A Naturalistic Model of Abstracting

Brigitte Endres-Niggemeyer, Polytechnic of Hannover, Germany

Abstract: A process model of abstracting has been developed, integrating the experiential knowledge of six expert abstractors. The model comprises knowledge about the experts' process organization, a toolbox containing their thinking tools, and a set of natural working steps by which the model is grounded in the experts' professional practice. In addition to its empirical value, the model prepares an implemented simulation system.

1. Introduction

Abstracting, or abstract writing, is a professional specialization of text summarization for bibliographic information systems, dealing with long documents which are scientific or technical most of the time. The process model of abstracting presented here is expected to explain how abstractors go about (Hutchins, 87). With respect to its information quality it is a naive qualitative model of expert performance (Norman 83; Weld and de Kleer, 90).

2. Modelling Methods

Modelling the abstracting process means to develop a grounded theory and a naturalistic model (Glaser and Strauss, 67; Diesing, 71; Lincoln and Guba, 85) or, following the KADS terminology (Wielinga et al., 91), a conceptual model of abstracting. Fortunately, abstracting can be explored in a "follow-up" modelling style, using an applicable domain-specific model of text understanding and summarizing (Kintsch and van Dijk, 83) and of text production (Hayes and Flower, 80). The text understanding model globally tells us that summarizing means to reconstruct the macrostructure of a text using cooperating strategies, whereas the text production model provides an overall process structure of text production, including planning, translating, and revising phases.

36 abstracting processes of 6 experts from Germany and the United States have been recorded on tape via thinking-aloud protocols, transcribed and interpreted (Hayes and Flower, 80; Ericsson and Simon 80, 84). Model-driven data analysis has expanded the initial model of the abstracting process provided by (Kintsch and van Dijk, 83) and (Hayes and Flower, 80) with personal, mostly experiential know-how of expert abstractors.

3. The Naturalistic Model of Abstracting

As a result of empirical modelling, we know better how expert abstractors organize their working process, which intellectual tools (standard strategies) they use, and how successful natural working contexts (steps, moves) look like.

3.1 Organizational Principles of Abstracting Processes

1. Expert abstractors work step by step. In every step several strategies cooperate. A leading strategy defines the overall goal of the working step (e.g., to acquire information, or to fornu-
late a text unit) and links the step to the overall working plan. All strategies fit into an ordered intellectual toolbox.

2. The working processes have a general pattern of knowledge processing: Knowledge items from the original text are isolated and either dropped from processing or moved into the target representation form, passing through intermediate products. Beyond this basic pattern of knowledge processing, experts follow a loosely coupled overall working plan that integrates experiential case-bound know-how. Typical units of this plan are the exploration of a table of content, or the formulation of a statement.

3. Professional expertise is embedded into general skills. The experts invest metacognitive activity (general monitoring and self-steering - Flavell, 81; Gamer, 87), they draw heavily on general literacy skills (reading, writing, and thinking), and they dispose of task-oriented control strategies for their working process.

4. Task oriented memory areas (schemata) structure the factual knowledge that is used during a working process. By their built-in views they impose on their data a structure that is appropriate for the intended type of processing, and exclude features that do not fit. Three source text oriented schemata are needed: surface text, document scheme, and document theme. Since output is smaller in size, one can do with one product area that contains different subareas.

<table>
<thead>
<tr>
<th>Planning, Control, and General Thinking Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>plan:</strong> State what you are going to do next.</td>
</tr>
<tr>
<td><strong>question:</strong> Formulate questions in order to answer them from input, e.g., from your document.</td>
</tr>
<tr>
<td><strong>inference:</strong> Infer implied knowledge.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>General Literacy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>underline:</strong> Underline a text passage.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Information Acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>start-explore:</strong> Begin to explore the document.</td>
</tr>
<tr>
<td><strong>explore:</strong> Get a document meaning item.</td>
</tr>
<tr>
<td><strong>hold:</strong> Keep an information unit.</td>
</tr>
<tr>
<td><strong>first:</strong> Look at the beginning of text organization units.</td>
</tr>
<tr>
<td><strong>browse:</strong> Explore the document by normal reading and understanding.</td>
</tr>
<tr>
<td><strong>read:</strong> Read sequentially.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relevance Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>relevant-say:</strong> State a topic item in your own words.</td>
</tr>
<tr>
<td><strong>relevant-unit:</strong> An information item is recommended as relevant by its position (beginning or end).</td>
</tr>
<tr>
<td><strong>relevant-topic-sentence:</strong> Topic sentences are relevant.</td>
</tr>
<tr>
<td><strong>relevant-call:</strong> Meaning items that are called by the text theme are relevant.</td>
</tr>
<tr>
<td><strong>relevant-texthint:</strong> Exploit textual hints to decide about the relevance of information items.</td>
</tr>
</tbody>
</table>

Fig. 1: A Choice of Abstracting Strategies
3.2 The Intellectual Toolbox

With its 453 abstracting strategies (see sample in fig. 1), the intellectual toolbox represents an abstractor's methodological expertise. As elsewhere, this expertise is composed of different types of skills (Steels, 90). At a very general level, four classes of skills are observed:

- metacognitive skills of self-monitoring and steering,
- control activities that drive the lengthy intellectual working process of abstracting,
- the general intellectual skills of reading, writing, and thinking,
- the genuine abstracting expertise.

We find two main groups of professional abstracting strategies: strategies for dynamic information acquisition, and presentation strategies. The latter include both tools for informational upgrading and for task-oriented presentation.

3.3 An Individual Working Step

Whereas the intellectual toolbox shows how the expertise of abstractors is composed, individual abstracting steps help to understand how strategies cooperate successfully in natural contexts. Fig. 2 displays a typical abstracting step. In the top windows of the display, input data is represented, the process description figures in the middle, output is presented at the bottom. The process description area in the middle comprises the current segment of the thinking-aloud protocol, and a tree-like structure that characterizes how abstracting strategies cooperate.

Let's observe what happens in a typical abstracting step (for strategy definitions, see fig. 1). Our step is the third of the respective working process. From the two preceding steps the abstractor knows the document title, and he has remarked an outline at the beginning of the paper (cf. fig. 2, memory areas in top of the display). Now he finds out what the document is about.

The abstractor begins the working step with a clear start signal, stating his plan and asking his standard question "Let me see what the article is about." (strategies plan and question). Then he starts out to explore the document (strategy start-explore) and begins the first exploration step (strategy explore). Reading behaviour is basic (strategies browse and read), from the thinking-aloud protocol we learn what has been read. The abstractor is aware of his position in the very beginning of the document (strategy first). He underlines "a series of points" (strategy underline, see external output area in fig. 2) and states "This is his sets." (strategies relevant-say and inference). We conclude from both underlining and rephrasing that the abstractor keeps for later use what he has read (strategy hold). Four frequent relevance assessment strategies account for his reasons:

- relevant-unit, because the abstractor knows that the beginning of a unit is a favourite place of topic sentences
- relevant-texthint, because the current statement is introduced by a typical indicator phrase of topic sentences ("This paper is concerned with...")
- relevant-call, because the abstractor is able to attach the current statement to the document theme ("This is his sets.")
- relevant-topic-sentence, because topic sentences are notoriously relevant for abstractors.

In addition, the abstractor uses his own words to reformulate the statement under consideration (strategy relevant-say).

The newly acquired topic sentence shows up in the document theme area. It has been linked to the title with a paraphrase relation.
This paper is concerned with the widespread tendency to present arguments, or to formulate insights or conclusions, in a series of points.
An empirical model is not only useful for its own sake, it gives us some specification of a simulation system as well. In the KADS methodology, the "conceptual model" prepares the "design model" that guides implementation. An inductive system design according to KADS starts out by a careful reconstruction of observational units, i.e., intellectual tools, working steps, and whole processes, under the roof of an appropriate system model. Since the units of empirical observation are organized in their natural context, it must be possible to win an overall architecture for intact natural units.

Empirical observation supports the following system features:
- working steps as basic units of activity, with a planning and control structure behind
- rather independent small-scale agents in considerable quantities, the intellectual tools
- a blackboard-type communication structure
- a dynamic text representation that emphasizes document structure

Figure 3 presents our sample abstracting step in an implementation-oriented blackboard view (cf. Englemore and Morgan, 88; Jagannathan et al., 89). There, intellectual tools show up as agents that assemble around dedicated blackboards. E.g., the relevance assessment agents use data from the document representation blackboard and discuss their judgement on the relevance board; information acquisition agents define and check the input of interest on the input board; agents of planning, control, and metacognition dip into the working step through the planning board. As usual, agents may activate each other by messages. While the document board and the planning board keep information beyond a working step, other boards are strictly local.

Fig. 3: Blackboard View of the Sample Abstracting Step
On the document board, the three tasks-specific general views of the document are assembled during the working process. The scheme view and the thematic view use RST-like relations (Mann and Thompson, 88) to structure meaning units. Individual agents may develop dynamic task-oriented text views from document board representations. For instance, the strategy relevant-textthint analyses an interesting text passage into an indicator phrase and its scope.

The above sketch allows an explorative implementation. In spite of sharing theoretical background (Kintsch and van Dijk, 83) and some features (e.g., the relevance strategies) with the SUSY system (Fum et al., 82 and 85), the "grounded" system design proposed here differs strikingly from the SUSY design. Most of the difference is due to empirical observation.

References


