From Searching to Browsing through Multimodal Documents Linking

Dalila Mekhaldi, Denis Lalanne and Rolf Ingold
Université de Fribourg, Chemin de Musée 3
1700 Fribourg, Switzerland
{dalila.mekhaldi, denis.lalanne, rolf.ingold}@unifr.ch

Abstract

Relationships that link static documents discussed during meetings to the corresponding speech transcripts can be of various kinds. The most important ones, thematic links, quotations and references are presented in this paper. Thematic links are detected via a thematic alignment process. However, quotations extraction is based on the detection of segments of documents that are quoted in the speech transcript. References, made by speakers to documents, are performed via a matching process between referring expressions detected in the speech transcript, and corresponding documents logical blocks. Finally, a framework that combines these links and an evaluation of the links complementarity are presented.

1. Introduction

The construction of links between related documents in collections has been widely exploited for organizing and archiving the increasing amount of documents available today [1, 2, 3]. Two levels of links exist between related documents. In the first level, called the global level, documents are considered as a whole and are linked each to other through various mechanisms (e.g. similarity, citations, hyperlinks, etc). This level is generally used for searching documents. In the second level called local level, documents are segmented into homogeneous blocks, and links are established between these blocks. While the first level has been greatly studied in information retrieval, for searching and retrieving related documents, few works focus on the second level [4]. However, the local level may be more useful and efficient than the global level for applications such as scientific conferences data archiving and indexing [5] or meetings analysis [6]. In this kind of applications, browsing interfaces are necessary for fully understanding the structure and the content of the event (e.g. conference presentation, meeting). And for this reason, local linking is required in order to align documents and play them synchronously. This alignment makes possible the linking of related parts in each independent data (static documents, slideshows and presenters audio/video recordings), transported by various media types (see figure 1). In this article, we present a case study of the local linking approach that focuses on meeting documents and the corresponding speech transcript. The linking between documents and slideshows is actually under progression. Our work is based on press reviews meetings, in which participants discuss french newspapers front pages that are composed of many heterogeneous articles. The meeting dialogs are currently transcribed manually. However, the integration of an automatic speech recognizer is under work. We will see later in this paper that, in this application, many kinds of semantic relationships

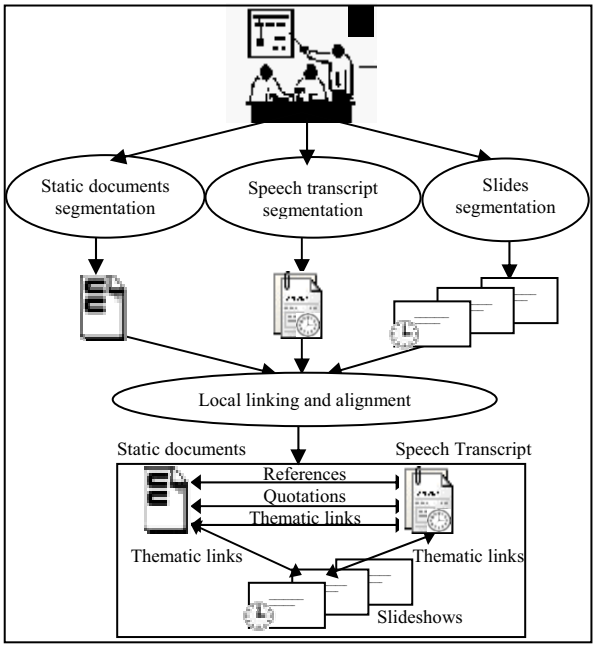

Figure 1: Conference data local linking

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE

might be built between documents and speech transcript, such as thematic links, quotations and references. Section 2 of this paper summarizes some methods of global linking approach. In section 3 we propose our local linking approach which bridges the gaps between printed documents and speech transcript, and three methods for detecting thematic links, quotations and references. Finally, in the last section, a framework that combines all the link types is presented along with a comparative evaluation.

2. Global linking: documents indexing

Many research areas, such as the classification and archiving of documents, use documents linking as a way to connect related documents together. The methods generally used for this task are citations detection, content or structural based techniques. The citation approach [3] is specific to academic and scientific documents. Two documents are related if one cites the other (citation), or if both of them are cited in the same document (co-citation method), or if they cite the same document (bibliographic method). The hyperlinked citation approach [2] is derived from the citation approach. It has been used for retrieving background readings for researchers, through hyperlinked citations in a given document. The content-based approach is appropriate for any kind of documents (academic, literature, etc). It is founded on the comparison of documents content via similarity metrics, generally computed using the frequency of document terms co-occurrences. The structural approach is used to classify documents into their various types (magazines, letters, journal articles, etc), via comparison of their structures, i.e. the layout level. In the content-based structural approach [1, 3] both content and structural information are considered. In the next section, the local linking approach that connects segments from various documents is presented, and is illustrated with documents/speech alignment in the context of meeting analysis. These local links help aligning the related documents, i.e. synchronizing them.

3. Local linking: documents alignment

3.1. Thematic relationship

Finding thematic links between meeting dialogs and static documents consists in finding the parts that are similar thematically, a process called thematic alignment [6]. After stop words removing and stemming processes, the similarities are detected using

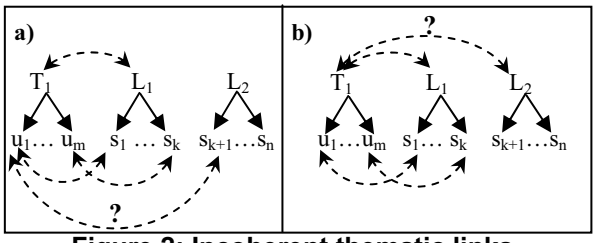

Figure 2: Incoherent thematic links

similarity metrics (*Cosine*, *Dice*, etc). The structures considered in this algorithm are the logical and syntactical structure (decomposition into sentences) for documents, and the segmentation into speakers' turns and into utterances for speech transcript. For reason of compatibility of segments, the compared units should have equivalent size, i.e. compare logical blocks with turns, and sentences with utterances. The thematic alignment process takes the two textual files, computes the similarities between their respective segments, and generates two symmetrical alignment files where all best similarities are considered (*multiple* alignments).

3.1.1 Levels grouping

Due to the difference in the structures of levels (turns/logical blocks and utterances/sentences), the results generated by the algorithm are different. Thus, some established thematic links might be incoherent between the two levels. In figure 2.a, a turn T_1 is aligned only with a logical block L_1, indeed its utterances (u_1,.., u_m) are aligned with the sentences (s_1,.., s_k) contained in L_1. However, there is a link between u1 and s_{k+1} that belongs to L_2, which causes an incoherence between the two levels. Figure 2.b shows another case of incoherence, where a link between a turn T_1 and a logical block L_2 is not validated by their parts. To check and correct this incoherence between levels, we propose an approach of validation and correction, based on grouping levels. It consists in the superposition of each turn and each logical block with their utterances and sentences respectively, and then the detection and the correction of the incoherent links.

3.1.2 Evaluation

The levels grouping algorithm has been tested on 8 meetings, which corresponds to 561 speakers' turns with 927 utterances, and 116 document logical blocks with 529 sentences. The evaluation consists in measuring the change in the values of *recall*/*precision* metrics and their combination *F*. Due to the complexity of evaluating the utterances/sentences level (small size of units and multiplicity of links), only the

turns/logical blocks level has been evaluated. Since the thematic links process is symmetric, only the alignment from speech to documents has been considered. Thus, a manual ground truth has been prepared for the 8 speech transcripts. Our first idea to correct the incoherent links was based on the deletion of those established between turns and logical blocks that are not validated by their parts (e.g. figure 2.b). The initial average values for the *recall/precision* and *F* were (55, 75, 63)%. After the grouping/correction process, the *precision* improved to 85%, while the *recall* decreased to 49%, which may be explained by the deletion of some correct links from turns composed of only one utterance. Given a turn T_i, one of these small turns. T_i is linked with a similar logical block L_j, but its single utterance may not be aligned with any sentences contained in L_j. According to the correction criteria, the link between T_i and L_j should be removed, even if it is correct. To avoid deleting these correct links, the *membership* value of the corresponding turns in their similar logical blocks should be checked, before removing their link. This metric measures the percent of terms of a turn T_i that are present in a similar logical block L_j. If its value overcomes a defined threshold Th1, then the link between T_i and L_j is preserved; otherwise, it should be removed. In order to observe the effect of the threshold *Th1*, its value has been changed from 0 to 0.3 (figure 3.a). We have remarked that its ideal value was 0.12, where the *F* metric reached its maximal value 64%, the *recall* was stable at 55%, and the *precision* increases to 78%. The increase of the *precision* while the *recall* was stable indicates that many false links were removed, and all the correct links are preserved. Afterward, our concern was the improvement of the *recall.* Our idea consisted in adding some correct links between turns and logical blocks that are not linked, according to the weight of their parts links. If the *membership* value of an utterance u_i in a similar sentence s_j overcomes a threshold *Th2*, then a new link should be built between their parents. This second threshold has been changed from 0 to 1 (figure 3.b) while *Th1*=0.12 (to preserve the existent correct links). When *Th2*=0.34, all added links were correct, which increased the *recall* value from 55% to 61%, and the *F* metric reached its maximal value 67% (table 1). With these two thresholds (*Th1=0.12, Th2=0.34*), the proposed grouping/correction approach is very efficient for correcting and improving the thematic links results. Moreover, the superposition of structures generates two new hierarchical structures representing complete information about the thematic links. In the next section, another kind of relationship is described, the quotations of static documents parts.

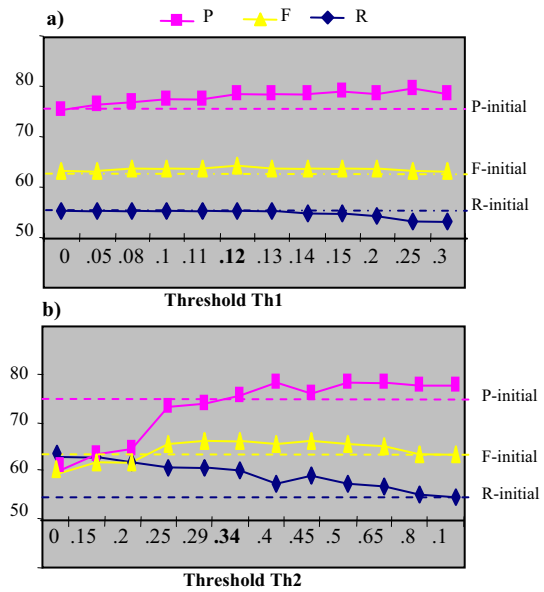

Figure 3: Thresholds effect on recall, precision and F metrics: a) Th1; b) Th2

3.2. Quotations relationship

A quotation can be defined as a sequence of terms in a document that refers to another document. In the context of documents-centric meeting recordings, a quotation corresponds to a segment (paragraph, sentence or part of sentence) that is quoted from the document. We used two rules to find quotations; first, the terms order should be the same in both quotation and quoted sequences. Second, a quotation should contain at least three terms, after stop words removing and stemming. The algorithm has been applied on 9 meetings, with 993 speakers' utterances and 606 documents sentences. Detected quotations have been evaluated with *recall*/*precision* and *F* metrics, in comparison with a manual ground truth. The average values of *recall*/*precision* are both of 95%. 5% of quotations have been missed due to many reasons, such as:

1. Some ambiguities exist due to a lexical similarity between some terms and stop words, e.g. the term "*été*" (*summer*) is confused with the stop word "*été*" (*been*). This causes the deletion of these ambiguous terms from the vectors representing the units to compare. If the number of the remaining terms, after stop words elimination, is less than three terms, then this quotation is neglected.
2. Our algorithm does not consider the lemmatization. For this reason, some terms derived from the same lemma were considered as different, e.g. the verb "*servire*" (to *serve*) and "*sert*" (*serves*).

… <utterance id="13"> Here is the article about <**er id=”10”**> “les radios généralistes” </er>, but there is nothing important to say.. Let’s go to the <**er id=”11”**> last article, “une apocalypse aveugle” </er>
</utterance >
<utterance id="14"> so the content of <**er id=”12”**> this article </er> is about … </ utterance >…
<References>…
<ref id="10" utter-id="13" logicalBlock-id="5" doc="file.xml"/>
<ref id="11" utter-id="13" logicalBlock-id="6" doc="file.xml"/>
<ref id="12" utter-id="14" logicalBlock-id="6" doc="file.xml"/>..
</ References >

Figure 4: Extract from a references file

3. Sometimes when a speaker quotes a segment from the document, he might not pronounce a given word correctly, or he pronounced just a part, before repeating it, e.g. the term “ba” in the extract “.. *this report is ba.. based on discovering*..”. This breaks the quotation into two, one or no sub quotation, depending on the size of generated segments.
4. The overlapping of quotations is not allowed. Two quotations *(q1, q2*) are overlapped, if they have a common part. With our algorithm, no more quotations are detected (*q2*), as long as the identification of the current one (*q1*) is not accomplished. Thus, *q2* is neglected, but a sub-quotation *q3* of *q2* may be considered.

$$\underbrace{t_1, t_2, \ldots, t_{k-1}}_{q_1}, \overbrace{t_k, t_{k+1}, \ldots, t_{n-1}}^{q_2}, \underbrace{t_n}_{q_3}$$

The quotations relationship is very strong, its detection is deterministic and thus its results can be fully trusted. Hence, it may be used to reinforce and correct other kinds of relationships. The third kind of relationships between meeting dialogs and documents is presented in the next section. It consists in the references that are made to documents parts.

3.3. References relationship

Ref2doc is an algorithm for detecting references made by the participants in meetings [7], to documents parts, which can be represented by logical blocks such as title, articles, authors, etc. *Ref2doc* requires a segmentation of the speech transcript into utterances, and the documents into logical blocks. This algorithm follows two main steps:

1. The detection of the referring expressions, in the various speakers’ utterances. For example: “*this first article*”, “*the author*”, “*the second title*”, etc.
2. Based on context information, the referring expressions are matched with the corresponding documents logical blocks.

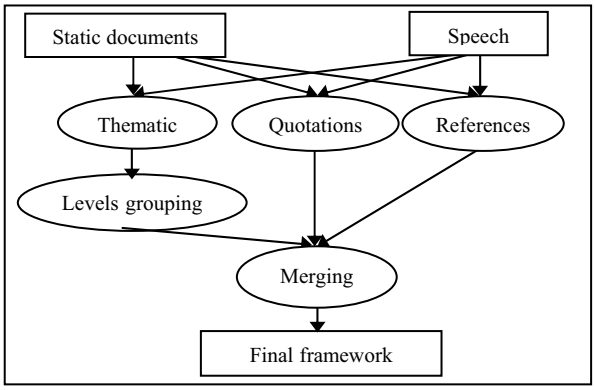

Figure 5: The merging process

Figure 4 shows an extract from an XML file generated by the *Ref2doc* algorithm. The referring expressions are highlighted within their utterances parents. Complete information about them is available, such as the corresponding logical blocks indexes. We have done an over flight on three kinds of relationships that link static documents with meeting dialogs. In the next section, an approach that combines these three relationships together, in order to measure their complementarity, is presented.

3.4. Combining the relationships

In this section, we are interested in putting together all the identified links, thematic links, quotations and references. Only the speech transcript is concerned with this combining process, as a source modality for quotations and references. After thematic links detection, the consideration of quotations and references presents added values and complementary information. The thematic links detection, as seen in section 3.1, is applied at two levels: utterances/sentences and turns/logical blocks. The level required for quotations is the utterances/sentences, and for references is the utterances/logical blocks level. After running the three algorithms independently, our merging process combines all these links (figure 5) according to their source unit, turn or utterance. An extract from the final structure is shown in figure 6. The quotations relationship is a special kind of the thematic relationship, and should be coherent with it. When both of them were combined, the *F* metric increased from 67% to 68%. The references links do not have inevitably to correspond to the thematic links. In figure 4, utterances 13 and 14 are linked thematically with sentences of the 6th logical block, even if there is a reference to the 5th logical block. This is why the *precision* value decreased from 76% to 72%, when combining the references with the thematic links

(*recall* increased from 61% to 67% and *F* from 67% to 68%).

Average	Initial	After grouping thematic links levels	After combining all links
R	55%	61%	67%
P	75%	76%	72%
F	63%	67%	68%

Table 1: Effect of levels grouping and links combining on the thematic links.

Nevertheless, the references may add other information to the user, such as, how the speakers chained up the various documentarticles. After combining the three relationships, thematic links, quotations and references, the final *recall/precision/F* values (table 1) obtained are (67, 72, 68)% . In order to visualize the framework generated by the combining process, an SVG tool has been implemented (figure 7). Within this tool, the document and the speech transcript are considered as axes. The thematic links are represented as circles at the intersection of the corresponding utterances and sentences, and the quotations as diamonds. The references are represented as rectangles, where the height depends on the size of the referred logical block (i.e. number of sentences).

4. Conclusion and future work

Several local linking mechanisms have been proposed in this article, in order to align and synchronize documents carried by various media (newspapers, slideshows, speech transcript, etc). In particular, three relationships, which cover all the links that may exist between static documents and meeting

```
<Turn id="1">
<Thematic with="logic">
        <logical_block id="22" similarity="0.15"/>  …
</Thematic>
<utterances>
<utter id="1">
        Didier will discuss the first article which is about
        "surprises du procès Elf", then….
<Thematic with="sentences">
          <sent-id="85" similarity="0.16"/>          …
</Thematic>
<Quotations>
  <quotation id="1" utter-id="2" sent-id="67" doc=" file.xml"/>
</Quotations>
<References>
 <er id="2" logicalBlock-id="1" doc="file.xml"> first article </er>
…
</References>
</utter > …
</utterances> …</Turn >..
```

Figure 6: Final structure after the merging

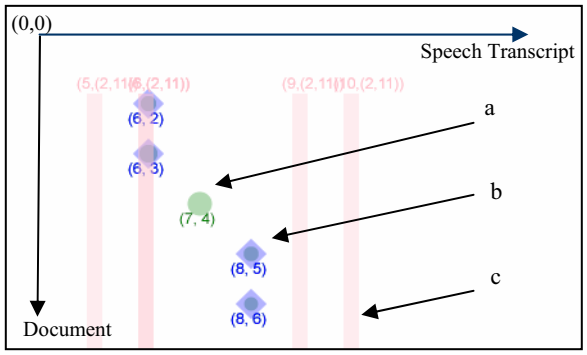

Figure 7: a) a thematic link; b) a quotation and a thematic link; c) a reference.

dialogs, have been studied. As shown in the evaluation, combining these relationships in a common framework validates each one of them, and reinforces the existing links between the two documents. In the future, our local linking methods will be applied on scientific conferences data (articles, audio recordings and slideshows), in order to enrich the corresponding browsing interfaces.

5. References

[1] L. Denoyer, J.N. Vittaut, P. Gallinari, and S. Brunessaux, "Structured multimedia document classification", ACM Symposium on Document Engineering, France 2003.

[2] S. Kim, and E.J. Whitehead, Jr., "Properties of academic paper references", 15th ACM Conference on Hypertext and Hypermedia, USA 2004.

[3] B. Zhang, M. André, P. Calado, and M. Cristo, "Combining structural and citation-based evidence for text classification", 13th Conference on Information and Knowledge Management CIKM, USA 2004.

[4] M. Hasan "Cross-language information retrieval, document alignment and visualization – A study with Japanese and Chinese", Doctor's Thesis, Nara Institute of Science and Technology, 2001.

[5] O. Abou Khaled, R. Scheurer, D. Lalanne, R. Ingold, and J.Y. Lemeur, "S.M.A.C, Smart Meeting Archives for Conferences", Flash Informatique FI 2/05, Ecole Polytechnique Fédérale de Lausanne, 2005.

[6] D. Mekhaldi, D. Lalanne, and R. Ingold, "Thematic alignment of recorded speech with documents", ACM Symposium on Document Engineering, France 2003.

[7] A. Popescu-Belis and D. Lalanne "Reference resolution over a restricted domain: references to documents". ACL Workshop on Reference Resolution and its Applications, Spain 2004.