

Intelligent Information Access

Petri Myllymäki, Ph.D., Professor
Complex Systems Computation Group
Helsinki Institute for Information Technology
University of Helsinki & Helsinki University of Technology
<http://www.hiit.fi/petri.myllymaki/>

2 / 17

The CoSCo group

- Mission
 - to develop intelligent solutions to hard problems found in realistic settings
 - requires both deep theoretical work and creative methodological innovations
- Research areas
 - probabilistic and information-theoretic predictive modelling / data analysis, *intelligent information access*
 - active collaboration with UCB (Michael Jordan), CWI (Paul Vitanyi, Peter Grünwald), Tsinghua University (Lizhu Zhou),...
- Researchers
 - Director: Petri Myllymäki (Henry Tirri)
 - Wray Buntine, Jorma Rissanen, Huizhen Yu, +15 others
- Funding
 - EU, Academy of Finland, Tekes, Industry, Universities

P. Myllymäki
03.05.2006



Outline

- Current status
- Problems
- Potential solutions
- Predictions for future trends
- Conclusions

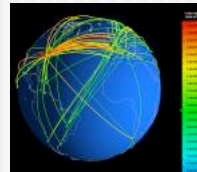
More information:

“Challenges of Future Search Engines” Consultation Workshop,
September 15th , 2005, Brussels

http://www.cordis.lu/ist/audiovisual/neweve/e/ws150905/ws150905_2.htm

The Internet today

- 10 Bn pages accessible through search engines like Google
- 85% of Internet traffic induced by search engines
- Billions of searches every month in US alone (4.8 Bn in July 2005)
- Search engine advertising: 8 Bn € today → 22 Bn € in 2010



All is well?



IA2006: International Workshop on Intelligent Information Access

http://www.helsinki.fi/ia2006/

July 4-8, 2006
Marteni Congress Center, Helsinki, Finland

IIIA-2006: International Workshop on Intelligent Information Access

Co-Chairs: Wray Buntine, *IIIT*
Henry Tim, *Nokia Research Center*
Local Arrangements: Pentti Mäyllymäki, *University of Helsinki*
Date: July 4-8, 2006
Venue: Marteni Congress Center, Helsinki, Finland

CFP: The CFP is here: [http://www.helsinki.fi/ia2006/cfp/](#)

Areas of interest: statistical methods, machine learning, probabilistic and information theoretic models, and other approaches to uncertainty and complexity in all areas of Information Access, including but not limited to:

- search and information retrieval,
- cross-language access,
- question answering,
- peer-to-peer support,
- semantic annotation,
- information extraction,
- information and reputation analysis, and
- support tools for machine translation,
- personalization,
- multimedia aspects of information access

Included speakers (developing):

Fernando Pereira *University of Pennsylvania*
Soumen Chakrabarti *MIT Boston*
Doug Cutting *Natasha-Lucene*

Workshop format: The workshop will consist of three days of talks, panels, informal breakout sessions, and a boat trip to local Helsinki islands. The number of workshop attendees is limited with a preference given to authors and presenters of contributed papers, and to a number of invited speakers chosen among the top experts of this field.

Dates:

Submission of abstracts or extended abstracts: April 15th
Notification of invitation to attend the workshop: May 15th
Deadline for final booking: May 31st
Workshop dates: July 4th-8th

Arrangements: The workshop will be held at the **Marteni Congress Center** in downtown Helsinki overlooking the harbour. A block of rooms has been reserved for workshop attendees until May 31. The special workshop rate of 80 euros/night for a single room includes a full breakfast buffet, and is available for those presenting an IIR or IIR-related work that will be used to extend their stay. (This work has abstracts are accepted to present their presentation directly with the IIRs on the IIRs association website that will be used to extend their stay.)

P. Mäyllymäki
03.05.2006

...perhaps not quite?



hotel lancaster amsterdam - Google-haku

http://www.google.fi/search?q=hotel+lancaster+amsterdam&start=0&ie=utf-8&oe=utf-8&client=firefox-a&rlz=org.mozilla.en-US-official

Google

Haku: kakilla web-sivulla / suomenkielisiä sivuja / sivuja maasta: Suomi

Tulokset 1 - 10 noin 1 910 000 osuman joukosta haille **hotel lancaster amsterdam**. (0,28 sekuntia)

Lancaster Hotel Amsterdam
www.hotelsRus.com Up to 60% off the Lancaster Hotel in Amsterdam. Breakfast Included

250 Hotellia - Amsterdam
www.bookings.nl Varaa täällä, säästät jopa 75%! Sivusto englanniksi.

BW Lancaster Hotel Amsterdam DISCOUNTS! on Hotels.nl x Amsterdam ...
BW Lancaster Hotel Amsterdam 3* Plantage Middenlaan 48 Amsterdam (City centre, next to Artis Zoo). The Lancaster Hotel is located on the Plantage ...
www.hotels.nl/amsterdam/lancaster - 22k - Viikossa - Samankaltaisia sivuja

BW Lancaster Hotel Amsterdam Angebote auf Hotels.nl x Hotelführer ...
BW Lancaster Hotel Amsterdam 3* Plantage Middenlaan 48 Amsterdam (Zentrum, neben dem Tierpark Artis). Gegenüber dem Artis Zoo und neben dem Westpark ...
www.hotels.nl/de/amsterdam/lancaster - 22k - Viikossa - Samankaltaisia sivuja

Best Western Lancaster Hotel - Amsterdam - Best Western Lancaster ...
Best Western Lancaster Hotel, Amsterdam: Visit TripAdvisor to find the best unbiased reviews, candid photos, and deals for Best Western Lancaster Hotel in ...
www.tripadvisor.com/Hotel_Review-g188590-d229170-Reviews-Best-Western_Lancaster_Hotel-Amsterdam_Noord_Hol... - 70k - Viikossa - Samankaltaisia sivuja

Best Western Lancaster Hotel - Amsterdam - Best Western Lancaster ...
El Hotel Lancaster está situado en Plantage Middenlaan, justo enfrente del zoo Artis, una de las zonas más verdes de Amsterdam. Cercano a las principales ...
www.bookings.nl/hotel/lancaster.es.html - 114k - Viikossa - Samankaltaisia sivuja

Amsterdam Hotels - Best Western Lancaster Amsterdam
Best Western Lancaster Amsterdam - Amsterdam The Lancaster Hotel is located on the Plantage Middenlaan, opposite the Artis Zoo, in the greenest part of ...
www.amsterdam.info/hotels/best-western_lancaster_amsterdam - 10k - Viikossa - Samankaltaisia sivuja

Bestwestern.com, the World's Largest Hotel Chain
Plantage Middenlaan 48, Amsterdam, The Netherlands, NL-1018DH, NL ... The Best Western Lancaster Hotel is situated in a tranquil and green residential area ...
book.bestwestern.com/bestwestern/productinfo.do?propertyCode=92659 - 67k - Viikossa - Samankaltaisia sivuja

Hotels Amsterdam, Hotel Best Western Lancaster, Hotels Holland ...
Hotels Amsterdam, Hotel Best Western Lancaster, Hotels Holland, Hotels The Netherlands. www.hotels-holland.com/amsterdam/lancaster.htm - 8k - Viikossa - Samankaltaisia sivuja

250 Hotels in Amsterdam
Book your Amsterdam hotel online. Good availability and great rates! www.booking.com

Amsterdam hotel
Hytel hotellit hotelluoneita. Varaa nyt ja säästät jopa 70%! www.hotels.com

Lancaster Hotel Amsterdam
Great Deals on Hotels in Amsterdam. Book Now Online for Huge Savings! www.Amsterdam-Bookings.com

Amsterdam Hotels 75% off
Up to 75% discount on selected hotels on www.hotels.nl www.hotels.nl

Amsterdam Hotels direct
Free online hotel reservations here representing most Amsterdam Hotels www.HotelAdvice.com

Hotels in Amsterdam from €35
New - Book directly at the hotel! For better rates and great service www.hotellers.com

Hotel Guide Amsterdam & Holland
All Hotels Up to 75 % Discount www.hollanddriveleweb.com

Hotel Amsterdam
Welcome bonus 50 Euro Book your room now with us! www.flashbooking.com

Lisää sponsorien linkkejä »

P. Mäyllymäki
03.05.2006

Problems

- Information explosion
 - The world produces 250MB of information every year, for every man, woman and child on Earth
 - 85% of the information is unstructured
 - people spend 30% of their time in searching for relevant information
- In the future the information is (probably) out there, **the key problem is how to find it**
 - **intelligent information access**
- Problems with the current approaches
 - Technological problems
 - Ethical and political problems

P. Myllymäki
03.05.2006

Technological problems

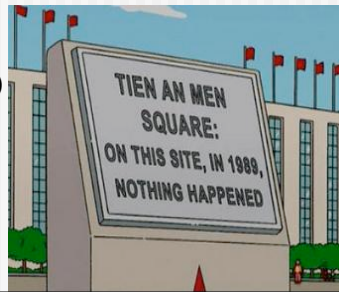
- Do I really want millions of answers to a simple query?
- More information in → less information out
- Popular documents vs. rare documents
- Moving away from “standard Internet”
 - new weblogs 40 000 / day, blogosphere doubles every 5 months
 - photos, music, video, ...
 - mobile devices
 - 2Bn mobile phones
 - GPS data, context, and other “sensor” data



P. Myllymäki
03.05.2006

Ethical and political problems

- Two US companies, Google and Yahoo, totally dominate the search market of today
- USA vs. Europe
 - Google as the “GPS” of the Internet
- Proprietary vs. open software
 - Google as the “Microsoft” of the Internet
- “Don’t be evil”
 - Google China
 - Google’s “death penalty” (BMW Germany)
- Language issues
 - Quearo: French-German initiative
 - Alvis EU project: www.alvis.info
 - Aino (aino.hiit.fi): a Finnish search engine
- Cultural issues



P. Mylly
03.05.2006

The challenge

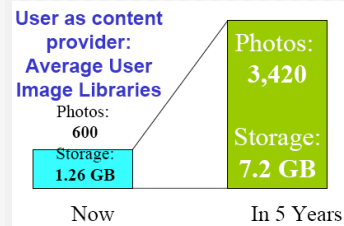
- “Search will eventually be as good as having 1000 human experts who know your tastes scanning billions of documents within a split second” (Gary Flake, Microsoft)



P. Myllymäki
03.05.2006

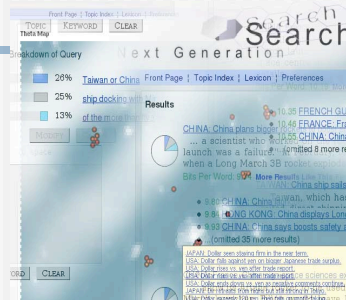
Semantic Web?

- Forcing structure in the unstructured world
- A distributed effort
- Problems:
 - global compliance
 - manual tagging work effort
 - limitations of logic-based search / document matching



Statistical approaches

- *Content-based search*
- Requires (statistical) modeling of the content
- State of the art:
 - sophisticated probabilistic “topic” models (discrete PCA)
 - simplifying assumptions (bag of words)
 - www.opensourcesearch.org
- Next step: audio-visual search



Intelligent Information Access

Collaborative search



Localized search



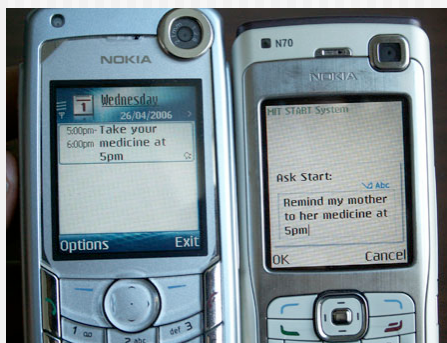
Personalized search



Topic-specific search

P. Myllymäki
03.05.2006

Natural language input



Using the MobileStart system on the phone on the right, you can remind your mother to take her medication. The phone on the left shows a new calendar event created in "mom's" phone by MobileStart. (Courtesy of Boris Katz and Federico Mora, MIT.)

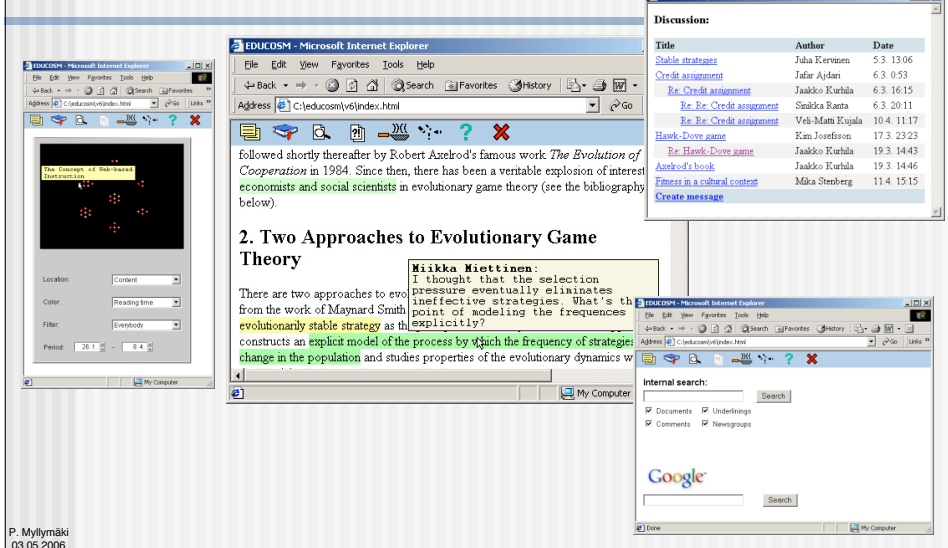
P. Myllymäki
03.05.2006

Implicit input

- Prima: Proactive Information Retrieval by Adaptive Models of Users' Attention and Interests
- Information retrieval enhanced with relevance data and context-sensitive information
- Eye tracking + probabilistic modeling

Kalle Palanderin kisat alkoivat täydellisesti
 Variksenpelottimella ja kymmenhännänsä.
 Kalle Palander palasi armeijan harnaisiin
 Vuosittainen runonpöytästä Washingtonin linjat
 David Beckhamista tuli miesten roolimalli
 Laskia Harkimo keskittyy hyvällä mielellä toisiin
 Herra Herala esiintyi paljastavassa vartolomaalissa
 Kalle Palander karaistui voittajaksi
 Kalle Palander yritti immaa
 Heikkinen sukasi tuonut jassissa ruukassa
 Suomalaiset rikkarisivät Itäkarasinnassa
 Suomalaiset osinkonsa

Collaboration & personalization



The screenshot displays the EDUCOSM web application interface, which includes a game simulation on the left, a discussion forum in the center, and a search bar at the bottom right.

Game Simulation (Left): A small window titled "The Concepts of the Star-based Evolutionary Game" showing a simulation of a game with various parameters like Location, Color, Filter, and Period.

Discussion Forum (Center): A section titled "2. Two Approaches to Evolutionary Game Theory" with a sub-section "Theory". It contains a paragraph about Maynard Smith's work and a quote from Mikko Miettinen. Below the text is a "Create message" button.

Discussion Table (Right): A table listing discussion topics, authors, and dates.

Title	Author	Date
Stable strategies	Juha Kervinen	5.3.13.06
Credit assignment	Jafar Ajdari	6.3.0.53
Re: Credit assignment	Jaakko Kurhila	6.3.16.15
Re: Re: Credit assignment	Sinikka Ranta	6.3.20.11
Re: Re: Credit assignment	Veli-Matti Kujala	10.4.11.17
Hawk-Dove game	Kari Josefsson	17.3.23.23
Re: Hawk-Dove game	Jaakko Kurhila	19.3.14.43
Axelrod's book	Jaakko Kurhila	19.3.14.46
Pluses in a cultural context	Mika Stenberg	11.4.15.15

Search Bar (Bottom Right): A Google search bar with a "Search" button.

Summary

- **Control**
 - who controls information access, controls the information
 - distributed vs. centralized search
- **Multimedia search**
 - images, audio, video, 3D,...
- **Content-based search**
 - semantic web vs. machine intelligence
- **Context-sensitive search**
 - who? where? when? why?
 - adaptation / personalization
- **Novel interfaces for interaction**
 - natural language (text/speech), gestures, eye movements, ...
 - implicit feedback vs. explicit input