Supporting Web Searches by Intelligent Local Agents

....because Google isn't future. Really not!

Herwig Unger
FernUniversität in Hagen, Faculty of Mathematics and Computer Science
Phone: +49 2331 9871155, +66 9797922070; Fax: +49 2331 987353
eMail: Herwig.Unger@gmail.com
Google became the all dominating empire.

and: we even don't know much how it works!
An example: Harry.

We get:
163 million results on 16.3 million pages!!!!!

→ no structured presentation
→ no independent ranking
→ no evaluation of trustworthiness
→ no support for the user
→ no use of user evaluation

!! Google only offer search for text or pictures or similar pictures !!
Our Approach to look at communication networks:

- Consider the mutual influence between content, users and user activities as well as network with its parameters and configuration.
So what we can do?
80% of all information in the WWW is given in a textual form!

- big challenge to filter relevant information
- usually 2-3 keywords are a weak description of what the users are looking for
- the typically received 10000+ search result overload the user and normally only the first 30 results are considered
- 3 til 6 words may return more precise results but it is hard to find words with high selection rate

The task is to find out **fastly** what the user is looking for and support him in this process.
Idea 1: Locality
Another example: “christmas tree“

- What does Google offer?
So what about “christmas tree“

This of course:

But also this:

An assembly of control valves, fittings, pressure gauges and pipes at the top of a well to control the flow of oil and gas after the well has been drilled and completed.

→ Looks like a decorated christmas tree (with some imagination).
Motivation / Problem statement 1: Disambiguation

- Disambiguation (also called word sense disambiguation or text disambiguation) is the act of interpreting an author's intended use of a word that has multiple meanings or spellings.

- Word sense disambiguation (WSD) is the task of selecting the appropriate senses of a word in a given context.
  - e.g. mouse (animal, comp.) cube (maths, car)
  - christmas tree (biol, oil), Harry (sev.names)
Already in 1911 the expression "Use a picture. It's worth a thousand words." appears in a newspaper article by Arthur Brisbane discussing journalism and publicity.

The roots of that phrase are even older and have been expressed by earlier writers.

The Russian writer Ivan Turgenev wrote (in *Fathers and Sons* in 1862), "A picture shows me at a glance what it takes dozens of pages of a book to expound."
Example: Harry

Search for Images: Harry

Image Results for: Harry

Select/deselect all images for analysis

Search for more similar images  Use keyword translation

Search with selected images
Idea 3: Systematic presentation of search results
Idea 4: User based evaluation of search results
Motivation / Problem statement 2

- Searching the WWW…
  - Manual query formulation is a tedious and error-prone task

- Evaluating large result sets is time-consuming

So why not let the computer read and find useful web documents for you?
Idea 5: Documents as queries

Concept:

- Use documents as the only initial search parameter while browsing

Technically:
- extract web (DocAnalyser) or local document’s (FxResearcher) main topics
- search for topical sources (important inherent, influential aspects / basics)
- use them as search words (query terms)

- Find similar and related content or track topics in real time (on-line) or when the user is off-line
DocAnalyser - Find Similar and Related Web Documents

What is DocAnalyser?

DocAnalyser is a new service that offers you a novel way to search for similar and related web documents and to track topics without the need to enter search queries manually. You just need to provide a web content to be analysed. DocAnalyser then extracts its main topics and their sources (important inherent, influential aspects / basics) and uses them as search words.

Installing DocAnalyser

In order to be able to use DocAnalyser, please drag and drop one or both of the following bookmarks to your bookmarks toolbar of your favourite web browser:

- **Analyse this content** (analyse currently shown/selected web content)
- **Analyse a web content** (analyse another web content)

*Remark:* The following bookmarklet should only be used when analysis errors occur using the first bookmarklet: **Analyse this content**
Idea 6: Decentralised search engines (see also YaCy and Faroo)
Summary and outlook. First ideas…

...to personalise search in the web much more.

- By local intelligence
- Considering user behavior
- Take history into account
- New services
- New interfaces
- Much less NSA

Thank you for your time! Q&A.