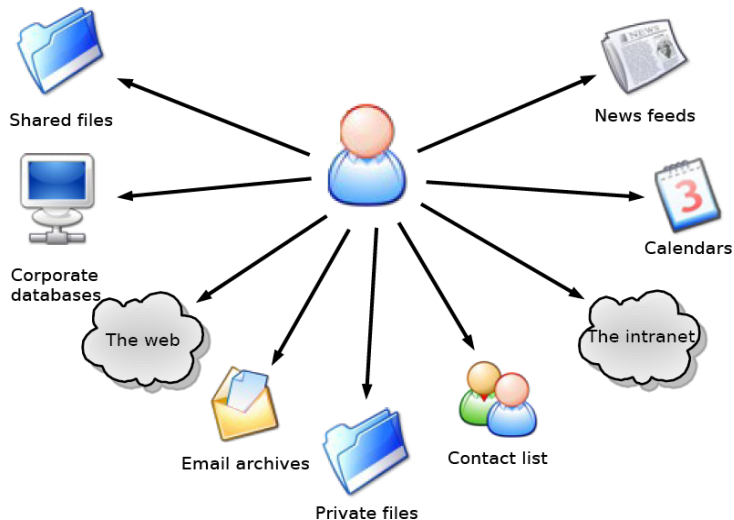


Personal metasearch

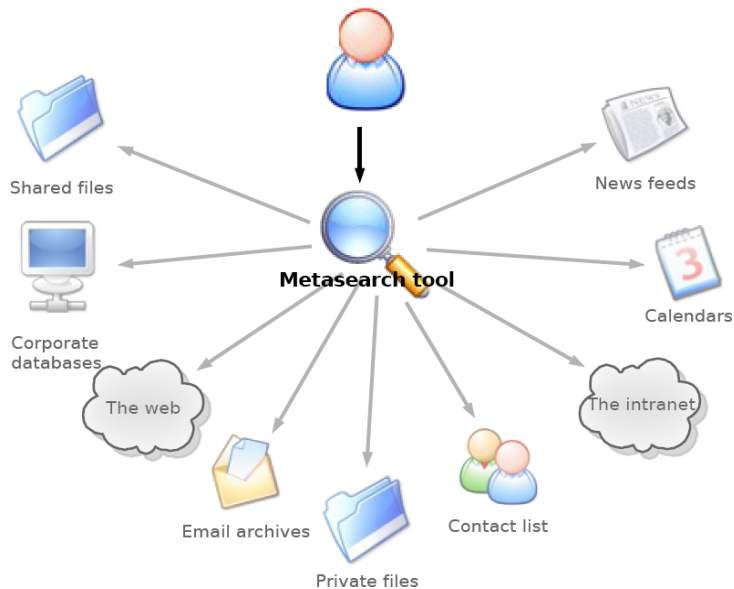
Paul Thomas (ANU)
David Hawking (CSIRO)

HCSNet, 21–22 September 2006

Personal metasearch



Personal metasearch



We can't index everything: there's too much, it's changing too fast, and we can't get access anyway.

This suggests *metasearch* techniques, offering a single interface to a set of existing search services.

- ▶ Scales well; can make use of search engine smarts; easier to keep up-to-date.

Some outstanding problems

We're focusing on **things that are different**:

- ▶ Discovering and selecting very different data sources;
- ▶ Combining results;
- ▶ Designing a user interface;
- ▶ Making use of personal or task context;
- ▶ Measuring success or failure;
- ▶ ...

Server characterisation

Build very simple models of the language used in each server's documents; use this to inform parts of the metasearch process.

- ▶ Experiments with sampling techniques to build these models: standard techniques don't seem to work.

Estimate the scale of each server's holdings; use this to inform parts of the metasearch process.

- ▶ Experiments with estimation methods: some simple techniques do surprisingly well, some more sophisticated techniques not so well.

Server selection

Given a query and a set of servers which may include relevant documents, choose which servers to forward the query to.

- ▶ Experiments selecting between (up to) six “servers”: standard techniques don’t work very well.

PIS, a prototype Personal Information Searcher

The screenshot displays the PIS interface with a search bar containing 'metasearch'. The results list includes:

- Metadata Principle 2**
http://sts.anu.edu.au/demetrius/standards/metadata_2.php?p=1, from ANU: e
- Re: More notes towards a proposal**
... While a large body of work exists on metasearch, the search proble
Fri, 21 May 2004 17:16:43 +1000, from Mail archives; estimated quality 69
- Meta-Search Engines**
<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/MetaSearch.htm>
<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/MetaSearch.html>, from
- Models for Metasearch**
Proc. ACM SIGIR 2001 W. Bruce Croft and David J. Harper and Donald
aslam01models, from References; estimated quality 100
- More notes towards a proposal**
... While a large body of work exists on metasearch, the search proble
Thu, 20 May 2004 10:57:16 +1200, from Mail archives; estimated quality 100
- metasearch**
metasearch, from Wikipedia; estimated quality 100
- ANU Library: InFlite**
<http://infinite.anu.edu.au/04/04b2c.htm>, from ANU; estimated quality 100
- Metasearch.com - The Original & Best Since 1995**
The original way to search the search engines. Images, Video, and Au
<http://metasearch.com/>, from Google; estimated quality 100
- Architecture of a Metasearch Engine that Support**
Proc. CIKM 1999 210-216 1999 ACM Press ...
glover99architecture, from References; estimated quality 96
- User-specified query types in Inquirus**
... it. It's described in Eric J. Glover et al. Architecture of a Metasearch
Tue, 8 Feb 2005 12:25:04 +1100, from Mail archives; estimated quality 94
- Fwd: Re: More notes towards a proposal**
... While a large body of work exists on metasearch, the search proble
Tue, 25 May 2004 09:08:31 +1200, from Mail archives; estimated quality 93
- Re: Martin Lucina**
I'm not sure where I want to go exactly, but then that's the point of

The inset window, titled "User-specified query types in Inquirus - KMail", shows an email from Paul Thomas to Dave Hawking:

User-specified query types in Inquirus

From: Paul Thomas <pault@cs.anu.edu.au>
To: Dave Hawking <David.Hawking@csiro.au>
Date: 2005-02-08 12:25

Hi Dave,

I mentioned this on Friday and you said you hadn't come across it. It's described in: Eric J. Glover et al. /Architecture of a Metasearch Engine that Supports User Information Needs/, Proc. CIKM 1999. I've got a paper copy, if that's easier.

The idea was just to select search engines (for metasearch) based on the user-specified query type -- "home page", "research paper", etc -- and to modify the query in some cases too. Seems like a good idea, although selecting the type of query each time would add substantially to the work a user has to do.

-- paul

Questions

Questions for the future:

- ▶ What sort of data sources do people use day to day? Which are useful, and for what type of need?
- ▶ What sort of queries do people issue? What relation is there between information need and query?
- ▶ How can we decide where to look for results?
- ▶ How can we merge (or rank) very different results when we get them?
- ▶ What use can we make of knowledge about the context of a query? How can we get this knowledge in the first place?
- ▶ ...