REPRESENTING AUTHOR'S INTENTIONS OF SCIENTIFIC DOCUMENTS

Kanso Hassan[1,2], Soulé-Dupuy Chantal[1,2] and Tazi Said[1,3]

[1]*Université de Toulouse 1, 2 rue Du Doyen Gabriel Marty 31042 Toulouse, cedex 9 France*
[2]*IRIT, CNRS Université Paul Sabatier118, route de Narbonne, 31062 Toulouse cedex 4, France*
[3]*LAAS, CNRS, 7 Avenue Colonel Roche 31077 Toulouse cedex 4, France*

Keywords: Intentions of communication, Text segmentation, Intentional structure, Electronic document, Experimentation.

Abstract: The existing structures of documents are not ample for nowadays user's needs in terms of search and processing. The Intentional Structure (IS) is a model that maps author's intentions to the segments of documents. It is defined to enhance documents process in terms of goals, means and reasons. The main objective of this work is to provide a methodology of recognizing intentions of communication of scientific documents associated to segments. This article focuses on the representational aspect of the author's intention, by providing a graphical representation of intentions.

1 INTRODUCTION

Processing electronic document is constantly becoming more difficult and more complex, mainly for volumetric and heterogeneity reasons such as the huge varieties of topics, and the complexity of structures. Indeed, the mass of produced and exchanged electronic documents has continued to grow; searching and finding relevant information in this mass is compared to the proverbial needle in a haystack. However, documents are till today generally processed in terms of their logical and physical structures, with the main objective being to represent and treat such documents in terms of their hierarchical organization. The use of these structures has a major interest in order to facilitate the processing of documents such as composing, storing and finding; although the are other concepts for the structure documents, such as rhetoric, semantics and particularly intention are not yet used to help authors to be more explicit and readers to reach knowledge more easily. We assume that if the communicative intentions of documents are made explicit following an appropriate model, the intelligibility and processing of these documents should be enhanced. Indeed, in any rational context, human actions are directed by intentions, i.e. by mental states which represent knowledge related to the desires and beliefs and to the context of the actions. Written communication between humans, in scientific, professional and pedagogical contexts, is also governed by intentions.

Document processing systems do not give authors and users the opportunity to express their intentions explicitly. Current models cannot at the moment represent the author's intentions. Despite wide diversity in the concept of document structure, (e.g. logical structure, layout structure, syntactic and semantic structures), only a few investigations have been carried out on authors' intentions concerning the document structures (Grosz, Sidner, 1990). However, much research has focused on the intentions of dialogue structures and only a few articles concern relations between written documents and intentions. The word intention in this context signifies the effects authors intend to have on their readers.

Intentions are generally implicit, and users (both authors and readers) are sometimes unaware of them. The main idea defended here, is that if authoring systems recognize and represent intentions in such a manner as to make them explicit, they might contribute to render a text more intelligible to readers. Moreover, it would be easier to find and process documents in a corpus by searching in terms of the author's intentions.

Hassan K., Chantal S. and Said T. (2007).
REPRESENTING AUTHOR'S INTENTIONS OF SCIENTIFIC DOCUMENTS.
In *Proceedings of the Ninth International Conference on Enterprise Information Systems - ISAS*, pages 489-492
DOI: 10.5220/0002357604890492

This paper deals with a first step towards the recognition of intention within scientific publications in the computer science field. The idea is that given a model of intention defined in (Al-Tawki, 2002), we aim at finding a methodology of matching a text to this model. At the progress state of this work, the recognition is made "by hand"; we have found certain regularities in the studied corpus which constitute the main new result presented here. The objective of the research conducted here is to formalize these findings and to implement algorithms to automate as much as possible the process of segmentation.

This paper is organized as follows. The next section presents an overview related to the concepts of intention in written communication, and to the field of text segmentation. In section 3, we present the model of intentional structure. Finally, the article finishes by a discussion of future developments.

2 STATE OF THE ART

The research undertaken here aims at analyzing authorial intentions. Each text is segmented into fragments that correspond to the recognized intentions associated with these fragments. This work is related to two main fields: intention and text segmentation.

2.1 Intention

The concept of intention is omnipresent in any human action and is particularly important in communication. Several works attempt to account for the relations between an action undertaken by a human being and the mental state which guides this action. Searle remains a main reference on the matter (Searle, 1983). He distinguishes between two types of intention: intentions in the course of action and former, or pre-formulated, intentions. Former intentions correspond to the representation of the initial goal fixed before the beginning of the action whereas the intention in the course of action accompanies the action during its execution. This distinction makes it possible to treat only intentional actions, and not the "micro actions", or the movements which are not inevitably intentional. Intentions in the course of action are those which represent these intentions, whereas the former intentions represent a condition of satisfaction of the intention (Pacherie 2003). Writing is an intentional action; its characteristic is that it represents two types of action, the physical actions of using a medium to transcribe thought by writing and the actions which aim at modifying the mental state of the reader, by transmitting information, knowledge, advice or orders to him. This second type of action can be accomplished or not depending on the receiver of the written text: the reader. The concept of associating intentions with segments of document was initiated by Grosz and Sidner in (Grosz, Sidner, 1986). This concept consists in describing the intentions of the author for each segment of the document. This description will help to read and consult the document in terms of the intentions of the author. These intentions are added in the form of annotations to the documents marked out with XML tags for example.

According to the theory of Grosz and Sidner (Grosz, Sidner, 1986), the intentional structure makes it possible to represent the structure of the goals. The subjacent objective allows the recognition of the intentions of the author by the reader. These authors identified two structural relations between intentions, fundamental for the analysis of the structure of the discourse at a basic level: the relation of dominance and the relation of satisfaction precedence: an intention I1 dominates an intention I2 if the satisfaction of I2 contributes to that of I1 and an intention I1 precedes (the satisfaction of) I2 if I1 must be satisfied before I2. It is not certain if these two relations are sufficient, on a pragmatic level, to describe the production process of a discourse effectively, because what is interesting in this case, is to be able to associate a finer direction to the relations between various parts of discourse. However, the two relations between intentions suggested by Grosz and Sidner do not account for the large variety of these intentions and may imply loss of semantics. On the other hand, this theory is built so as to depend neither on the domain, nor of the type of the discourse. Indeed, the studies on the modeling of the intention derive from the causal theories of the action. To describe an intention is to find a rational explanation of the action which was caused by this intention. This explanation depends on the context in which the action can be performed. The concept of association of the intentional structures is a concept which consists in describing the intentions of the author for each segment of the document. This description will be able to help to browse the document in terms of authorial intentions.

2.2 Segmentation of Text

Usually, segmentation is defined as determining the positions at which topics change in a stream of text or speech. This is determined by computing word distribution in text with similarity-based or feature-based algorithms. Research from the discourse processing field, inspired by the model of Sidner (Grosz, Sidner, 1986), has investigated the relation between the intentions and the spans of utterances referred to as discourse segments. Segmenting text or multimedia data into coherent regions would have a number of immediate practical uses such as information retrieval or text summarization.

Our research is motivated by enhancing document processing and exploiting intentional structures as a new paradigm. Our goal is to segment texts in terms of author's intentions (Passonneau, Litman, 1993), i.e. to distinguish segments from a text, as being a set of utterances that defines a sub-goal of the author. The author of a written document has a goal when he or she composes a document and particularly in scientific publications. To facilitate the comprehension of the document, the author organizes his ideas as a plan that achieves this goal. Each goal is then a set of sub-goals; the expression of sub-goals in the document is made through utterances as parts of the textual document. Thus on the level of the discourse, a segment is a set of utterances that expresses communicative goals. The structure of a text segment makes it possible to apprehend the sense of this text beyond the sense of each word which composes it. In the analysis of a discourse, the description of its structure consists in cutting out the text as a set of segments (also called fragments), and in identifying the relations which link these segments. An intention corresponds to an action that has a goal; the action is performed thanks to a means and it is justified by arguments that we call reason. A fragment of text is a textual unit that corresponds to a part of intention and it can be a means, a goal or a reason.

The segmentation we focus on corresponds to determining the positions of segments that represent parts of intentions such as what expresses the action, the means and the reason. We suppose that the structure of intentions corresponds to plans of resolution process according to shared plan theory defined by Lochbaum (Lochbaum, 1996), (Grosz, Kraus 1996).

3 OUR INTENTIONAL STRUCTURE MODEL

We propose a new concept which enables us to treat a document in terms of the intentions of its authors. Our objective is to have a representation of the intention through the relations between its constituents. By definition our representation of intention is:

I (A, G, M*, R*)

Where:
I represents the intention carried out by action A;
A is an action which expresses what the author of the intention wants to do;
G represents the goal to achieve by performing the action;
M represents the means to express how the action is accomplished; * to mean that we can have no means or multiple means.
R represents the reason to express why the author chooses this action and for which reasons, * to mean that we can have no reason or multiple reasons.

Intentions can be depicted as a graph as Figure 1 shows.

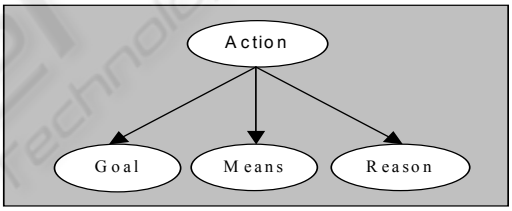

Figure 1: Graph of an intention.

3.1 The Intentional Structure model

The Intentional structure is a hierarchical composition of elementary intentions. An elementary intention corresponds to an action that can not be divided. Thus if we combine the intentions we obtain an intentional structure. A generalized global schema such as the one shown in Figure 2 depicts the intentional structure in a general case.

Each bloc in this model represents an intention; each intention is composed by an action, goals, a means and a reasons. Each Means and each Reason may be considered as an intention or as a final element of the tree. In the Figure 2 the first blocs composed by an action, a goal and a means and a reason that are considered as two intentions. We can take the means or reasons as a new intention bloc and develop it again recursively, but we cannot

develop the goal as an intention because it is considered by definition as a final element of the intentional structure tree.

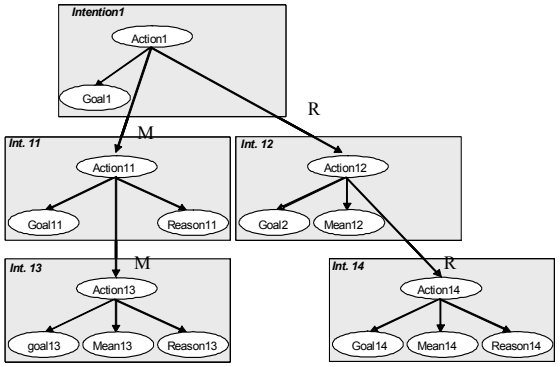

Figure 2: Representation of an intentional structure.

4 DISCUSSION AND CONCLUSION

The utility of our research is to improve the performances of information retrieval systems. Or in other words, to create an organization for these documents in order to facilitate the access to information in complement with traditional Information Retrieval models. This analysis should make it possible to establish a model of intentional structure and to propose a representation of this model. The choice of this manual cutting was done in conformity with competences and the representation of the concept of intention which was developed by Tazi et al (Tazi, 2001).

In our model, we used the ontology to define the whole collections of concepts (Action, Goals, Means, Reasons) which give us a structured of the intentions. Our model of recognition for the intentional structure of a document is based on the segmentations of texts to represent the components of the intentional structure. We chose a corpus of small size initially, to be able to analyze in a qualitative way, and not only in a quantitative way, the various stages of our methodology of representation of intentional structure. Indeed, this analysis becomes more difficult if we choose a corpus of documents of great size.

Our future work is to enhance the structure of the intention and to continue the experimentation for the caracterisation of the relations between intentions, and between the concepts of the Intentional Strcuture. We are working also to build a semi-automatic methodology to recognize intentions of scientific documents to help us to make an automatic segmentation. A building a support tools of assistance to the writing and the reading of documents based on discovered their intentions.

REFERENCES

Barbara J. Grosz, and Sarit Kraus. 1996. Collaborative plans for complex group action. Artificial Intelligence, 86(2):269-357B.

Barbara J. Grosz, Candace L. Sidner - Attention, Intentions and the Structure of Discourse, Computational Linguistics, 12(3), pp. 175-204 1986.

Grosz, Barbara J. and Candace L. Sidner. 1990. Plans for discourse. In P. R. Cohen, J. L. Morgan, and M. E. Pollack, editors, Intentions in Communication. MIT Press, Cambridge, MA, pages 417--444.

Lochbaum, Karen E. 1994. Using Collaborative Plans to Model the Intentional Structure of Discourse. Ph.D. thesis, Harvard University. Available as Technical Report TR-25-94, Center for Research in Computing Technology, Division of Applied Sciences.

Pacherie 03. Dynamic Intentions and Intention Action, Dialogue, XLII, 3, 2003, pp. 447-480.

Passonneau R., J., and Litman D., J., intention-based segmentation: Human reliability and correlation with linguistics cuesÿ, in proceedings of the 31st Meeting Of the association for computational linguistics, pp. 148-155, Columbus, Ohio, June 1993.

Searle, J. "Intentionality". Cambridge: Cambridge University Press, 1983.

Tazi S. et Evrard F., Intentional Structures of Documents, ACM Hypertext' proceedings, University of Arhus, Arhus, Denmark. August 14-18, 2001.

Y. Al-Tawki., Création par réutilisation de documents décrits par les intentions de l'auteur Doctorat de l'Université de Toulouse 1, Avril 2002.

http://protege.stanford.edu/ visited on July 2006.