

## Current Research Topics



by:

Jacob Palme

e-mail:

jpalm@dsv.su.se

web: <http://www.dsv.su.se/jpalme>

Current Research Topics page 1

## Finding information

General search engines:

- Google
- Yahoo/Inktomi

Specialized search engines and dictionaries

Directories:

- Yahoo
- The open directory
- The Sunet web index

Intelligent answering

- Ask Jeeves (not any more)
- The semantic web

Current Research Topics page 2

## Google versus site index



[Nätet](#) [Bilder](#) [Grupper](#) [Kategori](#)

site:eu.int ...query...

Google-sökning

Jag har tur

[Avancerad sökning](#)  
[Inställningar](#)  
[Språkverktyg](#)

Current Research Topics page 3

## Google Page Rank

Google orders matches for queries using:

1. The Page rank of the matching page.
2. The relevance of the matching page (found in title, links to this page, body text)

This may mean that found pages are more good than relevant.

### Google Page Rank:

Exact algorithm secret.

External links to a page raises its page rank.

Links from good and relevant pages give more boost to the page rank.

Current Research Topics page 4

## Human organization and evaluation

Note: What Google does is a clever way of collecting organization of information done by lots of people when producing external links to pages.

This is essential: Only humans can do good organization of information.

### Compare: Amazon Books recommendation system:

Other people who bought the books you bought also bought...

## QuickAsk (Eriks Sneiders, DSV)

1. Producing templates for each existing document.
2. Match queries against these templates.

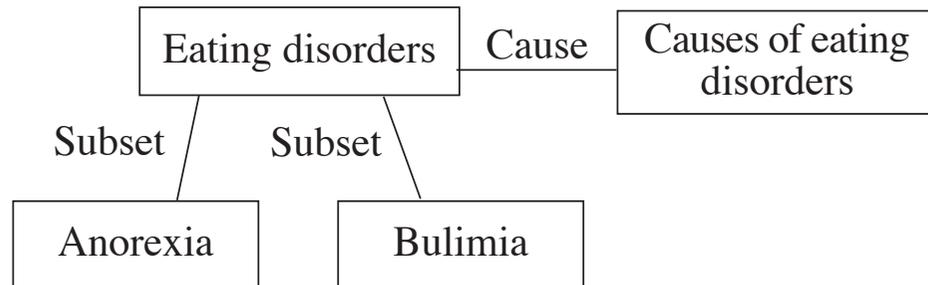
Example: What are the causes of anorexia?

Required: [ caus\* reason\* explanat\* basis\* origin\* motiv\* root\* why # \$eatingdisorder [\$eat # much] ]

Priority: caus\*

Optional: disord\*

## The Semantic Web

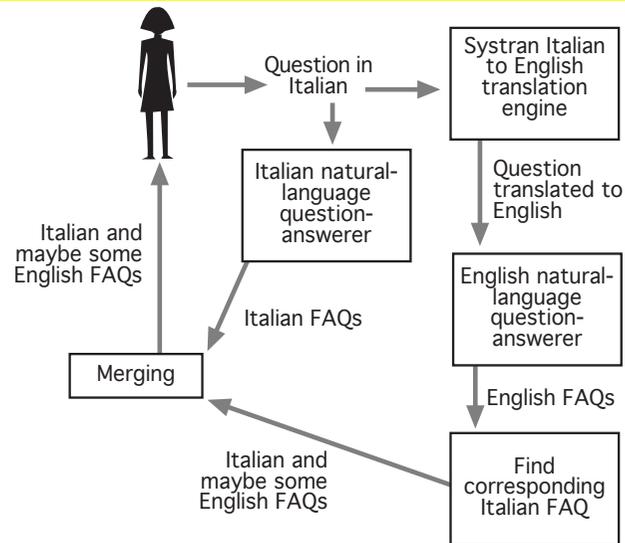


### Ontology

QuickAsk: Local ontology for each document.

The Semantic Web: Global ontology of all human knowledge.

## Cross-Lingual Question-Answering



Current Research Topics page 9

## 1970-s: The three truths of the 1970-s

- ⇒ Impoverishment of work tasks
- ⇒ Invasion of privacy
- ⇒ Industrial competitiveness

### WHAT I LEARNT:

The only way to get people to understand that computers can aid people instead of controlling them is to show them practical applications.

Current Research Topics page 10

## 1974: The General Public Information System Allmänhetens informationssystem

<<http://info.dsv.su.se/~jpalme/reports/general.html>>

Computers can be used to control people, or to give people more information and control over their own lives.

We must try to find uses of computers which give anyone easy access to lots of information, computers should open up information, not restrict information.

Current Research Topics page 11

## 1970: Gordon: P.E.T.: Parent Effectiveness Training

A book about how parents should solve family problems and train their children to become good people

### WHAT I LEARNT:

- Listen to each other
- Understand each other's views
- Show that you understand each other
- Find win-win-solutions

Current Research Topics page 12

## 1973: Swedish Data Act

- It is illegal to store personal information in computers without permission by the Data Inspection Agency
- Every field type must have a specification of what kind of personal information can be stored in that field
- Storage of information about religious and political beliefs (illnesses, sex life, etc.) is only allowed under special circumstances
- Exception: Use of computers for typesetting

### WHAT I LEARNT:

Politicians are not willing to see farther than their noses

Current Research Topics page 13

## 1970-s: The three truths of the 1970-s

- ⇒ Impoverishment of work tasks
- ⇒ Invasion of privacy
- ⇒ Industrial competitiveness

### WHAT I LEARNT:

The only way to get people to understand that computers can aid people instead of controlling them is to show them practical applications.

Current Research Topics page 14

## 1974: Arne Grip: ADB-system och kommunikation (Hermodsstudentlitteratur 1974)

What I learnt (not all of this taken from the book):

- All computer usage is communication between humans
- Running a program is communication between its designer and its user, just like reading a book is communication between the author and a reader
- Computers can be used to control and regulate communication, by specifying *what* information *can* and *cannot* be transferred, by *whom*, to *whom*.
- In his book 1984, Orwell invented a new language, *newspeak*, in which forbidden utterances could not be stated. Compare rules about improper language in human communication.
- People feel strongly about language, because language is their tools for controlling their living conditions. Taking away language from people is taking away their power and security.

Current Research Topics page 15

## 1975: Seeligman: Helplessness (Freeman, San Francisco, 1975)

A study of human depression showed that depression is strongly correlated by a feeling of helplessness, a feeling of not being able to control your own life.

### WHAT I LEARNT:

Computer can cause this feeling, if the user feels that the computer (in reality, its designer) restricts the freedom of the user.

Current Research Topics page 16

**1988**

Donald Broady: *Kulturens fält. Om Pierre Bordieus sociologi* (in *Masskommunikation och kultur*, Nordicom 1988) and Desmond Morris's books

Human beings like to belong to groups with similar language, culture and values as themselves. Such groups give them a feeling of security and acceptance. Members of such groups tend to favour each other and to belittle and disparage the cultural utterings of people outside their own group.

### WHAT I LEARNT:

Those who have control of computers tend to favour their own, those who are threatened by not having control of computers tend to disparage computer-oriented cultures.

Current Research Topics page 17

**1975:** Murray Turoff comes to Sweden

*“A computer is like a book with white pages. Any user can write what they like on these pages, and any other user can read what has been written.”*

(This may sound self-evident today, but it was not at all self-evident when Murray Turoff first said it.)

Murray Turoff introduced us to the *conference system paradigm* of software design.

Current Research Topics page 19

**1975:**

C.A.R. (Tony) Hoare: *Software Design — a Parable* (*Software World*, vol. 5, no. 9-10)

1. Computer software grows by adding more and more facilities.
2. This process is often strongly controlled by the present users of that software.
3. Through long experience with using the software, they master its usage and all its commands.
4. But computers by nature are restrictive, they only allow what their software is designed for.
5. Thus, the experienced users ask for new commands and facilities, and the developers give this to them.
6. Old software, which has gone through many such stages of evolution, tend to be very complex, with lots of commands, difficult to master for beginners.
7. Thus, strong user influence on software development can create a culture which frightens novice users away.

Current Research Topics page 18

**1975-1990:** EIES, Murray Turoff and Starr Roxanne Hiltz

The first software in this category was EMISARI invented by Murray Turoff in 1969, and another well-known software was Forum-Planet, invented by Jaques Vallee in 1971.

The EIES system, invented and set up by Murray Turoff in 1975, was very influential in developing the new ideas. Around EIES grew a community of users and a community of developers who invented and tried lots of novel ideas of organizing human communication. EIES was also the basis of much groundbreaking research on the effects of CMC.

Current Research Topics page 20

**1977-1978:** STU (today Vinnova): Forum-Planet, Consumer Information Systems

Tomas Ohlin was a person who strongly believed in the new ideas. He was responsible for getting Murray Turoff to Sweden, he arranged for Forum-Planet to be installed in Sweden, he arranged for a number of prototypes to be developed to show how computers could be used for giving more information to more people, for example in the area of information for consumers.

Current Research Topics page 21

**1978:** Turoff and Hiltz: The Network Nation

This revolutionary book describes a future society in which computer networks plays a central way in opening up new vistas for human-human communication and information exchange. It is a description of what Internet is beginning to become today, written more than a decade earlier.

Current Research Topics page 22

**1978:** KOM forbidden by the Swedish Data Inspection Agency

The Swedish Data Act, as it looked like in 1978 (and to a large extent still today) makes almost all storage of plain text information about humans in computers illegal. I noted this in a debate with Jan Freese, the director of the Data Inspection Agency and the creator of the Swedish Data Act. He said that these problems should be solved, and I should apply for permission.

After several talks to people in the Data Inspection Agency, I applied for permission to use the KOM system, and stated in my application that we could not, in advance, prescribe exactly what people would want to write about in their messages to each other.

Result: Our system was forbidden by the Data Inspection Agency.

Current Research Topics page 23

**1979:** KOM allowed with humiliating conditions

My employer at that time (FOA) did not dare to appeal the decision. Instead, they negotiated a settlement which allowed KOM to be started on the following conditions:

- All messages must be deleted after two years.
- Personally addressed messages must be deleted after one month.
- No search is allowed on personal information in message texts.
- No messages may contain information about political or religious beliefs, or information about a person's health, sexual behaviour, etc.
- FOA was to make a study of the effects of KOM on its users, and make a continued application for permission after two years.

Current Research Topics page 24

**1982:** First Swedish indirect connections to the Internet (KOM, Usenet News)

**1984:** KOM for hobby computer users

SE-banken sponsored our very expensive transatlantic transfer of messages between the Internet and KOM.

QZ established a special, low rate for use of KOM by hobby computer users after 20:00 in the evening.

Current Research Topics page 25

## Spamming

Freedom of speech is a freedom to say things.

It is not a right to force people to listen to what you have to say.

### WHAT I LEARNT:

Balance of control between sender and recipient.

Anti-spamming: Rule-based.

Anti-spamming: Bayesian.

**Note:** Both rely on human structuring!

Current Research Topics page 27

**1980-s:** Automatic software not liked

Case 1: Sorting of conferences in a personal priority order.

Case 2: Recognition of commands in ordinary text (get encouragement, get scolding)

Case 4: Get user description in 10th session.

Case 3: Basic and advanced mode.

### WHAT I LEARNT:

- Do not try to make computer software more intelligent than it can be.
- People want computer software to be predictable.

Current Research Topics page 26

## How Search Engines Order Web Pages

1. Many external incoming links from pages which themselves have a high ranking.
2. As exact match as possible for the query phrase:
  - a. Primarily in the title.
  - b. Also in text in links pointing to this page.
  - c. Also elsewhere in readable text in the body, repeated 2-3 times, if the phrase is repeated more than 2-3 times, there is a risk that the search engine will regard it as spamming.
  - d. Some engines also look at meta-description and meta-keywords.
3. Manual classification for the most common search phrases.

Current Research Topics page 28

## Search Engine Optimization

Method	Why it works	How we used it
High quality web site	Search engine rank web sites	Our main goal
Get links from external sites	Search engines count links to web site.	We asked a number of high quality sites to set links to our site.
Submission to search engines and portals	Users come from them, and search engines rank by them.	Yahoo, The Open Directory, the Sunet Directory.
Search phrase optimization	Search engines ranks by count of phrases, especially in titles.	We found popular search phrases for 78 of our best FAQs, and optimized them.
Relevant internal linking	More relevant key words refer to a page.	Yes, a lot.

Current Research Topics page 29

## Dubious/Unethical Methods of SEO

Method	Why it works	How we used it
Doorway pages	More links, links with different texts	A little
Cloaking	Giving search engine spiders different pages	No
White text on white background	Does not work, search engines have learned to recognize it	No
Lots of lots of keywords	Does not work, search engines have learned to recognize it	A little
Link farms	Does not work, because farms get low ranking	No

Current Research Topics page 30

## Example of phrase optimization

FAQ identifier	child-trauma-symptoms
Original title	How children react to trauma
Optimized title 1	Effects of child abuse
Optimized title 2	How children react to trauma and effects of child abuse

Current Research Topics page 31

No.	Keyword	KEI Analysis	Count	24Hrs	Competing
1	<a href="#">"effects of child abuse"</a>	51.124	187	70	684
2	<a href="#">"child abuse statistics"</a>	43.210	402	150	3740
3	<a href="#">"child sexual abuse trauma symptoms"</a>	16.000	4	1	0
4	<a href="#">"signs of child molestation"</a>	9.846	16	6	26
5	<a href="#">"articles on child abuse"</a>	9.074	82	31	741
6	<a href="#">"symptoms of child molestation"</a>	6.750	18	7	48
7	<a href="#">"child molestation law"</a>	4.500	15	6	50
8	<a href="#">"information on child abuse"</a>	3.531	161	60	7340
9	<a href="#">"child abuse laws"</a>	3.522	127	47	4580
10	<a href="#">"child molestation statistics"</a>	3.411	18	7	95

Current Research Topics page 32

No.	Keyword	KEI Analysis	Count	24Hrs	Competing
11	"child molestation"	0.671	213	80	67600
12	"child abuse prevention"	0.564	241	90	103000
13	"child abuse cases"	0.445	112	42	28200
14	"child sexual abuse"	0.431	373	139	323000
15	"child abuse and neglect"	0.109	168	63	258000
16	"molestation"	0.058	121	45	252000
17	"child trauma"	0.002	3	1	5900

So we put "Effects of child abuse" in <TITLE>, <H1> and once in the text of the FAQ.

## Ethics of Search Engine Optimization

White text on white background

Obviously unethical

Cloaking = showing special page to search engines

Dangerous, ethics depend on how you use it

Doorway pages = special pages to attract visitors

Depends how you use it, we have one or two such pages

## How people find answers in Web4Health

Method of finding answer	Percentage
Browsing our static pages	60 %
Using Google and other search engines fromewhere outside our site to get directly to one of our answers	25 %
Browsing our forums and ask-the-expert areas	8 %
Using our own natural-language question-answering system	5 %
Other methods	2 %

Note: The reason for this is of course that many more people use Google than our natural-language question-answering system as a web search tool, because they do not know of our system and do not have bookmarks to it. Probably also that people do not understand how good our natural-language question-answering system is.

## Usenet News "Cancel" command

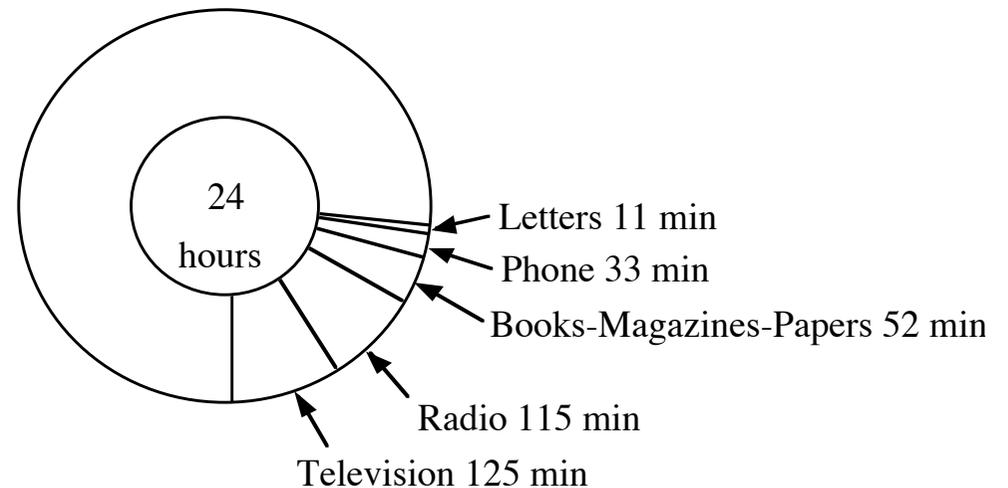
1. Originally: Only the author could cancel his own messages.
2. Anti-spam robots do automatic cancels on spam.
3. Scientology church does automatic cancels on messages it does not like.
4. Auto-resumbitters submit what has been cancelled.

## The future

1. Structure and organization is created by humans.
2. Structuring and organization work done by many people must be combined.
3. Structring and organization done by spammers must be handled with caution.

Current Research Topics page 37

## 1980 Human communication via technology



Current Research Topics page 38

## My research interests

1. The influence (power) of different actors on information transfer.
2. How should, and can, control of information transfer be handled to optimize legitimate needs of recipients and producers of information.
3. User influence on software design.

Current Research Topics page 39

## References

- Current Research Topics Overheads  
<http://www.dsv.su.se/jpalme/other-courses/crt-talk-ohs.pdf>
- Can Computers Decide what is Right and Wrong?  
<http://www.dsv.su.se/~jpalme/reports/right-wrong.html>
- A personal history of CMC  
<http://www.tema.liu.se/people/chrga/yw65/pdf/palme.pdf>
- Web4Health web site  
<http://web4health.info/>
- QuickAsk  
<http://www.askology.com/quickask.htm>
- Web4Health Search Engine Optimization (SEO)  
<http://web4health.info/documentation/D-6-2-first-promotion.pdf>

Current Research Topics page 40