

# Thematic Alignment of Documents with Meeting Dialogs

Dalila Mekhaldi

Département d'Informatique  
Chemin de musée 3, CH-1700 Fribourg  
+41 26 429 66 78  
Dalila.Mekhaldi@unifr.ch

Denis Lalanne

Département d'Informatique  
Chemin de musée 3, CH-1700 Fribourg  
+41 26 429 65 96  
Denis.Lalanne@unifr.ch

## ABSTRACT

The primary goal of this PhD thesis is to align printable documents with meetings' dialogs. This bi-modal alignment consists in bridging thematic links between documents' content and speech transcripts' content. An obvious application is a system that automatically link document parts with audio-video extracts of a meeting. Further, this bi-modal alignment is considered for thematically segmenting both meeting dialogs and documents discussed during this meeting.

## Categories and Subject Descriptors

H.3.1 [Content Analysis and Indexing] indexing methods

H.3.3 [Information Search and Retrieval] Clustering; Search process

## General Terms

Algorithms, Measurement, Performance, Experimentation

## Keywords

Multimodal thematic alignment, meeting thematic segmentation.

## 1. INTRODUCTION

The Document/Speech thematic alignment, proposed in this thesis, consists in detecting thematic links between documents content and speech transcript content. This alignment will enrich documents with temporal indexes related to spoken parts, and further bridges the gap between documents and temporal media such as audio and video. For this reason, this alignment will allow building document-based meeting browsing interfaces, that facilitates navigation and retrieval of multimedia data through documents, e.g. retrieving the part of the speech transcript that corresponds to a given document article.

In this paper, a thematic alignment method is briefly introduced along with the results of an evaluation. Then this bi-modal alignment is employed in order to discover the thematic structures of both documents and meeting dialogs. Finally, this method is evaluated through a comparison with two mono-modal segmentation methods.

Copyright is held by the author/owner(s).  
MM'04, October 10-16, 2004, New York, New York, USA.  
ACM 1-58113-893-8/04/0010.

## 2. MEETING DATA

Our experiments are based on press reviews recordings, during which 3-6 speakers discuss and debate around the cover page of various French-speaking newspapers. The main reason for choosing this type of meetings comes from the heterogeneity of the newspaper articles. Nevertheless, other documents types (e.g. agenda, slides, articles) will be considered in the future.

## 3. THEMATIC ALIGNMENT

The thematic alignment of documents with meeting dialogs first requires a preliminary segmentation of both modalities in various units: sentences and logical blocks for the document, utterances and turns for the speech transcript. Considering document and speech units as bags of weighted words, and after stop-words' removal and proper stemming, similarities between documents' units and speech transcript's units is then computed using a combination of various state-of-the-art metrics (Cosine, Jaccard, Dice). The relevant pairs are chosen according to two strategies: the *one-best* alignment and the *multiple* alignments. In the *one-best* alignment, the most similar target unit for each source unit is chosen. In the *multiple* alignments, all the similar target units, in respect to a fixed threshold are chosen. This generates symmetrical alignments.

8 meeting recordings have been experimented, of roughly 15 minutes each, with the *one-best* alignment strategy, with a total of 1409 documents sentences, 572 utterances, and 156 speaker turns. Indeed, only the relevant pairs are tested, i.e. when the source unit size is smaller than or equal to the target unit size (e.g. document sentence with speech turn). Very satisfactory results were obtained using the *Recall/Precision* metric. For instance, the *Cosine* method generates a *Recall/Precision* value of (.87,.51) for the sentences/utterances alignment, (.78,.60) for sentences/turns, (.83,.71) for the utterances/sentences and (.85,.84) for the turns/logical units. However, this *one-best* strategy presents some limits. The first limit is that the alignability is influenced by the units' size, which requires the consideration of other parameters when aligning the units (e.g. the *membership*, the *ownership* and the units' size). The second limit is that in many cases, a source unit has to be aligned with more than one target unit (e.g. a document sentence can be discussed many times during the meeting). This limit is solved by the second alignment strategy: the *multiple* alignments. For this strategy, other evaluation methods are required, due to the subjectivity and complexity of building a proper ground truth. In the next section we will see that the *multiple* alignments strategy results can be employed for discovering the thematic structure of both documents and speech transcript.

## 4. THEMATIC SEGMENTATION

The *multiple* alignments provide symmetrical results; from documents to speech transcript, and vice versa. This property is used for simultaneously segmenting documents and speech transcript. The alignment pairs are represented on a 2D space, where the  $X$  and  $Y$  axes are respectively document and speech transcript, and similarity links are nodes. The denser groups of nodes, probably meeting topics, are extracted using an extended *K-means* clustering method. The final clusters are then projected on each axis in order to obtain the thematic segments for both documents and speech transcript.

### 4.1 Evaluation

22 meeting recordings have been experimented, with a total of 3173 speech utterances and 2936 documents sentences. The speakers follow two scenarios, the stereotyped meetings, where they have very few debates (9 meetings). The average is about 69 utterances per 26 turns per (ratio 2.5), and an average duration of 497 seconds per meeting. The second category is the non-stereotyped meetings, i.e. meetings with numerous debates (13 meetings). In this category, there are 178 utterances per 127 turns (ratio 1.5), with an average duration of 745 seconds per meeting. Regarding the documents, the meetings are classified into two categories, 18 mono-document meetings, where only one document is discussed, and 4 multi-documents meetings where more than one document is discussed.

Our bi-modal segmentation method has been compared to two other mono-modal methods: *TextTiling* for both documents and speech transcript, and a baseline method for each modality. The speech baseline is a *speaker-turns* based segmentation method and the document baseline is a *reflexive* method (i.e. aligning the document with itself and then clustering of the results). The *Beeferman* metric  $P_k$  has been used in respect to a manual ground truth. This metric measures the ratio of pair units, at a distance of  $k$  units apart, that are badly classified in respect to the ground truth. For a perfect segmentation the  $P_k$  value is 0.

	Speech transcript		Documents	
	Stereotyped	Non-stereotyped	Mono-document	Multi-documents
Bi-modal	0.42	0.44	0.39	0.38
TextTiling	0.53	0.69	0.62	0.64
Baseline	0.44	0.61	0.46	0.57

**Table1: bi-modal thematic segmentation, comparing to other mono-modal methods.**

For the speech segmentation, the comparison using the  $P_k$  metric has shown promising results. The two mono-modal methods generate many extra segments. Our method is thus more efficient, especially for non-stereotyped meetings (i.e. with numerous debates), mainly because the documents provide a limited thematic context and constrain the segmentation. Moreover, thematic similarities between speech transcript segments can be detected easily with our method, indeed the corresponding clusters are aligned vertically (the speech transcript is the  $Y$  axis). Concerning the document segmentation, our bi-modal is more powerful in the two meeting categories (mono-document and multi-documents). Furthermore, our bi-modal can detect any thematic similarity between the non-adjacent document articles, when many generated clusters are aligned horizontally (the document is the  $X$  axis). Despite the satisfactory results obtained in these tested meetings, our bi-

modal method segments only partially documents when they are not fully discussed during the meeting. Finally, a recent filtering of the isolated nodes within the clusters, has improved our bi-modal method for the speech transcript segmentation in the case of stereotyped meetings ( $P_k = .42 \rightarrow .24$ ).

### 4.2 Analysis

The vertical or the horizontal alignment of the clusters through the 2D representation can carry to an overlapping of the thematic segments, during the projection step. When one of the two overlapped segments is contained in the other, this can be explained by the thematic similarity between them. But if they are partially overlapped, this can highly affect the segmentation results. The resolution of this overlapping is based on the elimination of the isolated nodes in each cluster. In the near future, the similarity values should be considered during the clustering process, which may help eliminating the un-trusted similarity links.

## 5. CONCLUSION

The problem of aligning static documents with speech recordings is presented in this paper. A thematic alignment method has been first described for thematically linking meeting documents with speech transcript. Further, a bi-modal thematic segmentation method has been proposed, which simultaneously segment documents and speech recordings. The results clearly illustrate the strong relationship that exists between the two processes, i.e. thematic alignment and thematic segmentation. These results tend to prove that combining modalities improves considerably segmentation scores and that documents considerably helps structuring meetings.

With the progression of our work, many alignments categories (citation, reference and thematic alignments) and many levels (sentences/utterances, utterances/logical blocks, etc.) have appeared [1,2,3]. For instance, during meetings, speakers often refer to documents or parts of document. To solve these references to documents, it is necessary to find links between each referring expression and the corresponding document element [3]. The goal of this thesis is to combine all these alignments in a single and robust framework, with respect to the various levels hierarchy. Merging the multiple levels may fortify the elementary alignments, e.g. the sentences/turns alignment may enrich the sentences/utterances alignment results. Other issues are to be tackled in the future, such as the documents that are partially discussed, or the documents with a poor textual content, such as meetings agenda or projected slides.

## 6. REFERENCES

- [1] Lalanne D., Mekhaldi D. and Ingold R. "Talking about documents: revealing a missing link to multimedia meeting archives". Document Recognition and Retrieval XI, IS&T/SPIE's, International Symposium on Electronic Imaging, USA, 2004, p.82-91.
- [2] Mekhaldi D., Lalanne D. and Ingold R., Thematic Alignment Of Recorded Speech With Documents, Proceedings of the ACM symposium on Document engineering, France, 2003, p.52-54.
- [3] Popescu-Belis A., Lalanne D., References to Documents (Ref2doc): Reference Resolution Over a Restricted Domain, ACL Workshop on Reference Resolution and its Applications, Spain, 2004.