UNIVERSITY OF TAMPERE

Information Search Processes in Complex Tasks

Pertti Vakkari
University of Tampere
Finland

COMS FACULTY OF COMMUNICATION SCIENCES

UNIVERSITY OF TAMPERE

Outline

- Definitions of basic concepts
  - Task complexity, knowing as changes in knowledge structures, search process
- Information search as a part of complex tasks
  - Search formulation, source selection, working with sources and task outcome
  - Usefulness of information (Information gain) as a criterion for evaluation
- The effect of search process on task outcome
- Conclusions

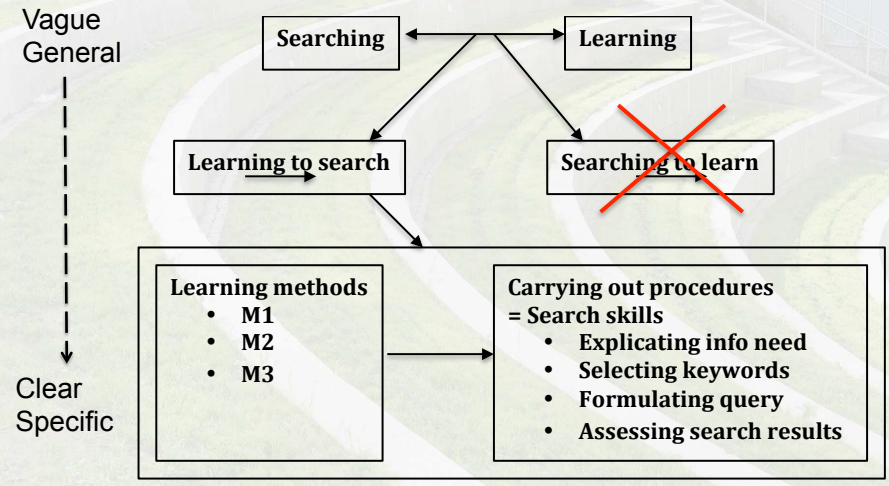
Definitions

- A larger task is an activity to be performed to accomplish a goal, that generates information searching (e.g. work tasks, hobbies) (Vakkari 2003)
  - Information search as a sub-task of a larger task
  - A search task is a set of activities aiming at finding useful (contributing) information for a larger task
- Task complexity
  - Pre-determinability of task outcome, process and information requirements (Byström & Järvelin 1995)
  - Complex tasks are vague and ill-structured typically generating multiple search session
- In this talk focus on complex, information intensive tasks with multiple search sessions

Knowledge structures in human mind and texts

- Knowledge represented as knowledge structures
- Knowledge structures consists of concepts and their relations representing some phenomenon
  - “Searching”, “learning”
  - “Searching to learn” or “Learning to search”
- Structures can be called as mental models or schemas
- Learning means changing or reinforcing existing knowledge structures

Changes in knowledge structures



Changes in knowledge structures

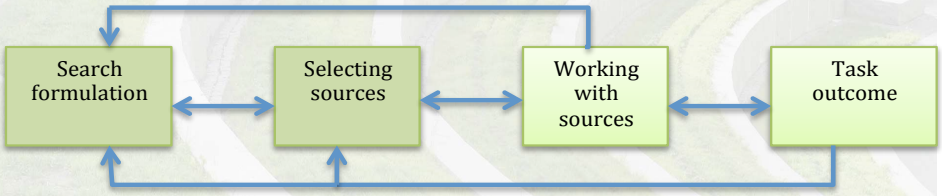
(Piaget 1976, Rumelhart & Norman 1981, Zhang & Soergel 2014)

- *Assimilation* = addition of information into existing knowledge structures
  - *Accretion* = gradual addition of factual information into existing knowledge structures, i.e. no conceptual change
- *Accommodation* = changes in existing knowledge structures
  - *Restructuring* = changing and replacing concepts and their relations in the knowledge structure
  - *Tuning* does not include replacing concepts or their relations in the structure, but merely tuning the scope or meaning of concepts

UNIVERSITY OF TAMPERE

Information search process

Järvelin et al. 2015



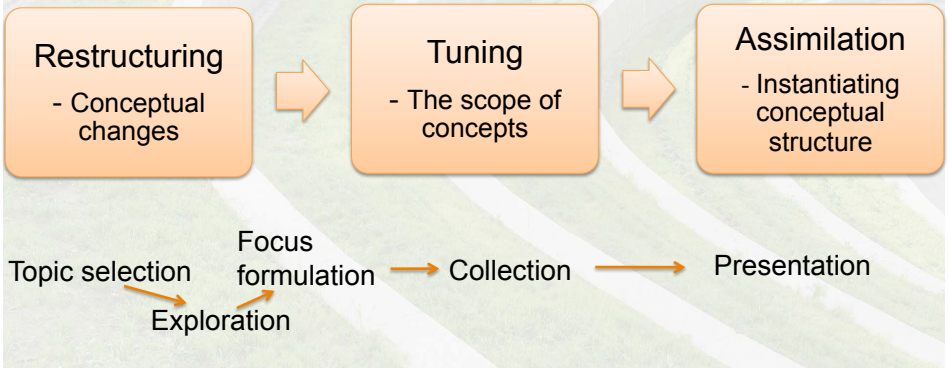
UNIVERSITY OF TAMPERE

Knowledge structures and searching

- My point of departure is the restructuring of knowledge structures in a task process
  - A person is not familiar with the topic of task
  - She is in an ASK (Belkin 1980), or in the beginning of ISP model (Kuhlthau 1993)
- In the course of task performance process the revised knowledge structures stabilize and learning changes from restructuring to tuning and assimilation
- The process resembles Kuhlthau's ISP model
  - *Searching is a thinking and learning process*
  - Searching changes from exploratory to lookup search over several sessions (Marchionini 2006)

Stabilization of knowledge structures

Task performance process



Articulating information need (Vakkari 2010)

- In the beginning of task users seek to articulate information need, to make sense of task topic
  - Lack of sufficient concepts and their relations for representing the topic
  - How to structure the topic-> search of conceptual structures for replacing an insufficient structure (cf. Klein et al. 2006, Qu & Furnas 2008; Russell et al. 1993, Vakkari 1999)
  - Categorize and relate concepts
  - There are no ready-made structures available, but persons have to create them incrementally – working with sources
  - No single right solution, but many options available how to proceed and to specify the structure (cf. Campbell 1988)
- Initial queries are based on a vague conceptual structure
  - Poorly defined search goals

UNIVERSITY OF TAMPERE

Query formulation as conceptual exploration

- Users explore the conceptual space of retrieved documents in relation to their own conceptual structure of the topic (cf. Qu & Furnas 2008)
  - Recognizing potential extended conceptual structures
  - Identifying additional search terms for querying
  - Limiting the extension of topic by adding concepts (structural change) in a query
  - Specifying existing concepts (tuning) in the query
- Vary tactics (replacing concepts) (Bates 1979; Xie & Joo 2012)
  - A,B -> A,C -> A,D
- Increase in the # and specificity of concepts (Vakkari 2010)

UNIVERSITY OF TAMPERE

Intrinsically diverse search (IDS)

- Vary tactics and exploration as means to articulating and structuring task topic resembles IDS
- IDS (Raman et al. 2013)
  - User requires information about multiple aspects of a topic
  - Longer and more complex, exploratory search sessions
  - Query suggestions representing new aspects of the topic, which the user is likely to issue later in the session ≈ *vary tactics*
  - Query suggestions specifications of the original query
- Query suggestions likely help searchers to explore the conceptual (aspect) space and to structure it
  - Help in categorizing a topic, not relating categories
- Do ID Query Suggestions help to identify more aspects?

Search formulation by the changes in knowledge structure

(Vakkari 2001, Vakkari 2002, Vakkari 2016)

Query formul.	The modification of knowledge structures		
	Restructuring	**Tuning**	**Assimilation**
Terms	• Few general terms • A few synonyms	→ →	• Increase in # and specificity of terms • Synonyms
Terms from results	• Much new terms from results • Terms with associative relations	→ Specifying terms	Less new terms from results
Query length	A few terms	→	More terms
Query reformulation	• Vary tactics • Much reformulation	→	Less reformulation
ID query suggestions	Useful, new aspects	→	Not useful
Query similarity	Low	→	High
Session time	Long sessions	→	Short sessions

Source selection

- In *restructuring* person's conceptual structure is vague
  - Difficulties in recognizing useful sources (with potential new structures) (Spink et al. 1998, Tang & Solomon 1998, Vakkari & Hakala 2000, Gwizdka 2014)
  - Vague selection (relevance) criteria
  - Many (potentially useful) documents selected
- In *tuning* the structure is stabilized
  - Easy to recognize useful sources matching with the stabilized conceptual structure
  - Stricter selection criteria -> fewer documents selected

UNIVERSITY OF TAMPERE

Source selection: SERP behaviour

Source selection	The modification of knowledge structure		
SERP Behaviour	**Restructuring**	**Tuning**	**Assimilation**
Usefulness criteria	Vague	→	Clear
Orientation	Recall	→	Precision
Level of usefulness	Possible useful	→	Useful
# Snippets viewed	Large	→	Small
Time on a snippet[1)]	Long	→	Short
Click trough rate	High	→	Low
SERP browsing	Deep	→	Shallow

1) cf. Gwizdka 2014, Smucker & Jethani 2013

UNIVERSITY OF TAMPERE

Source selection: document behaviour

Source selection	The modification of knowledge structure		
Document behaviour	**Restructuring**	**Tuning**	**Assimilation**
# Documents viewed	Large	→	Small
Dwell time/document • **Decision time** • **Usage time**	Varies by utility Long Varies by utility	→ → →	Varies by utility Short Varies by utility
# Documents selected	Large	→	Small
Selected/viewed	Large	→	Small
Used/selected	Small	→	Large
Information type needed	• Background • Conc. & Theories • Procedural	• Procedural • Specific	• Specific • Factual

The usefulness of documents

- Task represented as a collection of terms T from the text a user is working with (Budzig & Hammond 2000) before and after information in a document is used
  - $T_{t2} - T_{t1} = T_{new} = [(\sum \#t_d)/docwords] + e$ = Usefulness
    - $t_{d1,..,n}$ = a new term in text identified in the document
    - # = the frequency of new term in the text
    - e = new terms not identified in the document
      - Influenced by the document > include all new terms = T_{new}
      - Similarity with the new terms identified in the document
- Information about the new terms for task profile to re-rank results or for query reformulation (diversification)

The effect of information search on task outcome

- Kuhlthau (1993), Hersh (1996), Wildemuth (1995) pioneers
- Findings highly varied due to differences in tasks
  - Precision or recall has no impact (Hersh 2003), has a negative effect (Vakkari & Huuskonen 2012) or positive effect (Wildemuth et al. 1995) on task outcome
  - Effort in search process improves task outcome (Vakkari & Huuskonen 2012) or has no effect (Bron et al. 2012, Liu et al. 2012)
- The more effort in result inspection instead of querying, and in working with documents instead of result inspection, the better to the process of construction and task outcome (Liu & Belkin 2012, Butcher et al. 2011, Vakkari & Huuskonen 2012)

Conclusions

- In task performance information search processes patterned due to changes in actors' knowledge structures
- In complex tasks, in addition to querying and result examination, search process consists of working with sources for creating a task outcome
  - The latter stages influence significantly the former ones
  - Studying also information use in documents would deepen our knowledge about querying and result examination, not to speak about the whole search process -> e.g. personalization
  - In addition to accessing information, actors need help in making sense of search results for progressing in the task (Kuhlthau 1993)
- Tools for both accessing and making sense of information (Budzig & Hammond 2000, Ahn et al. 2008, Villa et al. 2009)

Empirical evidence of hypotheses proposed

- Empirical ground insufficient
  - A few longitudinal studies (e.g. Kuhlthau, Wildemuth, Wang, Vakkari, Liu & Belkin)
  - Supported by findings of some studies in pedagogics (e.g. Butcher et al. 2011, Cho et al. 2017) or on the role of topic knowledge in searching (e.g. White et al. 2009, Zhang et al. 2015)
  - Framed with ideas and findings mostly in cognitive psychology (e.g. Gavin 1998, Robertson 2001) or organization science (e.g. Campbell 1988)
- More studies to test the validity of these hypotheses
- Search patterns in successive sessions
  - Personalization between and within sessions

Usefulness as information gain

- Complex search tasks require also other indictors of success than perceived usefulness of documents
  - Search tasks based on cognitive complexity (Kelly et al. 2015)
  - The contribution of information in documents for progressing in the (search) task
- Information gain = $(\sum \#t_d)/docwords + e$
  - An indicator of information derived from a document into the evolving text
  - An objective measure, that can be generated from log data without interfering with user activities

Thank you!