MD-TRACE: A Cognitive Model of Multiple Document Use

Jean-François Rouet
Centre National de la Recherche Scientifique
Poitiers, France

Seminar
Computational Cognitive Modeling of Web Navigation

Acknowledgements

• Anne Britt, Charles Perfetti, Eduardo Vidal-Abarca
• Ludovic Le Bigot, Mônica Macedo, Thierry Olive, Christine Ros, Nicolas Vibert
• Natl Center for Scientific Research (CNRS), University of Poitiers, French Ministry of Education, Natl Research Agency, France Telecom R&D
• Herre van Oostendorp and colleagues at UU for inviting me!
Overview

• Introduction
• Complex document use: research background
• The MD-TRACE model:
  – Overview
  – Resource model
  – Process model
• Heuristic value of the model: some illustrations
• Discussion and perspectives

Introduction (and disclaimer)

• My research interests:
• The acquisition of functional literacy skills.
  "...understanding, using, and reflecting on written texts."
  (OECD's Programme for International Student Assessment)
• How readers interact with texts in functional situations.
  – How people search complex printed or electronic texts.
    Influence of aging, schooling and individual differences.
  – How people remember information from multiple documents.
• Cognitive and language skills needed for successful document use, how to teach them.

= Very little in terms of computational modeling so far 😞

But I expect I'll know a lot more by Sunday 😊
Understanding complex document use:
A sample scenario (Rouet & Britt, subm.)

Lisa, a 10th grader, is working on a science project on the issue of «human activities and climate change». She has found three interesting links on the Web:

- (D1) A web page on "human activities and climate change" designed by a group of high school students (750 words plus various illustrations).
- (D2) A scholarly article reporting on a scientific study posted by a climate research institute on their web site (2500 words plus tables and graphs).
- (D3) The online version of a news magazine commenting on the release of the study (250 words plus one picture).

For the next hour or so, Lisa will study these and other related materials, navigating across pages, copying some of the contents, and eventually sketching a draft or her essay that she then emails to her teacher for feedback.

Complex document use:
Issues for cognitive research

- What kind of prior knowledge and know-how does Lisa need in order to be able to handle her assignment?
- What are her decisions to select, study and retain document information based on?
- In the end, what does she learn about the topic?
- Does it matter that she is working in a digital compared to a printed environment?
- And what if Lisa was a 7th grader, a graduate student or an old lady? Would she then act in a different way and why?
Research background

• Theories of text comprehension (e.g., Kintsch, 1998) explain how people build up mental representations from the sustained reading of short passages.
  ⇒ little to say about selective reading or the integration of information across multiple texts.

• Theories of information search (e.g., Fu & Pirolli, 2007; Guthrie, 1988; Juvina & van Oostendorp 2008; Rouet & Tricot, 1995) focus on selection and localization processes.
  => little to say (?) about the comprehension and actual use of information in a problem-based situation.

Overview

• Introduction
• Complex document use: research background
• The MD-TRACE model:
  – Overview
  – Resource model
  – Process model
• Heuristic value of the model: some illustrations
• Discussion and perspectives
The MD-TRACE model of document use (Rouet & Britt, subm.)

- **External resources**
  - External task specifications
  - Search devices
    - Sources
    - Text organizers
  - Document contents
  - Transient resources

- **Internal Resources**
  - Prior knowledge
  - Reading skills
  - Transient resources
  - Self-Regulation skills

**Processing steps**

**Activity outcome**
TRACE « TASK-BASED… »

External resources
- External task specifications
- Search devices
- Sources
- Text organizers
- Document contents

(1) Create / Update task model

Internal resources
- Task model
- Prior knowledge
- Reading skills
- Self-Regulation skills

TRACE « …RELEVANCE ASSESSMENT… » (1/2)

External resources
- External task specifications
- Search devices
- Sources
- Text organizers
- Document contents

(1) Create / Update task model
(2) Information needed?

Internal resources
- Task model
- Prior knowledge
- Reading skills
- Self-Regulation skills
TRACE « …RELEVANCE ASSESSMENT… » (1/2)

External resources

External task specifications
Search devices
Sources
Text organizers
Document contents
Reader-generated products

(1) Create / Update task model
(2) Information needed?
No
(4) Create / Update task product
(5) Product meets goals?
Yes
Done

TRACE « …RELEVANCE ASSESSMENT… » (2/2)

External resources

External task specifications
Search devices
Sources
Text organizers
Document contents

(1) Create / Update task model
(3a) Select document
Yes
(2) Information needed?
TRACE « AND CONTENT EXTRACTION »

External resources

External task specifications

Search devices
Sources
Text organizers

Document contents

(1) Create / Update task model

(3a) Select document

(3b) Process text contents

(2) Information needed?

Internal resources

Task model

Prior knowledge

Reading skills

Text model

Self-Regulation skills

(Rouet, 2006)

MD-TRACE « MULTIPLE-DOCUMENT TRACE »

External resources

External task specifications

Search devices
Sources
Text organizers

Document contents

(1) Create / Update task model

(3a) Select document

(3b) Process text contents

Yes

(3c) Create / Update Documents model

(2) Information needed?

Internal resources

Task model

Prior knowledge

Reading skills

Text model

Documents model

Self-Regulation skills
The role of sources in integrating information from multiple documents

(D1) A web page on "human activities and climate change" designed by a group of high school students. Conclude that pollution is bad but global warming is natural.

(D2) A scholarly article reporting on a scientific study posted by a climate research institute on their web site. Data strongly suggest human cause of warming.

(D3) The online version of a news magazine commenting on the release of the study. Conclude that humans have an impact.

Students argue « no impact »

Britt et al., 1999;
Britt and Rouet, in prep.
Perfetti et al., 1999

News magazine argues « impact »

Evidence support

Scholarly article argues « impact »

The MD-TRACE model: REGULATION (1/2)
The MD-TRACE model: REGULATION (2/2)

Overview

- Introduction
- Complex document use: research background
- The MD-TRACE model:
  - Overview
  - Resource model
  - Process model
- Heuristic value of the model: some illustrations
- Discussion and perspectives
HEURISTIC VALUE OF MD-TRACE:
I. COGNITIVE RESOURCES AND RELEVANCE DECISIONS

External resources

- External task specifications
- Search devices
  - Sources
  - Text organizers
- Document contents
- Reader-generated products

(1) Create / Update task model
(2) Information needed?
(3a) Select document

Yes

Internal resources

- Task model
  - Prior knowledge
  - Reading skills
  - Self-Regulation skills

Awareness of information needs develops gradually during childhood

• Primary school children (ages 8-11)…
  – Typically can’t generate search questions based on a topic or scenario (Kobasigawa, 1983; Dinet et al., 2004).
  – Can’t always tell if a question will require them to look up an external resource (Moore, 1995; Raphael, 1985; Rouet & Coutelet, 2008)

• In the context of Web-based tasks, their selection of information depends
  – On the visual saliency of the materials (list order, font, keywords)
  – And on their prior elaboration of the search topic
Students’ relevance decisions on Web-like topic lists
(Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, in press, Learning and Instruction)

Topic
« Highest mountains in the World »
vs. « All about HIGH MOUNTAINS »

Selections as a function of grade, relevance and cueing
(Rouet, et al., in press)

Relevant: "All about HIGH MOUNTAINS"
Irrelevant: "HIGHEST railway of the WORLD"
Pre-search cognitive elaboration improves relevance assessment in good readers

HEURISTIC VALUE OF MD-TRACE:
II. INTERFACE ISSUES

External resources
- External task specifications
- Search devices
  - Sources
  - Text organizers
- Document contents
- Reader-generated products

Internal resources
- Task model
- Prior knowledge
- Reading skills
- Text model
- Documents model
- Self-Regulation skills

(3b) Process text contents
(3c) Create / Update Documents model
(4) Create / Update task product
(5) Product meets goals?
Effects of parallel windows on the effectiveness of online notetaking

Simultaneous display (e.g., large screens)
- Faster cycling (approx. 5 vs. 1 per min).
- Higher writing fluency
- Notes closer to original content (exp. 2)

Alternate display (simulates smaller screens)

HEURISTIC VALUE
III. The role of source information in document integration
Inexpert source evaluation: Naïve attributions and heuristics

- Web page written by a medical doctor has to be credible (11th grader, Brem et al. 2001)
- Campers' blog on UFOs more credible than astronomer since they saw it (6th grader, Rouet et al., in prep.)
- Passage from a "science" Web site looks harder to comprehend (undergraduates, Eastin, 2001; Le Bigot et al., 2007)
- Can't challenge the views expressed by the President (US undergraduate, Britt et al., 1999)
- American authors always try to justify US action (French graduate, Rouet et al., 1997)
- See also Braten, Stromso, and Britt, 2009; Braasch, et al. in press; Stadtler and Bromme, 2008

Conclusions and implications

- A model of Web navigation has to take into account users’ representation of the task and their awareness of available resources
- The Web user’s ability to sustain/update the task model throughout the activity is key to effective document processing.
- Sources play a critical role in the intertextual integration of information.
- Implications for the cognitive modeling of Web navigation
  - Model has to include (a) the task context (b) the user’s cognitive resources and (c) structural features of the environment
  - Need to perform a detailed analysis of tasks, both formal and subjective
  - Need to assess users’ search strategies and awareness of Web features
  - Need to model the dynamic updating of memory representations as a function of the ongoing integration of information
Some further research issues

- Mapping cognitive resources onto processing steps
- When/how is the task model updated?
- How accurate is one’s feeling of information need?
- Does relevance monitoring apply to all steps?
- How do users balance surface and deeper relevance cues?
- Is the place of the information encoded as part as source information? (semantic vs. episodic cues in users’ memory for navigation experience?)

Some bibliographical references