Exploiting User Context and Preferences for Intelligent Web Search

Ana G. Maguitman - Carlos I. Chesñevar
Carlos M. Lorenzetti - Fernando M. Sagui
Guillermo R. Simari
Web Search Engines: Limitations

Web Search Engines address the problem of information overload by providing search results.

Limitations

• Unable to perform **qualitative inference** on user queries.
• Unable to deal with the **defeasible nature of user’s preferences**.
• Unable to provide **explanations**: trustworthiness issues!
Web Search Engines: Limitations

How to find out about any topic of special interest?

Exploiting contextual information to automatically retrieve resources relevant to the user’s task.

Existing search engines can’t reflect rich contexts:
- Length limits on queries may preclude including sufficient contextual information
- Variation in vocabularies may hinder successful retrieval
(1) Context-Specific Terms

**Descriptors:** Terms that occur *often* in the topic.
- Good topic descriptors are those terms that answer to the question: "*What is this topic about?*" (Recall)

**Discriminators:** Terms that occur *only* in the topic.
- Good topic discriminators are those terms that answer to the question: "*What are the best terms to build a query?*" (Precision)
(2) Context-based search system

Application incremental search to find novel but relevant material.

Apply clustering techniques to obtain new topics.

Fill in term-document matrix used for computing descriptors and discriminators.

Artificial Topics

Term-document matrix

Search Engine
(3) Extended DeLP Program

Facts are extracted from Context-based search results

Defeasible Rules

Strict Rules

Facts

S1

author(s1, jim_liar)
address(s1, “nyt.com/…”)
date(s1, 20031003)

S2

author(s3, jane_truth)
address(s3, “nyt.com”)
date(s3, 20031003)

S3

author(s4, bob_beak)
address(s4, “mynews.com/”)
date(s4, 20031003)
Defeasible rules

Strict rules

Facts

DeLP Program P

DeLP Interpreter

Abstract Machine

User Query

?- relevant(search_result)

Possible Answers to Query h

• **YES** (there exists a warranted argument <A,h> )

• **NO** (there exists a warranted argument for <A,~h>)

• **UNDECIDED** (none of the above cases hold).
After analysis based on implicit knowledge

(5) Prioritizing search results

Warranted as rel.
Warranted as rel.
Undecided
Warranted as ~rel.

author(s2, jim_liar)
address(s2, “nyt.com/...”)
date(s2,20031003)
rel(s2)

trust(jim_liar)
not faked_news(jim_liar)

~ biased(“nyt.com/...”)
american(“nyt.com/...”)
domain(“nyt.com/.; “nyt.com”) (“nyt.com”= “nyt.com”)
faked_news(jim_liar)

~ rel(s2)
address(s1; “nyt.com/...”) biased(“nyt.com/...”)

american(“nyt.com/...”)
This work has described ongoing research on exploiting the information in the user context to refine Web search queries. In addition it proposes a novel approach for enhancing Web search technologies through the use of qualitative, argument-based analysis. We have used DeLP for carrying on that analysis. Current research trends show that the combination of quantitative and qualitative analysis of the user context and preferences will play a major role in Web search technology.