

Evaluating Digital Library Search Systems by using Formal Process Modelling

Christin Katharina Kreutz¹, Martin Blum², Philipp Schaer¹, Ralf Schenkel³, Benjamin Weyers³

¹ TH Köln - University of Applied Sciences

² Schloss Dagstuhl LZI, dblp computer science bibliography

³ Trier University

27th June 23

Technology
Arts Sciences
TH Köln

IR
GROUP



дблп
computer science bibliography



UNIVERSITÄT
TRIER

How do we currently evaluate digital library systems?



How do we currently evaluate digital library systems?

User evaluations with **quantifiable** measures:



How do we currently evaluate digital library systems?

User evaluations with quantifiable measures:

- ▶ Query correctness
- ▶ Time required to satisfy information need
- ▶ Query size
- ▶ Number of clicks
- ▶ Subjective feedback via questionnaire



How do we currently evaluate digital library systems?

User evaluations with quantifiable measures:

- ▶ Query correctness
- ▶ Time required to satisfy information need
- ▶ Query size
- ▶ Number of clicks
- ▶ Subjective feedback via questionnaire

Qualitative user evaluations:



How do we currently evaluate digital library systems?

User evaluations with quantifiable measures:

- ▶ Query correctness
- ▶ Time required to satisfy information need
- ▶ Query size
- ▶ Number of clicks
- ▶ Subjective feedback via questionnaire

Qualitative user evaluations:

- ▶ Think-aloud protocols
- ▶ Query log analysis
- ▶ Open-ended questions for users
- ▶ Interviews with domain experts



That's good, isn't it?

That's good, isn't it?

YES, but not always enough!

Problem and Possible Solution

Problem:

- ▶ No comparison between user-specific information seeking behaviour and a digital library
- ▶ → Overlooking user needs, system requirements

Problem and Possible Solution

Problem:

- ▶ No comparison between user-specific information seeking behaviour and a digital library
- ▶ → Overlooking user needs, system requirements

Our suggestion:

Problem and Possible Solution

Problem:

- ▶ No comparison between user-specific information seeking behaviour and a digital library
- ▶ → Overlooking user needs, system requirements

Our suggestion:

- ▶ Formalise users' task solution strategies
- ▶ Compare with capabilities of system

Problem and Possible Solution

Problem:

- ▶ No comparison between user-specific information seeking behaviour and a digital library
- ▶ → Overlooking user needs, system requirements

Our suggestion:

- ▶ Formalise users' task solution strategies
- ▶ Compare with capabilities of system

How can we compare users' conceptions of search tasks in a digital library with capabilities of such a system?

- Motivation
- **Concept**
- Evaluation
 - Questions and Tasks
 - Study
- Observations and Dataset
 - Strategies
 - Findings
- Conclusion

Idea

How can we compare users' conceptions of search tasks in a digital library with capabilities of such a system?



Idea

How can we compare users' conceptions of search tasks in a digital library with capabilities of such a system?

- ▶ Compare ideal task strategies with corresponding capabilities of a DL



Idea

How can we compare users' conceptions of search tasks in a digital library with capabilities of such a system?

- ▶ Compare ideal task strategies with corresponding capabilities of a DL
- ▶ Compare them with what a person would do using the specific DL



Ideal Strategy



- ▶ vIMM: verified Ideal Mental Model
- ▶ Person's description of them ideally solving a task
- ▶ Independent of DL, fuzzy

Ideal Strategy



- ▶ vIMM: verified Ideal Mental Model
- ▶ Person's description of them ideally solving a task
- ▶ Independent of DL, fuzzy
- ▶ *How would a person **ideally** solve a task if they were free to do it any way they wanted?*

Ideal Strategy → Specific Digital Library



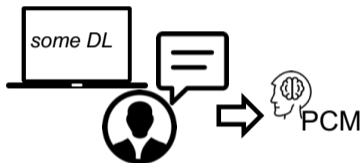
- ▶ vPGM: verified Process Gold Model
- ▶ Translation of a person's ideal strategy to actual DL
- ▶ Hypothetical model constructed by expert

Ideal Strategy → Specific Digital Library



- ▶ vPGM: verified Process Gold Model
- ▶ Translation of a person's ideal strategy to actual DL
- ▶ Hypothetical model constructed by expert
- ▶ *How could a person's strategy be realised using one specific system?*

Specific Digital Library



- ▶ PCM: Process Conduction Model
- ▶ Strategy shown by person actually using one DL

Specific Digital Library



- ▶ PCM: Process Conduction Model
- ▶ Strategy shown by person actually using one DL
- ▶ *How does a person's strategy **actually** look like using one specific system?*



- ▶ vIMM: User's general or ideal strategy to solve a task using their usually preferred systems
- ▶ vPGM: Translation of a user's ideal task solution to the DL
- ▶ PCM: User's actual strategy solving the task by using the specific DL which is evaluated



- ▶ vIMM: User's general or ideal strategy to solve a task using their usually preferred systems → Oral description
- ▶ vPGM: Translation of a user's ideal task solution to the DL
- ▶ PCM: User's actual strategy solving the task by using the specific DL which is evaluated



- ▶ vIMM: User's general or ideal strategy to solve a task using their usually preferred systems → Oral description
- ▶ vPGM: Translation of a user's ideal task solution to the DL → Oral description
- ▶ PCM: User's actual strategy solving the task by using the specific DL which is evaluated



- ▶ vIMM: User's general or ideal strategy to solve a task using their usually preferred systems → Oral description
- ▶ vPGM: Translation of a user's ideal task solution to the DL → Oral description
- ▶ PCM: User's actual strategy solving the task by using the specific DL which is evaluated → Oral description + observation



- ▶ vIMM: User's general or ideal strategy to solve a task using their usually preferred systems → Oral description
- ▶ vPGM: Translation of a user's ideal task solution to the DL → Oral description
- ▶ PCM: User's actual strategy solving the task by using the specific DL which is evaluated → Oral description + observation

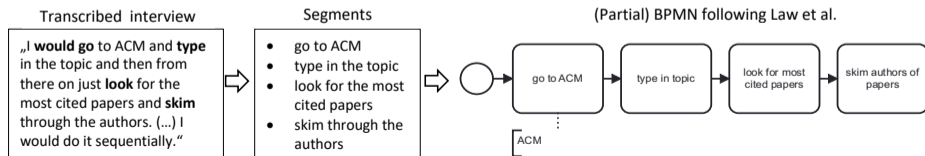
Make strategies comparable by formalising them

Formalisation with BPMN

- ▶ BPMN: Business Process Model Notation
- ▶ Variant for unstructured, process-oriented thinking-aloud interviews by Law et al. ['23]
- ▶ Do not incorporate **modeller's perspective**, only focus on capturing **user's perspective**
- ▶ Approach:
 - ▶ Cut interview in segments (by verbs or time-related phrases)
 - ▶ Classify segments into classes **setting**, **annotation**, **task**, **event**, **condition**, **other**
 - ▶ Classes correspond to specific BPMN elements

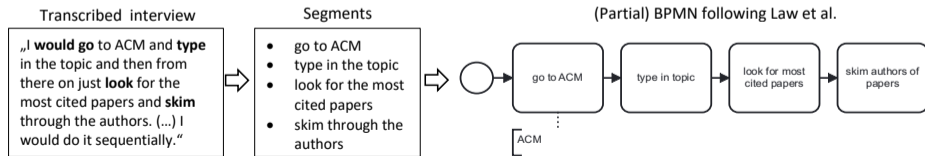
Formalisation with BPMN

- ▶ BPMN: Business Process Model Notation
- ▶ Variant for unstructured, process-oriented thinking-aloud interviews by Law et al. ['23]
- ▶ Do not incorporate **modeller's perspective**, only focus on capturing **user's perspective**
- ▶ Approach:
 - ▶ Cut interview in segments (by verbs or time-related phrases)
 - ▶ Classify segments into classes **setting**, **annotation**, **task**, **event**, **condition**, **other**
 - ▶ Classes correspond to specific BPMN elements



Formalisation with BPMN

- ▶ BPMN: Business Process Model Notation
- ▶ Variant for unstructured, process-oriented thinking-aloud interviews by Law et al. [’23]
- ▶ Do not incorporate **modeller’s perspective**, only focus on capturing **user’s perspective**
- ▶ Approach:
 - ▶ Cut interview in segments (by verbs or time-related phrases)
 - ▶ Classify segments into classes **setting**, **annotation**, **task**, **event**, **condition**, **other**
 - ▶ Classes correspond to specific BPMN elements



Comparable formalisations of strategies

Comparable formalisations of strategies

Comparable formalisations of strategies

... now what?

Comparable formalisations of strategies

... now what?

We can evaluate, if these formalisations help us evaluate DL interfaces

- Motivation
- Concept
- **Evaluation**
 - Questions and Tasks
 - Study
- Observations and Dataset
 - Strategies
 - Findings
- Conclusion

Research Questions

How can we compare users' conceptions of search tasks in a digital library with capabilities of such a system?

- RQ₁ What are users' preferences, which **components** of digital libraries **are usually used** for the predefined tasks?
- RQ₂ How do users utilise the **example system**, which **components** are used for the specific predefined tasks?
- RQ₃ What are the **limitations** of the example DL system? Which components or functions were ignored or missed?
- RQ₄ Is the example system usable for **advanced DL tasks**?
- RQ₅ What are the **discrepancies between the ideal task conduction models of users and their actual task conduction**, how are **models adapted** to solve the predefined tasks?

Tasks

- ▶ **Expert search task:** Find two experts on a topic of your liking.
- ▶ **Paper search task:** Find relevant papers from a topic of your liking which appeared after 2017.

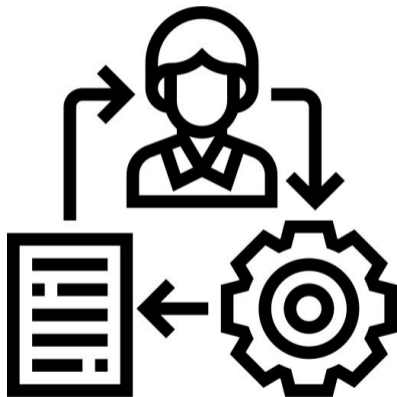
Participants

- ▶ 13 computer/information scientists, differing expertise in using DLs for research tasks:
 - ▶ 2 Master's students
 - ▶ 6 PhD students (first to last year)
 - ▶ 1 industry researcher
 - ▶ 1 dblp staff member
 - ▶ 1 postdoc
 - ▶ 2 professors
- ▶ Code names for anonymity, e.g. *green_deer*



Steps

- i) Pre-study questionnaire
- ii) Interview
- iii) Modelling I
- iv) Verification
- v) Tasks
- vi) Post-task questionnaire
- vii) Modelling II



SchenQL

[Home](#)
[Colors](#)
[Example Queries](#)

schenQL



Hint: Press tab for autocompletion

[WITH DBLPKEY](#)
[WITH DOI](#)
[WITH ISBN](#)
[WITH TITLE](#)
[WRITTEN BY](#)
[WRITTEN BY ANY](#)
[EDITED BY](#)
[PUBLISHED WITH](#)
[ABOUT](#)
[WITH YEAR](#)
[APPEARED IN](#)
[WITH VOLUME](#)
[CITED BY](#)
[REFERENCES](#)
[WITH OPEN ACCESS](#)
[AND](#)
[OR](#)
[NOT](#)
[WITH](#)

[Investigate Query](#)
Title ▲▼

Year ▲▼ **Type** ▲▼

Neural Networks and Deep Learning - A Textbook

2018 citext



ArnetMiner: extraction and mining of academic social networks.

2008 citext



Ideal Strategy (vIMM)

*How would a person **ideally** solve a task if they were free to do it any way they wanted?*



- ▶ Participants describe how to ideally solve tasks
- ▶ Audio-recording, transcription, Law et al. [’23]’s method → **I**deal **M**ental **M**odel (IMM) as BPMNs
- ▶ Participants verify/modify IMM after it has been modelled → **v**erified **I**MM (vIMM)

Ideal Strategy → Specific Digital Library (vPGM)

*How does a person's strategy **actually** look like using one specific system?*



- ▶ vIMM is given to expert in DL
- ▶ Expert translates vIMM to using DL → **P**rocess **G**old **M**odel (PGM) as BPMN
- ▶ Second expert in DL verifies PGM → **v**erified **P**GM (vPGM)

Specific Digital Library (PCM)

*How does a person's strategy **actually** look like using one specific system?*

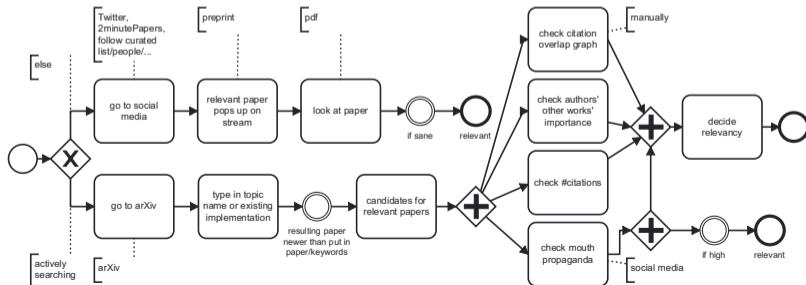


- ▶ Person uses DL to solve task, thinks aloud
- ▶ Screen + audio recording, transcription, Law et al. ['23]'s method + annotations from screen recording → **P**rocess **C**onduction **M**odel (PCM)

- Motivation
- Concept
- Evaluation
 - Questions and Tasks
 - Study
- **Observations and Dataset**
 - Strategies
 - Findings
- Conclusion

Ideal Strategy (vIMM) - *green_deer*

How would a person *ideally* solve a task if they were free to do it any way they wanted?



Ideal Strategy (vIMM) - Observations

Expert search:

- ▶ Features: Keyword search, authors of popular/good papers = experts, # citations, affiliations, references
- ▶ System switch (9)
- ▶ Using Google Scholar (7) or Google search (7)
- ▶ Multiple starting points (2)

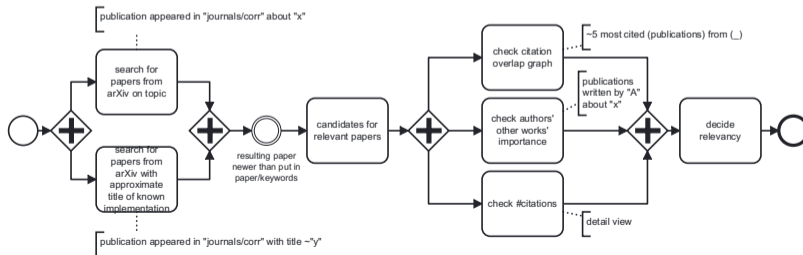
Paper search:

- ▶ Features: Keyword search, following references, related terms, asking others
- ▶ System switch (13)
- ▶ More varied systems compared to expert search
- ▶ Multiple starting points (5)



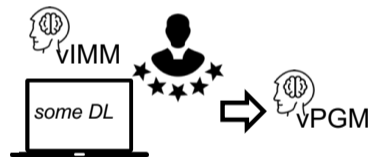
Ideal Strategy → Specific Digital Library (vPGM) - *green_deer*

How *could* a person's strategy be realised using one specific system?



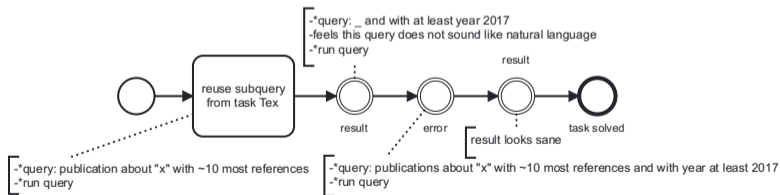
Ideal Strategy → Specific Digital Library (vPGM) - Observations

- ▶ Inability to translate specific parts of processes:
 - ▶ Getting help from person
 - ▶ Google keywords for overview or ranking
 - ▶ Data/Information used not contained in dataset (e.g., publisher)
- ▶ Logical reordering segments
- ▶ Eliminate asking others for information



Specific Digital Library (PCM) - *green_deer*

How does a person's strategy *actually* look like using one specific system?



Specific Digital Library (PCM) - Observations

Expert search:

- ▶ Search for publications about topic (13)
- ▶ Example queries (9), documentation (5)
- ▶ Check person profile (5)

Paper search:

- ▶ Initial query publications about topic (12)
- ▶ Check publication detail view (9)



Specific Digital Library (PCM) - Observations

Expert search:

- ▶ Search for publications about topic (13)
- ▶ Example queries (9), documentation (5)
- ▶ Check person profile (5)

Paper search:

- ▶ Initial query publications about topic (12)
- ▶ Check publication detail view (9)

... and errors in the example system



Research Questions (again)

How can we compare users' conceptions of search tasks in a digital library with capabilities of such a system?

- RQ₁ What are users' preferences, which **components** of digital libraries **are usually used** for the predefined tasks?
- RQ₂ How do users utilise the **example system**, which **components** are used for the specific predefined tasks?
- RQ₃ What are the **limitations** of the example DL system? Which components or functions were ignored or missed?
- RQ₄ Is the example system usable for **advanced DL tasks**?
- RQ₅ **What are the discrepancies between the ideal task conduction models of users and their actual task conduction, how are models adapted to solve the predefined tasks?**

RQ₅: What are the discrepancies between the ideal task conduction models of users and their actual task conduction, how are models adapted to solve the predefined tasks?

Expert in example DL and expert in DLs discuss vIMMs/vPGMs and PCMs

- ▶ Few participants followed vIMMs in PCM (1 expert, 2 paper search)
- ▶ Little overlap in vIMMs and PCMs, seem to have forgotten or do not know how to translate
- ▶ If no syntax problems with example DL, then stuck to vIMM
- ▶ Following models in beginning, then simplification on possibly most important part
- ▶ „*What can I do with this tool?*“ instead of applying usual strategy
- ▶ „*I strongly idealized my search behaviour. (...) my real search behaviour is much simpler*“
→ Overmodelling vIMMs

Discussion

General:

- ▶ Dataset on Zenodo
- ▶ BPMNs suitable to discuss processes with participants, discussions between experts
- ▶ Law et al. ['23]'s method alone limited suitability for PCMs

RQs:

- ▶ Evaluated all 5 RQs
- ▶ People willingly use multiple systems
- ▶ Users heavily rely on examples, entry point for exploration is paper search on topic
- ▶ Ideal processes overmodelled, focus on portion in actual exploration

- Motivation
- Concept
- Evaluation
 - Questions and Tasks
 - Study
- Observations and Dataset
 - Strategies
 - Findings
- **Conclusion**

Recap:

- ▶ BPMNs to depict users' task solution strategies using DLs
- ▶ Compare what users want to do, a systems capabilities and what they are actually doing
- ▶ Find discrepancies between users' perceptions and DL's capabilities → Room for improvement




Future Work

- ▶ User models for simulation
- ▶ Find reasons for users switching systems
- ▶ Design DL with problems in mind

Thank you for your kind attention!

 @kreutzch

 christin.kreutz@th-koeln.de