

Intelligent Systems

Petri Myllymäki
Complex Systems Computation Group
Department of Computer Science
University of Helsinki, Finland

<http://www.cs.helsinki.fi/petri.myllymaki/>

Intelligence?

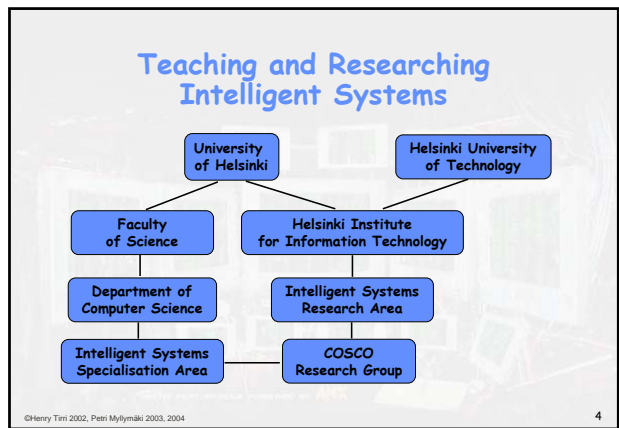
- Computer science is considering ways to **make things automated**
- Intelligent Systems studies the question **how can intelligence be automated?**
- Intelligent **behavior** can be characterized by **learning/adaptation**
- In order to learn one needs to **remember** and **generalize** (world is never repeating itself!)
- In order to generalize one needs **models**

©Henry Tiri 2002, Petri Myllymäki 2003, 2004

Name of the game?

- Artificial intelligence
- Science of uncertainty
- Adaptive and intelligent systems
- Computational intelligence
- Soft computing
- Real-world computing
- Complex systems computation
- Deep computing
- ...
- **Keywords:** knowledge representation, reasoning logic, expert systems, machine learning, data mining, Bayesian networks, neural networks, fuzzy systems, evolutionary computing, artificial life, robotics, planning, optimization...

©Henry Tiri 2002, Petri Myllymäki 2003, 2004



<http://www.cs.helsinki.fi/alykkaat/>

Main courses:

- Artificial Intelligence
- Three concepts: probability
- Three concepts: information
- Three concepts: utility
- Graphical models

Related courses:

- Machine learning
- Data mining

©Henry Tiri 2002, Petri Myllymäki 2003, 2004

Complex Systems Computation Group CoSCo

- Hannes Wietzig presents the paper "Bayesian Analysis of Online Newspaper Log Data" at the 2003 International Symposium on Applications and the Internet (Orlando, USA, January 2003) → [20.01.2003](#)
- Professor Henry Tiri gives talk "Adaptive modeling of Internet Users" at the concluding seminar of USIX technology programme → [15.01.2003](#)
- Mikko Miettinen gives talk "User Profiling and Visualization in new-based learning environments" at the concluding seminar of USIX technology programme → [15.01.2003](#)
- Professor Henry Tiri gives invited talk "Computational World" at Europe 2003 seminar, arranged by Finnish Society for Information Services
- Professor Henry Tiri gives an invited talk "Personalized Adaptive Interfaces" at Media and Everyday Life seminar arranged by StoreEnso and AlmaMedia
- Professor Henry Tiri gives talk "Individual theory - Mathematics of Personalization in Computer Networks" at Finnish Society → [12.01.2003](#)

©Henry Tiri 2002, Petri Myllymäki 2003, 2004

Our mission

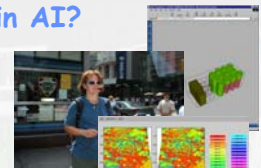
- Develop intelligent solutions to hard problems found in realistic settings
- Requires both deep theoretical work and creative methodological innovations
- Complex Systems Computation Group:
 - Petri Myllymäki, Henry Tirri
 - Wray Buntine, Jorma Rissanen, Huizhen Yu (MIT)
 - around 20 researchers, mostly quite experienced
 - a rare combination of theoretical competence with top-level programming skills
 - active collaboration with
 - » UCB (Michael Jordan)
 - » CWI (Paul Vitányi, Peter Grünwald)
 - » Tsinghua University (Lizhu Zhou)



©Henry Tirri 2002, Petri Myllymäki 2003, 2004

Student: So what is it like to do research in AI?

- Great fun!
- Some **hacking**
- some **math**
- Some **gaming**
- Some **technology "freaking"**
- Some **science fiction**



©Henry Tirri 2002, Petri Myllymäki 2003, 2004

8

"On the shoulders of Giants"



Andrey Nikolaevich Kolmogorov



Rev. Thomas Bayes

©Henry Tirri 2002, Petri Myllymäki 2003, 2004

9

CoSCo Research areas

- Probabilistic and information-theoretic modeling in sciences and business
 - Information-theoretic modeling approaches
 - Bayesian and Causal Networks
 - Finding the position of mobile devices
 - Personalization
 - Next generation information search
 - Tools and theory for E-learning
- Stochastic optimization in complex domains

©Henry Tirri 2002, Petri Myllymäki 2003, 2004

10

mdl-research.org

Minimum Description Length on the Web



©Henry Tirri 2002, Petri Myllymäki 2003, 2004

11

Bayes vs. MDL

Under regularity conditions $-\log P_{NML}(x^n | M) =$
 $-\log P(x^n | \hat{\theta}_i(x^n)) + \frac{k}{2} \log \frac{n}{2\pi} + \log \int \sqrt{\det I(\theta)} d\theta + o(1)$

Under regularity conditions $-\log P_{Bayes}(x^n | M) \approx$
 $-\log P(x^n | \hat{\theta}_i(x^n)) + \frac{k}{2} \log \frac{n}{2\pi} - \log w(\hat{\theta}) + \log \sqrt{\det I(\hat{\theta})} + o(1)$

If we take Jeffrey's prior

$$w(\theta) = \frac{\sqrt{\det I(\theta)}}{\int \sqrt{\det I(\theta)} d\theta} \dots \odot$$

©Henry Tirri 2002, Petri Myllymäki 2003, 2004

12

B-course Data Analysis Server (<http://b-course.hiit.fi>)



©Henry Tiri 2002, Petri Myllymäki 2003, 2004

13

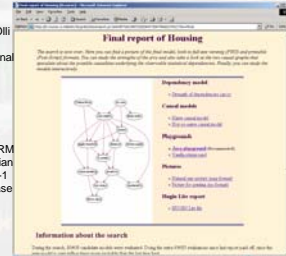
B-course Data Analysis Server (<http://b-course.hiit.fi>)

- commercial-quality ASP tool for research and teaching purposes
- 10 000 users world-wide in 2 years

Sampa Hautaniemi, Henrik Edgren, Petri Vesanen, Majja Wolf, Anna-Kaarina Jarvinen, Olli Yli-Harja, Jaakko Astola, Olli Kallioniemi and Outi Monni, A Novel Strategy for Microarray Quality Control Using Bayesian Networks. To appear in Journal of Bioinformatics.

Jarvis, Smith, Wada, Rivas, McElroy, Smulders, Caminci, Hayashizaki, Dietrich, Wu, McConnell, Yu, Wan, Hartemink, Lin, A framework for integrating the songbird brain. J Comp Physiol A (2002) 188: 961-980.

K DeForche, K Van Laethem, A Abecasis, J Snoeck, AP Carvalho, I Derdelincx, P Gomes, J Cabanas, MA Soares, RM Brindeiro, A Tanuri, R Carnacho and AM Vandanime, Bayesian Network Reveals Linkage for Mutations at Position 89 of HIV-1 Protease to other Protease Codons and Therapy with Protease Inhibitors. Proceedings of the 2nd European HIV Drug Resistance Workshop, March 11th-13th, 2004, Rome.



©Henry Tiri 2002, Petri Myllymäki 2003, 2004

Quick Summary

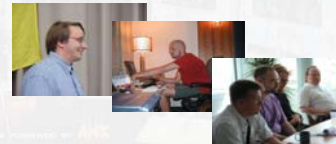
- Designed for dependency analysis with graphical models
- ASP architecture (works with most browsers)
- inference of Bayesian networks (and elementary causal networks)
- "tutorial style" user interface
- no user modifiable parameters
- interactive tool for inference
- extendible platform (v 2.0 classification)

©Henry Tiri 2002, Petri Myllymäki 2003, 2004

15

So what "science" was needed?

- Theory, heavy theory
- Empirical work with data sets ...
- Multidisciplinary co-operation
- Brilliant hacking (B-course had predecessors: D-Side, BAYDA)

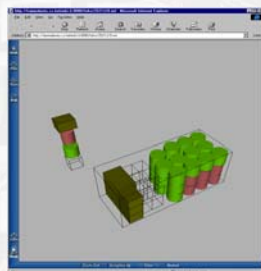


©Henry Tiri 2002, Petri Myllymäki 2003, 2004

16

Koptimi

- Intelligent algorithms for constrained bin-packing
- In fielded use at StoraEnso since 2000



©Henry Tiri 2002, Petri Myllymäki 2003, 2004

17

BayMiner™

- Commercial ASP tool for data analysis and visualization
- HS Vaalikone



©Henry Tiri 2002, Petri Myllymäki 2003, 2004

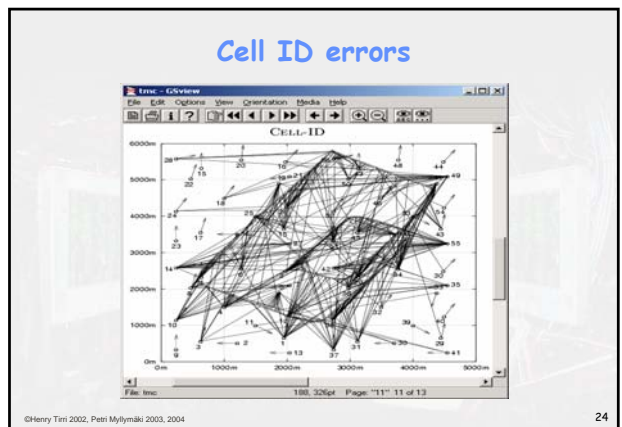
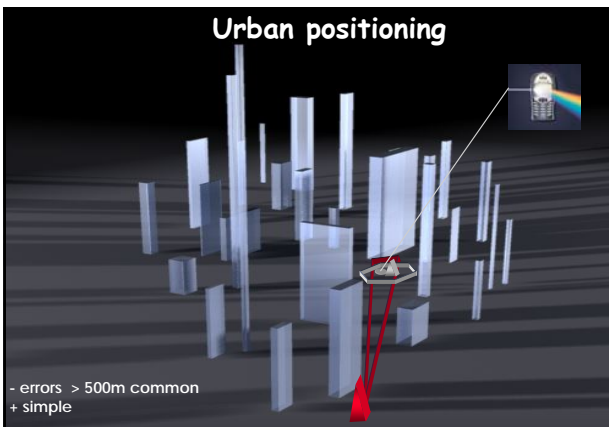
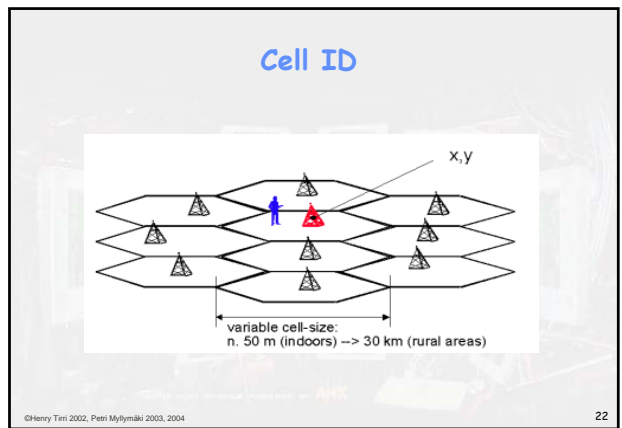
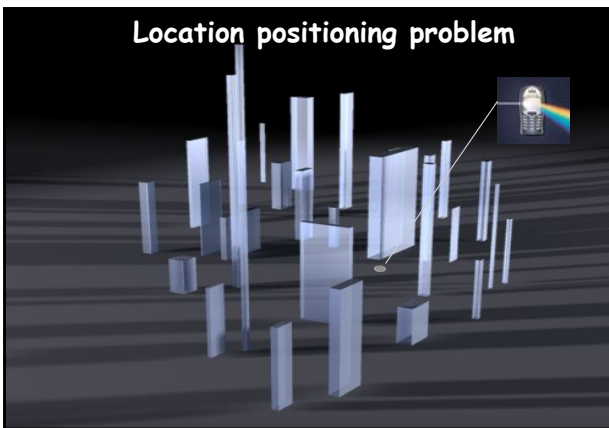
18

Prima: Proactive Information Retrieval by Adaptive Models of Users' Attention and Interests

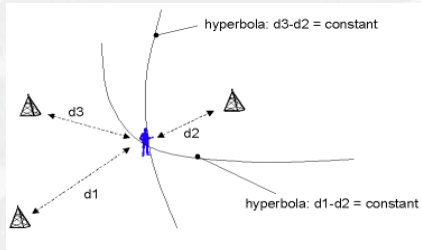
- Basic research project funded by the Academy of Finland (2003-2005)
- Joint work with the Neural Networks Research Centre (NNRC) and Center for Knowledge and Innovation Research (CKIR)
- Information retrieval enhanced with relevance data and context-sensitive information

(Note: The slide contains a list of Finnish text items, many of which are crossed out with red lines, including names like Kalle, David, and Nokia.)

©Henry Tiri 2002, Petri Myllymäki 2003, 2004 19



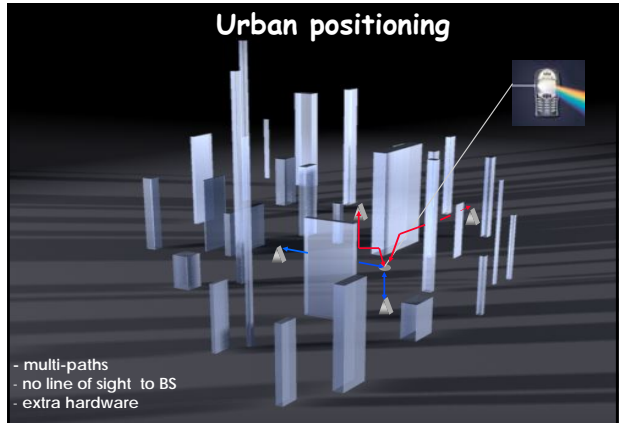
Enhanced Observed Time Difference (E-OTD)



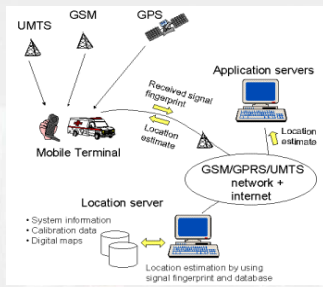
©Henry Tiri 2002, Petri Myllymäki 2003, 2004

25

Urban positioning



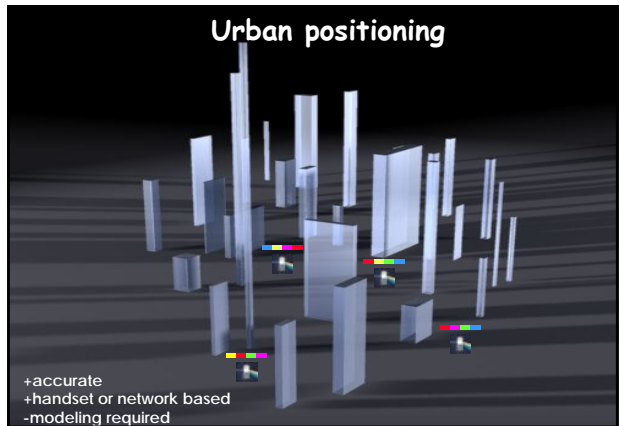
Modeling approach



©Henry Tiri 2002, Petri Myllymäki 2003, 2004

27

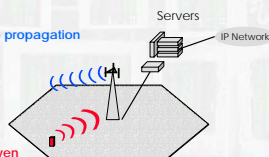
Urban positioning



Modeling alternatives

"Downlink modeling: the propagation models (GM level 1)"

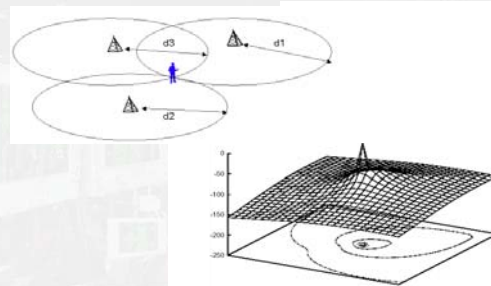
"Uplink modeling: data driven models (GM level 2)"



©Henry Tiri 2002, Petri Myllymäki 2003, 2004

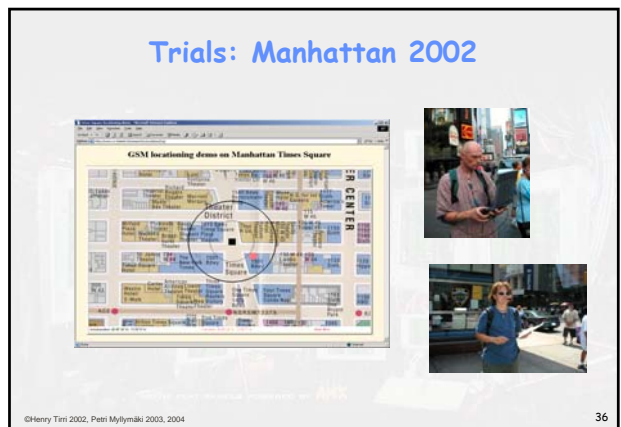
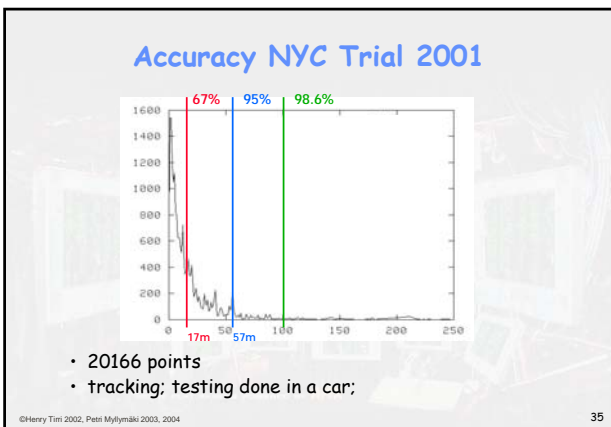
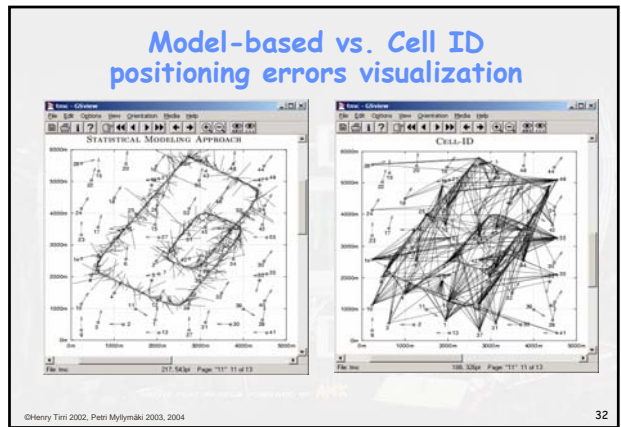
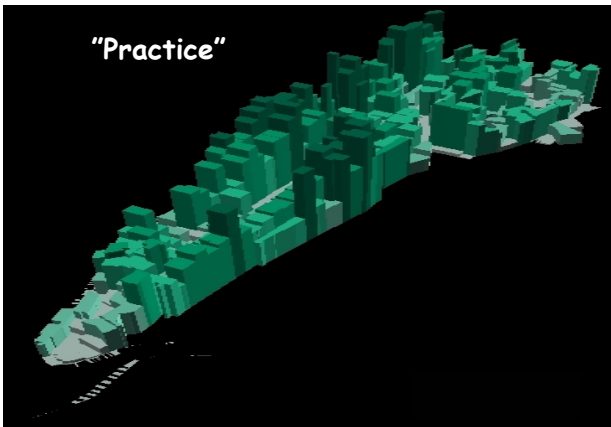
29

"Theory"

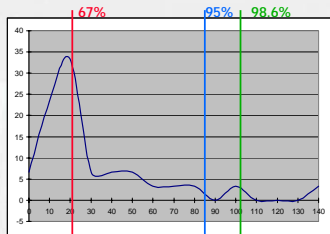


©Henry Tiri 2002, Petri Myllymäki 2003, 2004

30



Accuracy NYC Trial 2002

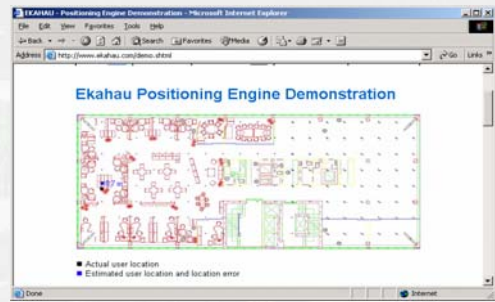


- 30 points
- static; testing done by walking;

©Henry Tiri 2002, Petri Myllymäki 2003, 2004

37

Generalization: WLAN



©Henry Tiri 2002, Petri Myllymäki 2003, 2004

38

Ekahau Positioning Engine™

- Software for locating devices in Wi-Fi networks
 - European Union: The European Information Society Technology Prize 2002.
 - Technology Marketing Corporation (TMC): Best product of the year 2002.
 - Planet PDA, the Global Summit on Enterprise & Custom Volume Handheld Computing: Best of show.
 - Software Industry Summit: Best commercialized innovation in Finland in 2002.
 - SearchNetworking.com: Bronze medal, best product of the year 2003 (enterprise wireless applications and systems).
 - Wi-Fi Planet 2004: Best of Show.



©Henry Tiri 2002, Petri Myllymäki 2003, 2004

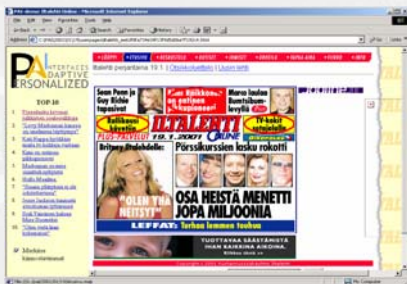
39

Personal WEB

Personalization and Intelligent Search



Personalized, adaptive interfaces

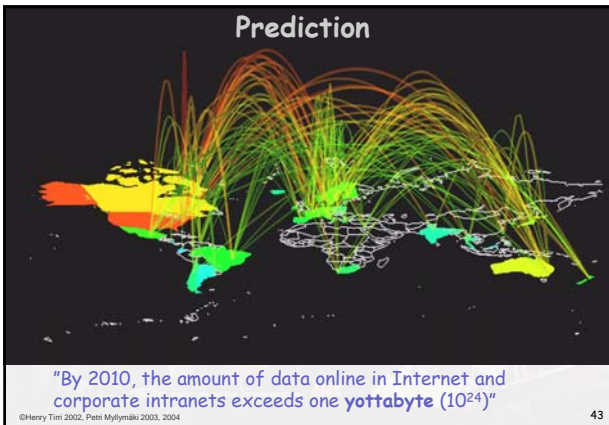


©Henry Tiri 2002, Petri Myllymäki 2003, 2004

41

Future of Search



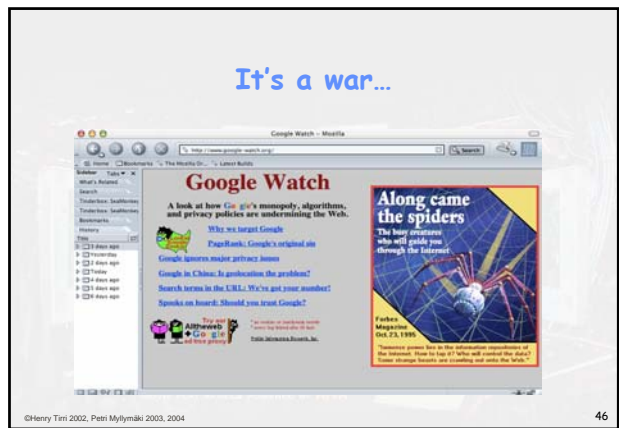


Knowledge workers spend 35% of their productive time searching for information online, while 40% of the corporate users report they cannot find the information they need to do their jobs on the intranet.

--Working council of CIOs, Business Wire, February 2001

©Henry Tiri 2002, Petri Myllymäki 2003, 2004 44

- ### Problems with Google
- Google's immortal cookie (2038)
 - Google records all they can
 - Google retains all data indefinitely
 - Google ignores privacy policy questions
 - Google hires spooks (NSA)
 - Google's toolbar is spyware
 - Google's cache is problematic
 - Google's database as a data mining resource (150 million searches/day)
- ©Henry Tiri 2002, Petri Myllymäki 2003, 2004 45



Next Generation Information Management Research

Collaborative search

Search-ina-Box

YDIN

Mobile search

Topic-specific search

©Henry Tiri 2002, Petri Myllymäki 2003, 2004 47

Emerging key technologies

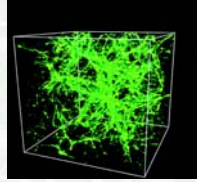
"Chelsea Football Club",
"Clinton DeLeon" and
"Chelsea Clinton"

- language processing
- (probabilistic) statistical modeling
- personalization
- collaborative filtering

©Henry Tiri 2002, Petri Myllymäki 2003, 2004 48

What is NOT there

- no major search engine performs any sophisticated language processing (scalability)
- no open source engines provide language processing
- non-interfering and (search) integrated personalization
- no quality corporate-wide intranet search systems (several sites)



©Henry Tiri 2002, Petri Myllymäki 2003, 2004

49

Related projects

Scalable Probabilistic Methods for the Next Generation Search Engine (PROSE)



Search-In-a-BOX



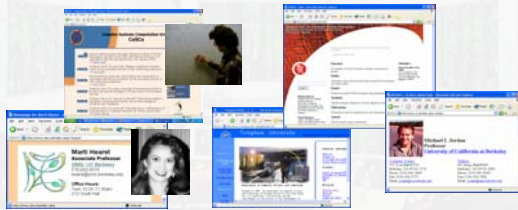
ALVIS (FP6 STREP)

©Henry Tiri 2002, Petri Myllymäki 2003, 2004

50

SIB research core competencies

- probabilistic modeling for embedded language models, automatic content analysis and personalization
- advanced user interface design



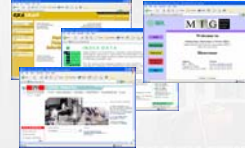
©Henry Tiri 2002, Petri Myllymäki 2003, 2004

51

Superpeer Semantic Search (ALVIS)



ALVIS (FP6 STREP)



- HIIT
- MIG-INRA (France)
- EPFL (Switzerland)
- Lund University (Sweden)
- JSI (Slovenia)
- Index Data (Denmark)
- Exalead (France)
- DTV (Denmark)
- Université Paris Nord

©Henry Tiri 2002, Petri Myllymäki 2003, 2004

52

モウラチェウ search

- Open Source
- Advanced language model (hidden from user)
- Probabilistic query models
- Integrated personalization
- Superpeer architecture



©Henry Tiri 2002, Petri Myllymäki 2003, 2004

53

モウラチェウ search



©Henry Tiri 2002, Petri Myllymäki 2003, 2004

54



The New York Review of Books - Mozilla

search

Component '64' Summary

Component Number: 64
 Empirical Proposition: 0.00218
 Directed Proposition: 0.001187
 Expected No. Of Lexemes: 1289.36
 Expected No. Of Documents: inf

Typical Lexemes in Component (sorted by frequency):

Acme (0.000), Japanese (0.016), Buffy (0.016), series (0.014), episode (0.012), Vol (0.011), characters (0.011), Robinson (0.011), DVD (0.010), Ben (0.008), by (0.008), movie (0.007), animation (0.007), Sailor Moon (0.006), English (0.006), Jason (0.006), Kinnaman (0.006), Vampire Slayer (0.004), Oxy (0.004), manga (0.003), story (0.004), Fall (0.004), Gardner (0.004), Miyazaki (0.004), Tazaki (0.004), Ramona (0.003), Slayer (0.003), job (0.003), watch (0.003), set (0.003), base (0.003), Aina (0.003), Sisy (0.003), series (0.003), Sams (0.003), volume (0.003), Algea (0.003), Robinson (0.003), Collection (0.003), Chrono Star Z (0.003), Neil (0.003), New (0.003), cartoon (0.003), site (0.003), Gates (0.003), get (0.003), Video (0.002), TV (0.002), show (0.002), Gemini (0.002), original (0.002), database (0.002), seen (0.002), Algea (0.002), Robinson (0.002), word (0.002), Takahashi (0.002), and (0.002), Pioneer (0.002), Hsu (0.002), set (0.002), box (0.002), Oxy (0.002), Advances (0.002), color (0.002), Blood Sails (0.002), Showtime (0.002), Iron (0.002), error (0.002), action (0.002), Sisy (0.002), voice (0.002), Download (0.002), the (0.002), Scimitar Away (0.002), Puck (0.002), Corcoran (0.002), a (0.002), Neon Cinema Evangelion (0.002), and (0.002), Cowboy Bebop (0.002), Simpson (0.002), Linux (0.002), best (0.002), 1982 (0.002), Lost Time (0.002), watching (0.002), comic (0.002), adventure (0.002), information (0.002), Road (0.002), Dragooz (0.002), Escapade (0.002), collector (0.002), dub (0.002), Sailor (0.002), Super (0.001), Master (0.001), Charles (0.001), Brown (0.001), science (0.001), Earth (0.001), Iron (0.001), robot (0.001), features (0.001), B (0.001), suggests (0.001), Top (0.001), Baracus (0.001), network (0.001), Skips (0.001), version (0.001), Translucent (0.001), China (0.001), Maki (0.001), Vain (0.001), Genset (0.001), Seneca (0.001), Quaker King (0.001), cantina (0.001), chud (0.001), Trench Mags (0.001), Iron (0.001), Chud (0.001), Kix (0.001), ending (0.001), Processes, Monoclonal (0.001), 1152 (0.001), Kix (0.001), Ramon (0.001), saga (0.001), Desert (0.001), computer (0.001), Exclamation (0.001), Army (0.001), saga (0.001), creator (0.001), hat (0.001), here (0.001), Tolson (0.001), funding (0.001), Female (0.001), language (0.001).

©Henry Tiri 2002, Petri Myllymäki 2003, 2004 55

"Ubisearch"

©Henry Tiri 2002, Petri Myllymäki 2003, 2004 56

Search-Ina-Box concept

©Henry Tiri 2002, Petri Myllymäki 2003, 2004 57

http://cosco.hiit.fi/search/

Next Generation Information Search - We search Internet Lexemes

Overview

Research and Development in the Next Generation Search Group covers the following areas:

Modeling Principles

Search-Ina-Box

European Search Engine

©Henry Tiri 2002, Petri Myllymäki 2003, 2004 58

EDUCATIONAL TECHNOLOGY

Personalized Adaptive Interfaces

Intelligent Tools for E-learning

©Henry Tiri 2002, Petri Myllymäki 2003, 2004 59

EDUFORM

EDUFORM - Microsoft Internet Explorer

Learning Experiences and Motivation

©Henry Tiri 2002, Petri Myllymäki 2003, 2004 60

EDUCO

©Henry Tiri 2002, Petri Myllymäki 2003, 2004 61

OurWeb

File	Author	Date
Chalkboard	John Korman	2.3.2004
Chalkboard	Federico	4.3.2004
Chalkboard	Josko	4.3.2004
Chalkboard	John Korman	4.3.2004
Chalkboard	John Korman	4.3.2004
Chalkboard	John Korman	4.3.2004
Chalkboard	John Korman	4.3.2004
Chalkboard	John Korman	4.3.2004
Chalkboard	John Korman	4.3.2004
Chalkboard	John Korman	4.3.2004

©Henry Tiri 2002, Petri Myllymäki 2003, 2004 62

and other things 😊

©Henry Tiri 2002, Petri Myllymäki 2003, 2004 63