# 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

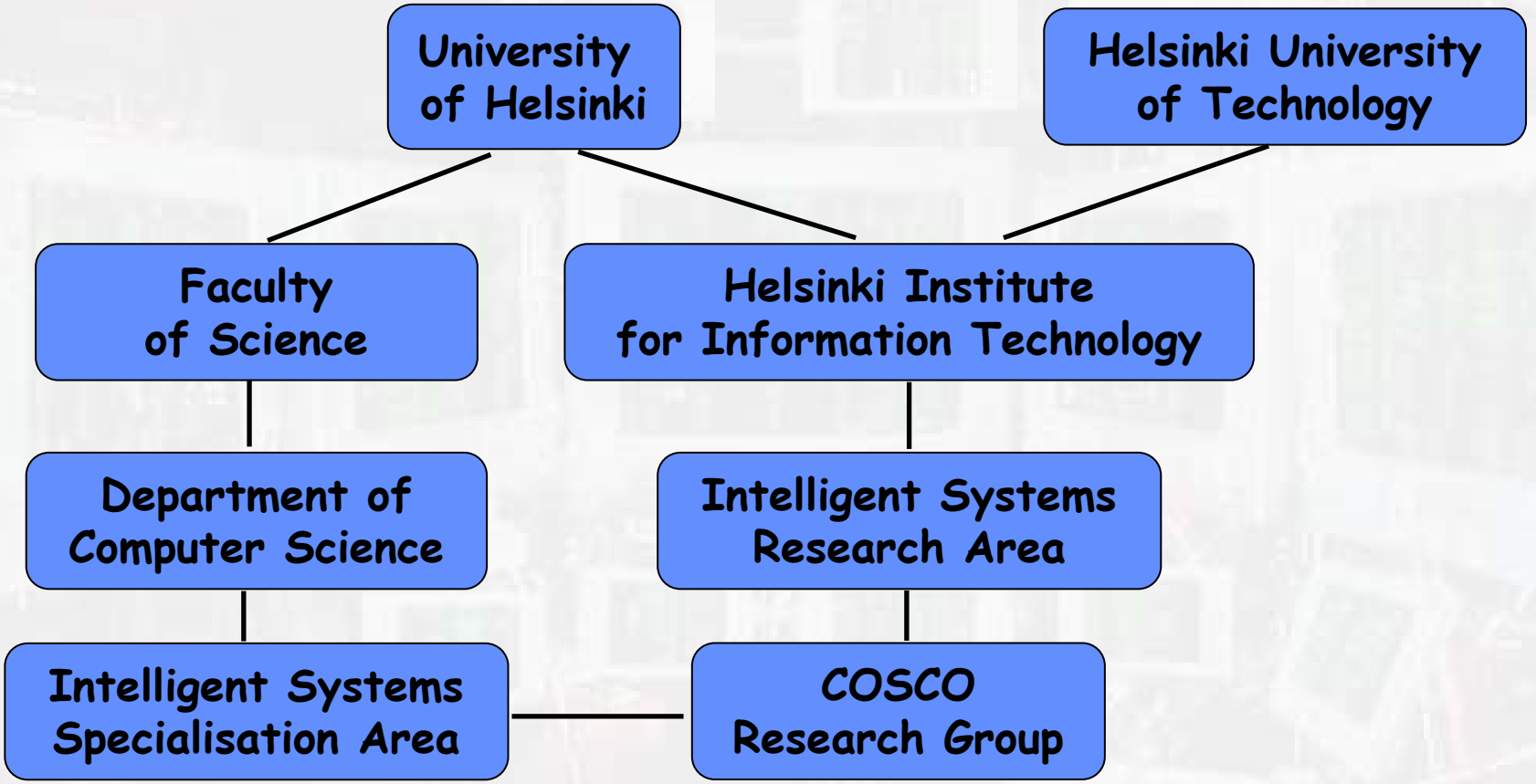
©Henry Tirri 2002, Petri Myllymäki 2003

# 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...

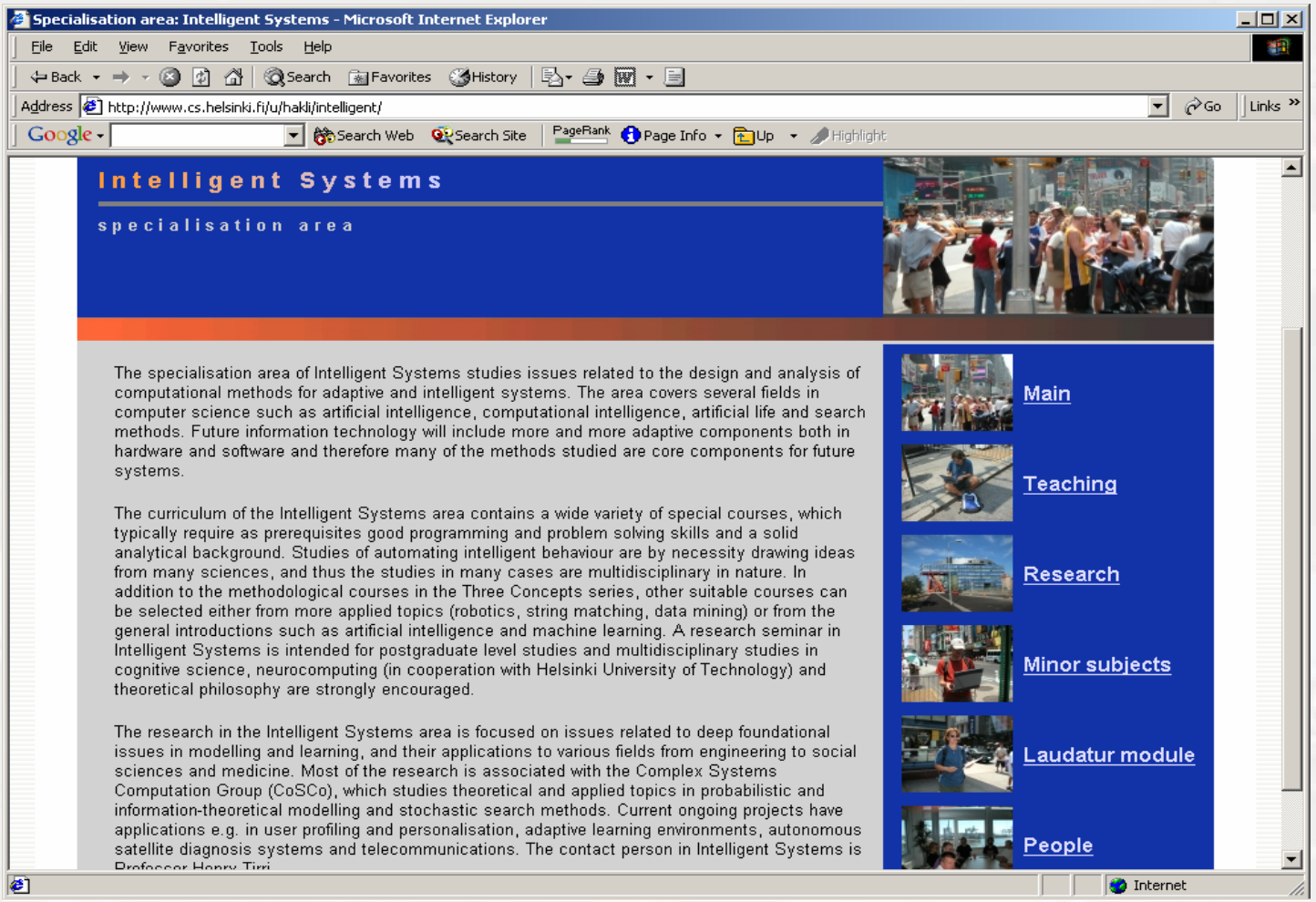
©Henry Tirri 2002, Petri Myllymäki 2003

# Teaching and Researching Intelligent Systems




©Henry Tirri 2002, Petri Myllymäki 2003

# http://www.cs.helsinki.fi/alykkaat/




©Henry Tirri 2002, Petri Myllymäki 2003

CoSCo - Complex Systems Computation Group - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://cosco.hiit.fi/
Go Links

# Complex Systems Computation Group
# CoSCo

27-31.1.2003 Hannes Wettig presents the paper "Bayesian Analysis of Online Newspaper Log Data" at the 2003 International Symposium on Applications and the Internet(Orlando, USA, January 2003).
→ SAINT 2003

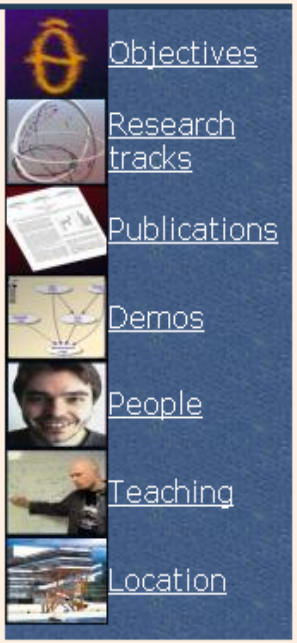
28.1.2003 Professor Henry Tirri gives talk "Adaptive modeling of Internet users" at the concluding seminar of USIX technology programme
→ USIX 2003

28.1.2003 Miikka Miettinen gives talk "User Profiling and Visualization in www-based learning environments" at the concluding seminar of USIX technology programme
→ USIX 2003

21.1.2003 Professor Henry Tirri gives invited talk "Computational World" at Europe 2020 seminar arranged by Finnish Society for Information Services

15.1.2003 Professor Henry Tirri gives an invited talk "Personalized Adaptive Interfaces" at Media and Everyday Life seminar arranged by StoraEnso and AlmaMedia

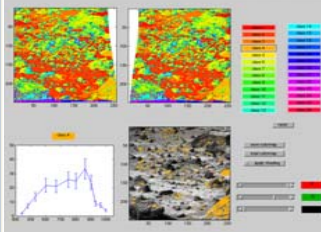
12.1.2003 Professor Henry Tirri gives talk "Individual theory - Mathematics of Personalization in Computer Networks" at Finnish Science Days
→ Tieteen päivät





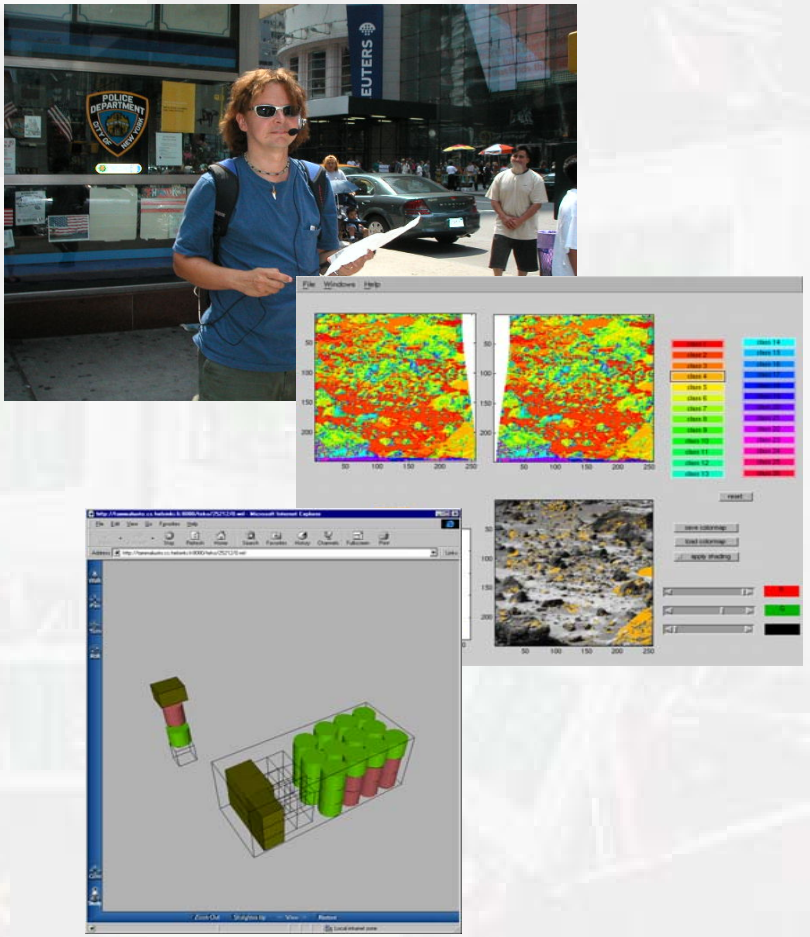
B-Course

Internet
start Tablet PC Tuto... presentations 4 Microsoft P... berkeley.doc - ... CoSCo - Compl... 7:59


# 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

# "On the shoulders of Giants"

**Andrey Nikolaevich Kolmogorov**

**Rev. Thomas Bayes**

©Henry Tirri 2002, Petri Myllymäki 2003

# 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

mdl-research.org
Minimum Description Length on the Web


©Henry Tirri 2002, Petri Myllymäki 2003

# Bayes vs. MDL

Under regularity conditions $-\log P_{NML}(x^n \mid M) =$

$$-\log P(x^n \mid \hat{\theta}_i(x^n)) + \frac{k}{2}\log\frac{n}{2\pi} + \log\int\sqrt{\det I(\theta)}d\theta + o(1)$$

Under regularity conditions $-\log P_{Bayes}(x^n \mid M) \approx$

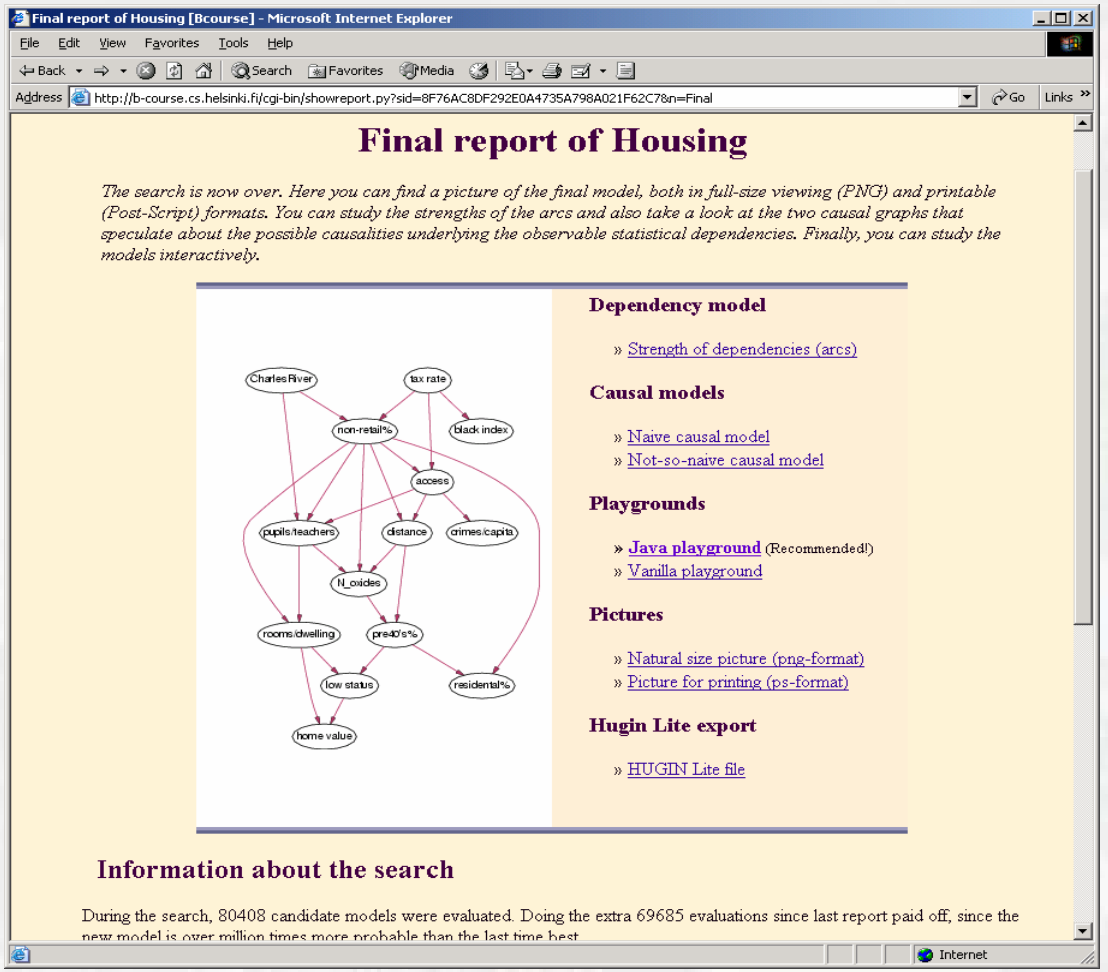
$$-\log P(x^n \mid \hat{\theta}_i(x^n)) + \frac{k}{2}\log\frac{n}{2\pi} - \log w(\hat{\theta}) + \log\sqrt{\det I(\theta)}d\theta + o(1)$$

If we take Jeffrey's prior

$$w(\theta) = \sqrt{\det I(\theta)} \Big/ \int_{\theta}\sqrt{\det I(\theta)}d\theta \quad \text{..... ☺}$$

©Henry Tirri 2002, Petri Myllymäki 2003

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




©Henry Tirri 2002, Petri Myllymäki 2003

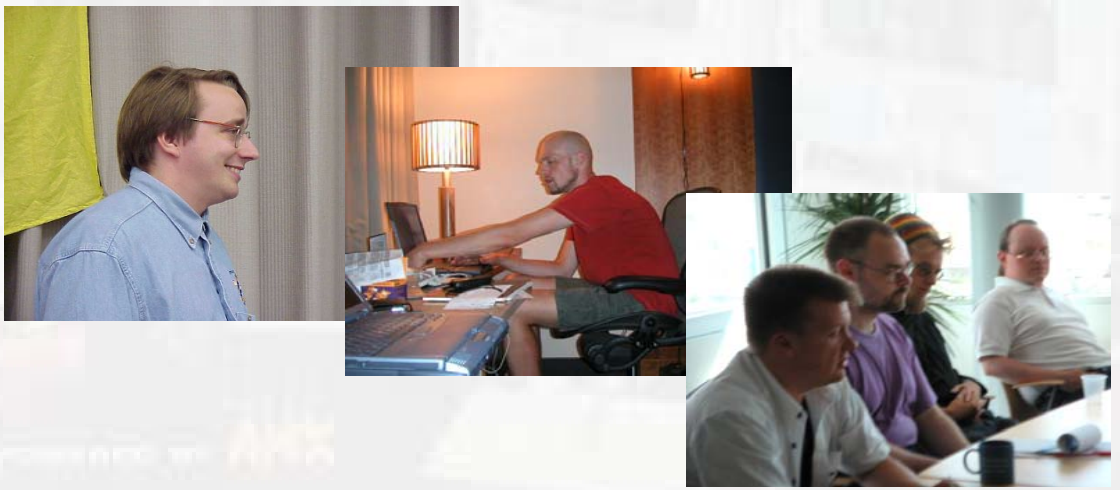
# Quick Summary

- 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 Tirri 2002, Petri Myllymäki 2003

# 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)

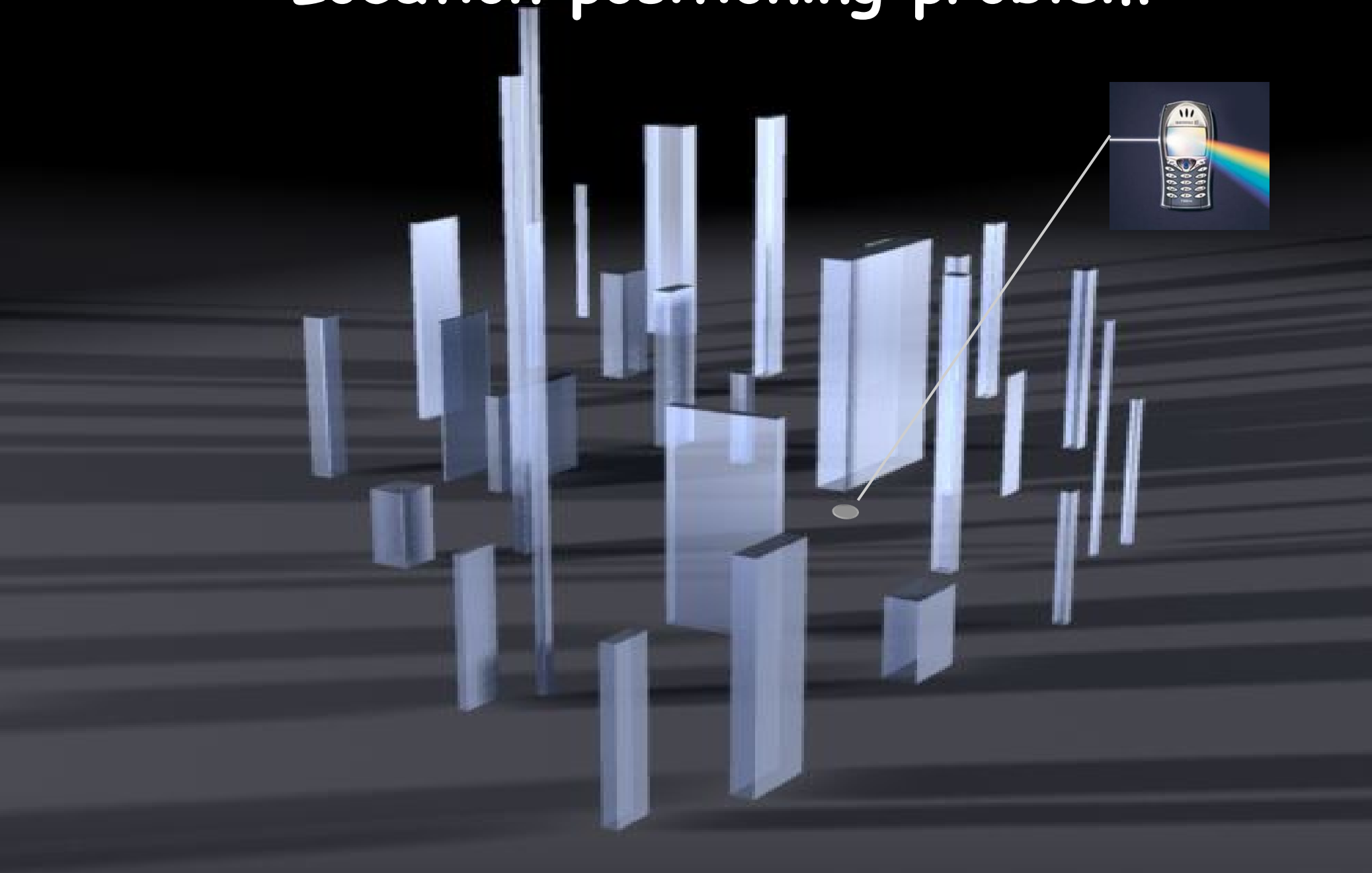


©Henry Tirri 2002, Petri Myllymäki 2003

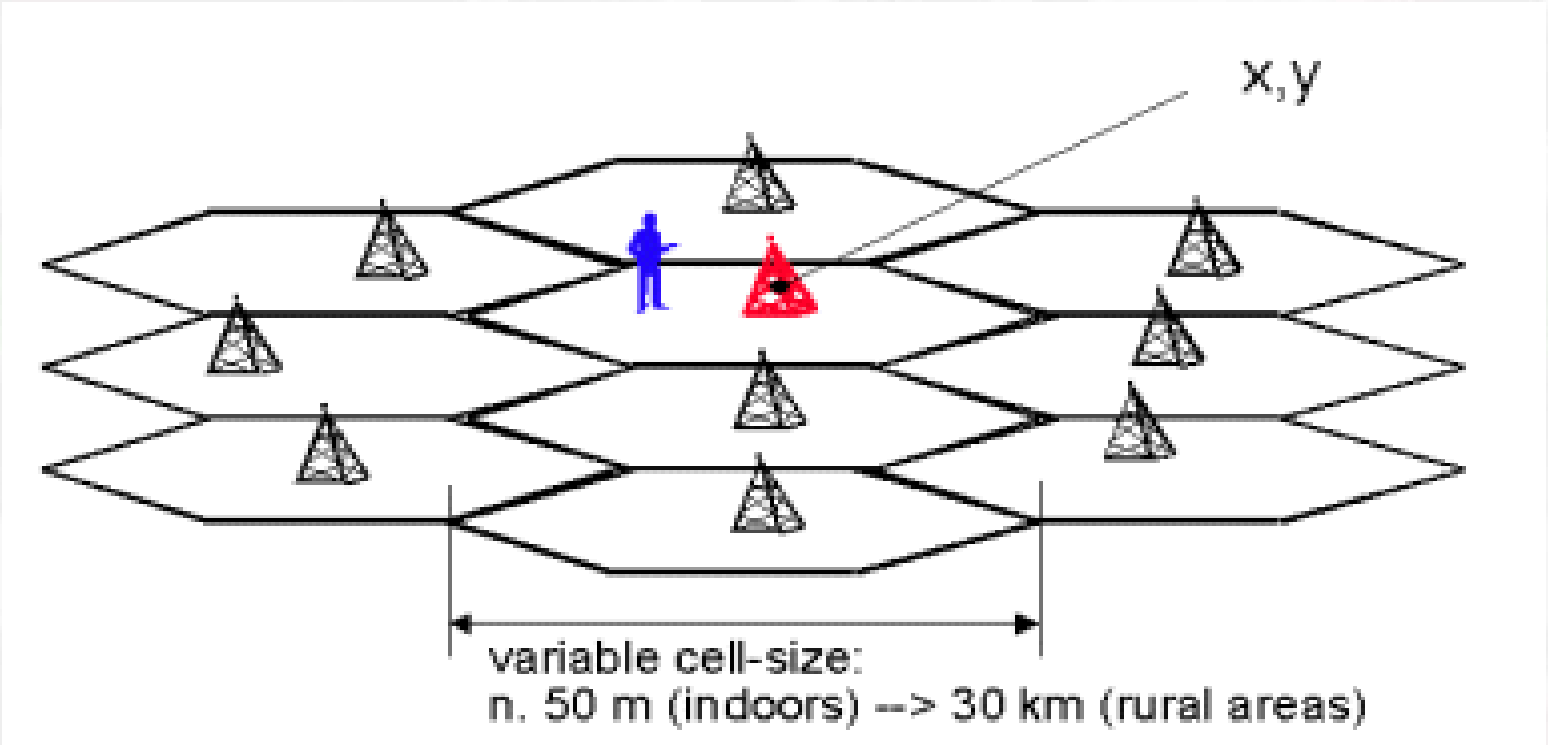

# Loclt

## Mobile Device Positioning

# Location positioning problem

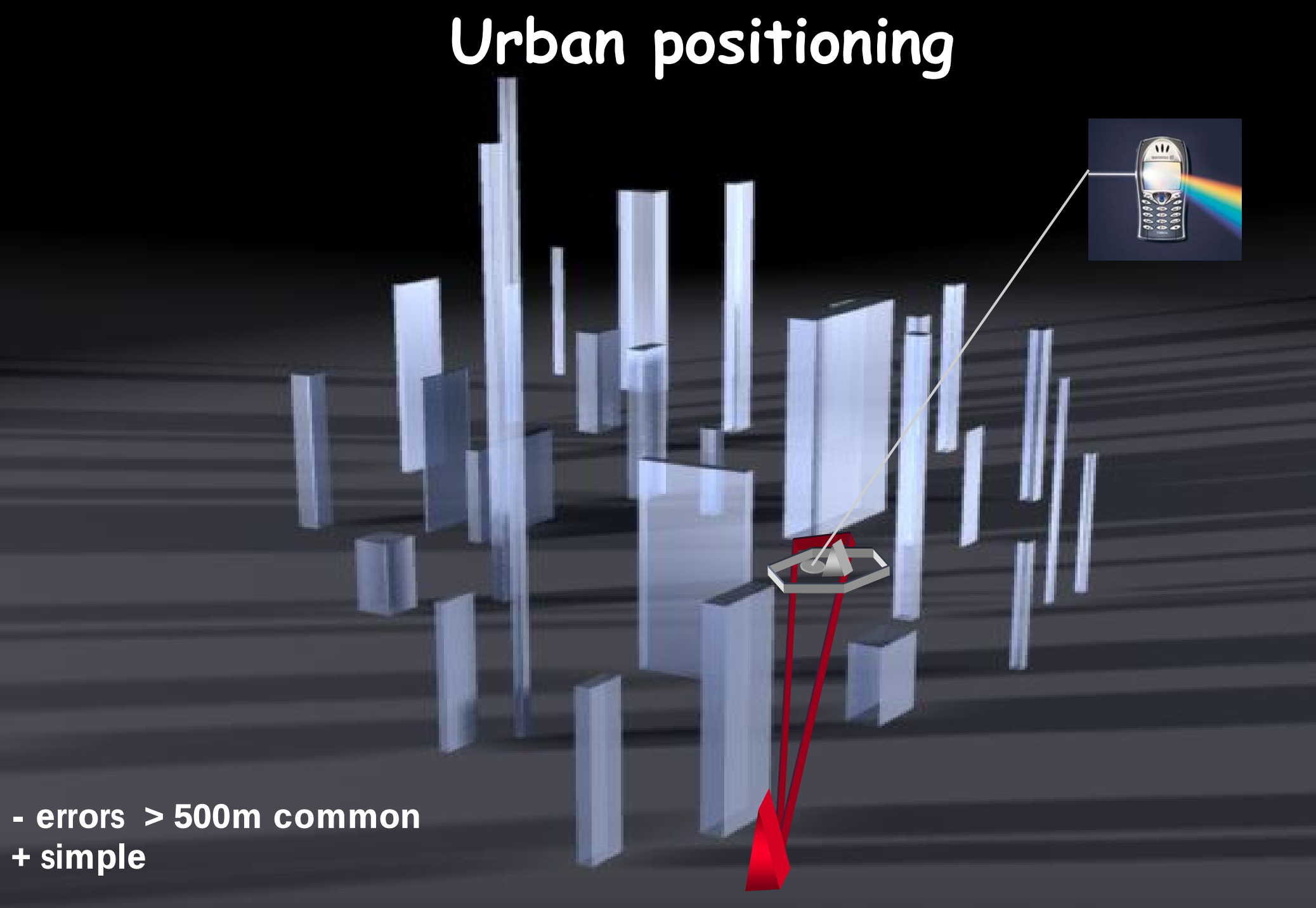


# Cell ID



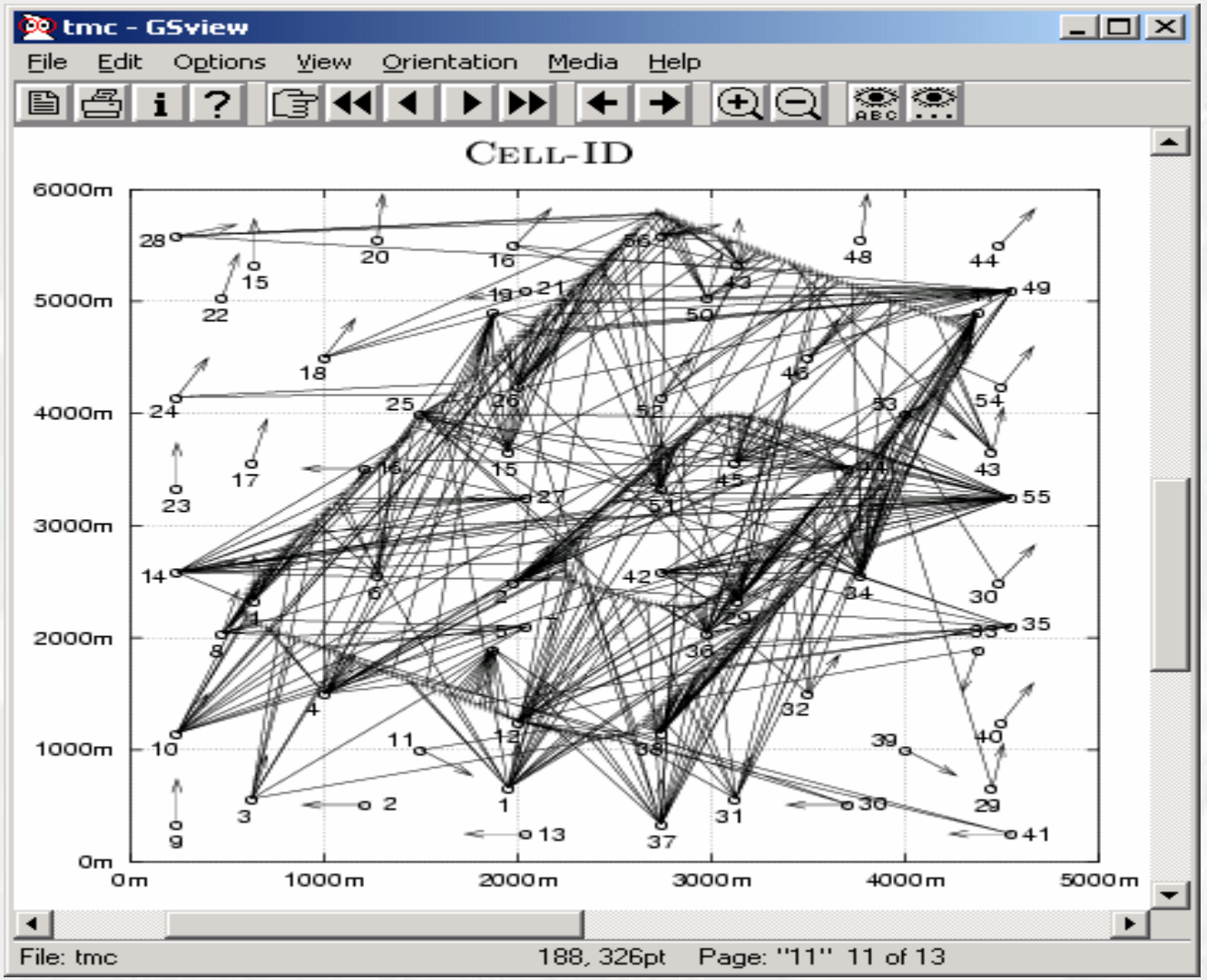

©Henry Tirri 2002, Petri Myllymäki 2003

# Urban positioning



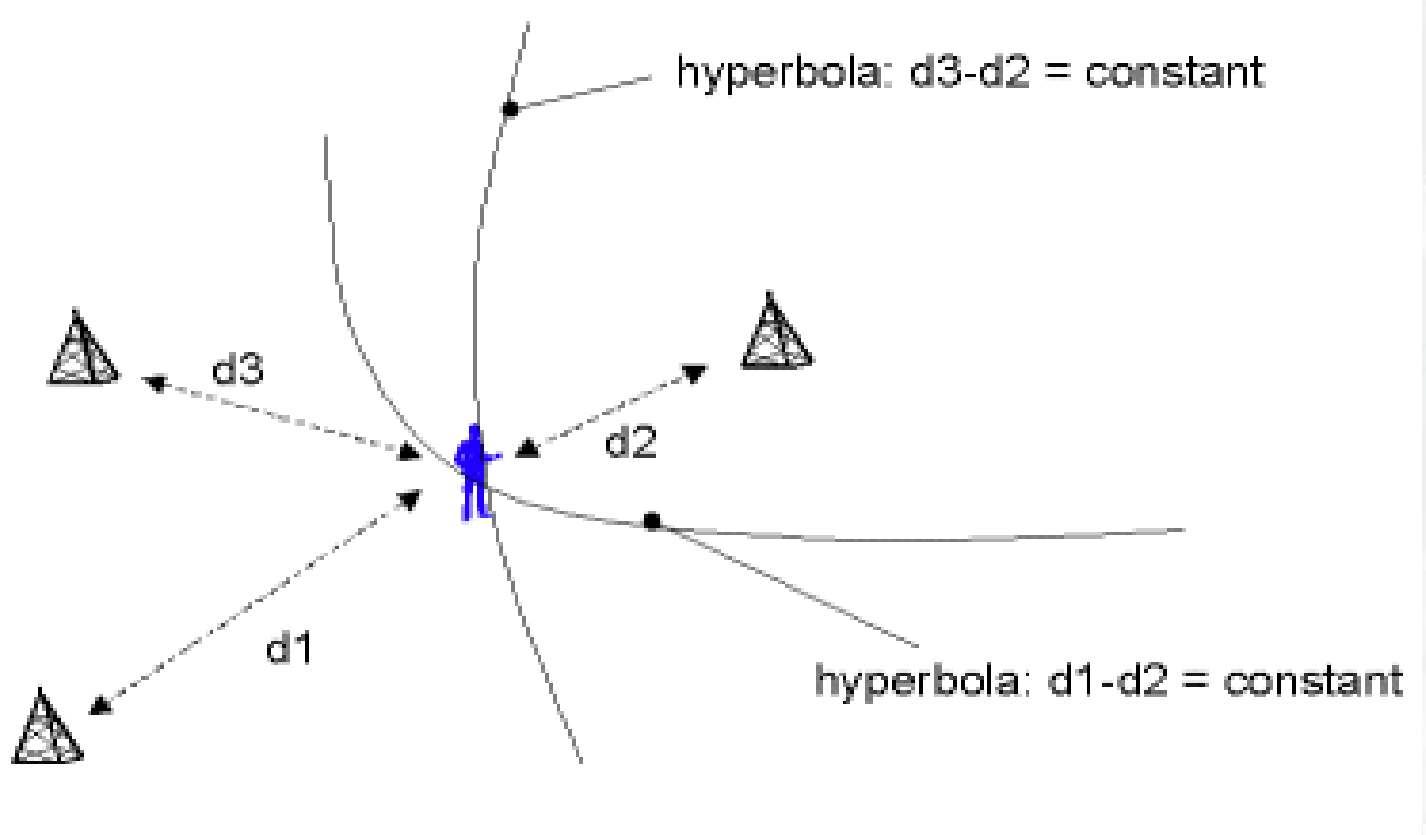
- errors > 500m common
+ simple

# Cell ID errors



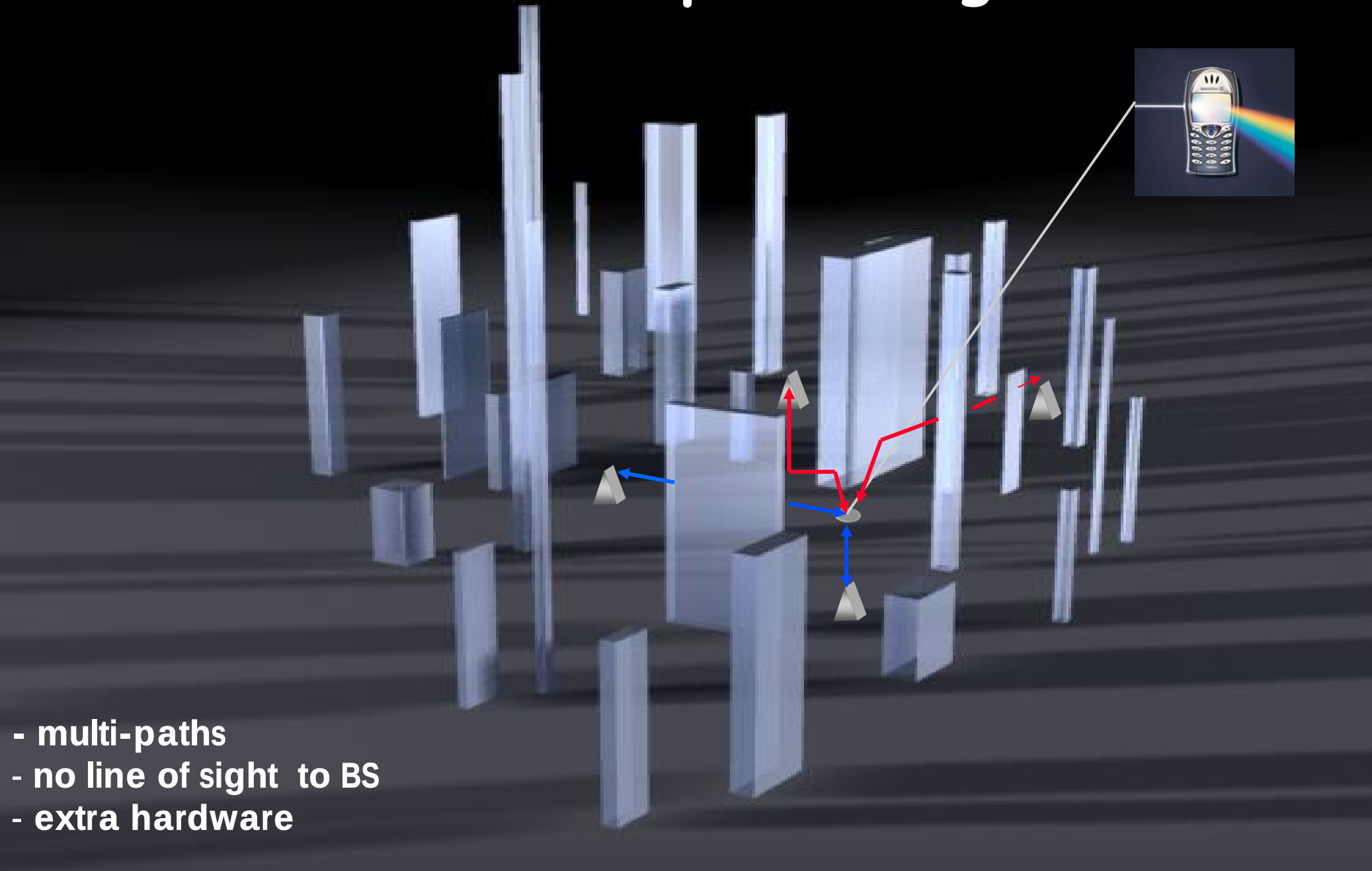

©Henry Tirri 2002, Petri Myllymäki 2003

# Enhanced Observed Time Difference (E-OTD)



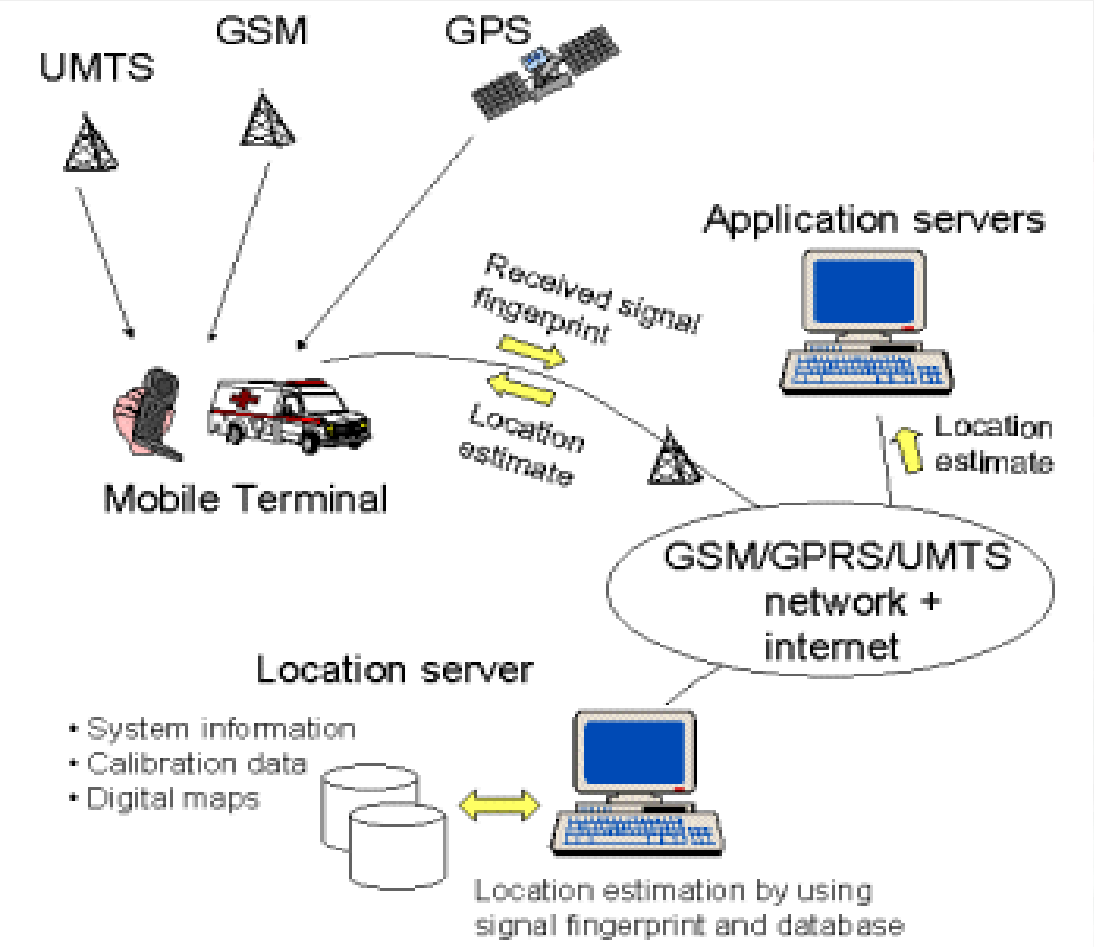

©Henry Tirri 2002, Petri Myllymäki 2003

# Urban positioning



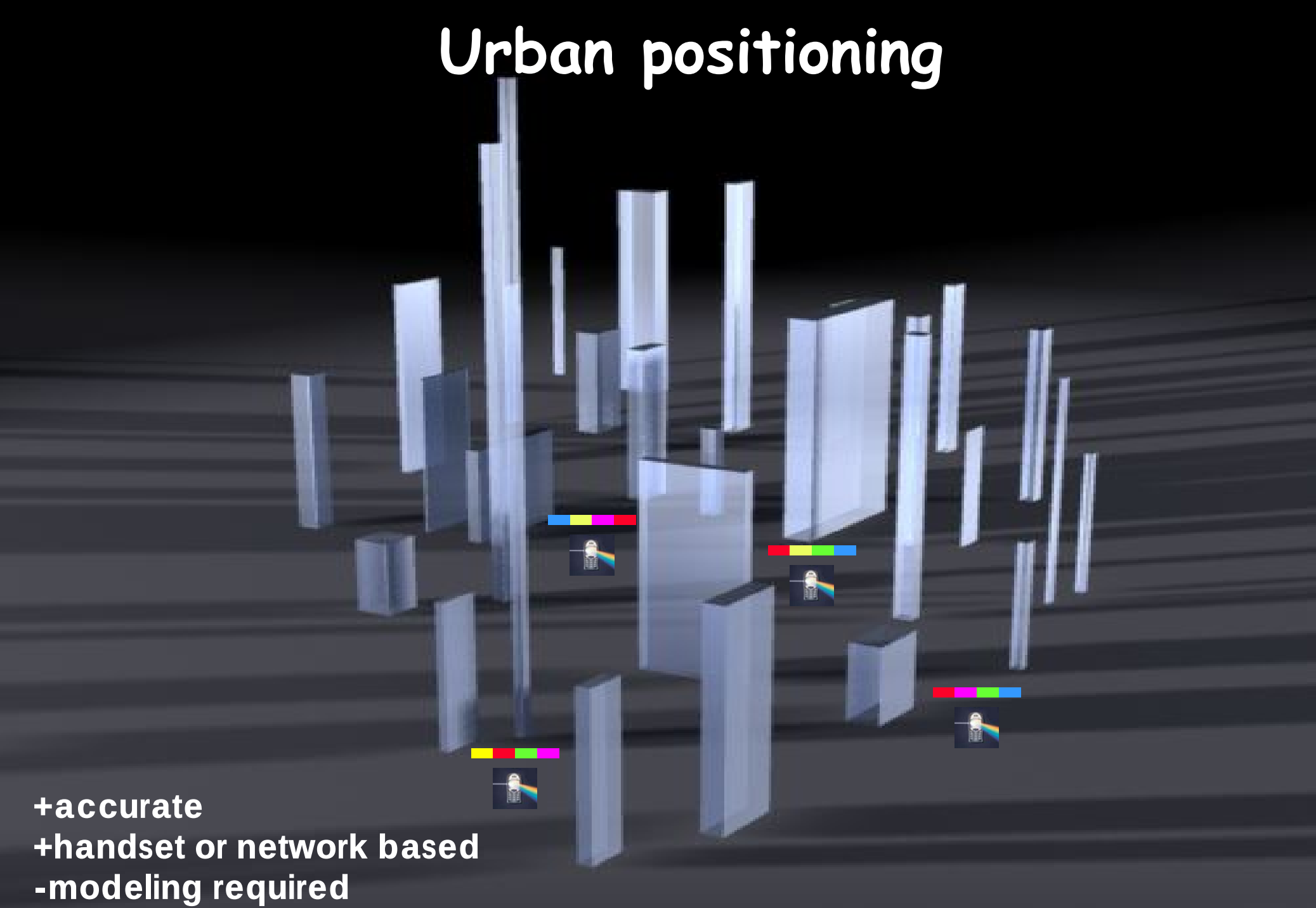
- multi-paths
- no line of sight to BS
- extra hardware

# Modeling approach




©Henry Tirri 2002, Petri Myllymäki 2003

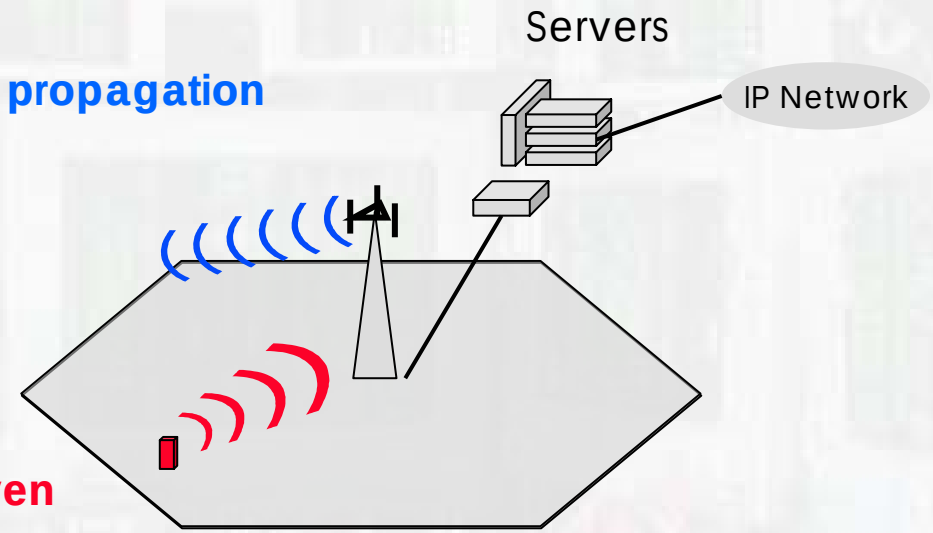
# Urban positioning



+accurate
+handset or network based
-modeling required

# Modeling alternatives

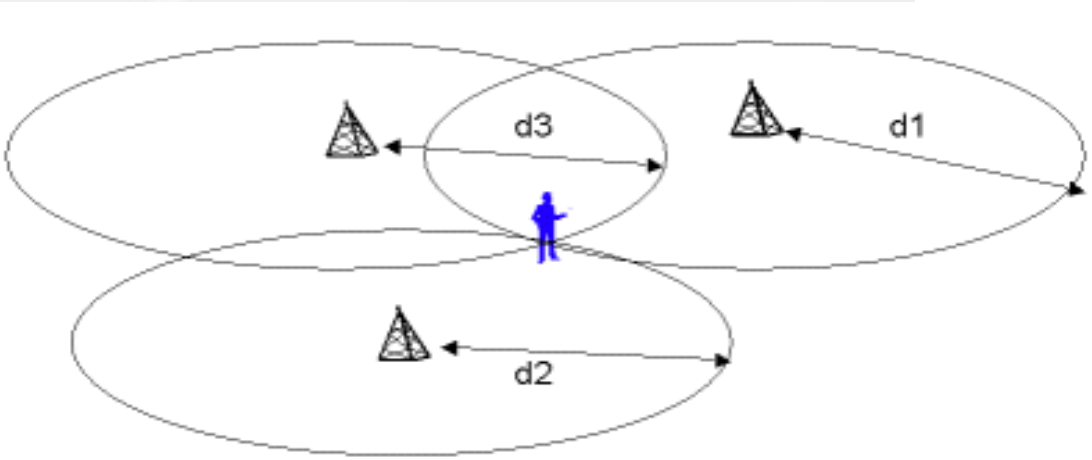
”Downlink modeling: the propagation models (GM level 1)”

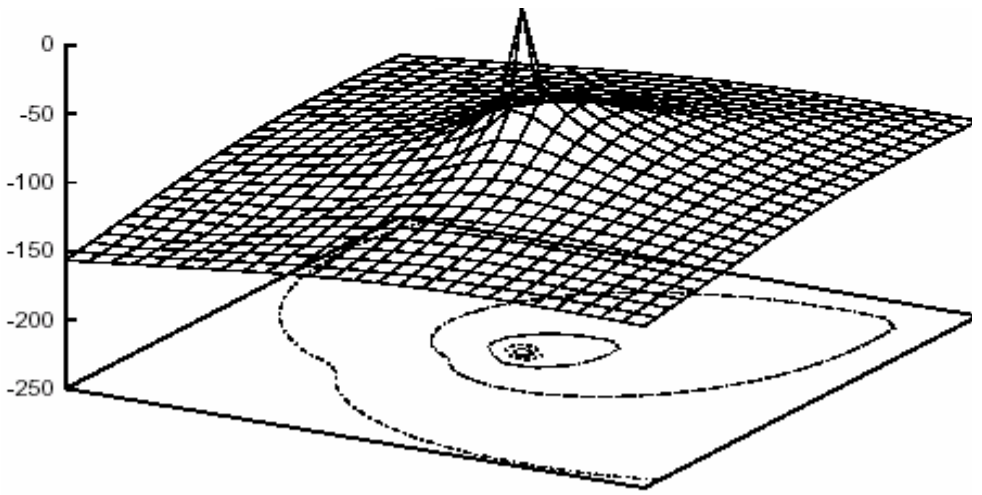



”Uplink modeling: data driven models (GM level 2)”

©Henry Tirri 2002, Petri Myllymäki 2003

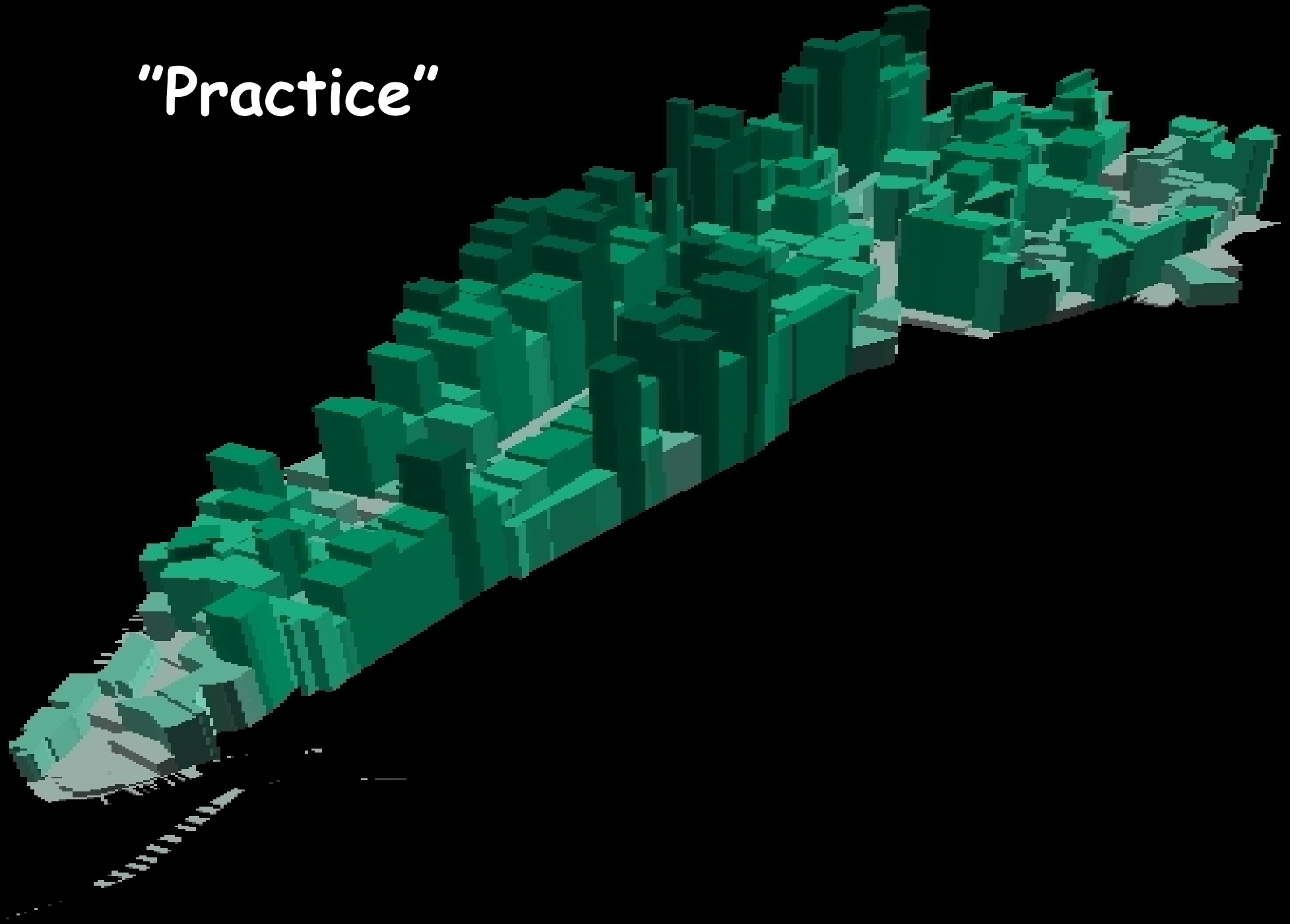
# "Theory"



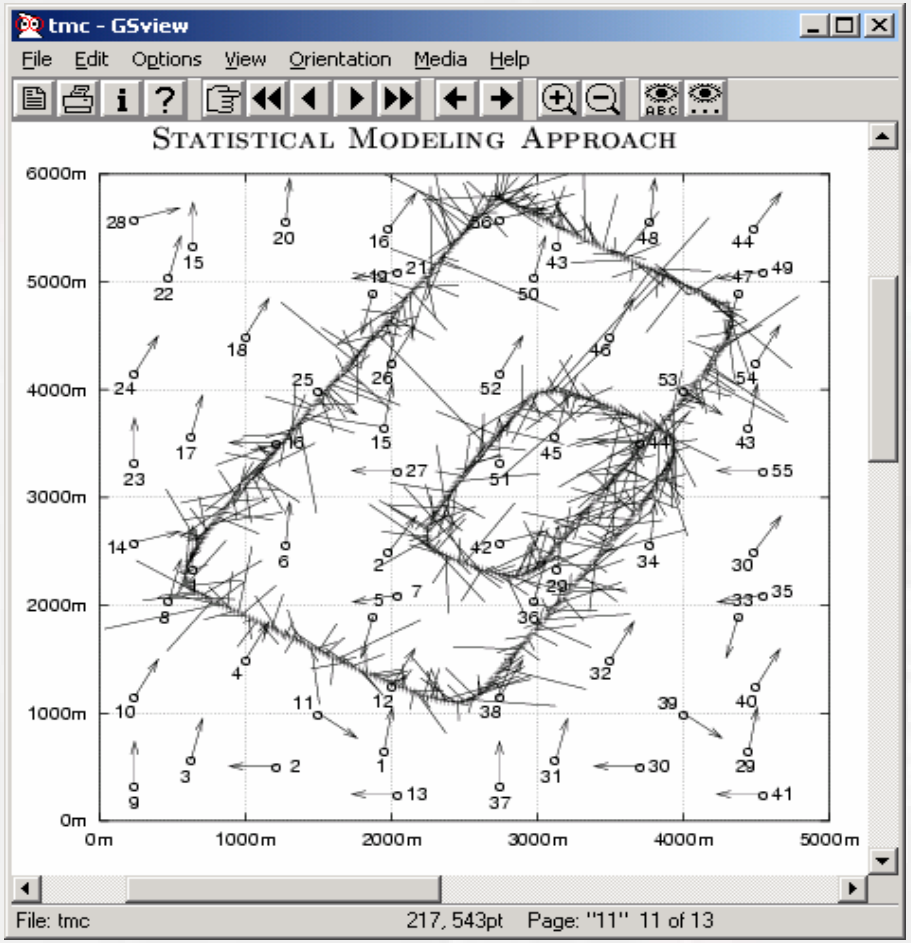



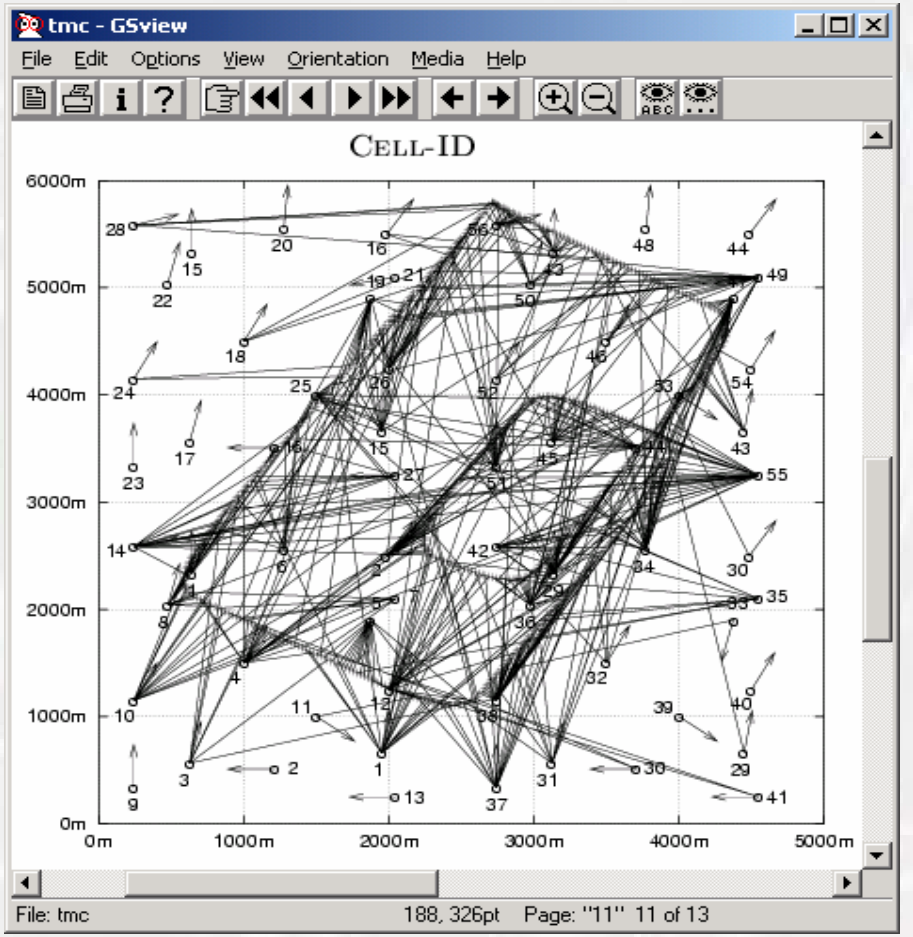

©Henry Tirri 2002, Petri Myllymäki 2003

# "Practice"



# Model-based vs. Cell ID positioning errors visualization

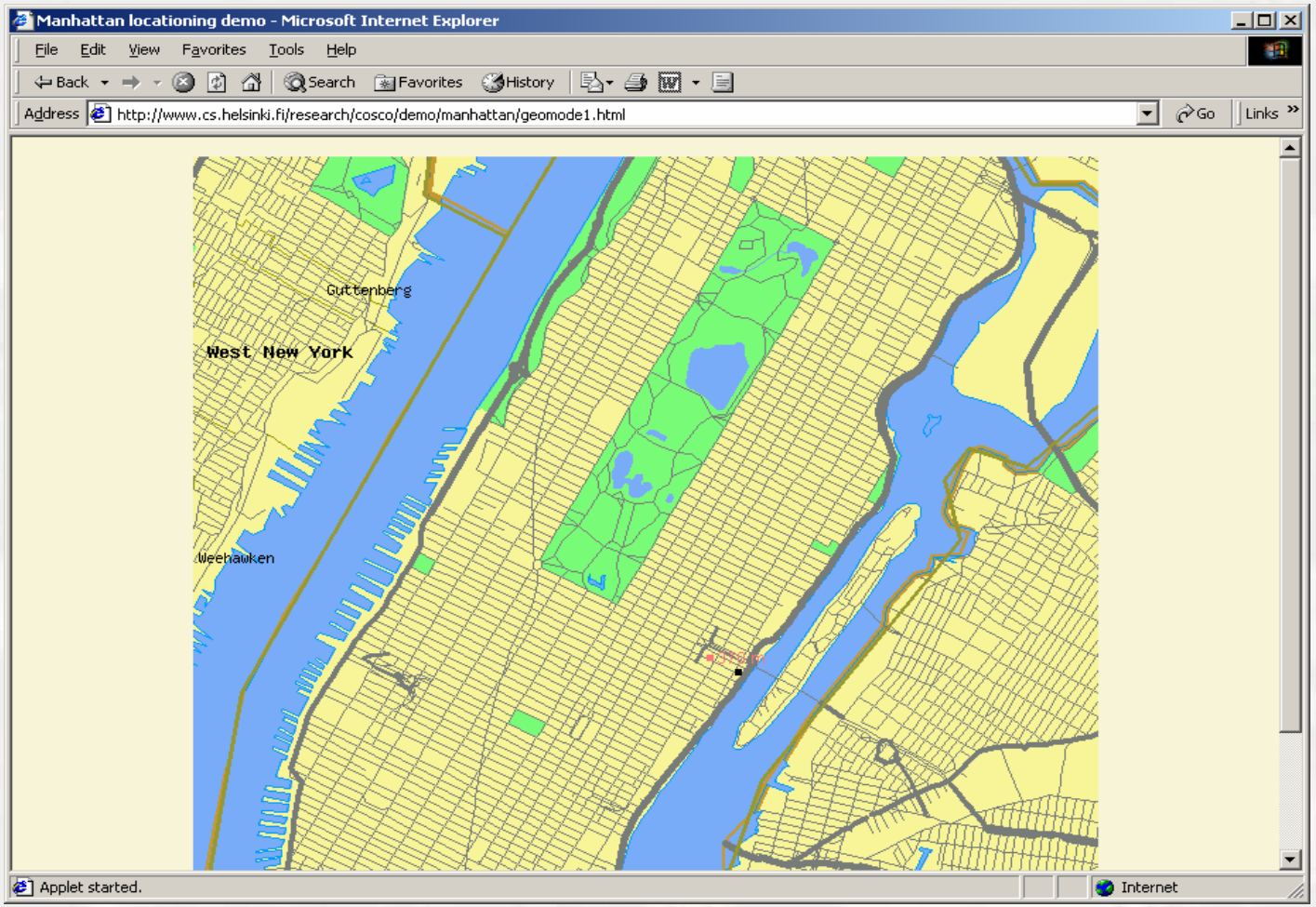



©Henry Tirri 2002, Petri Myllymäki 2003

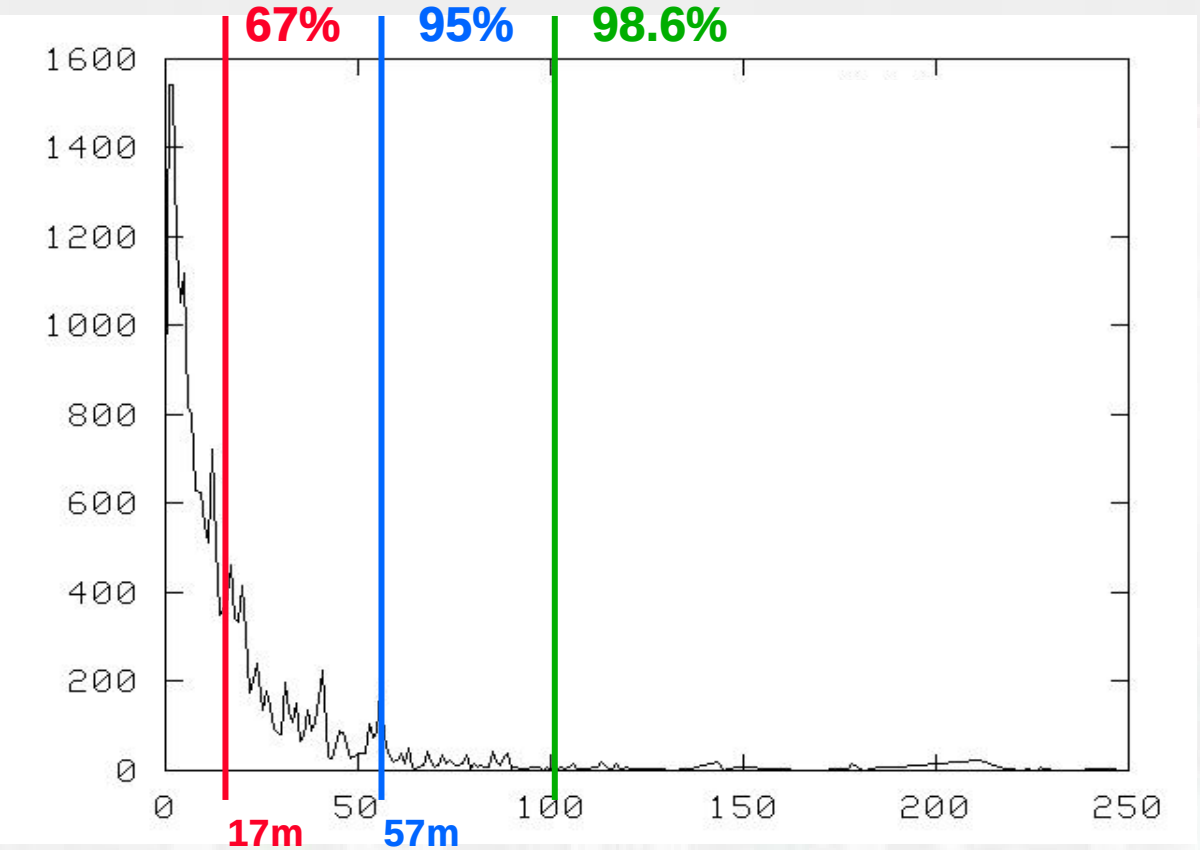
# Mobile Positioning: Manhattan Trials

# NYC Trial 2001




©Henry Tirri 2002, Petri Myllymäki 2003

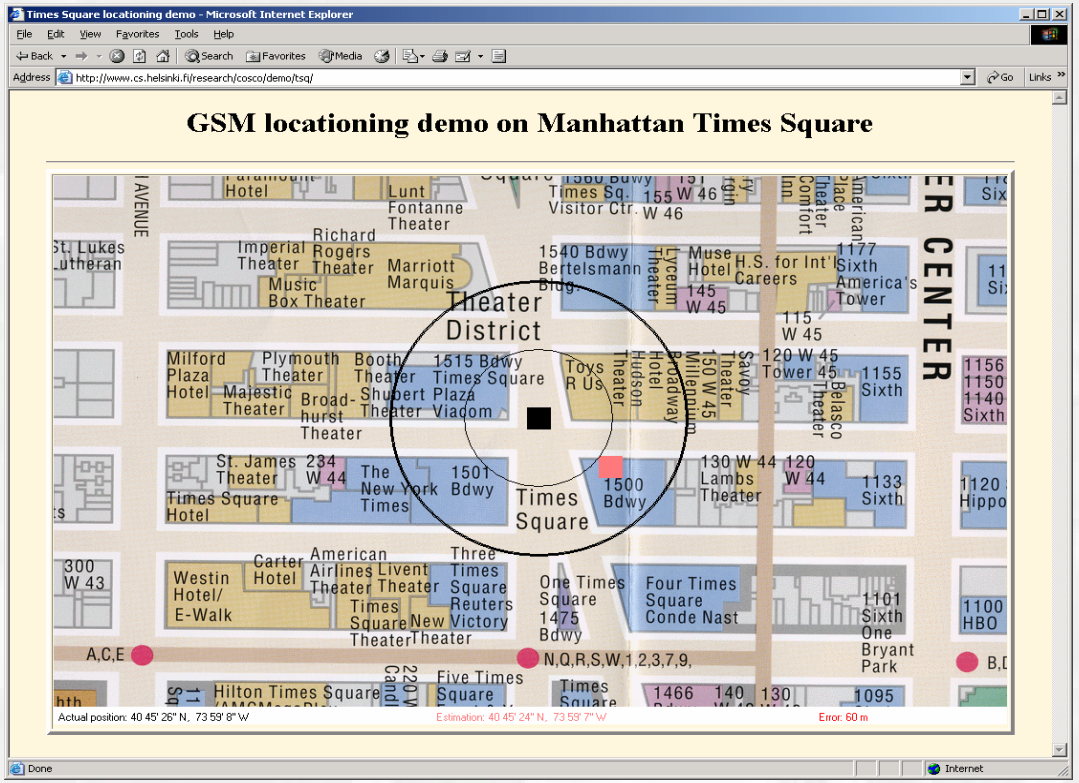
# Accuracy NYC Trial 2001




- 20166 points
- tracking; testing done in a car;

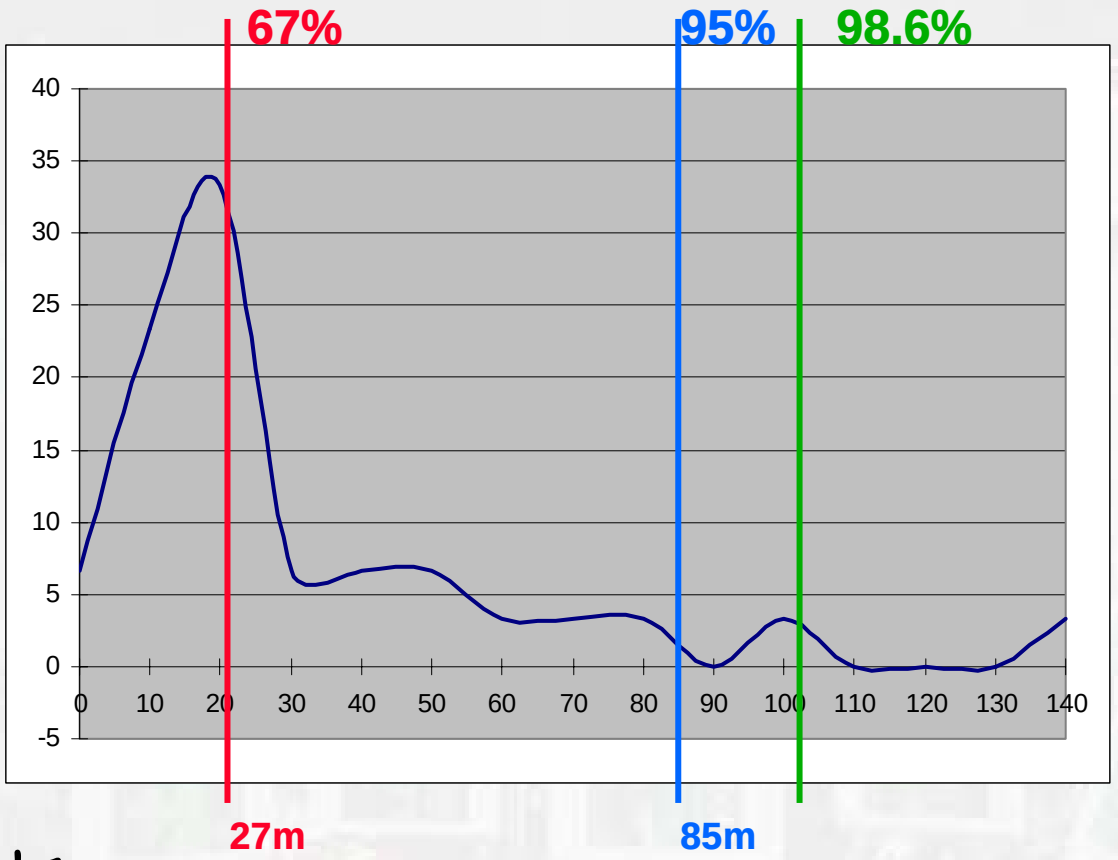
©Henry Tirri 2002, Petri Myllymäki 2003

# Trials: Manhattan 2002




©Henry Tirri 2002, Petri Myllymäki 2003

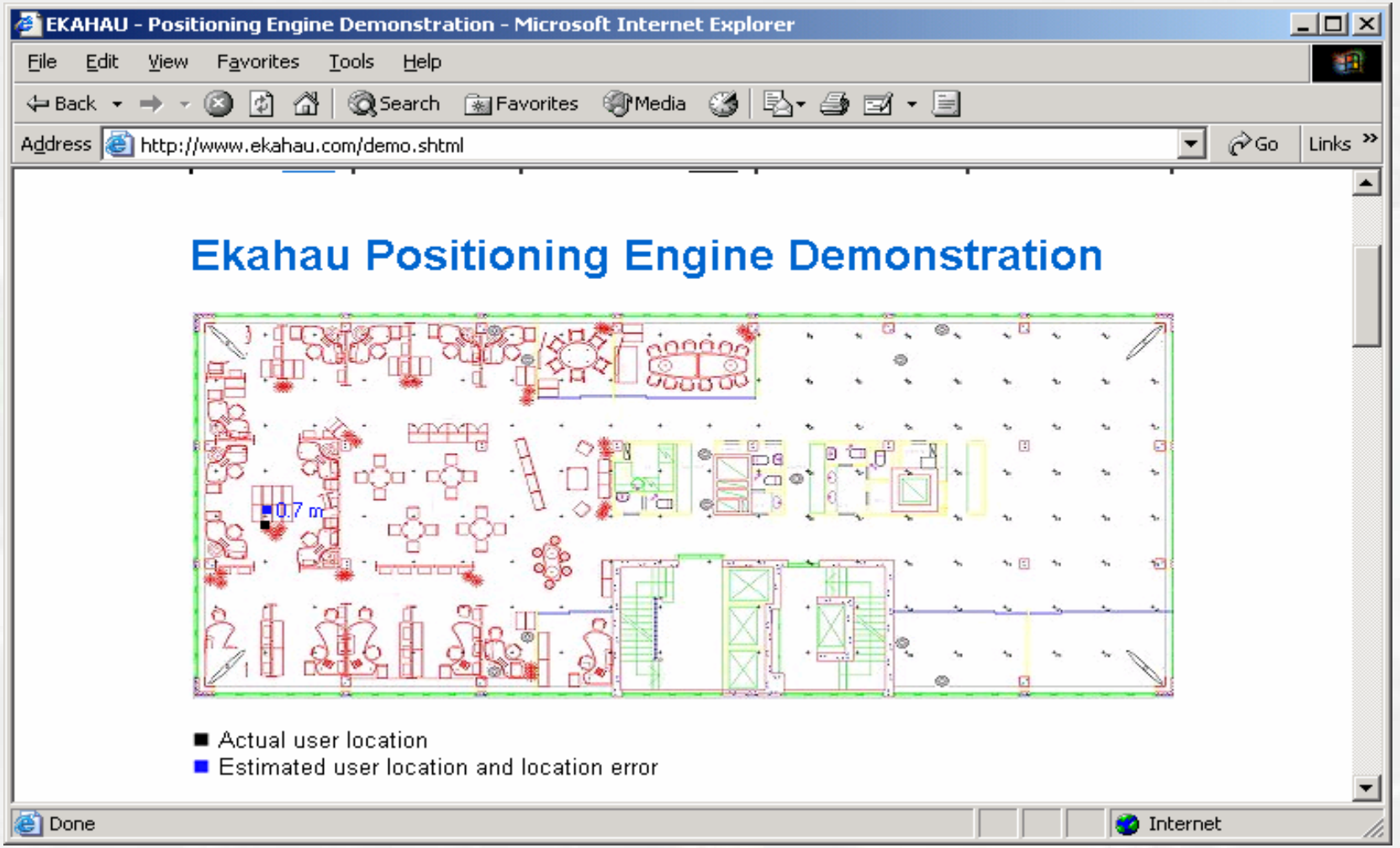
# Accuracy NYC Trial 2002




- 30 points
- static; testing done by walking;

©Henry Tirri 2002, Petri Myllymäki 2003

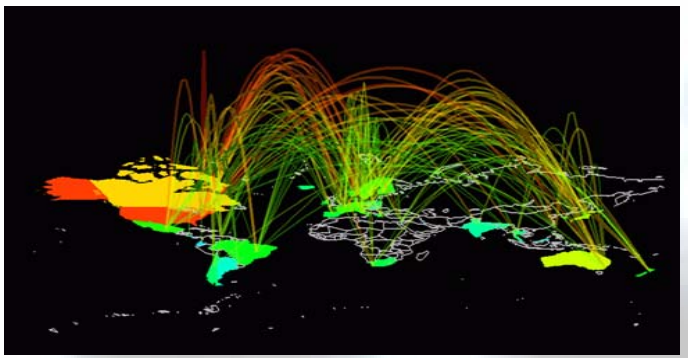
# Genaralization: WLAN



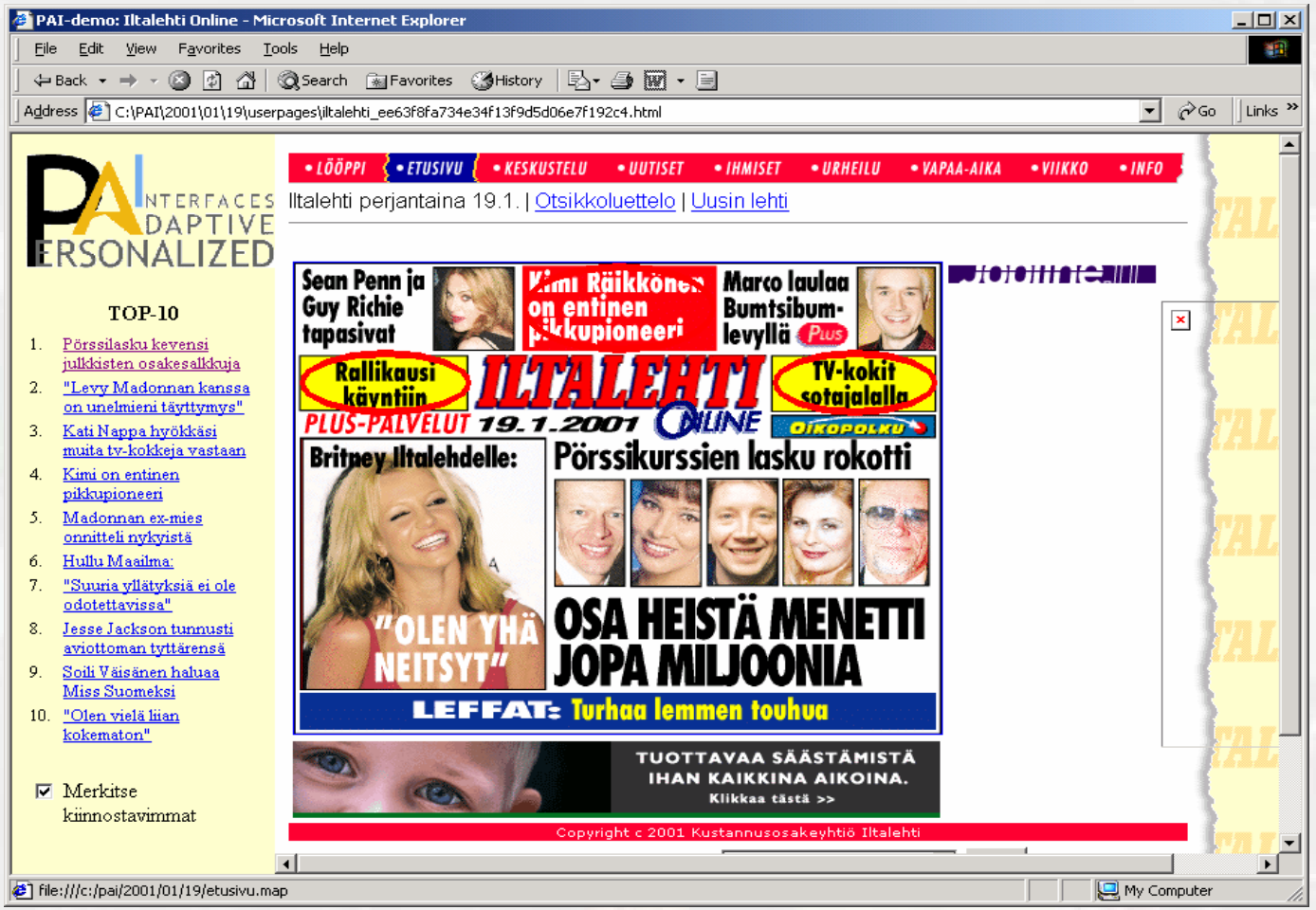

©Henry Tirri 2002, Petri Myllymäki 2003

# Personal WEB

## Personalization and Intelligent Search

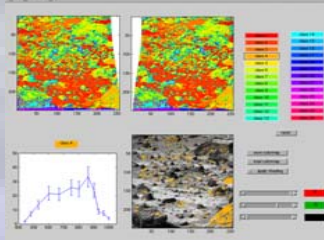


# Personalized, adaptive interfaces

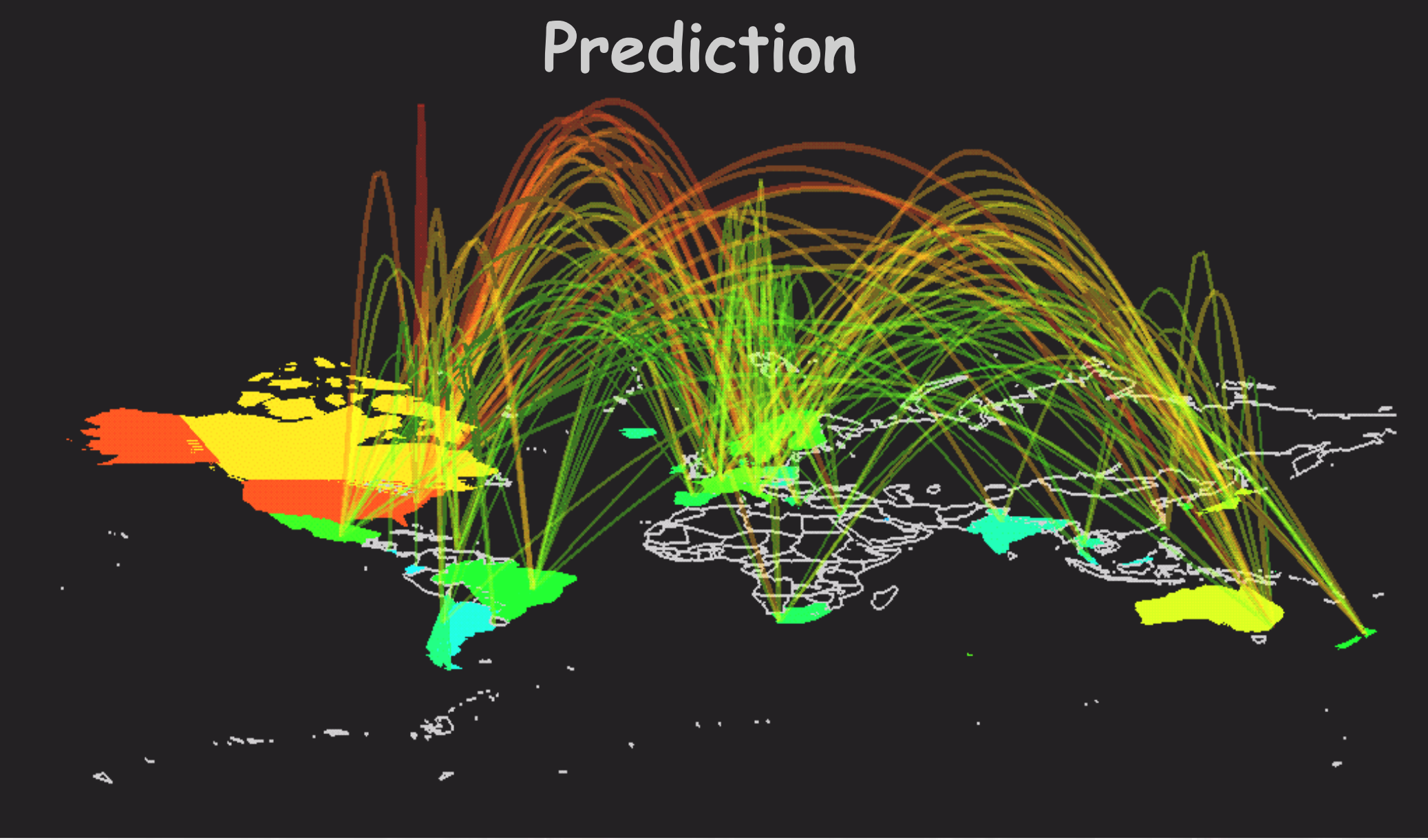



©Henry Tirri 2002, Petri Myllymäki 2003

# Future of Search


# Prediction



"By 2010, the amount of data online in Internet and corporate intranets exceeds one **yottabyte** ($10^{24}$)"

©Henry Tirri 2002, Petri Myllymäki 2003

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 Tirri 2002, Petri Myllymäki 2003

# 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 Tirri 2002, Petri Myllymäki 2003

# It's a war…


©Henry Tirri 2002, Petri Myllymäki 2003

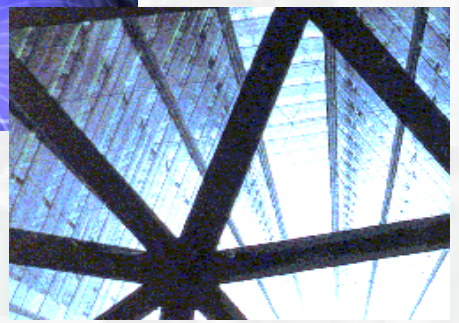
# Next Generation Information Management Research

Collaborative search


Search-Ina-Box


YDIN


Mobile search



Topic-specific search

©Henry Tirri 2002, Petri Myllymäki 2003

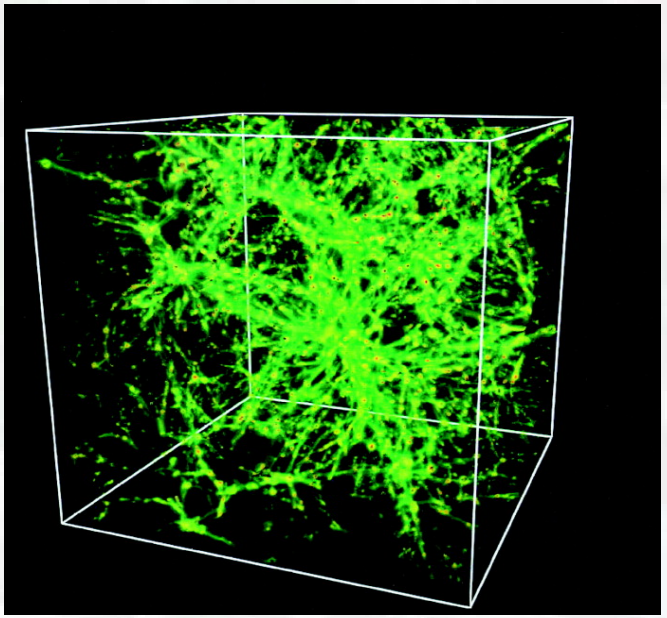
# Emerging key technologies

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

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

# 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 Tirri 2002, Petri Myllymäki 2003

# Related projects

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

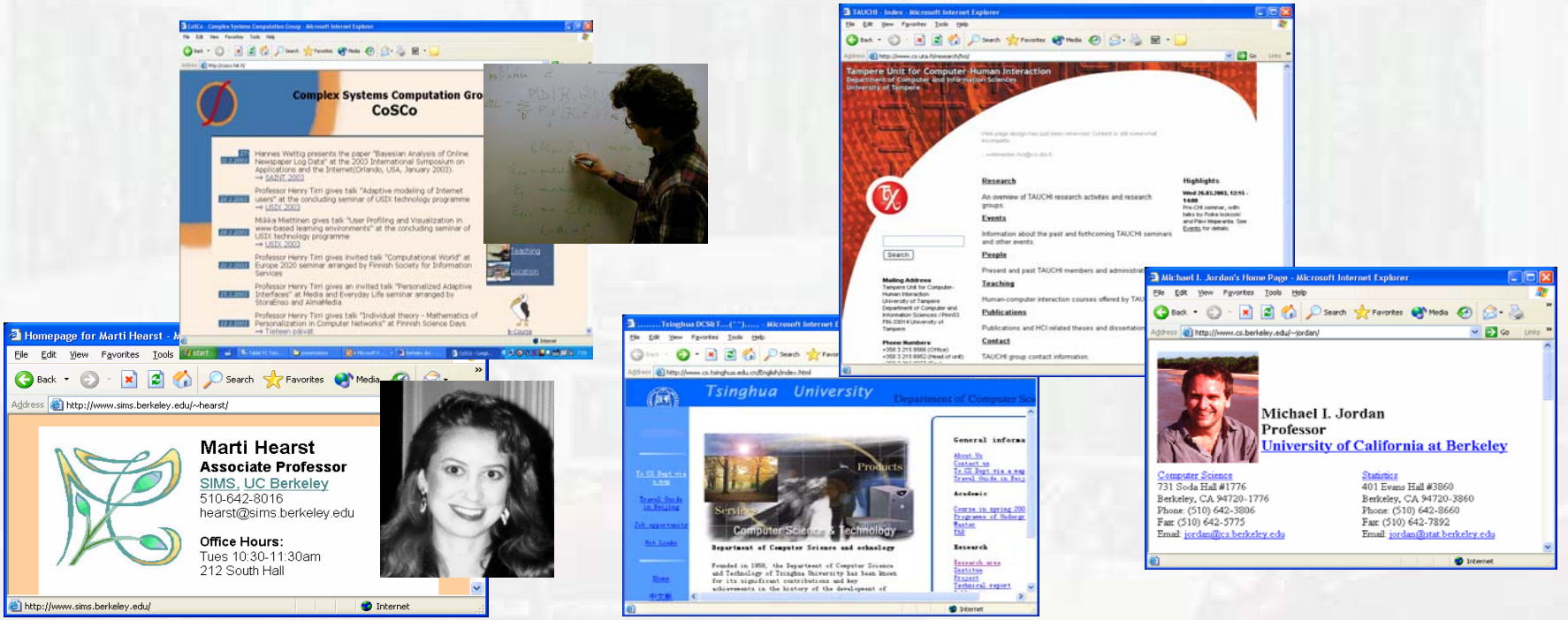

Search-Ina-BOX


ALVIS (FP6 STREP)

©Henry Tirri 2002, Petri Myllymäki 2003

# SIB research core competencies

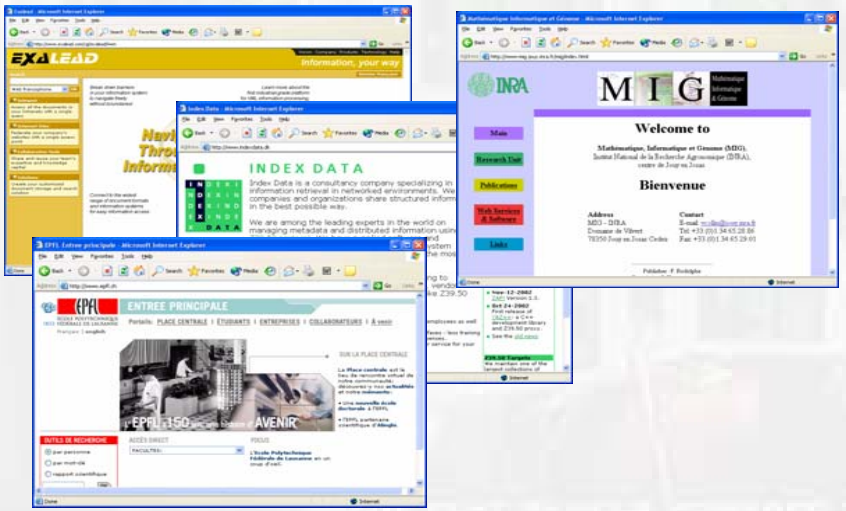
- probabilistic modeling for embedded language models, automatic content analysis and personalization
- advanced user interface design




©Henry Tirri 2002, Petri Myllymäki 2003

# Superpeer Semantic Search (ALVIS)

ALVIS (FP6 STREP)

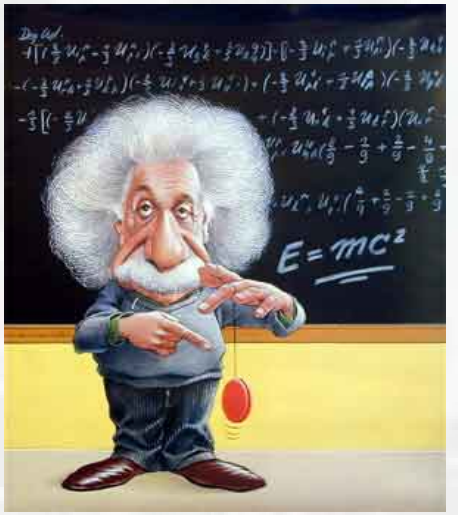



- HIIT
- MIG-INRA (France)
- EPFL (Switzerland)
- Lund University (Sweden)
- JSI (Slovenia)
- Index Data (Denmark)
- Exalead (France)
- DTV (Denmark)
- Université Paris Nord

©Henry Tirri 2002, Petri Myllymäki 2003

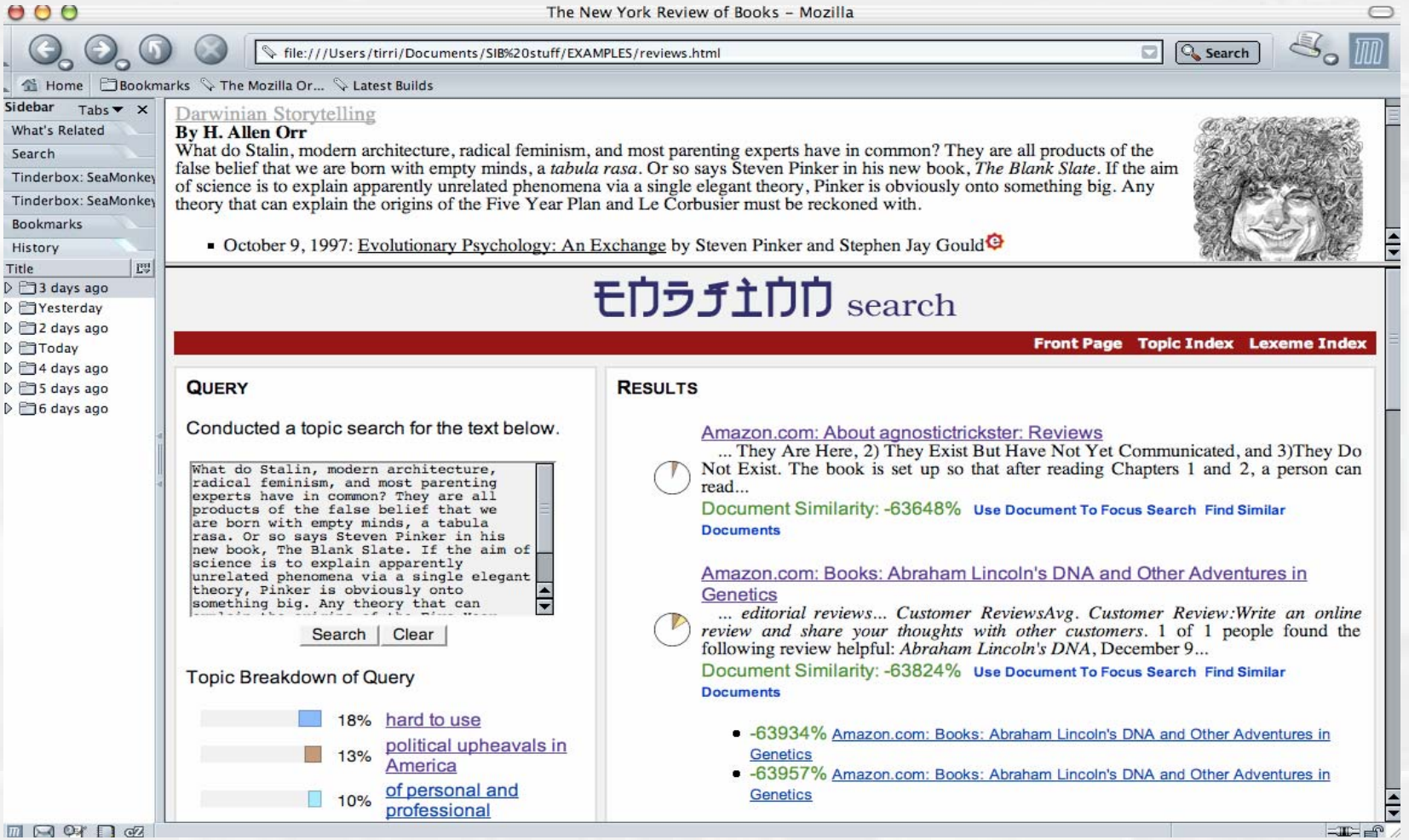
# モカラチュカカ search

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

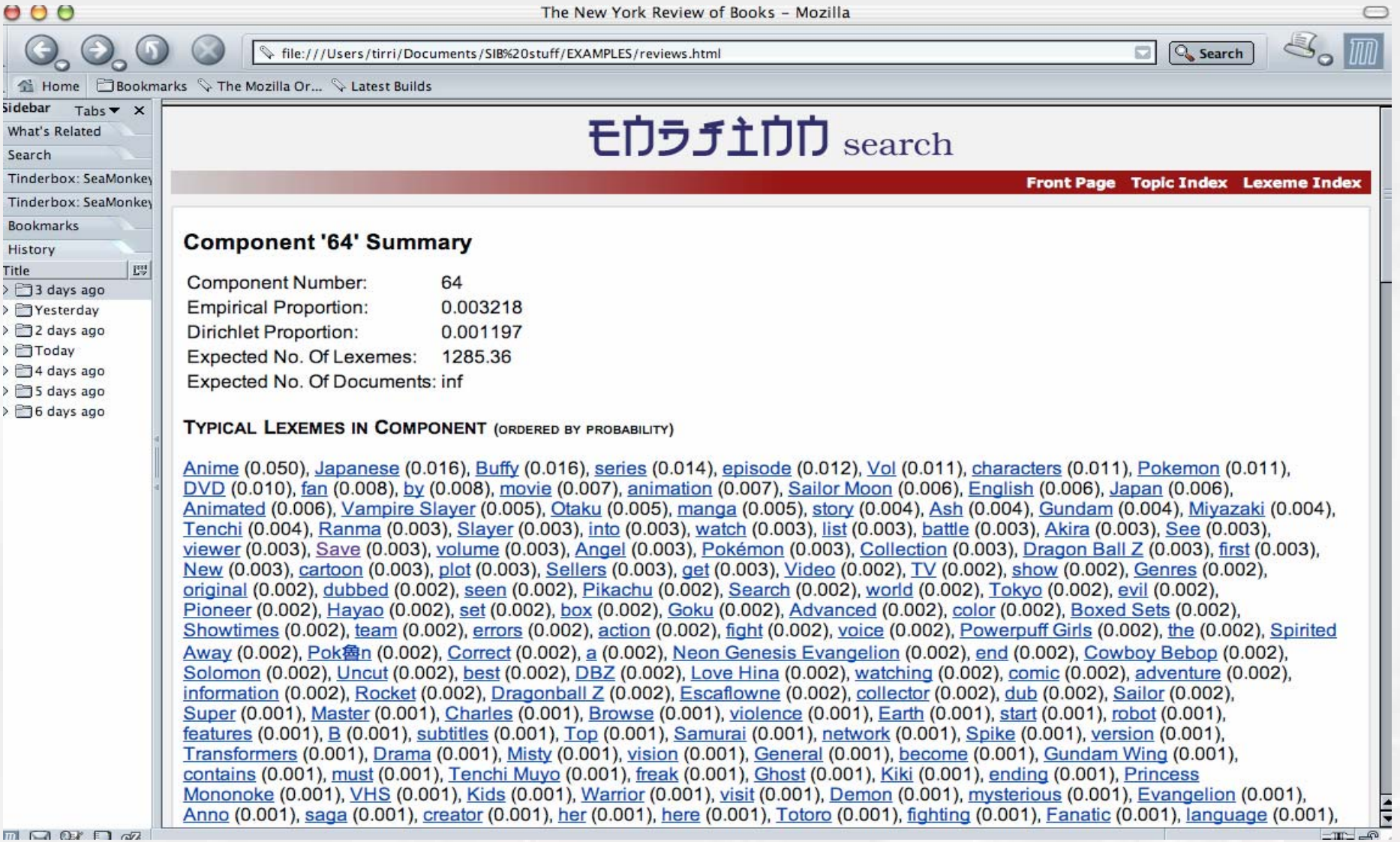



©Henry Tirri 2002, Petri Myllymäki 2003

# MOTHER search




©Henry Tirri 2002, Petri Myllymäki 2003

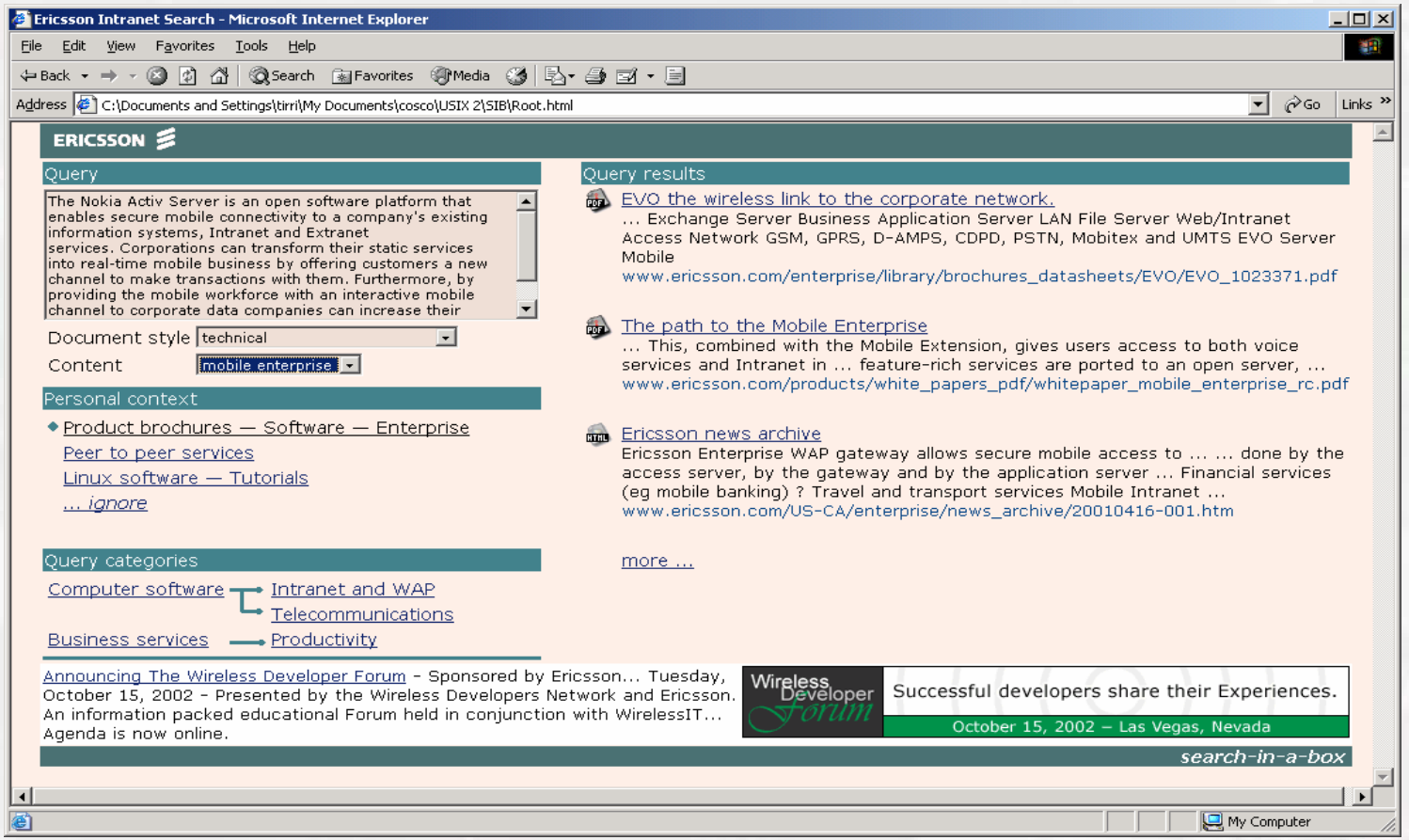



©Henry Tirri 2002, Petri Myllymäki 2003

# "Ubisearch"


©Henry Tirri 2002, Petri Myllymäki 2003

# Search-Ina-Box concept




©Henry Tirri 2002, Petri Myllymäki 2003

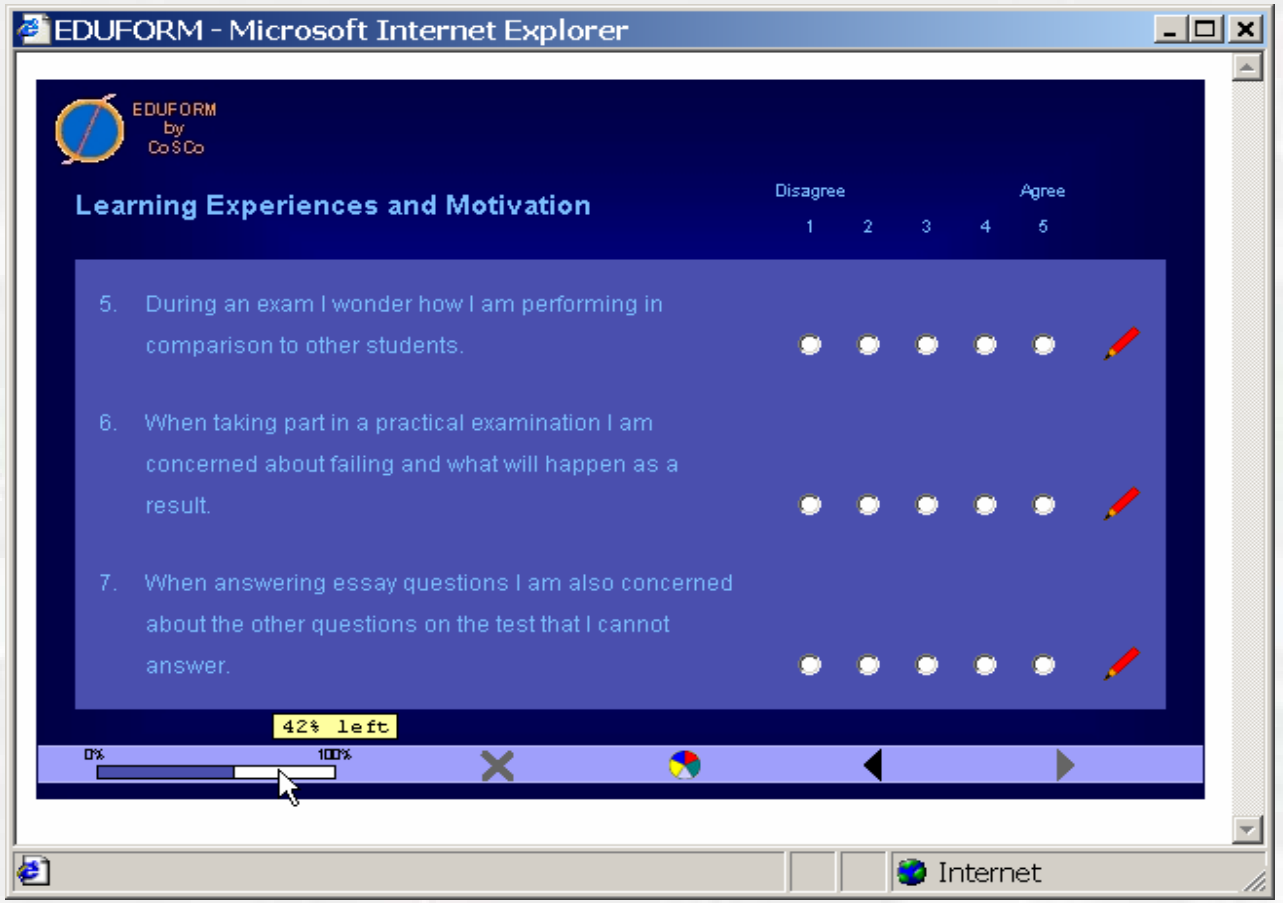
# http://cosco.hiit.fi/search/


©Henry Tirri 2002, Petri Myllymäki 2003


# Intelligent Tools for E-learning

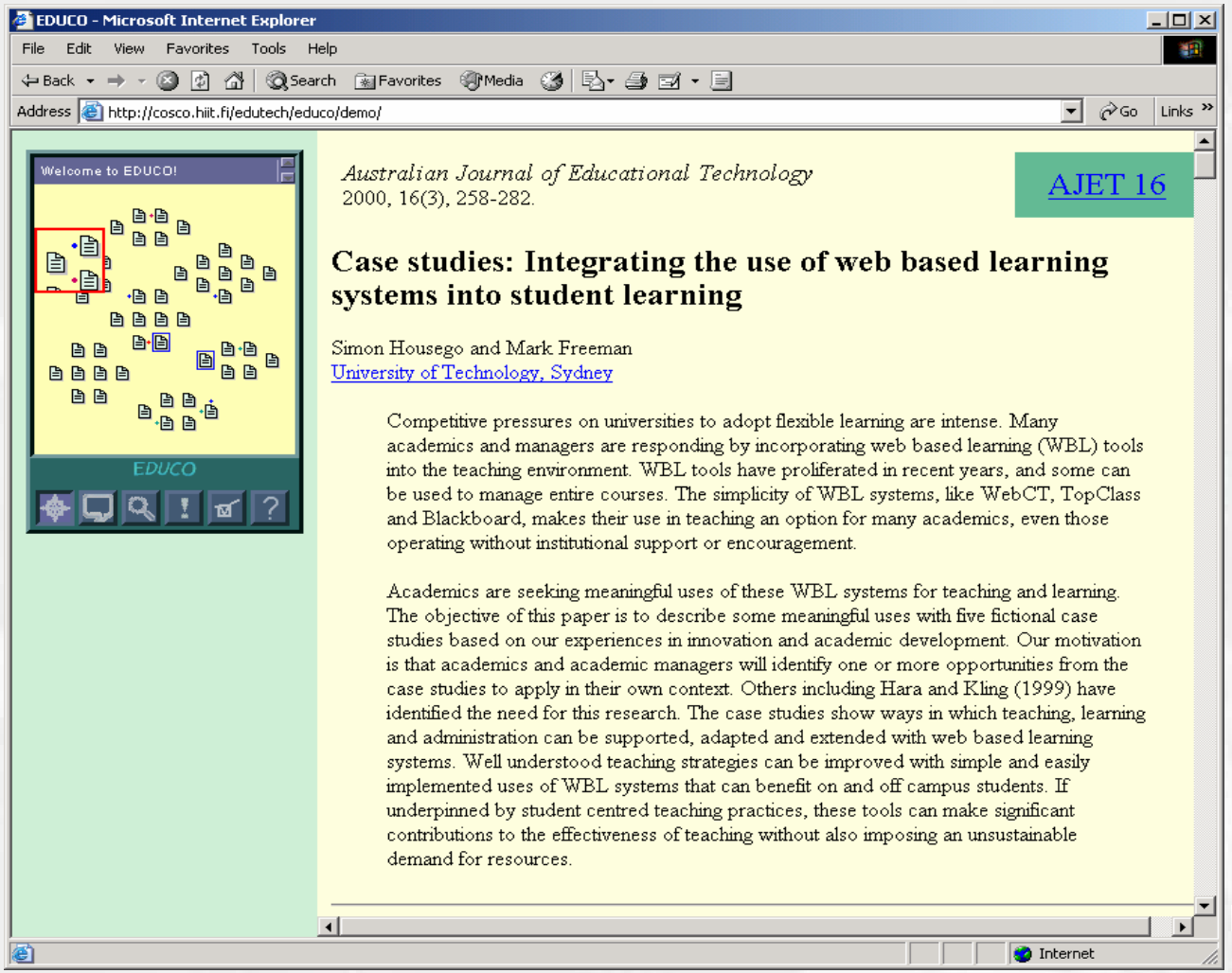
©Henry Tirri 2002, Petri Myllymäki 2003

# EDUFORM



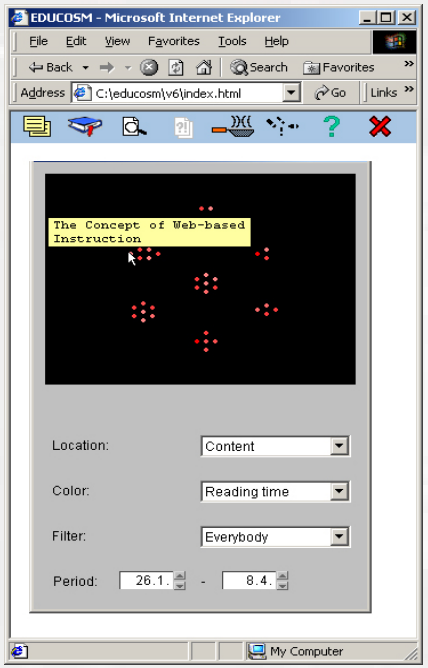

©Henry Tirri 2002, Petri Myllymäki 2003

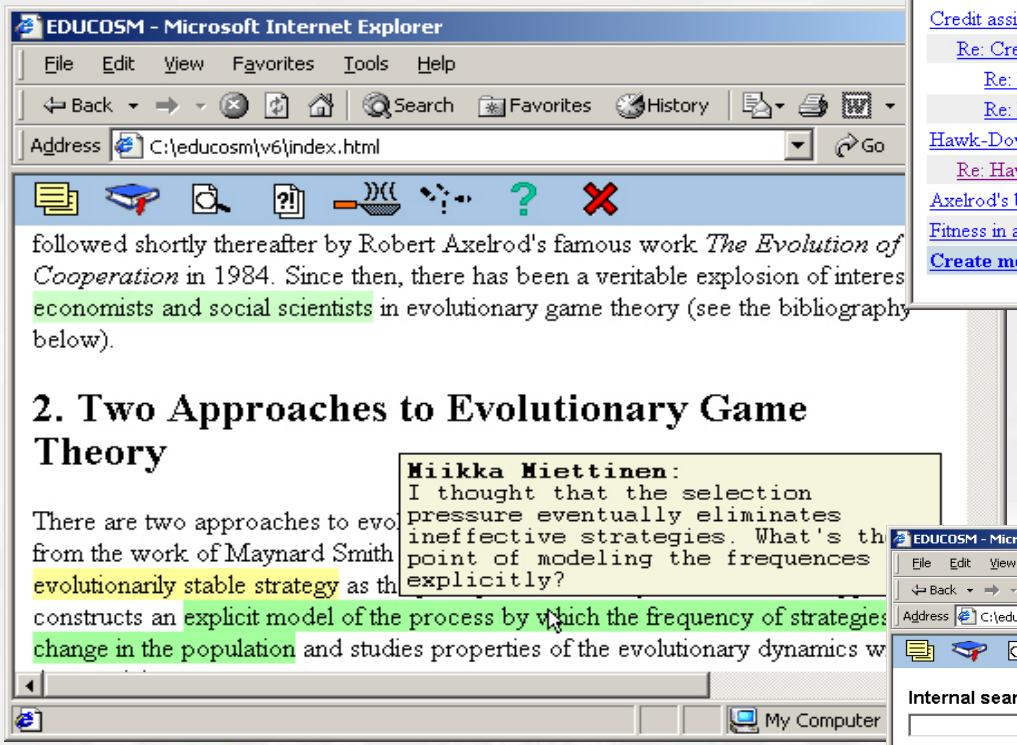
# EDUCO

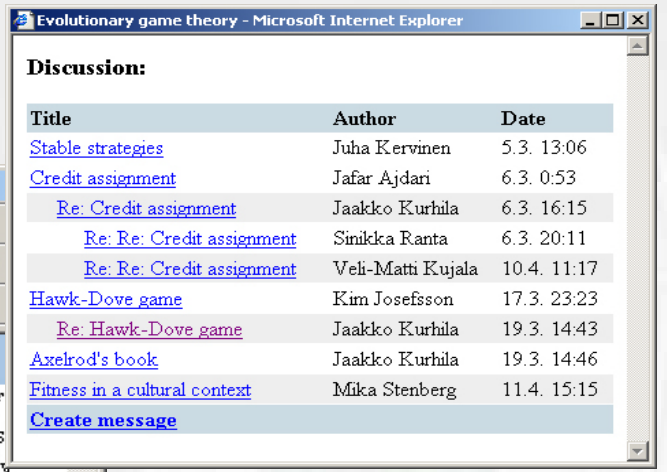


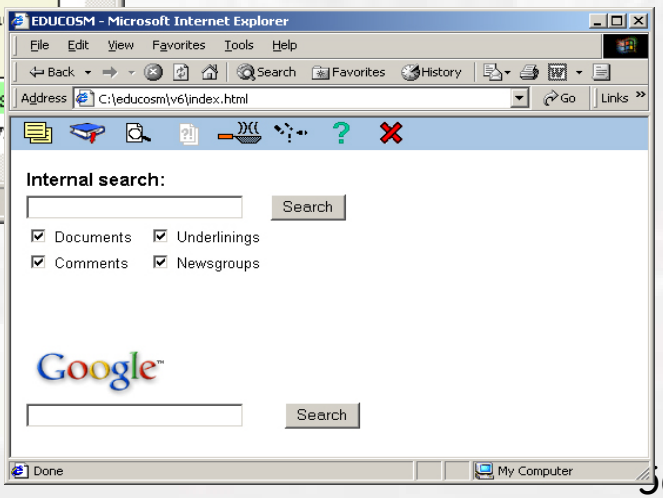

©Henry Tirri 2002, Petri Myllymäki 2003

# EduCosm



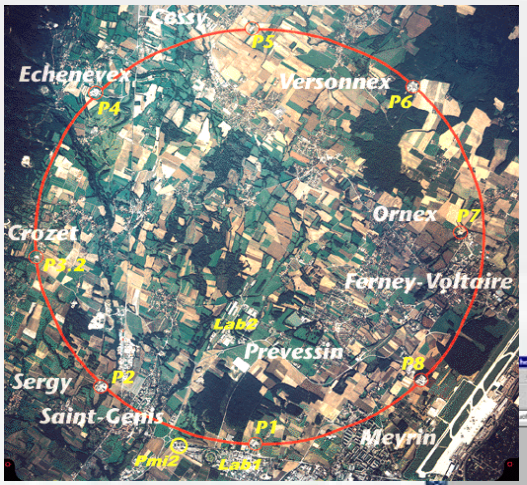



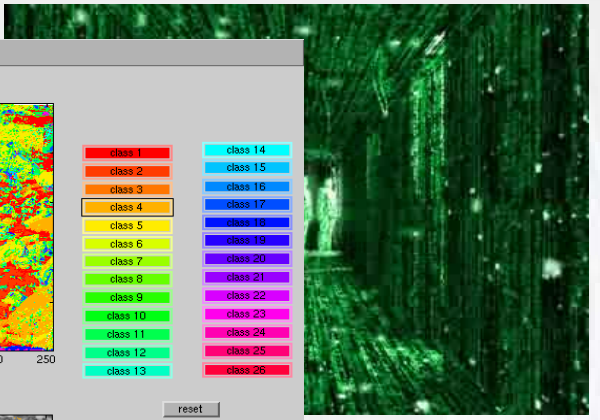



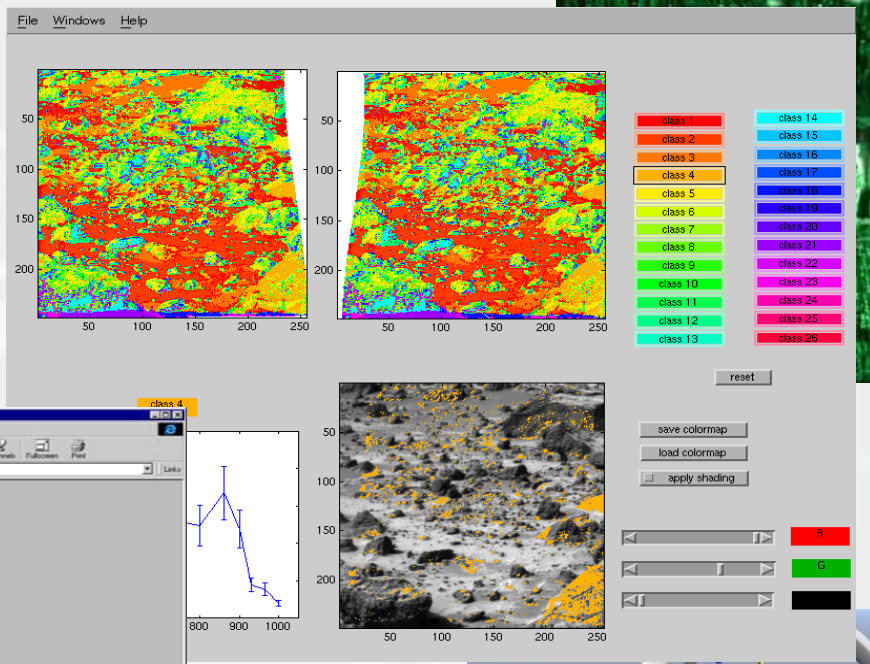



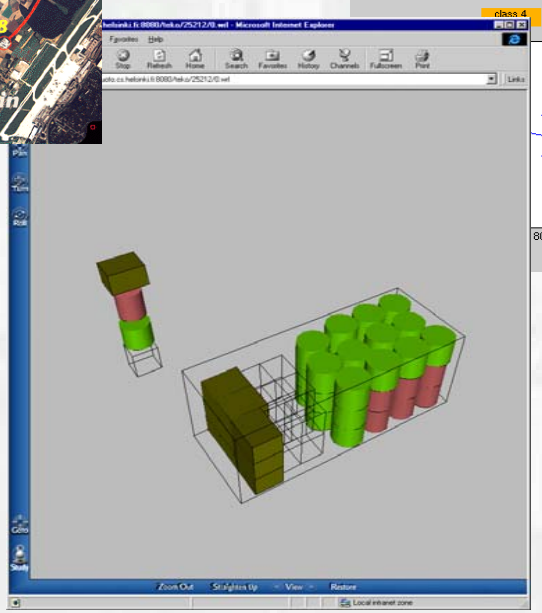

©Henry Tirri 2002, Petri Myllymäki 2003

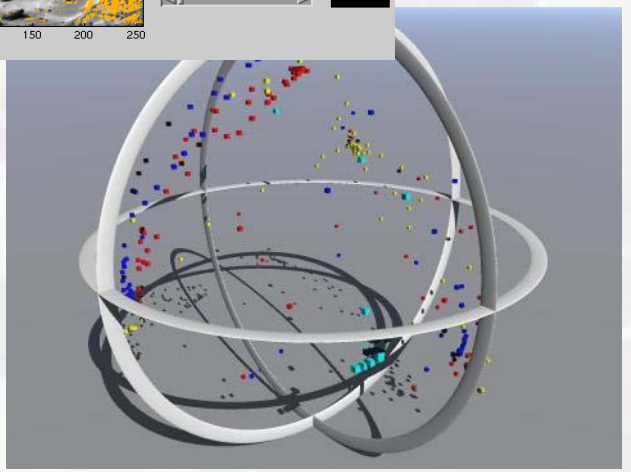
# and other things







©Henry Tirri 2002, Petri Myllymäki 2003