User Requirements Analysis for Meeting Information Retrieval Based on Query Elicitation

Vincenzo Pallotta
Department of Computer Science
University of Fribourg
Switzerland
Vincenzo.Pallotta@unifr.ch

Violeta Seretan
Language Technology Laboratory
University of Geneva
Switzerland
seretan@lettres.unige.ch

Marita Ailomaa
Artificial Intelligence Laboratory
Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
Marita.Ailomaa@epfl.ch

Abstract

We present a user requirements study for Question Answering on meeting records that assesses the difficulty of users questions in terms of what type of knowledge is required in order to provide the correct answer. We grounded our work on the empirical analysis of elicited user queries. We found that the majority of elicited queries (around 60%) pertain to argumentative processes and outcomes. Our analysis also suggests that standard keyword-based Information Retrieval can only deal successfully with less than 20% of the queries, and that it must be complemented with other types of metadata and inference.

1 Introduction

Meeting records constitute a particularly important and rich source of information. Meetings are a frequent and sustained activity, in which multi-party dialogues take place that are goal-oriented and where participants perform a series of actions, usually aimed at reaching a common goal: they exchange information, raise issues, express opinions, make suggestions, propose solutions, provide arguments (pro or con), negotiate alternatives, and make decisions. As outcomes of the meeting, agreements on future action items are reached, tasks are assigned, conflicts are solved, etc. Meeting outcomes have a direct impact on the efficiency of organization and team performance, and the stored and indexed meeting records serve as reference for further processing (Post et al., 2004). They can also be used in future meetings in order to facilitate the decision-making process by accessing relevant information from previous meetings (Cremers et al., 2005), or in order to make the discussion more focused (Conklin, 2006).

Meetings constitute a substantial and important source of information that improves corporate organization and performance (Corrall, 1998; Romano and Nunamaker, 2001). Novel multimedia techniques have been dedicated to meeting *recording*, *structuring* and content *analysis* according to the metadata schema, and finally, to *accessing* the analyzed content via browsing, querying or filtering (Cremers et al., 2005; Tucker and Whittaker, 2004).

This paper focuses on *debate meetings* (Cugini et al., 1997) because of their particular richness in information concerning the decision-making process. We consider that the meeting content can be organized on three levels: (i) *factual level* (what happens: events, timeline, actions, dynamics); (ii) *thematic level* (what is said: topics discussed and details); (iii) *argumentative level* (which/how common goals are reached).

The information on the first two levels is explicit information that can be usually retrieved directly by searching the meeting records with appropriate IR techniques (i.e., TF-IDF). The third level, on the contrary, contains more abstract and tacit information pertaining to how the explicit information contributes to the rationale of the meeting, and it is not present as such in raw meeting data: whether or not the meeting goal was reached, what issues were debated, what proposals were made, what alternatives were discussed, what arguments were brought, what decisions were made, what task were assigned, etc.

The motivating scenario is the following: A user

Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pages 1008–1015,
Prague, Czech Republic, June 2007.

needs information about a past meeting, either in quality of a participant who wants to recollect a discussion (since the memories of co-participants are often inconsistent, cf. Banerjee et al., 2005), or as a non-participant who missed that meeting. Instead of consulting the entire meeting-related information, which is usually heterogeneous and scaterred (audio-video recordings, notes, minutes, e-mails, handouts, etc.), the user asks natural language questions to a query engine which retrieves relevant information from the meeting records.

In this paper we assess the users' interest in retrieving argumentative information from meetings and what kind of knowledge is required for answering users' queries. Section 2 reviews previous user requirements studies for the meeting domain. Section 3 describes our user requirements study based on the analysis of elicited user queries, presents its main findings, and discusses the implications of these findings for the design of meeting retrieval systems. Section 4 concludes the paper and outlines some directions for future work.

2 Argumentative Information in Meeting Information Retrieval

Depending on the meeting browser type[1], different levels of meeting content become accessible for information retrieval. Audio and video browsers deal with factual and thematic information, while artifact browsers might also touch on deliberative information, as long as it is present, for instance, in the meeting minutes. In contrast, derived-data browsers aim to account for the argumentative information which is not explicitly present in the meeting content, but can be inferred from it. If minutes are likely to contain only the most salient deliberative facts, the derived-data browsers are much more useful, in that they offer access to the full meeting record, and thus to relevant details about the deliberative information sought.

2.1 Importance of Argumentative Structure

As shown by Rosemberg and Silince (1999), tracking argumentative information from meeting discussions is of central importance for building project memories since, in addition to the "strictly factual, technical information", these memories must also store relevant information about decision-making processes. In a business context, the information derived from meetings is useful for future business processes, as it can explain phenomena and past decisions and can support future actions by mining and assessment (Pallotta et al., 2004). The argumentative structure of meeting discussions, possibly visualized in form of argumentation diagrams or maps, can be helpful in meeting browsing. To our knowledge, there are at least three meeting browsers that have adopted argumentative structure: ARCHIVUS (Lisowska et al., 2004b), ViCoDe (Marchand-Maillet and Bruno, 2005), and the Twente-AMI JFerret browser (Rienks and Verbree, 2006).

2.2 Query Elicitation Studies

The users' interest in argumentation dimension of meetings has been highlighted by a series of recent studies that attempted to elicit the potential user questions about meetings (Lisowska et al., 2004a; Benerjee at al., 2005; Cremers et al., 2005).

The study of Lisowska et al. (2004a), part of the IM2 research project[2], was performed in a simulated environment in which users were asked to imagine themselves in a particular role from a series of scenarios. The participants were both IM2 members and non-IM2 members and produced about 300 retrospective queries on recorded meetings. Although this study has been criticized by Post et al. (2004), Cremers et al. (2005), and Banerjee et al. (2005) for being biased, artificial, obtrusive, and not conforming to strong HCI methodologies for survey research, it shed light on potential queries and classified them in two broad categories, that seem to correspond to our argumentative/non-argumentative distinction (Lisowska et al., 2004a: 994):

- "elements related to the interaction among participants: acceptance/rejection, agreement/disagreement; proposal, argumentation (for and against); assertions, statements; decisions; discussions, debates; reactions; questions; solutions";

[1] (Tucker and Whittaker, 2004) identifies 4 types of meeting browsers: audio browsers, video browsers, artifacts browsers (that exploit meeting minutes or other meeting-related documents), and browsers that work with derived data (such as discourse and temporal structure information).

[2] http://www.im2.ch

- "concepts from the meeting domains: dates, times; documents; meeting index: current, previous, sets; participants; presentations, talks; projects; tasks, responsibilities; topics".

Unfortunately, the study does not provide precise information on the relative proportions of queries for the classification proposed, but simply suggests that overall more queries belong to the second category, while queries requiring understanding of the dialogue structure still comprise a sizeable proportion.

The survey conducted by Banerjee et al. (2005) concerned instead real, non-simulated interviews of busy professionals about actual situations, related either to meetings in which they previously participated, or to meetings they missed. More than half of the information sought by interviewees concerned, in both cases, the argumentative dimension of meetings.

For non-missed meetings, 15 out of the 26 instances (i.e., 57.7%) concerned argumentative aspects: what the decision was regarding a topic (7); what task someone was assigned (4); who made a particular decision (2); what was the participants' reaction to a particular topic (1); what the future plan is (1). The other instances (42.3%) relate to the thematic dimension, i.e., specifics of the discussion on a topic (11).

As for missed meetings, the argumentative instances were equally represented (18/36): decisions on a topic (7); what task was assigned to interviewee (4); whether a particular decision was made (3); what decisions were made (2); reasons for a decision (1); reactions to a topic (1). The thematic questions concern topics discussed, announcements made, and background of participants.

The study also showed that the recovery of information from meeting recordings is significantly faster when discourse annotations are available, such as the distinction between discussion, presentation, and briefing.

Another unobtrusive user requirements study was performed by Cremers et al. (2005) in a "seminatural setting" related to the design of a meeting browser. The top 5 search interests highlighted by the 60 survey participants were: *decisions made*, participants/speakers, topics, agenda items, and *arguments for decision*. Of these, the ones shown in italics are argumentative. In fact, the authors acknowledge the necessity to include some "functional" categories as innovative search options. Interestingly, from the user interface evaluation presented in their paper, one can indirectly infer how salient the argumentative information is perceived by users: the icons that the authors intended for emotions, i.e., for a emotion-based search facility, were actually interpreted by users as referring to people's opinion: *What is person X's opinion?* – positive, negative, neutral.

3 User Requirements Analysis

The existing query elicitation experiments reported in Section 2 highlighted a series of question types that users typically would like to ask about meetings. It also revealed that the information sought can be classified into two broad categories: argumentative information (about the argumentative process and the outcome of debate meetings), and non-argumentative information (factual, i.e., about the meeting as a physical event, or thematic, i.e., about what has been said in terms of topics).

The study we present in this section is aimed at assessing how difficult it is to answer the questions that users typically ask about a meeting. Our goal is to provide insights into:

- how many queries can be answered using standard IR techniques on meeting artefacts only (e.g., minutes, written agenda, invitations);
- how many queries can be answered with IR on meeting recordings;
- what kind of additional information and inference is needed when IR does not apply or it is insufficient (e.g., information about the participants and the meeting dynamics, external information about the meeting's context such as the relation to a project, semantic interpretation of question terms and references, computation of durations, aggregation of results, etc).

Assessing the level of difficulty of a query based on the two above-mentioned categories might not provide insightful results, because these would be too general, thus less interpretable. Also, the complex queries requiring mixed information would escape observation because assigned to a too general class. We therefore considered it necessary to perform a separate analysis of each query instance, as this provides not only detailed, but also traceable information.

3.1 Data: Collecting User Queries

Our analysis is based on a heterogeneous collection of queries for meeting data. In general, an unbiased queries dataset is difficult to obtain, and the quality of a dataset can vary if the sample is made of too homogenous subjects (e.g., people belonging to the same group as members of the same project). In order to cope with this problem, our strategy was to use three different datasets collected in different settings:

- First, we considered the *IM2 dataset* collected by Lisowska et al. (2004a), the only set of user queries on meetings available to date. It comprises 270 questions (shortly described in Section 2) annotated with a label showing whether or not the query was produced by an IM2-member. These queries are introspective and not related to any particular recorded meeting.
- Second, we cross-validated this dataset with a large corpus of 294 natural language statements about existing meetings records. This dataset, called the *BET observations* (Wellner et al., 2005), was collected by subjects who were asked to watch several meeting recordings and to report what the meeting participants appeared to consider interesting. We use it as a 'validation' set for the IM2 queries: an IM2 query is considered as 'realistic' or 'empirically grounded' if there is a BET observation that represents a possible answer to the query. For instance, the query *Why was the proposal made by X not accepted?* matches the BET observation *Denis eliminated Silence of the Lambs as it was too violent*.
- Finally, we collected a new set of 'real' queries by conducting a survey of user requirements on meeting querying in a natural business setting. The survey involved 3 top managers from a company and produced 35 queries. We called this dataset *Manager Survey Set* (MS-Set).

The queries from the IM2-set (270 queries) and the MS-Set (35 queries) were analyzed by two different teams of two judges. Each team discussed each query, and classified it along the two main dimensions we are interested in:

- *query type*: the type of meeting content to which the query pertains;
- *query difficulty*: the type of information required to provide the answer.

3.2 Query Type Analysis

Each query was assigned exactly one of the following four possible categories (the one perceived as the most salient):

1. *factual*: the query pertains to the factual meeting content;
2. *thematic*: the query pertains to the thematic meeting content;
3. *process*: the query pertains to the argumentative meeting content, more precisely to the argumentative process;
4. *outcome*: the query pertains to the argumentative meeting content, more precisely to the outcome of the argumentative process.

Category	IM2-set (size:270)		MS-Set (size: 35)	
	Team1	Team2	Team1	Team2
Factual	24.8%	45.6%	20.0%	20.0%
Thematic	18.5%		20.0%	11.4%
Process	30.0%	32.6%	22.9%	28.6%
Outcome	26.7%	21.8%	37.1%	40.0%
Process+ Outcome	56.7%	54.4%	60.0%	68.6%

Table 1. Query classification according to the meeting content type.

Results from this classification task for both query sets are reported in Table 1. In both sets, the information most sought was argumentative: about 55% of the IM2-set queries are argumentative (process or outcome). This invalidates the initial estimation of Lisowska et al. (2004a:994) that the non-argumentative queries prevail, and confirms the figures obtained in (Banerjee et al., 2005), according to which 57.7% of the queries are argumentative. In our real managers survey, we obtained even higher percentages for the argumentative queries (60% or 68.6%, depending on the annotation team). The argumentative queries are followed by factual and thematic ones in both query sets, with a slight advantage for factual queries.

The inter-annotator agreement for this first classification is reported in Table 2. The proportion of queries on which annotators agree in classifying them as argumentative is significantly high. We only report here the agreement results for the individual argumentative categories (Process, Outcome) and both (Process & Outcome). There were 213 queries (in IM2-set) and 30 queries (in MS-set) that were consistently annotated by the two teams on *both* categories. Within this set, a high percentage of queries were argumentative, that is, they were annotated as either Process or Outcome (label AA in the table).

Category	IM2-set (size: 270)		MS-set (size: 35)	
	ratio	kappa	ratio	kappa
Process	84.8%	82.9%	88.6%	87.8%
Outcome	90.7%	89.6%	91.4%	90.9%
Process & Outcome	78.9%	76.2%	85.7%	84.8%
AA	117/213 = **54.9%**		19/30 = **63.3%**	

Table 2. Inter-annotator agreement for query-type classification.

Furthermore, we provided a re-assessment of the proportion of argumentative queries with respect to query origin for the IM2-set (IM2 members vs. non-IM2 members): non-IM2 members issued 30.8% of agreed argumentative queries, a proportion that, while smaller compared to that of IM2 members (69.2%), is still non-negligible. This contrasts with the opinion expressed in (Lisowska et al., 2004a) that argumentative queries are almost exclusively produced by IM2 members.

Among the 90 agreed IM2 queries that were cross-validated with the BET-observation set, 28.9% were argumentative. We also noted that the ratio of BET statements that contain argumentative information is quite high (66.9%).

3.3 Query Difficulty Analysis

In order to assess the difficulty in answering a query, we used the following categories that the annotators could assign to each query, according to the type of information and techniques they judged necessary for answering it:

1. *Role of IR*: states the role of standard[3] Information Retrieval (in combination with Topic Extraction[4]) techniques in answering the query. Possible values:
    a. *Irrelevant* (IR techniques are not applicable). Example: *What decisions have been made?*
    b. *successful* (IR techniques are sufficient). Example: *Was the budget approved?*
    c. *insufficient* (IR techniques are necessary, but not sufficient alone since they require additional inference and information, such as argumentative, cross-meeting, external corporate/project knowledge). Example: *Who rejected the proposal made by X on issue Y?*
2. *Artefacts*: information such as agenda, minutes of previous meetings, e-mails, invitations and other documents related and available before the meeting. Example: *Who was invited to the meeting?*
3. *Recordings*: the meeting recordings (audio, visual, transcription). This is almost always true, except for queries where Artefacts or Metadata are sufficient, such as *What was the agenda?, Who was invited to the meeting?*).
4. *Metadata*: context knowledge kept in static metadata (e.g., speakers, place, time). Example: *Who were the participants at the meeting?*
5. *Dialogue Acts & Adjacency Pairs*: Example: *What was John's response to my comment on the last meeting?*
6. *Argumentation*: metadata (annotations) about the argumentative structure of the meeting content. Example: *Did everybody agree on the decisions, or were there differences of opinion?*
7. *Semantics*: semantic interpretation of terms in the query and reference resolution, including deictics (e.g., for *how long, usually, systematically, criticisms; this, about me, I*). Example: *What decisions got made easily?* The term requiring semantic interpretation is underlined.
8. *Inference*: inference (deriving information that is implicit), calculation, and aggregation (e.g., for 'command' queries asking for lists of things – participants, issues, proposals). Example: *What would be required from me?*

[3] By standard IR we mean techniques based on bag-of-word search and TF-IDF indexing.

[4] Topic extraction techniques are based on topic shift detection (Galley et al., 2003) and keyword extraction (van der Plas et al., 2004).

9. *Multiple meetings*: availability of multiple meeting records. Example: *Who usually attends the project meetings?*

10. *External*: related knowledge, not explicitly present in the meeting records (e.g., information about the corporation or the projects related to the meeting). Example: *Did somebody talk about me or about my work?*

Results of annotation reported on the two query sets are synthesized in Table 3: IR is sufficient for answering 14.4% of the IM2 queries, and 20% of the MS-set queries. In 50% and 25.7% of the cases, respectively, it simply cannot be applied (irrelevant). Finally, IR alone is not enough in 35.6% of the queries from the IM2-set, and in 54.3% of the MS-set; it has to be complemented with other techniques.

IR is:	IM2-set		MS-set	
	all queries	AA	all queries	AA
Sufficient	39/270 = 14.4%	1/117 = 0.8%	7/35 = 20.0%	1/19 = 5.3%
Irrelevant	135/270 = 50.0%	55/117 = 47.0%	9/35 = 25.7%	3/19 = 15.8%
Insufficient	96/270 = 35.6%	61/117 = **52.1%**	19/35 = 54.3%	15/19 = **78.9%**

Table 3. The role of IR (and topic extraction) in answering users' queries.

If we consider agreed argumentative queries (Section 3.2), IR is effective in an extremely low percentage of cases (0.8% for IM2-set and 5.3% for MS-Set). IR is insufficient in most of the cases (52.1% and 78.9%) and inapplicable in the rest of the cases (47% and 15.8%). Only one argumentative query from each set was judged as being answerable with IR alone: *What were the decisions to be made (open questions) regarding the topic t1? When is the NEXT MEETING planned? (e.g. to follow up on action items).*

Table 4 shows the number of queries in each set that require argumentative information in order to be answered, distributed according to the query types. As expected, no argumentation information is necessary for answering factual queries, but some thematic queries do need it, such as *What was decided about topic T?* (24% in the IM2-set and 42.9% in the M.S.-set).

Overall, the majority of queries in both sets require argumentation information in order to be answered (56.3% from IM2 queries, and 65.7% from MS queries).

Category	IM2-set, Annotation 1			MS-set, Annotation 1		
	total	Req. arg.	Ratio	Total	Req. arg.	Ratio
Factual	67	0	0%	7	0	0%
Thematic	50	12	24.0%	7	3	42.9%
Process	81	73	90.1%	8	7	87.5%
Outcome	72	67	93.1%	13	13	100%
All	270	152	**56.3%**	35	23	**65.7%**

Table 4. Queries requiring argumentative information.

We finally looked at what kind of information is needed in those cases where IR is perceived as insufficient or irrelevant. Table 5 lists the most frequent combinations of information types required for the IM2-set and the MS-set.

3.4 Summary of Findings

The analysis of the annotations obtained for the 305 queries (35 from the Manager Survey set, and 270 from the IM2-set) revealed that:

- The information most sought by users from meetings is argumentative (i.e., pertains to the argumentative process and its outcome). It constitutes more than half of the total queries, while factual and thematic information are similar in proportions (Table 1);
- There was no significant difference in this respect between the IM2-set and the MS-set (Table 1);
- The decision as to whether a query is argumentative or not is easy to draw, as suggested by the high inter-annotator agreement shown in Table 2;
- Standard IR and topic extraction techniques are perceived as insufficient in answering most of the queries. Only less than 20% of the whole query set can be answered with IR, and almost no argumentative question (Table 3).
- Argumentative information is needed in answering the majority of the queries (Table 4);
- When IR alone fails, the information types that are needed most are (in addition to recordings): Argumentation, Semantics, Inference, and Metadata (Table 5); see Section 3.3 for their description.

IR alone fails	IM2-set												
Information types	IR insufficient 96 cases 35.6%							IR irrelevant 135 cases 50%					
Artefacts									x				
Recordings	x	x	x	x	x	x	x	x	x	x	x		
Meta-data			x		x			x		x		x	x
Dlg acts & Adj. pairs													
Argumentation	x	x	x	x	x	x	x	x	x		x		
Semantics	x	x	x	x	x			x	x	x	x	x	
Inference	x		x	x			x	x	x	x	x	x	
Multiple meetings				x								x	
External													
Cases	15	11	9	8	7	5	4	14	9	8	8	7	5
Ratio (%)	15.6	11.5	9.4	8.3	7.3	5.2	4.2	10.4	6.7	5.9	5.9	5.2	3.7

IR alone fails	MS-set					
Information types	IR insufficient 19 cases 54.3%				IR irrelevant 9 cases 54.3%	
Artefacts					x	x
Recordings	x	x	x	x		
Meta-data					x	x
Dlg acts & Adj. pairs						
Argumentation	x	x	x	x		
Semantics	x		x	x	x	
Inference	x	x		x	x	
Multiple meetings						
External				x		
Cases	6	4	2	2	2	2
Ratio (%)	31.6	21	10.5	10.5	22.2	22.2

Table 5. Some of the most frequent combinations of information required for answering the queries in the IM2-Set and in the MS-set when IR alone fails.

3.5 Discussion

Searching relevant information through the recorded meeting dialogues poses important problems when using standard IR indexing techniques (Baeza-Yates and Ribeiro-Nieto, 2000), because users ask different types of queries for which a single retrieval strategy (e.g., keywords-based) is insufficient. This is the case when looking at answers that require some sort of entailment, such as inferring that a proposal has been rejected when a meeting participant says *Are you kidding?*.

Spoken-language information retrieval (Vinciarelli, 2004) and automatic dialogue-act extraction techniques (Stolke et al., 2000; Clark and Popescu-Belis, 2004; Ang et al., 2005) have been applied to meeting recordings and produced good results under the assumption that the user is interested in retrieving either topic-based or dialog act-based information. But this assumption is partially invalidated by our user query elicitation analysis, which showed that such information is only sought in a relatively small fraction of the users' queries. A particular problem for these approaches is that the topic looked for is usually not a query itself (*Was topic T mentioned?)*, but just a parameter in more structured questions (*What was decided about T?*). Moreover, the relevant participants' contributions (dialog acts) need to be retrieved in combination, not in isolation (*The reactions to the proposal made by X*).

4 Conclusion and Future Work

While most of the research community has neglected the importance of argumentative queries in meeting information retrieval, we provided evidence that this type of queries is actually very common. We quantified the proportion of queries involving the argumentative dimension of the meeting content by performing an in-depth analysis of queries collected in two different elicitation surveys. The analysis of the annotations obtained for the 305 queries (270 from the IM2-set, 35 from MS-set) was aimed at providing insights into different matters: what type of information is typically sought by users from meetings; how difficult it is, and what kind of information and techniques are needed in order to answer user queries.

This work represents an initial step towards a better understanding of user queries on the meeting domain. It could provide useful intuitions about how to perform the automatic classification of answer types and, more importantly, the automatic extraction of argumentative features and their relations with other components of the query (e.g., topic, named entities, events).

In the future, we intend to better ground our first empirical findings by i) running the queries against a real IR system with indexed meeting transcripts and evaluate the quality of the obtained answers; ii) ask judges to manually rank the difficulty of each query, and iii) compare the two rankings. We would also like to see how frequent argumentative queries are in other domains (such as TV talk shows or political debates) in order to generalize our results.

Acknowledgements

We wish to thank Martin Rajman and Hatem Ghorbel for their constant and valuable feedback. This work has been partially supported by the Swiss National Science Foundation NCCR IM2 and by the SNSF grant no. 200021-116235.

References

Jeremy Ang, Yang Liu and Elizabeth Shriberg. 2005. Automatic Dialog Act Segmentation and Classification in Multiparty Meetings. In *Proceedings of IEEE ICASSP 2005*, Philadelphia, PA, USA.

Ricardo Baeza-Yates and Berthier Ribeiro-Nieto. 2000. *Modern Information Retrieval*. Addison Wesley.

Satanjeev Banerjee, Carolyn Rose and Alexander I. Rudnicky. 2005. The Necessity of a Meeting Recording and Playback System, and the Benefit of Topic-Level Annotations to Meeting Browsing. In *Proceedings of INTERACT 2005*, Rome, Italy.

Alexander Clark and Andrei Popescu-Belis. 2004. Multi-level Dialogue Act Tags. In *Proceedings of SIGDIAL'04*, pages 163–170. Cambridge, MA, USA.

Jeff Conklin. 2006. *Dialogue Mapping: Building Shared Understanding of Wicked Problems*. John Wiley & Sons.

Sheila Corrall. 1998. Knowledge management. Are we in the knowledge management business? *ARIADNE: the Web version,* 18.

Anita H.M Cremers, Bart Hilhorst and Arnold P.O.S Vermeeren. 2005. “What was discussed by whom, how, when and where?" personalized browsing of annotated multimedia meeting recordings. In *Proceedings of HCI 2005*, pages 1–10, Edinburgh, UK.

John Cugini, Laurie Damianos, Lynette Hirschman, Robyn Kozierok, Jeff Kurtz, Sharon Laskowski and Jean Scholtz. 1997. Methodology for evaluation of collaborative systems. Technical Report Rev. 3.0, The Evaluation Working Group of the DARPA Intelligent Collaboration and Visualization Program.

Michel Galley, Kathleen McKeown, Eric Fosler-Lussier and Hongyan Jing. 2003. Discourse Segmentation of Multi-Party Conversation. In *Proceedings of ACL 2003*, pages 562–569, Sapporo, Japan.

Agnes Lisowska, Andrei Popescu-Belis and Susan Armstrong. 2004a. User Query Analysis for the Specification and Evaluation of a Dialogue Processing and Retrieval System. In *Proceedings LREC 2004*, pages 993–996, Lisbon, Portugal.

Agnes Lisowska, Martin Rajman and Trung H. Bui. 2004b. ARCHIVUS: A System for Accesssing the Content of Recorded Multimodal Meetings. In *Proceedings of MLMI 2004*, Martigny, Switzerland.

Stéphane Marchand-Maillet and Eric Bruno. 2005. Collection Guiding: A new framework for handling large multimedia collections. In *Proceeding of AVIVDiLib05*, Cortona, Italy.

Vincenzo Pallotta, Hatem Ghorbel, Afzal Ballim, Agnes Lisowska and Stéphane Marchand-Maillet. 2004. Towards meeting information systems: Meeting knowledge management. In *Proceedings of ICEIS 2005*, pages 464–469, Porto, Portugal.

Lonneke van der Plaas, Vincenzo Pallotta, Martin Rajman and Hatem Ghorbel. 2004. Automatic keyword extraction from spoken text: A comparison between two lexical resources: the EDR and WordNet. In *Proceedings of the LREC 2004*, pages 2205–2208, Lisbon, Portugal.

Wilfried M. Post, Anita H.M. Cremers and Olivier Blanson Henkemans. 2004. A Research Environment for Meeting Behavior. In *Proceedings of the 3rd Workshop on Social Intelligence Design,* pages 159–165, University of Twente, Enschede, The Netherlands.

Rutger Rienks and Daan Verbree. 2006. About the Usefulness and Learnability of Argument–Diagrams from Real Discussions. In *Proceedings of MLMI 2006*, Washington DC, USA.

Nicholas C. Romano Jr. and Jay F. Nunamaker Jr. 2001. Meeting Analysis: Findings from Research and Practice. In *Proceedings of HICSS-34*, Maui, HI, IEEE Computer Society.

Duska Rosemberg and John A.A. Silince. 1999. Common ground in computer-supported collaborative argumentation. In *Proceedings of the CLSCL99*, Stanford, CA, USA.

Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema and Marie Meteer. 2000. Dialog Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computational Linguistics,* 26(3):339–373.

Simon Tucker and Steve Whittaker. 2004. Accessing multimodal meeting data: systems, problems and possibilities. In *Proceedings of MLMI 2004*, Martigny, Switzerland.

Alessandro Vinciarelli. 2004. Noisy text categorization. In *Proceedings of ICPR 2004*, Cambridge, UK.

Pierre Wellner, Mike Flynn, Simon Tucker, Steve Whittaker. 2005. A Meeting Browser Evaluation Test. In *Proceedings of CHI 2005*, Portand, Oregon, USA.