Available online at https://www.asjp.cerist.dz/en/PresentationRevue/134

RIST

Information Processing at the Digital Age Journal

Volume 27. No. 1 (2023)

A set of rhetorical relationships for educational multimedia document

Azze-Eddine Maredj[a], Madjid Sadallah[a]

[a] *Centre de recherche sur l'information scientifique et technique (CERIST), Alger, Algérie*

Abstract

In this paper, we propose a set of rhetorical relations to support applications such as automatic summary generation and content adaptation of a multimedia document. These relations have been proposed in the context of an educational environment. These rhetorical relations are integrated and handled as part of the logical dimension of the multimedia document. The proposal of these relations is motivated by the need to take into account the particularities inherent to: (1) the composition, editing and presentation of a multimedia document and (2) the educational context. Indeed, multimedia documents in an educational context are very different from textual documents, for which automatic analysis and generation have led to the proposal of a set of commonly used rhetorical relations, as described in the work of Mann and Thompson. The study of this now more common context, allowed us to go beyond the body of existing work to develop a more appropriate set of rhetorical relationships related to educational multimedia documents.

Keywords: multimedia document; educational multimedia documents; RST; rhetorical relationships;

* Corresponding author. Azze-Eddine Maredj. Tel : +213-661-925-659
E-mail address: amaredj@cerist.dz

1. Introduction

Multimedia documents are commonly represented by a model based on four dimensions, Roisin 1998: (1) the temporal dimension, which describes the sequence of entities in time; (2) the spatial dimension, which describes the spatial arrangement of entities, (3) the logical dimension, which decomposes a document into parts and the (4) hypermedia dimension, which defines the links between entities. Recent works have

suggested integrating annotations at the level of the logical dimension for specific application needs Maredj et al. 2021. Yet, for multimedia document application domains such as adaptation (i.e., when rendering is a function of user profile constraints) and automatic summary generation, the information contained in these four dimensions is insufficient to meet the application requirements. Other information is therefore required.

To address this issue, we propose in this work to define and integrate, at the level of the logical dimension of the multimedia document, a set of rhetorical relations that will contribute to support the applications mentioned above.

2. Rhetorical Structure Theory

2.1 Rhetorical structure theory

Rhetorical structure theory (RST) was first proposed by Mann and Thompson, 1988 as a comprehensive theory of textual organization. With more and more attention focused on this theory, RST has been applied to many high-level applications since Marcu's earlier works on RST parsing and its applications on text summarization, Marcu, 1996; 1998; 2000. RST is now one of the most popular discourse structure theories.

The core of RST is a set of rhetorical relations. A rhetorical relation holds between two neighboring textual units. Since RST explains coherence by postulating a tree structure of a text, rhetorical relations are also referred to as coherence relations. According to RST, there are two types of rhetorical relations: mononuclear relations and multi-nuclear relations. In the former ones, one of the text spans is more important than the other one, which play the role of nucleus and satellite respectively. All text spans are equally salient in multi-nuclear relations, all of which play the role of nucleus, Hou et al., 2020.

Broadly speaking, RST is described as having the following characteristics:

1. The abstract structure of a text is a tree-like structure; Following many literatures, we mark the rhetorical structure tree as RS-tree for short, whose leaves correspond to elementary discourse units (EDUs). The internal nodes of an RS-tree correspond to contiguous text spans, each of which is characterized by a rhetorical relation, such as Elaboration or Contrast. The children of each internal node play the role of nucleus or satellite.
2. RST is intentionally limited to a single document; According to RST, a well-formed document can be represented as an RS-tree. To the best of our knowledge, all the published research assumes that each RS-tree corresponds to one input document.
3. RST describes the deliberateness of the writer's writing style; Rhetorical relations capture what writers' intentions are, including what they let the reader recognize? What effect (such as positive regard, belief, etc.) that they want to make?
4. Rhetorical relations exist between two neighboring textual units. A left-to-right reading of the EDU of an RS-tree corresponds to the analyzed text unit in the same order, which guarantees that only consecutive textual units can be linked by rhetorical relations.

2.2 Coherence

Coherence is a metric that is intended to measure how well-formed a discourse is. In concrete terms, the coherence of the text measures the role that each discursive unit plays in relation to the whole. A discourse D is coherent if it can be described as consisting of a set of text spans in which every two spans Si and Sj which have communicative roles between them. Thus, a coherent text precisely answers the following two questions:

1. What can the reader recognize?
2. What does the writer want to express?

Text coherence is therefore defined as the ease with which a person understands a text. RST explains coherence by postulating the RS-tree of each text. Each leaf node of RS-tree corresponds to an elementary discourse unit Hovy, 1993; Mann and Thompson, 1988.

2.3 Elementary Discourse Unit

An elementary discourse unit (EDU) is functionally a simple sentence or a clause in a complex sentence. EDU has the following properties:

1. EDU is the minimal textual unit of RS-trees; EDU acts as the leaf node of each RS-tree, which means that it can't be split into smaller text spans.
2. EDU acts as a syntactic constituent that has independent semantics; In this sense, an EDU corresponds to a clause or a simple sentence, but not a phrase.
3. Some EDUs may contain parenthetical units. In other words, some EDUs may have embedded textual units, whose deletion doesn't affect their semantic meaning.

To obtain the tree structure of a text, according to RST, the first step is to segment this text into EDU sequences. Then the related and continuous EDUs are merged into higher textual spans iteratively until the whole tree structure is derived. Whether two continuous textual spans are related depends on whether there exists a rhetorical relation between them, Hou et al., 2020.

2.4 Rhetorical relation

Rhetorical relation holds between two non-overlapping text spans and reflects how they link together to be a whole. Each text span is called nucleus or satellite with regard to the rhetorical relation. Nucleus and satellite play different roles with respect to the writer's purpose. In general, what nucleus of a rhetorical relation expresses is more essential than what satellite expresses. The nucleus is comprehensible independent of the satellite, but not vice versa. A nucleus element may have more than one satellite element, and if a rhetorical relation does not have a particular element which is more central than the other, it is called a multinuclear relation.

The number of rhetorical relations defined in the original RST paper is 24, Mann and Thompson, 1988. Recent works have added more rhetorical relations for more expressive power. According to the intended effect, mononuclear relations can further be categorized into two types: subject matter relations and presentational counterparts. The writer's intention for using the former relation is to let the reader recognize the meaning of this relation. On the other hand, the effect that the writer wants to make is to increase some inclination on the part of the reader, Hou et al., 2020.

3. Proposed multimedia document model

The model we propose is based on the four dimensions of multimedia documents (figure 1). The logical dimension gives the structure of a document in the form of parts. It reflects the author's point of view. It is at this level that the author's specification of the rhetorical relations that link the entities of the document are integrated. The temporal dimension defines the sequence of entities in time in order to define a coherent temporal scenario. The spatial dimension describes the spatial placement, as specified by the document author, of document entities on a multimedia document rendering device. Finally, the hypermedia dimension expresses the navigational links between the document entities.

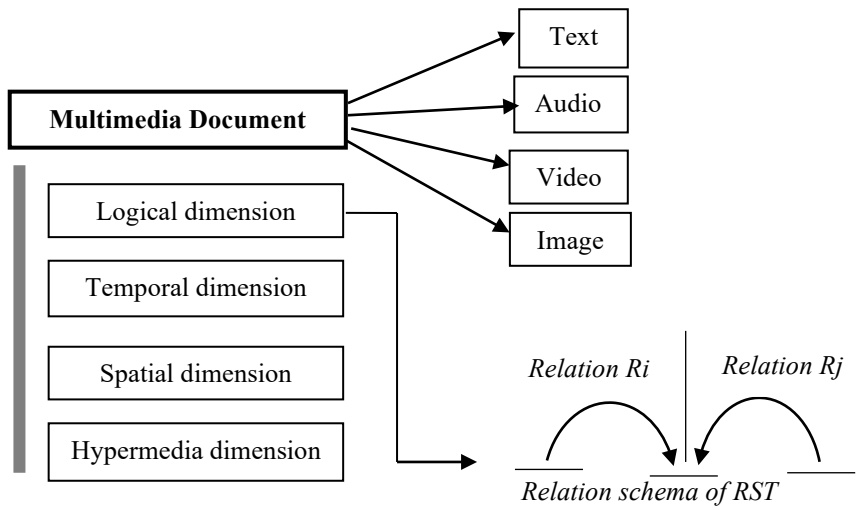

Fig.1. Model of multimedia document enriched by rhetorical information

3.1 The entities of a multimedia document

In a multimedia document, we refer to an entity as the equivalent of the unit concept introduced in RST; an entity can be a text, an image, an audio, or a video. We can distinguish two types of entities: (1) a simple entity which can be a text, an image, a video, an audio, etc. and, (2) a composite entity, which is a logical grouping of a set of simple entities that conveys a complete and coherent information (Figure 2).

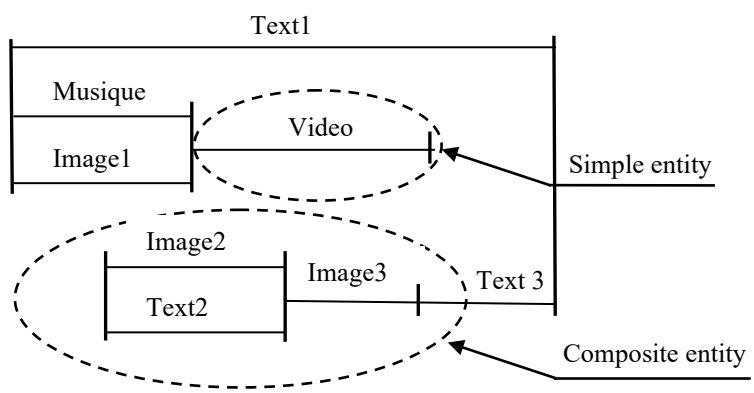

Fig. 2. Defined entities

3.2 Coherence of an entity

As with a unit in RST, the coherence of an entity is defined by the ease with which a person understands

the entity presented; it also measures the role that each entity plays in understanding the document.

A document d is coherent if: Ei and Ej have communication roles with each other, where Ei and Ej are entities in d.

4. Rhetorical relations for a multimedia document

Mann and Thompson's rhetorical relations have been proposed in the context of analysis and coherence control of discourse (text) where the construction (writing) of the units of a text is often of a very fine subtlety and nuance (use of implicit form, metaphors, allegories and other parables) which can make its comprehension (reading) complex and therefore its analysis and coherence control require a very elaborate set of rhetorical relations.

In the case of a multimedia document, on the other hand, the manipulated entities (text, image, video, audio) have a greater completeness character than the units of a text: an entity alone can convey more or less complete information. Consequently, its construction (edition) and its presentation (reading) are simpler. Therefore, the rhetorical relations defined for the analysis and the control of the coherence of a text are not totally adapted for a multimedia document: some relations are not intuitive, and for others, their semantic borders are not clearly established, which can lead to difficulties in their interpretation.

Taking this specificity into account, we adopted the following approach to propose a set of rhetorical relations for a multimedia document in a learning environment:

- Study of the other's intentions during the editing process and proposal of a set of rhetorical relations.
- Refinement of rhetorical relations.

4.1 The author's intentions

When editing a multimedia document, the author selects the information he wants to communicate and then chooses how to present it in a coherent way, so as to convey his message to readers. After an analysis, we have identified the following author's intentions:

1. Introduce, set the context, motivations, objectives of a situation (Introduction).
2. Reformulate or interpret an idea or concept (Clarification).
3. Give an example, pose a problem or exercise (Reinforcement).
4. Demonstrate or justify a situation (Demonstration).
5. Give the solution to a problem or exercise (Solution).
6. Contrast or highlight a similarity (Comparison).
7. Give the condition or anti-condition of a situation (Condition).
8. Give the causes or consequences of a situation or an event (Cause/Effect).
9. Define a composition (Hierarchy).
10. Give a list or sequence of items (Enumeration).
11. Give a title or a legend to an entity (Identification).
12. Add a comment, a warning or a notice (Annotation).
13. Summarize a situation (Summary).
14. Mark the end of an entity, deduce a consequence from a reasoning or an observation (Conclusion).

Table 1 below shows the different relationships retained.

Tableau 1. The proposed rhetorical relations

Rhetorical relationships for educational multimedia document						
Introduction	Clarification	Reinforcement	Demonstration	Solution	Comparison	Condition
Cause/Effect	Hierarchy	Enumeration	Identification	Annotation	Summary	Conclusion

4.2 Refinement of rhetorical relationships

In order to further clarify the author's intentions, as well as to allow for a better interpretation of the rhetorical relations, we have added qualifiers to the identified rhetorical relations. Table 2 summarizes these relations and their qualifiers.

Table 2. Rhetorical relations and their qualifiers

Rhetorical relations	**Qualifier**	**Rhetorical relations**	**Qualifier**
Introduction	Introduction Background Context Motivation Objective	Solution	---
Clarification	Interpretation Reformulation	Comparaison	Opposition Similarity
Reinforcement	Example Exercise Problem	Condition	Condition Anti-condition
Summary	General summary Summary of a part	Cause/Effect	Cause/ Effect
Conclusion	General conclusion Conclusion of a part	Enumeration	List Sequence
Demonstration	---	Identification	Title Caption
Annotation	Metadata Comment Enrichment Appreciation	Hierarchy	Composition Is part of Is a

5. Conclusion

The specificities of a multimedia document, compared to a discourse (text) and the context of a learning environment, led us to propose a set of rhetorical relations, integrated and managed at the level of the logical dimension of a multimedia document to contribute to the support of applications such as the automatic generation of summaries of multimedia documents, the adaptation of the content of multimedia documents. The same approach could be applied in another context. However, the analysis of these applications shows that their support globally results in the execution of entity deletion operations. Therefore, the perspectives of this work are:

1. Propose indicators to determine the importance of each entity in the document so as not to delete the most important ones. One way is to exploit the fact that the core is more important than the satellite.

2. Establish an order of importance between the types of entities in the document: a composite entity is more important than a simple entity (composite entity > simple entity).

3. Establish an order of importance between document entities: video > audio > text > image, i.e., video contains most of the information compared to other media and so on.

References

Carlson, L., Marcu, D., & Okurowski, M. E. (2003). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In Current and new directions in discourse and dialogue (pp. 85-112). Springer, Dordrecht.

Hou, S., Zhang, S., & Fei, C. (2020). Rhetorical structure theory : A comprehensive review of theory, parsing methods and applications. Expert Systems with Applications, 157, 113421.

Hovy, Eduard. 1993. Automated Discourse Generation Using Discourse Structure Relations. 63, 341–385

Isonuma, M., Mori, J., & Sakata, I. (2019, July). Unsupervised Neural Single-Document Summarization of Reviews via Learning Latent Discourse Structure and its Ranking. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 2142-2152).

Ji, Y., & Smith, N. A. (2017, July). Neural Discourse Structure for Text Categorization. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (pp. 996-1005).

Joty, S., Carenini, G., Ng, R., & Mehdad, Y. (2013, August). Combining intra-and multi-sentential rhetorical parsing for document-level discourse analysis. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (pp. 486-496).

Kraus, M., & Feuerriegel, S. (2019). Sentiment analysis based on rhetorical structure theory : Learning deep neural networks from discourse trees. Expert Systems with Applications, 118, 65-79.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Towards a functional theory of text organization. Text-interdisciplinary Journal for the Study of Discourse, 8(3), 243-281.

Marcu, D. (2000). The theory and practice of discourse parsing and summarization. MIT press.

Marcu, D. (1998). The rhetorical parsing, summarization, and generation of natural language texts. Toronto, University of Toronto (Doctoral dissertation, Thesis doctoral).

Marcu, D. (1996). Building up rhetorical structure trees. In Proceedings of the National Conference on Artificial Intelligence (pp. 1069-1074).

Maredj, A. E., Sadallah, M., & Hamouche, L. (2021). Une cinquième dimension pour les documents multimédia : La dimension annotation. Revue de l'Information Scientifique et Technique, 25(2), 12-20.

Roisin, C. (1998). Authoring structured multimedia documents. In International Conference on Current Trends in Theory and Practice of Computer Science (pp. 222-239). Springer, Berlin, Heidelberg.

Taboada, M., & Mann, W. C. (2006). Rhetorical structure theory : Looking back and moving ahead. Discourse studies, 8(3), 423-459.