

Narratives to preserve coherence in collaborative writing

Nishadi De Silva
School of Electronics and Computer Science
University of Southampton, UK
nhds03r@ecs.soton.ac.uk

Hala Skaf-Molli
LORIA-INRIA Lorraine
University Henri Poincaré, Nancy, France
skaf@loria.fr

ABSTRACT

During collaborative writing, shared documents are replicated on geographically distant sites. Each user works on an individual copy. This results in divergent copies. Merging techniques such as those proposed by the Operational Transformation (OT) approach reconcile the differences among the replicas and ensure their convergence. Although the merging techniques resolve conflicting syntax, they do not help preserve coherence which is an important aspect of an effective document. Therefore, we investigate the use of ideas from narrative-based writing to improve the coherence of the document during collaborative editing. Narrative-based writing is a new technique for planning documents that enhances the implicit story conveyed by a document to the readers; thereby improving coherence. This paper presents a discussion of this investigation.

Keywords

Collaborative writing, Merging documents, Narrative-based writing, Rhetorical Structure Theory (RST).

1. INTRODUCTION

Collaborative writing is becoming increasingly common; often compulsory in academic and corporate work. Writing journal papers, technical manuals and planning presentations are a few examples of common collaborative writing activities. In spite of this, it is surprising to find that computer support for collaborative activities is rather poor. Very often collaboration is seen as an optimistic replication problem where shared documents are replicated on geographically distant sites. These replicas are edited simultaneously giving rise to multiple divergent copies of the document. The existing merging approaches such as Operational Transformation (OT) [1] enable the syntactic convergence of these copies. The principle behind these merging methods is to integrate concurrent operations on the various sites so that they reach a state of convergence and resolve eventual conflicts.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...\$5.00.

Resolving conflicting syntax in the merged document is important. However, for a document to reach its full potential it has to also be coherent. Syntactic convergence does not guarantee a document's semantic coherence as demonstrated in the following example.

2. MOTIVATING EXAMPLE

Imagine two authors with the simple task of producing a three-line text to describe their research. Let us assume that both begin with the following set of sentences, called *narratI*:

1. In collaborative writing, a document can suffer from both syntactic discrepancies and a lack of coherence.
2. By integrating a merging algorithm with narrative-based writing,
3. both these issues can be addressed simultaneously.

Figure 1: Initial Text *narratI*

Each author then edits his copy of *narratI* as depicted in Figure 2.

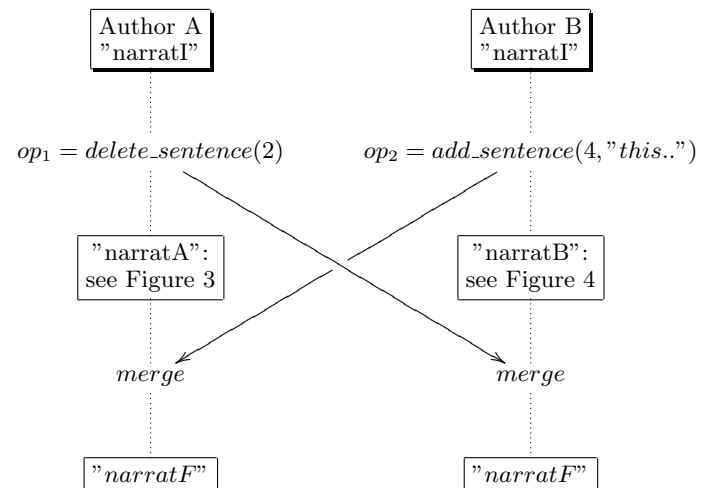


Figure 2: Convergence after integrating concurrent modifications

Author A decides to simplify the text by deleting the second sentence and obtains the text *narratA* (cf Figure 3).

1. In collaborative writing, a document can suffer from both syntactic discrepancies and a lack of coherence.
3. both these issues can be addressed simultaneously.

Figure 3: The text of author A *narratA*

At the same time, author B adds a new line to his copy of the initial text and obtains the text *narratB* shown in Figure 4.

1. In collaborative writing, a document can suffer from both syntactic discrepancies and a lack of coherence.
2. By integrating a merging algorithm with narrative-based writing,
3. both these issues can be addressed simultaneously.
4. This integration is a unique solution.

Figure 4: The text of author B *narratB*

A merging algorithm will ensure that both changes are taken into account and offer both authors the final text *narratF* (cf Figure 5).

1. In collaborative writing, a document can suffer from both syntactic discrepancies and a lack of coherence.
3. both these issues can be addressed simultaneously.
4. This integration is a unique solution.

Figure 5: The final text *narratF*

While the resulting text incorporates the syntactic changes made by the two authors, it does not convey a coherent story to the reader. The fourth line does not make much sense without the second line (in the initial text). In the above example, the defect is easily detectable. However, a similar situation in larger documents such as research papers or reports that are often written by multiple authors would not be so obvious.

The current merging algorithms do not address the problem of semantic coherence. In previous work [2], integrity constraints were used to ensure semantic coherence. However, it was not clear how to capture the semantics of a textual document through logical constraints. We realized that special relationships may be needed to define this coherence.

The notion of document coherence is subjective. Several factors such as language, style and the content can affect how coherent a document is. A new technique called narrative-based writing [3] attributes coherence to the story that a document conveys to the reader. This story is called a document narrative (DN).

A document is expected to be more coherent if its narrative is coherent. Rhetorical Structure Theory (RST) [4] ¹

¹More information about RST in:

is used to analyze the DN and to ensure its semantic coherence. RST is a well established discourse theory that explicates the coherence of a text by virtue of relationships between the segments in the text.

In this paper, we discuss the integration of narrative-based writing with merging techniques. This allows the production of documents that are both syntactically sound and coherent. In the next section, we briefly introduce narrative-based writing and Rhetorical Structure Theory. After that, we illustrate the integration of the two methods to further improve collaborative writing tools. Finally, we outline future work ideas and present our conclusions.

3. NARRATIVE-BASED WRITING

A narrative is a representation of events meaningfully connected in temporal and causal way [5]. For the purposes of this research, it is sufficient to think of a narrative as being analogous to a story. A *document narrative* (DN) is an explicit précis of the story a document conveys to the reader. A DN clarifies the authors' intentions and provides a coherent structure [6]. Narrative-based writing is a new approach proposed by De Silva and Henderson [3] that can organize and enhance the story conveyed by a document.

3.1 Process of narrative-based writing

The process of narrative-based writing can be summarized as follows:

Write the DN Ruminates on the key ideas for the document and formulate the story. For instance, the text in *narratB* in the motivating example can be seen as a short DN for this paper.

Analyze the DN Narratologists and linguists have developed several theories to analyse and synthesise narratives. From among them, Rhetorical Structure Theory (RST) was chosen to ensure that the DN is coherent. RST is explained briefly in the next section.

Write the Document Use the DN and the RST analysis to construct the document such that it implements the story and indicates the relationships.

More about this technique can be found at www.narratives-uk.com.

3.2 Rhetorical Structure Theory (RST)

When a narrative is analysed using RST, it is divided into text segments and relationships are defined between them ². The size of a text segment is arbitrary but each should have independent functional integrity (for example, a clause). A text segment assumes one of two roles in a relationship: the nucleus (N) or satellite (S). Nuclei are considered more essential to the understanding of the text, whereas satellites are secondary. In general, a text should still be comprehensible when its satellites are removed (like a synopsis of the original text).

In the original paper by Mann and Thompson [4], there are 23 RST relations. The four relations used in this paper are described in Figure 6. By defining these relations, the existence of each text segment is justified and it is easy to identify segments that are unnecessary or out of place.

www.sfu.ca/rst/04text_generation/index.html

²This can be done by using the free software tool, RSTTool <http://www.wagsoft.com/RSTTool/>

Relation	Nucleus	Satellite
Background	Text whose understanding is being facilitated	Text for facilitating that understanding
Elaboration	Basic information	Additional information
Justify	Text	Information supporting the writer's right to express the text
Motivation	An action	Information intended to increase the reader's desire to perform the action

Figure 6: Brief descriptions of the RST relations in this paper. Complete definitions in ([4])

Relations also have an associated effect on the reader which gives rise to text coherence. For example, in an Elaboration relation, the nucleus is expected to provide basic information while the satellite contains additional information about the nucleus. In a Motivation relation, the satellite is expected to contain some information that will motivate the reader to perform the action in the nucleus. If the relations in a text can be assembled to form a tree, the text is said to be coherent.

Figure 7 illustrates one possible RST analysis of the text in *narratB*. The figure shows three relations and the tree structure produced.

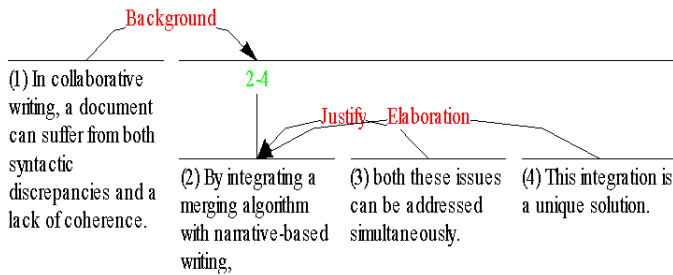


Figure 7: A possible RST diagram for the text *narratB*

In this diagram, labelled, curved arrows are used to show the relations. In relationships between a nucleus and a satellite, the arrow always points towards the nucleus. For instance, in the Elaboration relation (where the satellite is segment 4 and the nucleus is segment 2), the arrow points to segment 2. (Some relationships like Sequence exist between multiple nuclei.)

Text segments involved in a relationship can be grouped together to form a span which can, in turn, become part of another relationship (hence, recursive). This continues until all the segments can be joined by a relationship. For example, in Figure 7, the span formed by segments 2,3 and 4 (linked by relationships) are in a relationship with segment 1.

In [3], the authors propose a technique called narrative-based writing which uses RST to help technical authors enhance the coherence of documents and in [6], the authors introduce narrative-based collaborative writing for technical

authors and propose a simple business process model for it. In this paper, we explore the use of narrative-based writing in collaborative writing further, in particular its integration with merging techniques like OT.

4. COLLABORATIVE NARRATIVE-BASED WRITING

In collaborative writing, just deliberating on the DN and doing a RST analysis can improve coherence. However, if authors work with and change individual copies of the DN, document coherence is, once again, compromised (different authors will be writing according to different DNs). Therefore, we make use of a merging algorithm to converge the replicas of the DN and the corresponding RST analysis.

4.1 Motivating Example Revisited

We revert to the scenario in the motivating example. This time, let us assume that the two authors were using *narratI* as a DN for their joint paper. They did a RST analysis for it to make sure it was coherent and to understand the underlying relationships between parts of the DN (cf Figure 8).

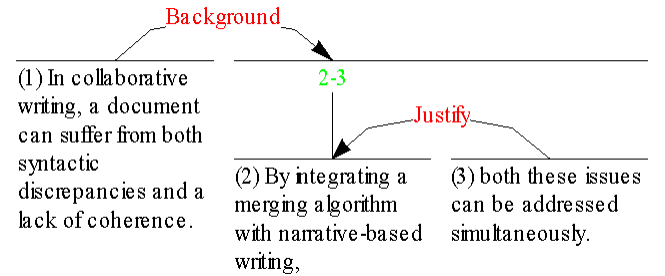


Figure 8: A possible RST diagram for the narrative *narratI*

Now each author starts writing sections of the paper such that it implements this DN and the RST relationships. During this process, authors may also decide to make minor changes to the DN or RST analysis as they see fit. For instance, author A changes the DN to produce *narratA* and author B produces *narratB*. Figures 9 and 7 show the RST analyses for these DNs. To simplify this discussion, we have concentrate only on the modifications to the DN and RST analysis (and do not show the corresponding changes to the paper).

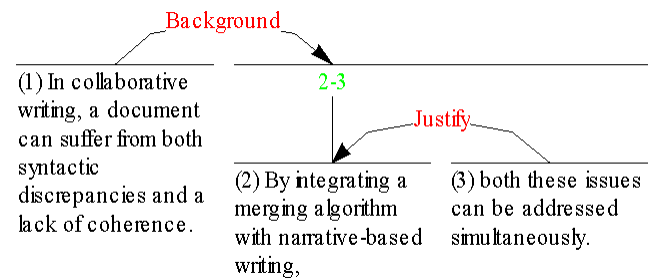


Figure 9: A possible RST diagram for the narrative *narratA*

A merge algorithm can make sure that these changes are

propagated accordingly. Therefore, it can produce a unique value of the DN (*narratF*) that contains both authors' modifications (as shown in the Motivating example). However, *narratF* does not implement a coherent story (i.e. it does not convey the story that the authors had intended for their paper). To further expose this incoherence, the authors can investigate the RST relations in *narratF* (cf Figure 10).

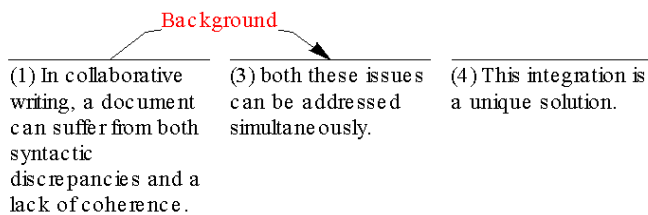


Figure 10: A possible RST diagram for the narrative *narratF*

In this diagram, there are several problems. For instance, segment 4 cannot be related to any other text segment. It is likely to be unnecessary or out of place. This indicates that the DN is incoherent [7]. At the same time, the existence of segment 4 in the merged text shows that at least one of the authors thought it was essential to the DN. However, in order to link it to the RST tree, it either needs to be revised or another segment needs to be added that explains it. A possible solution is to either delete segment 4 or to add a statement like segment 2. The most important aspect of this is that both authors are aware of the incoherent DN. This is highly beneficial when writing with multiple geographically dispersed authors, where a segment in the DN could correspond to sections or chapters in a large document.

The example above shows how we can apply narrative-based writing to enhance document coherence in collaborative writing. RST can help the authors to know that they made incompatible concurrent modifications. The added benefit is that inconsistent relationships can be detected and the authors notified.

4.2 Work in Progress

In collaborative narrative-based writing, there are three main concepts: the DN, the RST analysis and the eventual document that is expected to implement its DN. In the above scenario, we only showed how a merging algorithm can combine changes done to a DN. However, in order to make collaborative narrative-based writing operational, we need to also integrate changes to the RST tree and the document itself. In fact, authors can change only the RST tree without changing the document narrative. For example, imagine a scenario in which instead of deleting segment 2, author A realizes that the story is better if segment 3 is used to motivate the merging of the two methods. So he changes the Justify relation in the initial RST analysis (cf Figure 8) to a Motivation relation. This change will affect how that section of the paper is crafted (perhaps by including an example that motivates the reader to accept that a merging of these parallel strands of research (the action) is useful).

From among several merging algorithms, we chose the Operational Transformation (OT) approach because it allows to define a safe and generic merging tool independently of document type (text file, figure, XML) [8] and it is suitable

for hierarchical document structures as those proposed by RST.

In order to integrate narrative-based writing and the OT approach, we need to define the shared data types (to store the DN and RST analysis) with their operations and finally write and prove transformation functions for all combinations of operations.

We need to also identify possible violations of RST relations. This will allow us to provide the authors with valuable awareness which will help them to detect incompatibility in their concurrent modifications.

5. CONCLUSION

In collaborative writing, authors are often allowed to work on individual copies of the document. This results in discrepancies both in the syntax and the story the document conveys to the reader (or the DN). There are well established methods to deal with syntactic differences such as Operational Transformation approach. However, they do not remedy the semantic coherence.

Narrative-based writing provides a possible solution. It adds a dimension to collaborative writing tools that was not previously done. Just enabling DN and RST analysis features in collaborative writing tools will be a big step forward. Then, authors will not only know the syntactic changes made by the co-authors but also have an idea of the evolving story. This paper discusses our work to combine merging techniques and narrative-based writing to assist the production of documents that are both syntactically consistent and semantically coherent. We believe this will be a breakthrough in software that supports cooperative writing activities.

6. REFERENCES

- [1] Sun, C., Jia, X., Zhang, Y., Yang, Y. and Chen, D.: Achieving convergence, causality-preservation and intention-preservation in real-time cooperative editing systems. *ACM Transactions on Computer-Human Interaction - TOCHI* **5**(1) (1998)
- [2] Skaf-Molli, H., Molli, P. and Oster, G.: Semantic consistency for collaborative systems. In: *Fifth International Workshop on Collaborative Editing - ECSCW'2003, Helsinki* (2003)
- [3] De-Silva, N. and Henderson, P.: Narrative support for technical documents: Formalising rhetorical structure theory. In: *ICEIS*. (2005)
- [4] Mann, W.C. and Thompson, S.A.: Rhetorical structure theory: Toward a functional theory of text organization. *Text* **8**(3) (1988)
- [5] Onega, S., Landa, J.: Introduction. In Onega, S., Landa, J., eds.: *Narratology*, New York and London, Longman Group Ltd (1996)
- [6] Henderson, P. and De-Silva, N.: A narrative approach to collaborative writing - a business process model. In: *ICEIS* (4). (2006)
- [7] Taboada, M. and Mann, W.C.: Rhetorical structure theory: Looking back and moving ahead. *Discourse Studies* **8**(3) (2006)
- [8] Molli, P., Oster, G., Skaf-Molli, H. and Imine, A.: Using the transformational approach to build a safe and generic data synchronizer. In: *Conference on Supporting Group Work*, ACM Press (2003)