Inferato – Don't search, inferate!
Scenario Based Search

Ulrich Norbisrath/
Distributed Systems Group

University of Tartu, Estonia

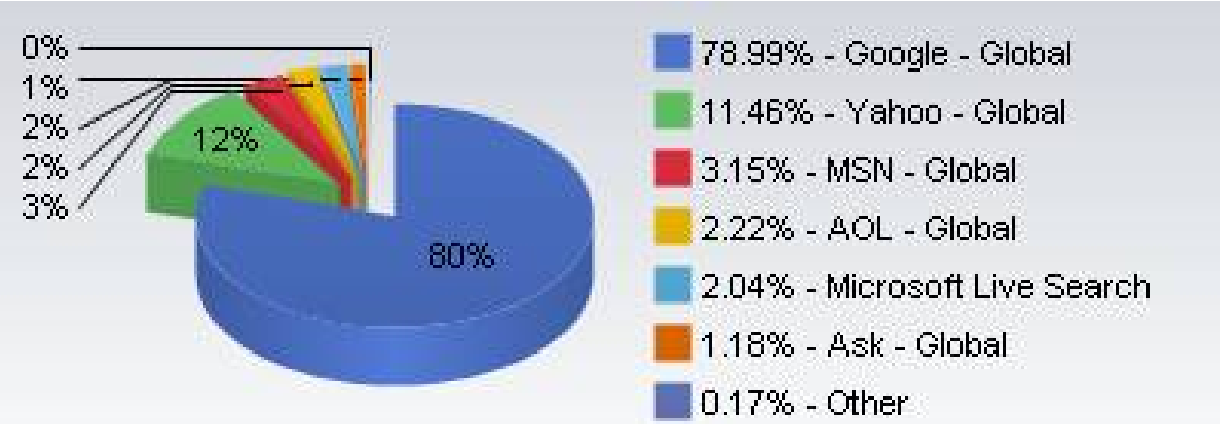
Presentation at
Theory Days, Kääriku
2009/01/31

Overview

- Search - Today
- Innovation in Search
- Graphs for Contexts
- Scenario Based Search
- Application: Inferato
- Summary/Questions

Search - Today

- Internet
  - Google (search, maps, etc.)/Yahoo/MSN
  - Craigslist/Ebay/Expedia/Kayak
  - PubMed/IEEE
  - Wikipedia
- Other means
  - Library (also net)
  - Newspaper (classifieds)
  - Bulletin Board (supermarket)



search engines market share 2008 – source: hitslink.com

Search Today in Practice

- Dance shoes in Arlington, Virginia
- Washington, DC Metroarea postdoc, adjunct, or assistant professor
- Real Estate in Austria
- Open Master theses in Computer Science at University of Tartu
- Dance event with not too loud music near my current location

Problems in Today's Search

- Too diverse
- Outdated
- Hard to aggregate, check all results (too much, too much time to spend to scan all)
- All parts of the page scanned, unreasonable combinations
- Information scattered via multiple web-pages, can't be found
- Search not specifiable

Innovation in Search

- Semantic Web
  - proposal Berners-Lee 2001
  - make the web better understandable by machines
  - RDF (Resource Description Framework): allow to annotate web pages and elements of web-page with facts (of form subject predicate object)
  - Use ontologies to mediate between pages
- Web 2.0 elements – Search 2.0
  - user rating of search results

Innovation in Search

- Vertical (domain specific) search, search 3.0:
  - channeled information for particular area
  - spiders instead of crawlers
  - spiders scrape the content of pages
- Examples:
  - Travel search sites
  - Professional hardware search sites (for doctors, machine engineers)
  - Media related sites (video, music, books)
  - Domain libraries (CS library, Math library)

Graphs for Contexts

- Berners-Lee on Semantic Web:
  - 2001: “Evolution of the existing Web to a Semantic Web”
  - et al., 2006: “This simple idea, however, remains largely unrealized.”
- Why not more adopted?
  - Deployment tedious, no direct benefit
  - Complex interrelation across multiple websites unlikely and difficult
  - Reach of the SPO-facts is very small
- Better: use graphs for description of knowlege

Software Engineering methods to Create Context Graphs

- Graphs
- Agile Software Development
  - Scenario and test driven
- Visual Prototyping

- Combine into development method
  → “Scenario Based Search”

Proposal for Scenario Based Search

- Find case
- Write user stories (scenarios)
  - search
  - edit
- Identify and analyze objects
- Transfer to object graph
- Analyze interesting patterns
- Try to identify synergies with existing cases
  - example: Wikipedia → locations
- Scrape content

Scenario Based Search by Example: User-Story

Ralph Smith, 33 years old, investment banker for Goldman Sachs, wants relocate within London. He is currently in his private office at his apartment in London/Maidavale. He opens his Webbrowser and selects from his bookmarks inferato.com.

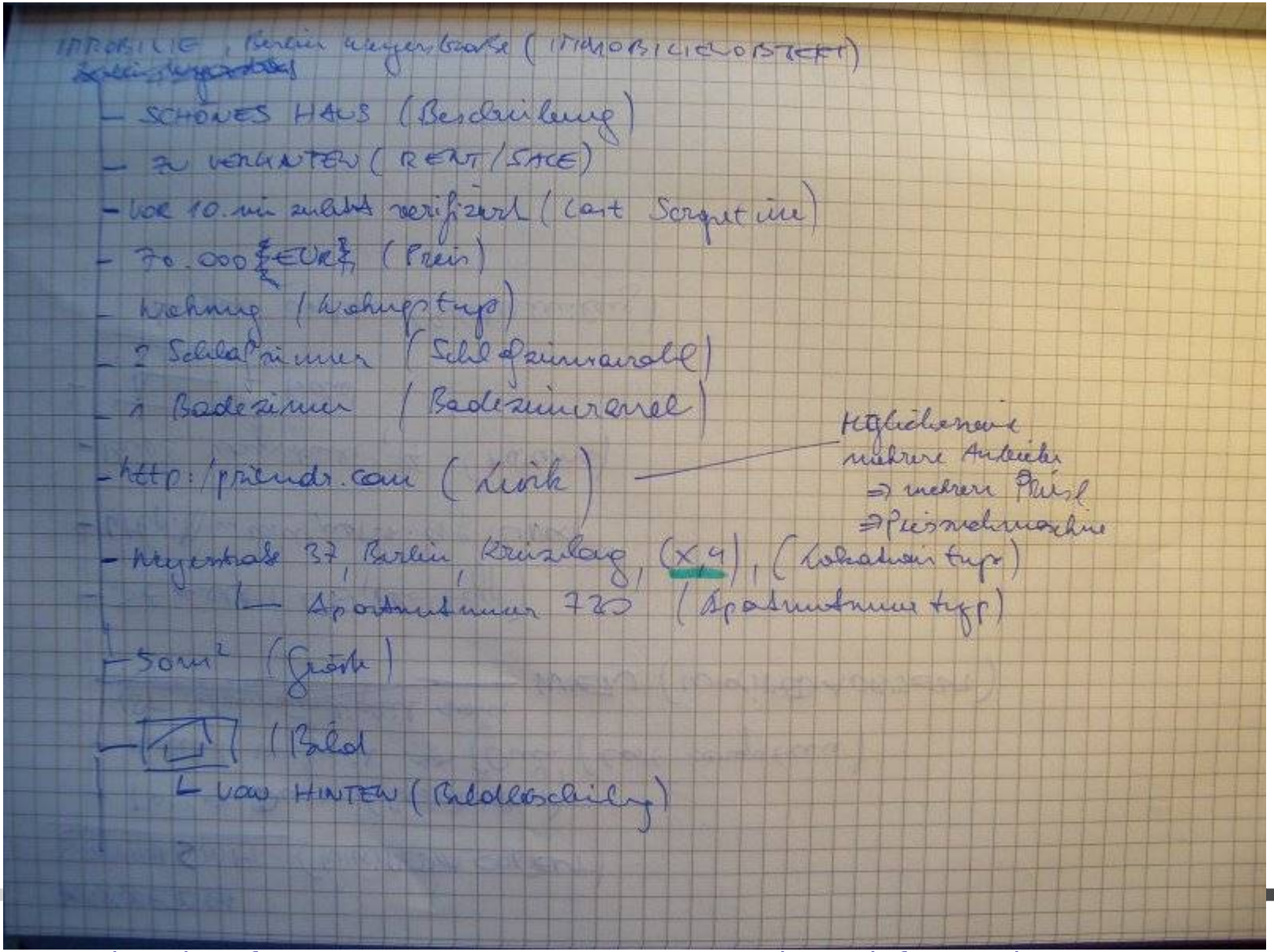
The initial webpage shows in the header: “You are currently located in London/Maidavale”. The left sidebar shows the categories “Housing, Events, Jobs, Other classifieds”. As Ralf has recently looked for buying a special electric guitar in the neighbourhood, the button “other classifieds” is pressed. He selects the Housing button, Other classifieds is automatically deselected and the first 10 search results show up. The first hit shows a description [...]
The second hit shows as description: “Nice quiet 3 bedroom apartment in Notting Hill with a perfect view and short ways to groceries and medical facilities.”
He clicks on details, a window pops up showing the id 9928299, that this apartment is for sale, has 3 bedrooms [...]

Scenario Based Search by Example: Object analysis

Ralph Smith, 33 years old, investment banker for Goldm
relocate within London. He is currently in his private office at his apartment in London/Maidavale. He opens his Webbrowser and selects from his bookmarks inferato.com.

The initial webpage shows in the header: “You are currently located in London/Maidavale”. The left sidebar shows the categories “Housing, Events, Jobs, Other classifieds”. As Ralf has recently looked for buying a special electric guitar in the neighbourhood, the button “other classifieds” is pressed. He selects the Housing button, Other classifieds is automatically deselected and the first 10 search results show up. The first hit shows a description [...]
The second hit shows as description: “Nice quiet 3 bedroom apartment in Notting Hill with a perfect view and short ways to groceries and medical facilities.”
He clicks on details, a window pops up showing the id 9928299, that this apartment is for sale, has 3 bedrooms [...]

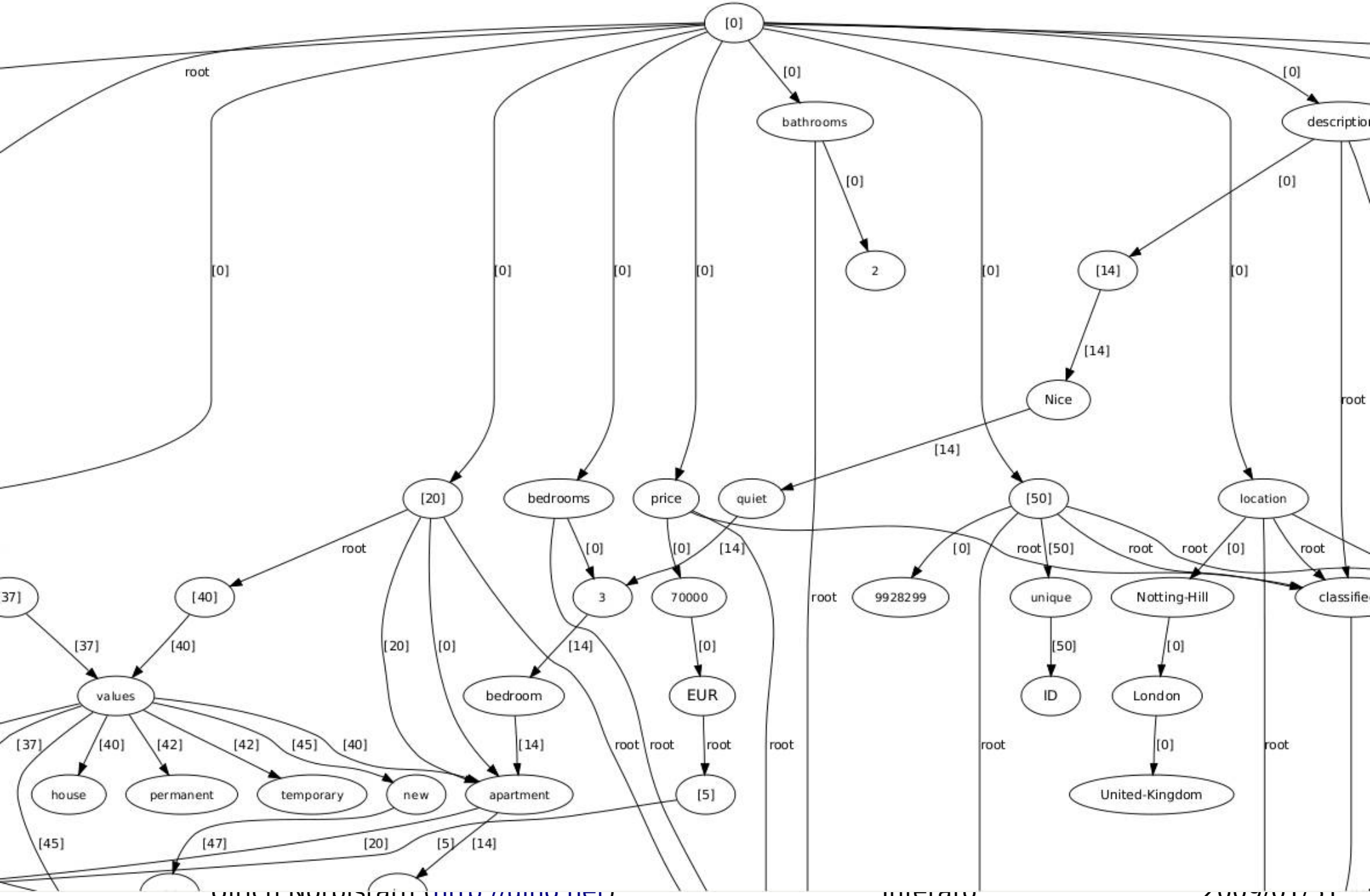
Scenario Based Search by Example: Object-Graph



Scenario Based Search by Example: Code

```
building, """
    {type property}
    |{description [Nice quiet 3 bedroom apartment in Notting
Hill with a perfect view and short ways to groceries and
medical facilities]}
    |{[rent type] sale}
    |{[unique ID] 9928299}
    |{price 70000 EUR}
    |{[apartment type] apartment}
    |{bathrooms 2}
    |{bedrooms 3}
    |{location Notting-Hill London United-Kingdom}
    """
```

Scenario Based Search by



Extensions and Visions

- Desktop / home user
  - look at your files
    - music collection
    - media collection
    - work/related files
  - not pure indexing like google-desktop, beagle
  - more like kde's semantic desktop: nepomuk
  - new storage paradigm
    - new document storage
    - new user interface, allowing complex queries
      - similar to directory but more like a graph
      - 3D browser with wiimote?

Extensions and Visions

- Business:
  professional Knowledge Management
  - consulting process
  - agile analysis method to create day to day search-cases
  - scrape the business data, add special data warehousing
  - new data input methods
  - per case user interfaces

Application: Inferato

- First industrial application (feasibility study in progress)
  - inferato, to inferate
  - infer → deduce information out of set of facts
- cases
  - real estate
  - jobs
  - classifieds
  - events

Summary/Questions

- Today's search reaches its limits
- Context is important but Semantic Web is not enough, infeasible
- Can graphs for modeling context combined with scenario based analysis and development help?

- What about using/learning bigger inference scenarios?
- Good Graph-DB, Graph-browsers?