C" are well-tested hypotheses, then the hypothesis that A causes C is strongly suggested
- d1: ”dietary fish oil causes specific blood changes”
- d2: “certain blood changes cause Raynaud's disease”
- q: “Might Raynaud's disease be caused by dietary fish oil”? 
  
  \[ d_1 \cup d_2 \rightarrow q \]

- combined evidence from multiple documents!
The logical view on IR

Search for document $d$ implying the query $q$

d = 'Dagstuhl seminar debating technologies'
q = 'debating technologies'
d = \{t_1, t_2, t_3, t_4\}  q = \{t_3, t_4\}

$t_1 \land t_2 \land t_3 \land t_4 \rightarrow t_3 \land t_4$
Ontologies and Logic-Based IR

- Subclasses as implication:
  \[ \text{Rioja} \sqsubseteq \text{Red\_Wine} \sqsubseteq \text{Wine} \]
  \[ \text{rioja} \rightarrow \text{red\_wine} \rightarrow \text{wine} \]

- Rules in ontologies
  \[ \text{parent}(X,Z) \land \text{parent}(Y,Z) \land \text{gender}(X,\text{female}) \rightarrow \text{sister}(X,Y) \]
Information Retrieval as *Uncertain* inference

- \( P(d \rightarrow q) := P(q|d) \)
IR based on probabilistic rules

0.3 causes(dietary_fish_oil, blood_changes);
0.2 causes(blood_changes, Raynaud's_disease);
0.2 causes(tar, lung_cancer);
0.1 causes(alcohol, headache);

contains(cigarette_smoke, tar); contains(wine, alcohol);
subclass_of(rioja, red_wine); subclass_of(red_wine, wine);

causes(X, Y):-contains(X, Z), causes(Z, Y);
0.5 causes(X, Z):-causes(X, Y), causes(Y, Z);

?- causes(X, Y);
0.03 causes(dietary_fish_oil, Raynaud's_disease);
0.06 causes(cigarette_smoke, cancer);
0.10 causes(rioja, headache);
Research challenges

• Argument representations suitable for IR
  – logic-based vs. more surface-oriented
  – proof trees across document borders
  – contradictions

• Deduplication of arguments
  – clustering, recognition of paraphrasing

• Importance + credibility of arguments
  – “page rank” for arguments?

• Personalized retrieval
  – “show me new arguments that I understand”

• Interactive argument retrieval
  – interacting with sets of arguments/proof trees