Using Sowa's Conceptual Graphs for enhancing hypertext readers performances

Saïd TAZI

LIS/ Université Toulouse I
Place A. France
31420 Toulouse cedex
Tel 61 63 35 64
Fax 61 63 37 98
Email tazi@cix.cict.fr

ABSTRACT

This paper describes a work in progress about the use of Sowa's Conceptual Graphs (CG) [Sowa□84] as a knowledge model of representing semantic of discourse (written text) in hypertext. A cognitive model is also used to represent structural knowledge of the hypertext. The major objective of this work is to improve hypertext documents usage by enriching their structures with these cognitive and conceptual models.
In the cognitive model presented here, we consider three levels in hypertext's structures: Microstructure, Macrostructure and Superstructure [Kintsch□78], [Vandijk□83].
* Microstructure of discourse is a collection of coherent basic units of text. Sentences for instance could be considered of a microstructure level .
* Macrostructure is a network derived from the microstructure by application of some semantic rules as generalization for instance.
* Superstructure is a schematic form that organizes the global meaning of the hypertext.
All these structures levels may be represented in terms of concepts and relations within the CG formalism. Among the interests of CG formalism: the possibility to represent graphically concepts and relations respectively as nodes and links in terms of hypertext concepts, and also its expressive power for representing different levels of hypertext's structure.
In the state of progress of the present work, a trial on what CG brings to the hypertext readers is on the way.

1.INTRODUCTION

The "Lost in Hyperspace" famous problem mentioned by Edwards et al [Edwards□89] is due to the weakness of structure specification [Nanard□91]. This structure can be modeled in terms of CG following the reader's cognitive model proposed by Vandijk [Vandijk□83]. Some solutions could be brought by answering to the question what is "to be lost in hyperspace"?
If someone wants to reach a precise target node while browsing a hypertext document, but can't because of one (or more) of the following reasons:
* He (she) doesn't know if the target exists (but it does).
* It's difficult to choose the right way among the numerous ones available from the current location.
* The distance between two nodes is too long and consequently uncertain.
* The target is not well defined. The reader has to explore the document through a blind research which project him (her) in different directions..
Among all these reasons, the three first ones illustrate either the weakness of semantic represented by nodes and links and/or poverty of the interface. The last one has a methodological source, like the absence of reading strategy.
The research project presented here includes studies on natural language semantic, methodological investigations on writing and reading, and also studies of human interface to access to hypertext. In this paper we will concentrate on how could we design and use a hypertext based on Conceptual Graphs of Sowa as a representation of semantic and the

structure of text. We based our process modeling on a three levels defined in cognitive science field . In the following Section we will present works on model of text comprehension resulting from cognitive studies [Vandijk□83] that specify the three levels of speech structure, we also present briefly the CG formalism. In Section 3, we present pieces of a simplified example from text interviews which can be used as information collection and show how the example is processed to be read as a CG based hypertext. A Discussion on this study with some conclusions are presented in Section 4.

2. MODELS OF SEMANTIC STRUCTURES

2.1 Cognitive model of semantic structure

It is now well admitted that readers of both linear text or hypertext construct a mental representation of document's semantic structure while reading. This representation depends on the nature of readers purposes, we refer here to Vandijk et al studies on model of text comprehension and production [Kintsch□78] The authors of this article proposed a model of discourse comprehension that involves readers analyzing the propositions of a text and forming a hierarchy of propositions. The semantic structure of a discourse is characterized at two levels called Microstructure and Macrostructure. The mental representation provides also information on organization of the most important topics within it, this organization is called Superstructure.

Microstructures

The semantic of hypertext must be described at the local micro level to specify the relations of texts units. The discourse at its surface structure is considered as a set of propositions. Propositions are connected by some semantic relations which can be explicitly expressed by cues in the text or not. Microstructure deal with the propositions whose connections are explicitly expressed in the discourse. Propositions are represented as Sowa's CG. The relations which are not expressed explicitly in the text are inferred by applying semantic rules defined by Sowa (matching, copy, restrict, join and simplify).

Macrostructures

Propositions of the microstructure are transformed into a set of propositions that represent the essential of the discourse. These new propositions are called macrostructure. They define connections between topics of discourse, and are defined from the microstructures by applying semantic rules that represent the process of interpretation with the help of various kinds of knowledge. Linking propositions in a local manner is not sufficient because the whole meaningful must be represented to give a global view of the discourse. Topics of discourse are represented as connected propositions. They are derived from the microstructure by applying the following rules [Kintsch□78] :

* Deletion: Propositions that aren't direct or indirect interpretation condition may be deleted.

* Generalization: Propositions set may be replaced by a proposition which generalizes them.

* Construction: To each sequence of propositions denoted by a set of facts, substitute a sequence of proposition denoting a global fact of which the facts denoted by the microstructure propositions are normal conditions, components or consequences.

In order to test the feasibility if this model and its impact on hypertext reading, the building macrostructures is made manually in the actual state of this study. The aims is to process this construction semiautomatically.

Superstructures

Superstructures play an important role in discourse comprehension. They are used to guide the reader while browsing in hypertext. They represent a schematic form that organizes the global meaning of the hypertext. In traditional paper text the writer uses outlines to sketch main principles he wants to write. Outlines are plans that permit to organize the whole structure of document. They play an important control function in the processing of the application of semantic rules (called by Kintsh macrorules), to construct macrostructures. We quote Kintsch□et al. " ...The macrorules are applied under the control of schema, which constrains their operation so that macrostructures do not

become virtually meaningless abstractions or generalizations ..."
These superstructures must be used by hypertext readers to help them to be aware of where they are.

2.2 Sowa's Conceptual Graphs

Sowa [Sowa□84] proposed a general model or representing knowledge One can distinguish between concrete and abstract concepts. Concrete concepts are elaborated by perception. An abstract concept is CG built (elaborated) from concrete concepts. Aristotle's idea that to every concept are attached a word or symbol and a set of world referential, is used in CG. We don't give here formal definition, but merely some examples to help to read the rest of this article. Concepts are represented between brackets, relations linking concepts are represented within parenthesis (for a more convenient notation). The interest of CG is their adaptability to be processed as nodes and links in hypertext systems. If we use a labeled graph as an interface to represent hypertexts, then concepts are represented as nodes and relations are the labels of links.

Examples
Referents denoting individuals, value or sets as mentioned in the world in the universe of discourse: 'John', {PARIS, TOULOUSE}, 1.78

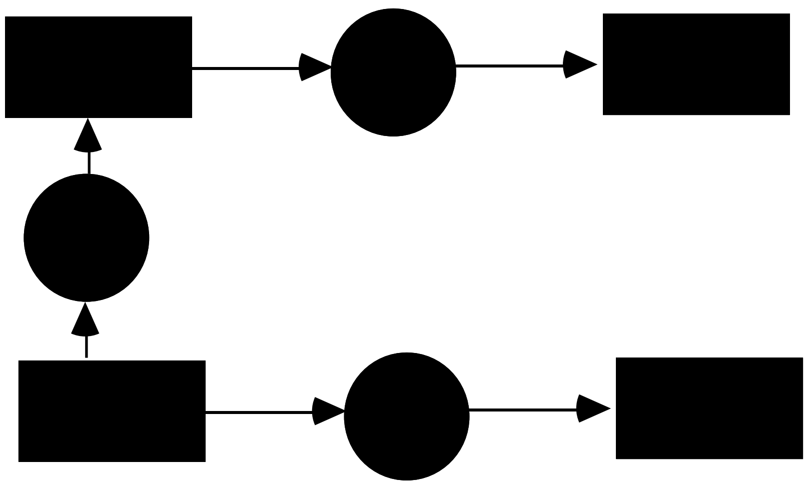
Generic concept:	a man	[MAN : *]
Individual concept	This man	[MAN : #458]
proper noun	John	[MAN : John]
Measure	height =1.78	[HEIGHT: @1.78].

(LOC), (AGT), and (INST) denote the conceptual relations "localization", "agentive", and "instrumental".

CG may represent propositions within concepts. Here is the CG representing the proposition
"The little cat eats a mouse"

[CAT :#25] ---> (SIZE) ---> [LITTLE : *]
(AGENT)
[EAT : *] ---> (OBJECT) ---> [MOUSE : #5]

Note that (AGENT) is the relation between [CAT :#25] and [EAT : *] concepts.
Which can be graphically represented as



Example of CG

3. EXAMPLE

In order to illustrate this work on what we can do with the models presented above in hypertext systems, let treat briefly a simplified example. During the phase of collecting information in order to specify users needs, the analysts must organize a document which is filled by collected information. The application aimed here is to develop a hypertext for collecting these information by constructing CGs of microstructure, macrostructure and superstructure. These CGs represent the hypertext of the document interviews.

Superstructure

First we assume that the engineer who makes interviews to collect future users needs, has a specific document outline which corresponds to the tasks made over the time and the space to realize interviews. The logical structure of the plan of this document will be considered as the superstructure, so nodes and links of this superstructure are first constructed. A piece of this structure is presented as follows:

```
[Interviewer] <--- (Agent) <--- [Interview] ---> (AtDate) ---> [Date]
                                 (Object)
                                  [User] ---> (Fct) ---> [Director]
                                                          (Head)
                                                   [Marketing Service]
```

This CG represent a piece of the superstructure in which one can read that the interview was made by the interviewer at a certain date .

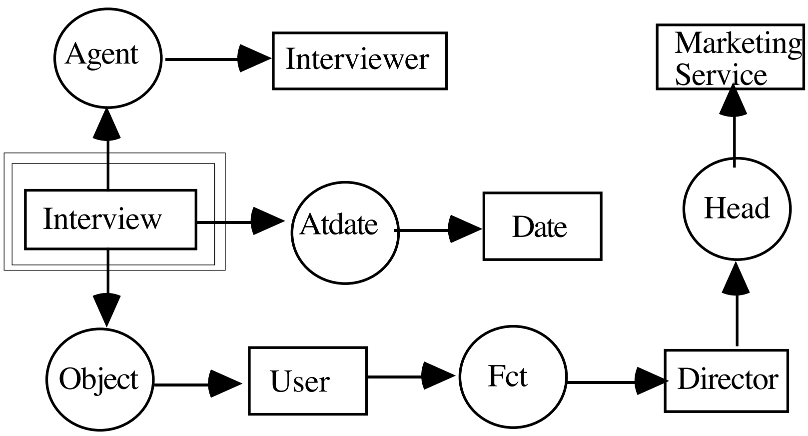


Figure 2

The CG of figure 2 represents a piece of the superstructure in which one can read that the interview was made by the interviewer at a certain date near the Director the Marketing service .

One remarks that in this example the superstructure corresponds to the plan of the interview task. In other type of document it could be as the following:

```
                                 [Section :#2] <--- (Entitle) <--- [title : #5]
                                   (contains)                        (precedes)
                                 [Paragraph : #1] <--- (Entitle) <--- [title : #6]
                                   (precedes)
[Section :#2] <--- (concludes) <---[Paragraph : #2] <--- (Entitle) <--- [title : Conclusion]
```

This simplified example expresses that the section #2 is entitled by title #5, it contains Paragraph #1 which is entitled by the title #6. The Paragraph #1 precedes the conclusion (paragraph #2) of the Section #2, the title of Paragraph #2 is the cue Conclusion.

4. DISCUSSION

We suggest to enrich hypertext with semantic to enhance user's performances. This semantic is represented within CG in three levels. Results from the earlier tests on a highly structured document seem to be significant for the performances of reading. However many questions stay open. The type of document is vital on the representation: Technical, scientific and highly structured documents could be more appropriate to be represented with CG The field of the document is also important, the universe of discourse could be too big in the case of large fields.

The three levels model (micro/macro/superstructure) of Vandijk et al. represents what happens in reader's mind while reading. It gives us a kind of methodology to design structures and interfaces to explore hypertexts depending on reader's purposes. If he (she) wants to get a general view of the document he (she) would navigate into a hypertext at the Macrostructure level. If he (she) wants to read into details all the document, then microstructure is well adapted. Superstructure is a higher level in which the reader can see, at any moment, the outline of the document. An other point must be quoted. The level of macrostructure is completely risky. Indeed if the macrostructure doesn't give access to opportune elements of the document, the problem will persist open. Our contribution to this question is proposed as a methodology point of view (a research on this way is undertaken).

The prototype we are implementing under Toolbook (for rapidity reason) confirms our presumption, that the reader of hypertext feels less lost with the ability of switching easily from a level to another. The work's next step is to implement algorithms to construct semiautomatically at different levels the knowledge base, taking the text base. as input. A deeper reflection is needed to help to specify what are boundaries of micro and macro structures.

5. REFERENCES

[Edwards□89] Edwards D & Hardman L. "Lost in hyperspace" cognitive mapping and navigation in a hypertext environment . In R. McAleese, Ed Hypertext: Theory into practice , pp. 105-125. Oxford Intellect.

[Kintsch□78] Van Dijk T. A. & Kintsch W. "Toward a Model of Text Comprehension and Production" Psychological Review, Vol 85, Num. 5, sept 1978.

[Nanard 90] J. Nanard & M. Nanard "Using Structured Types to incorporate Knowledge in Hypertext" in Hypertext'91 proceedings.

[Sowa□84]. Sowa J. F "Conceptual Structures: Information Processing in mind and Machine", Addison Wesley Publishing Co Reading, MA 1984.

[Vandijk□83] Van Dijk T. A. & Kintsch W. "Strategies of Discourse Comprehension" London: Academic Press 1983.