



OTTO VON GUERICKE  
UNIVERSITÄT  
MAGDEBURG

INF

FAKULTÄT FÜR  
INFORMATIK



Data & Knowledge Engineering Group

# How to Evaluate Exploratory User Interfaces?

SIGIR 2011 Workshop on "entertain me":  
Supporting Complex Search Tasks

Tatiana Gossen, Stefan Haun, Andreas Nürnberger

Email: [tatiana.gossen@ovgu.de](mailto:tatiana.gossen@ovgu.de)

# Agenda

- Introduction & Background
- Evaluation challenges
- Methodological shortcomings
- Benchmark evaluation
- Conclusion

- Complex Information Needs (CIN)
  - Creative discovery of information, i.e. relations between concepts in data sets
    - Simple example: build association chain between *amino acids* and *Gerardus Johannes Mulder*

- Complex Information Needs (CIN)
  - Creative discovery of information, i.e. relations between concepts in data sets
    - Simple example: build association chain between *amino acids* and *Gerardus Johannes Mulder*
    - Using Wikipedia as a document collection:
      - **Amino acids** are critical to life, and have many functions in metabolism. One particularly important function is to serve as the building blocks of **proteins**, which are linear chains of amino acids. Amino acids can be linked together in varying sequences to form a vast variety of proteins.
      - **Proteins** were first described by the Dutch chemist **Gerardus Johannes Mulder** and named by the Swedish chemist Jöns Jacob Berzelius in 1838.

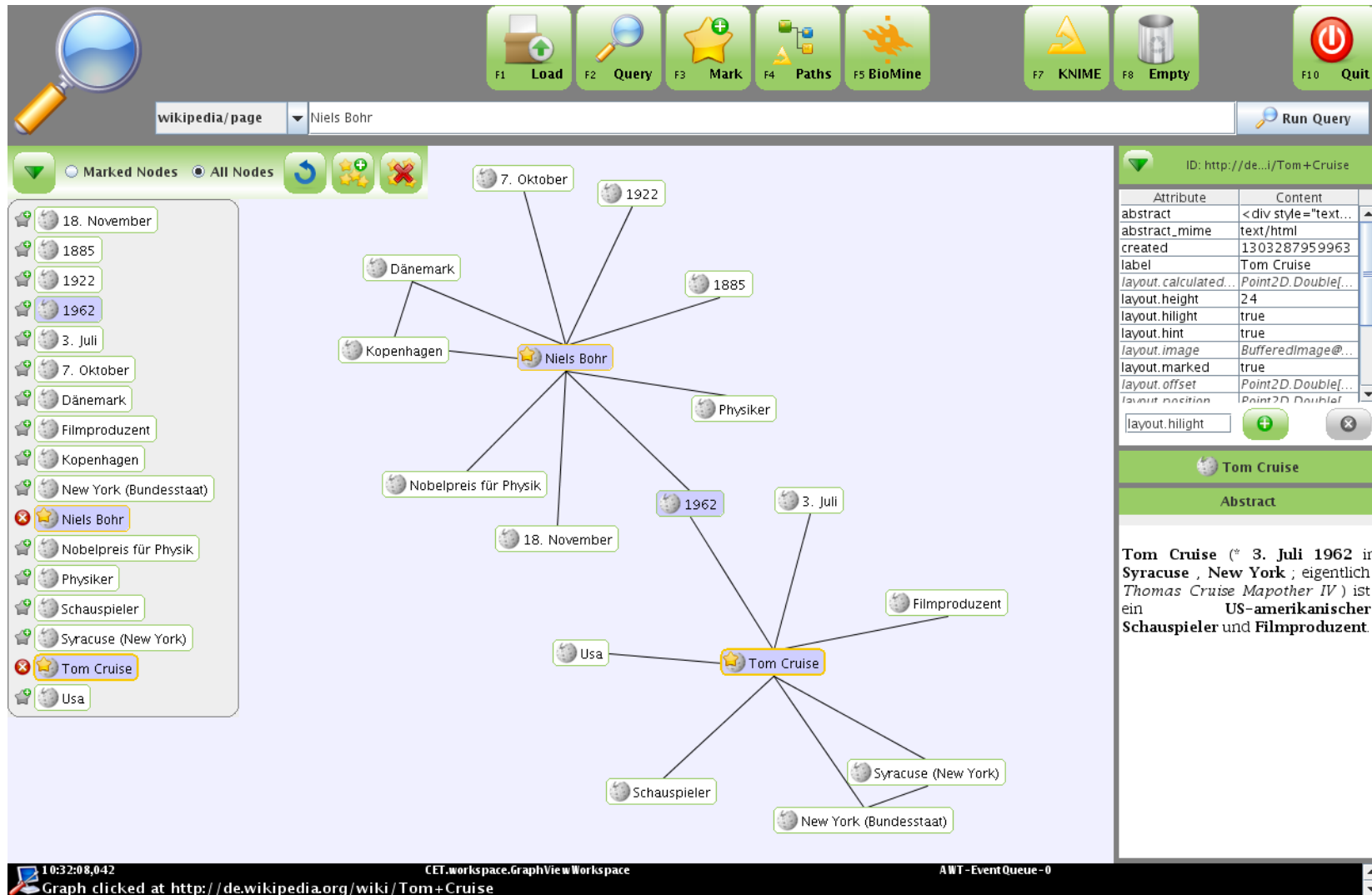
Doc 1

Doc 2

- Complex Information Needs (CIN)
  - Creative discovery of information, i.e. relations between concepts in data sets
    - Undirected search for relevant information within the data
    - *Scenario*: analysts explore collections of text documents to help investigators uncover stories, plots, and threats embedded.

# Introduction & Background

## Tools example



The screenshot displays the Creative Exploration Toolkit (CET) interface. At the top, there is a toolbar with icons for various functions: Load (F1), Query (F2), Mark (F3), Paths (F4), BioMine (F5), KNIME (F7), Empty (F8), and Quit (F10). Below the toolbar, a search bar contains the text "wikipedia/page" and "Niels Bohr". A "Run Query" button is visible on the right. The main workspace shows a network diagram with two central nodes: "Niels Bohr" and "Tom Cruise". "Niels Bohr" is connected to nodes for "7. Oktober", "1922", "Dänemark", "1885", "Kopenhagen", "Physiker", "Nobelpreis für Physik", and "18. November". "Tom Cruise" is connected to nodes for "1962", "3. Juli", "Usa", "Filmproduzent", "Schauspieler", "Syracuse (New York)", and "New York (Bundesstaat)". On the left side, there is a list of "Marked Nodes" and "All Nodes". The "Marked Nodes" list includes: 18. November, 1885, 1922, 1962, 3. Juli, 7. Oktober, Dänemark, Filmproduzent, Kopenhagen, New York (Bundesstaat), Niels Bohr, Nobelpreis für Physik, Physiker, Schauspieler, Syracuse (New York), Tom Cruise, and Usa. The "All Nodes" list includes: 18. November, 1885, 1922, 1962, 3. Juli, 7. Oktober, Dänemark, Filmproduzent, Kopenhagen, New York (Bundesstaat), Niels Bohr, Nobelpreis für Physik, Physiker, Schauspieler, Syracuse (New York), Tom Cruise, and Usa. On the right side, there is a panel for the selected node "Tom Cruise". It shows the ID: <http://de.../Tom+Cruise> and a table of attributes and content:

Attribute	Content
abstract	<div style="text...
abstract_mime	text/html
created	1303287959963
label	Tom Cruise
layout.calculated...	Point2D.Double[...
layout.height	24
layout.hilight	true
layout.hint	true
layout.image	BufferedImage@...
layout.marked	true
layout.offset	Point2D.Double[...
layout.position	Point2D.Double[...

Below the table, there is a section for "Tom Cruise" with the text: "Tom Cruise (\* 3. Juli 1962 in Syracuse , New York ; eigentlich Thomas Cruise Mapother IV ) ist ein US-amerikanischer Schauspieler und Filmproduzent."

At the bottom of the interface, there is a status bar showing the time "10:32:08,042", the workspace name "CET.workspace.GraphViewWorkspace", and the URL "Graph clicked at <http://de.wikipedia.org/wiki/Tom+Cruise>".

Screenshot of the Creative Exploration Toolkit (CET) [Haun, 2010]

Research question: how to evaluate such systems?

- Requires **collaboration with domain experts** for creating scenarios and participation
- CINS are usually vaguely defined and require **much user time** to be solved

# Methodological shortcomings

- Comparative evaluation
  - IR automated evaluation of ranking algorithms requires:

Available

- Set of test queries
- Document collections with labels according to relevancies (e.g. TREC)
- Measures (e.g. Average Precision)

- CIN exploration system user evaluation requires:

?

- Standardized evaluation methodology
- Benchmark data sets
- Benchmark tasks and standard solutions
- Evaluation measures



- Two parts:
  - “small” controlled experiment
    - Qualitative data, i.e. feedback
    - No explicit task
  - Large-scale study
    - Quantitative data
      - Time
      - Success rate
      - Interaction logs
      - Feedback
- Use *VAST* (Visual Analytics Science and Technology) benchmark data with an investigative task as benchmark data set, task and solution

- Evaluation measures – still open question:
  - How to judge creativity?
  - How to judge partially correct answers?
- Can we do automatic evaluation of exploration systems for CIN?
  - Reduce costs for participants?
- Can we model the user creativity process?

# Conclusion

- Evaluation of CIN exploration tools using
  - standardized evaluation methodology,
  - in combination with benchmark data sets,
  - tasks & solutions,
  - and measures
- Only then can discovery tools designers evaluate their tools more efficiently

