Towards Multi-Document Summarization of Scientific Articles:Making Interesting Comparisons with SciSumm

Nitin Agarwal
Language Technologies Institute
Carnegie Mellon University
nitina@cs.cmu.edu

Kiran Gvr
Language Technologies Resource Center
IIIT-Hyderabad, India
kiran_gvr@students.iiit.ac.in

Ravi Shankar Reddy
Language Technologies Resource Center
IIIT-Hyderabad, India
krs_reddy@students.iiit.ac.in

Carolyn Penstein Rosé
Language Technologies Institute
Carnegie Mellon University
cprose@cs.cmu.edu

Abstract

We present a novel unsupervised approach to the problem of multi-document summarization of scientific articles, in which the document collection is a list of papers cited together within the same source article, otherwise known as a co-citation. At the heart of the approach is a topic based clustering of fragments extracted from each co-cited article and relevance ranking using a query generated from the context surrounding the co-cited list of papers. This analysis enables the generation of an overview of common themes from the co-cited papers that relate to the context in which the co-citation was found. We present a system called SciSumm that embodies this approach and apply it to the 2008 ACL Anthology. We evaluate this summarization system for relevant content selection using gold standard summaries prepared on principle based guidelines. Evaluation with gold standard summaries demonstrates that our system performs better in content selection than an existing summarization system (MEAD). We present a detailed summary of our findings and discuss possible directions for future research.

1 Introduction

In this paper we present a novel, unsupervised approach to multi-document summarization of scientific articles. While the field of multi-document summarization has achieved impressive results with collections of news articles, summarization of collections of scientific articles is a strikingly different problem. Multi-document summarization of news articles amounts to synthesizing details about the same story as it has unfolded over a variety of reports, some of which contain redundant information. In contrast, each scientific article tells its own research story. Even with papers that address similar research questions, the argument being made is different. Instead of collecting and arranging details into a single, synthesized story, the task is to abstract away from the specific details of individual papers and to find the common threads that unite them and make sense of the document collection as a whole.

Another challenge with summarization of scientific literature becomes clear as one compares alternative reviews of the same literature. Each review author brings their own unique perspective and questions to bear in their reading and presentation of that literature. While this is true of other genres of documents that have been the target of multi-document summarization work in the past, we don't find query oriented approaches to multi-document summarization of scientific articles. One contribution of this work is a technical approach to query oriented multi-document summarization of scientific articles that has been evaluated in comparison with a competitive baseline that is not query oriented. The evaluation demonstrates the advantage of the query oriented approach for this type of summarization.

We present a system called SciSumm that summarizes document collections that are composed of lists of papers cited together within the same source article, otherwise known as a co-citation. Using the context of the co-citation in the source article, we generate a query that allows us to generate a summary in a query-oriented fashion. The extracted por-

Proceedings of the Workshop on Automatic Summarization for Different Genres, Media, and Languages, pages 8–15,
Portland, Oregon, June 23, 2011.

tions of the co-cited articles are then assembled into clusters that represent the main themes of the articles that relate to the context in which they were cited. Our evaluation demonstrates that SciSumm achieves higher quality summaries than the MEAD summarization system (Radev, 2004).

The rest of the paper is organized as follows. We present an overview of relevant literature in Section 2. The end-to-end summarization pipeline has been described in Section 3 . Section 4 presents an evaluation of summaries generated from the system. We end the paper with conclusions and some interesting further research directions in Section 5.

2 Literature Review

We begin our literature review by thinking about some common use cases for multi-document summarization of scientific articles.

First consider that as a researcher reads a scientific article, she/he encounters numerous citations, most of them citing the foundational and seminal work that is important in that scientific domain. The text surrounding these citations is a valuable resource as it allows the author to make a statement about her viewpoint towards the cited articles. A tool that could provide a small summary of the collection of cited articles that is constructed specifically to relate to the claims made by the author citing them would be useful. It might also help the researcher determine if the cited work is relevant for her own research.

As an example of such a co-citation consider the following citation sentence

Various machine learning approaches have been proposed for chunking (Ramshaw and Marcus, 1995; Tjong Kim Sang, 2000a; Tjong Kim Sang et al. , 2000; Tjong Kim Sang, 2000b; Sassano and Utsuro, 2000; van Halteren, 2000).

Now imagine the reader trying to determine about widely used *machine learning* approaches for *noun phrase chunking*. Instead of going through these individual papers, it would be more useful to get the summary of the topics in all those papers that talk about the usage of machine learning methods in chunking.

2.1 Overview of Multi-Document Summarization

An exhaustive summary of recent work in summarization is out of the scope for this paper. Hence, we review only the most relevant approaches in summarization to our current work. As most recent work in multi-document summarization has been extractive, and in our observation, scientific articles contain the type of information that we would want in a summary, we follow this convention. This allows us to avoid the complexities of natural language generation based approaches in abstractive summarization.

Multi-document summarization is an extension of single document summarization in which the thematically important textual fragments are extracted from multiple comparable documents, e.g., news articles describing the same event. The techniques not only need to address identification and removal of redundant information but also inclusion of unique and novel contributions. Various graph based (Mani et al., 1997) and centroid clustering based methods (Radev et al., 2000) have been proposed to address the problem of multi-document summarization. Both of these methods identify common themes present in a document collection using a sentence similarity metric.

2.2 Summarization of Scientific Articles

Surprisingly, not many approaches to the problem of summarization of scientific articles have been proposed in the past. One exception is Teufel and Moens (2002), who view summarization as a classification task in which they use a Naive Bayes classifier to assign a rhetorical status to each sentence in an article and thus divide the whole article into regions with a specific argumentation status (e.g. categories such as AIM, CONTRAST and BACKGROUND). In our proposed approach, we are trying to identify reoccurring topic themes that are common across the articles and may appear under a variety of rhetorical headings.

Nanba and colleagues (1999) argue in their work that a co-citation frequently implies a consistent viewpoint towards the cited articles. Similarly, for articles cited within different sentences, textual similarity between the articles is inversely proportional to the size of the sentential gap between the citations.

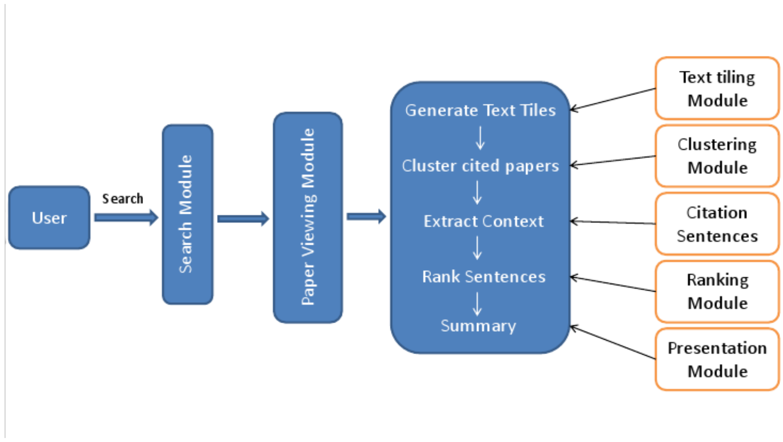


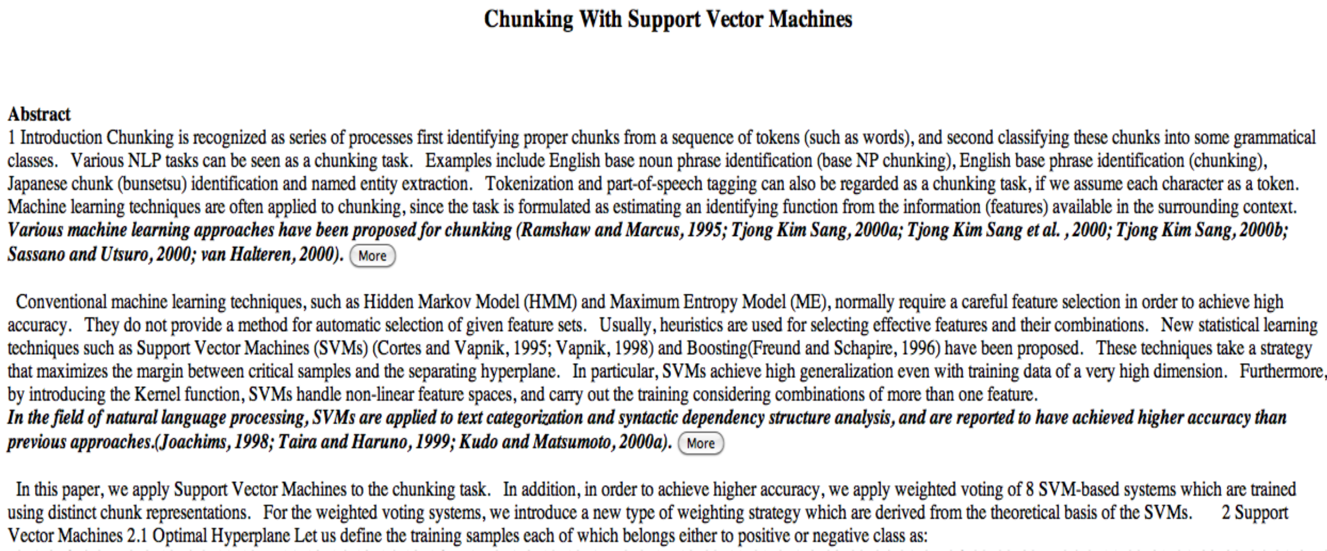
Figure 1: SciSumm summarization pipeline

In our work we make use of this insight by generating a query to focus our multi-document summary from the text closest to the citation. Qazvinian and colleagues (2008) present a summarization approach that can be seen as the converse of what we are working to achieve. Rather than summarizing multiple papers cited in the same source article, they summarize different viewpoints expressed towards the same article from different citing articles. Some of the insights they use in their work also apply to our problem. They used a clustering approach over different citations for the same target article for discovery of different ways of thinking about that article. Citation text has been already shown to contain important concepts about the article that might be absent from other important sections of an article e.g. an *Abstract* (Mohammad et al., 2009) . Template based generation of summaries possessing similar hierarchical topic structure as the *Related Work* section in an article has also been proposed (Hoang et al., 2010). In our work, we consider a flat topic structure in the form of topic clusters. More specifically, we discover the comparable attributes of the co-cited articles using Frequent Term Based Clustering (Beil et al., 2002). The clusters generated in this process contain a set of topically related text fragments called tiles, which are extracted from the set of co-cited articles. Each cluster is indexed with a label, which is a frequent term set present in the tile. We take this to be an approximation of a description for the topic represented by the cluster.

3 System Overview of the SciSumm Summarization System

A high level overview of our system's architecture is presented in Figure 1 . The system provides a web based interface for viewing and summarizing research articles in the ACL Anthology corpus, 2008. The summarization proceeds in three main stages. First, a user may retrieve a collection of articles of interest by entering a query. SciSumm responds by returning a list of relevant articles. The user can continue to read an article of interest as shown in Figure 2. The co-citations in the paper are highlighted in bold and italics to mark them as points of interest for the user. If a user clicks on a co-citation, SciSumm responds by generating a query from the local context of the co-citation and uses it to rank the clusters generated.

As an example consider the following citation sentence "Various machine learning approaches have been proposed for chunking (Ramshaw and Marcus, 1995; Tjong Kim Sang, 2000a; Tjong Kim Sang et al. , 2000; Tjong Kim Sang, 2000b; Sassano and Utsuro, 2000; van Halteren, 2000)". If the user clicks on this co-citation, SciSumm generates a list of clusters and ranks them for relevance. Most of the top ranked cluster labels thus generated are shown in Figure 3 along with the cluster content of the highest ranked cluster labelled as *Phrase, Noun*. The labels index into the corresponding cluster. An example of such cluster is displayed in Figure 4. The cluster has a label *Chunk* and contains tiles from two of the three papers discussing about a topic identi-



Chunking With Support Vector Machines

Abstract
1 Introduction Chunking is recognized as series of processes first identifying proper chunks from a sequence of tokens (such as words), and second classifying these chunks into some grammatical classes. Various NLP tasks can be seen as a chunking task. Examples include English base noun phrase identification (base NP chunking), English base phrase identification (chunking), Japanese chunk (bunsetsu) identification and named entity extraction. Tokenization and part-of-speech tagging can also be regarded as a chunking task, if we assume each character as a token. Machine learning techniques are often applied to chunking, since the task is formulated as estimating an identifying function from the information (features) available in the surrounding context. ***Various machine learning approaches have been proposed for chunking (Ramshaw and Marcus, 1995; Tjong Kim Sang, 2000a; Tjong Kim Sang et al. , 2000; Tjong Kim Sang, 2000b; Sassano and Utsuro, 2000; van Halteren, 2000).*** More

Conventional machine learning techniques, such as Hidden Markov Model (HMM) and Maximum Entropy Model (ME), normally require a careful feature selection in order to achieve high accuracy. They do not provide a method for automatic selection of given feature sets. Usually, heuristics are used for selecting effective features and their combinations. New statistical learning techniques such as Support Vector Machines (SVMs) (Cortes and Vapnik, 1995; Vapnik, 1998) and Boosting(Freund and Schapire, 1996) have been proposed. These techniques take a strategy that maximizes the margin between critical samples and the separating hyperplane. In particular, SVMs achieve high generalization even with training data of a very high dimension. Furthermore, by introducing the Kernel function, SVMs handle non-linear feature spaces, and carry out the training considering combinations of more than one feature. ***In the field of natural language processing, SVMs are applied to text categorization and syntactic dependency structure analysis, and are reported to have achieved higher accuracy than previous approaches.(Joachims, 1998; Taira and Haruno, 1999; Kudo and Matsumoto, 2000a).*** More

In this paper, we apply Support Vector Machines to the chunking task. In addition, in order to achieve higher accuracy, we apply weighted voting of 8 SVM-based systems which are trained using distinct chunk representations. For the weighted voting systems, we introduce a new type of weighting strategy which are derived from the theoretical basis of the SVMs. 2 Support Vector Machines 2.1 Optimal Hyperplane Let us define the training samples each of which belongs either to positive or negative class as:

Figure 2: Interface to read a paper. The sentences containing co-citations are automatically highlighted and contain a "More" button beside them letting the user elaborate on the sentence.

fied by this label. In this specific example the topic was not shared by tiles present in the third paper. The words highlighted are interesting terms which are either part of the label of the cluster or show a low IDF (Inverse Document Frequency) amongst the tiles generated from the co-cited papers. These words are presented as hyper-links to the search interface and can be further used as search queries for finding articles on related topics.

3.1 System Description

SciSumm has four primary modules that are central to the functionality of the system, as displayed in Figure 1. First, the Text Tiling module takes care of obtaining tiles of text relevant to the citation context. It uses the Texttiling algorithm (Hearst, 1997), to segment the co-cited papers into text tiles based on topic shifts identified using a term overlap measure computed between fixed-length blocks of text. Next, the clustering module is used to generate labelled clusters using the text tiles extracted from the co-cited papers. The labels provide a conveniently comprehensible and yet terse description of each cluster. We have used a Frequent Term Based Clustering algorithm (Beil et al., 2002) for clustering. The clusters are ordered according to relevance with respect to the generated query. This is accomplished by the Ranking Module. Finally, the summary presentation module is used to display the ranked clusters obtained from the ranking module. Alongside the clusters, an HTML pane also shows the labels of all the clusters. Having such a bird's-eye view of all the cluster labels helps the user to quickly navigate to an interesting topic. The entire pipeline is used in real-time to generate topic clusters which are useful for generating snippet summary and more exploratory analysis.

In the following sections, we discuss each of the main modules in detail.

3.2 Texttiling

The Text Tiling module uses the TextTiling algorithm (Hearst, 1997) for segmenting the text of each article. Each such segment obtained by the TextTiling algorithm has been referred as a text tile. We have used these text tiles as the basic unit for our summary since individual sentences are too short to stand on their own. Once computed, text tiles are used to identify the context associated with a co-citation. The intuition is that an embedded co-citation in a text tile is connected with the topic distribution of the tile. We use important text from this tile to rank the text clusters generated using Frequent Term based text clustering.

Figure 3: Clusters generated in response to a user click on the co-citation. The list of clusters in the left pane gives a bird-eye view of the topics which are present in the co-cited papers

3.3 Frequent Term Based Clustering

The clustering module employs Frequent Term Based Clustering (Beil et al., 2002). For each co-citation, we use this clustering technique to cluster all the of the extracted text tiles generated by segmenting each of the co-cited papers. We settled on this clustering approach for the following reasons:

- Text tile contents coming from different papers constitute a sparse vector space, and thus the centroid based approaches would not work very well.
- Frequent Term based clustering is extremely fast in execution time as well as and relatively efficient in terms of space requirements.
- A frequent term set is generated for each cluster which gives a comprehensible description of the cluster.

Frequent Term Based text clustering uses a group of frequently co-occurring terms called a frequent term set. Each frequent term set indexes to a corresponding cluster. The frequent term set has the property that it occurs at least once in each of the documents present in the cluster. The algorithm uses the first *k* term sets if all the documents in the document collections are clustered.To discover all the possible candidates for clustering, i.e., term sets, we used the *Apriori* algorithm (Agrawal et al., 1994), which identifies the sets of terms that are relatively frequent. We use entropy measure to score each frequent term set as discovered from the Apriori algorithm. The entropy overlap of a cluster C_i, $EO(C_i)$ is calculated as follows:

$$EO(C_i) = \sum_{D_j \epsilon C_i} -\frac{1}{f_j}.ln(\frac{1}{f_j})$$

where D_j is the jth document which gets binned in the cluster C_i, f_j is the number of clusters which contain D_j. A smaller value means that the document D_j is contained in few other clusters C_i. $EO(C_i)$ increases monotonically as f_j increases. We thus rank the clusters with their corresponding $EO(C_i)$ and then pick a cluster with the smallest entropy overlap $EO(C_i)$. Once a cluster is chosen to be included in the final clustering, we remove the documents present in chosen cluster from other candidate clusters. This results in a hard clustering of documents. We also remove term set corresponding to C_i from the list of candidate frequent term sets and then again recompute the $EO(C_i)$'s for the clusters. We continue this re-scoring and selecting a candidate cluster until the final clustering does not completely exhaust the entire document collection.

3.4 Cluster Ranking

The ranking module uses cosine similarity between the query and the centroid of each cluster to rank all the clusters generated by the clustering module. The context of a co-citation is restricted to the text of the tile in which the co-citation is found. In this way we attempt to leverage the expert knowledge of the author as it is encoded in the local context of the co-citation in our process of automatically ranking the clusters in terms of importance.

CHUNK, TAG, WORD

Ramshaw, Lance A.; Marcus, Mitchell P.1995 The same method can be applied at a higher level of textual interpretation for locating chunks in the tagged text, including non-recursive chunks. For this purpose, it is convenient to view chunking as a tagging problem by encoding the chunk structure in new tags attached to each word. In automatic tests using Treebank-derived data, this technique achieved recall and precision rates of roughly 92% for baseNP chunks and 88% for somewhat more complex chunks that partition the sentence.
Tjong Kim Sang, Erik F.2000 2 Approach Tjong Kim Sang (2000) describes how a systeminternal combination of memory-based learners can be used for base noun phrase (baseNP) recognition. The idea is to generate different chunking models by using different chunk representations. Chunks can be represented with bracket structures but alternatively one can use a tagging representation which classifies words as being inside a chunk (I), outside a chunk (O) or at a chunk boundary (B) (Ramshaw and Marcus, 1995). There are four variants of this representation. The B tags can be used for the first word of chunks that immediately follow another chunk (the IOB1 representation) or they can be used for every chunk-initial word (IOB2).
Tjong Kim Sang, Erik F.2000 Alternatively an E tag can be used for labeling the final word of a chunk immediately preceding another chunk (IOE1) or it can be used for every chunk-final word (IOE2).
Ramshaw, Lance A.; Marcus, Mitchell P.1995 In this study, training and test sets marked with two different types of chunk structure were derived algorithmically from the parsed data in the Penn Treebank corpus of Wall Street Journal 82 text (Marcus et al. , 1994). The source texts were then run through part-of-speech tagger (Brill, 1993c), and, as a baseline heuristic, chunk structure tags were assigned to each word based on its part-of-speech tag. Rules were then automatically learned that updated these chunk structure tags based on neighboring words and their part-of-speech and chunk tags. Applying transformation-based learning to text chunking turns out to be different in interesting ways from its use for part-of-speech tagging.

Top

Figure 4: Example of a summary generated by our system. We can see that the clusters are cross cutting across different papers, thus giving the user a multi-document summary.

4 Evaluation

In a typical evaluation of a multi-document summarization system, gold standard summaries are evaluated against fixed length generated summaries. Summarization conferences such as DUC have competitions where different summarization systems compete on a standard task of generating summaries for a publicly available dataset. The summaries generated using each individual summarization system are then evaluated against the summaries prepared by human annotators. Summarization of scientific article is a novel task and hence no test collection of gold standard summaries exist. Thus, it was necessary to prepare our own evaluation corpus, consisting of gold standard multi-document summaries for a set of randomly selected co-citations.

4.1 Experimental Setup

An important target user population for multi-document summarization of scientific articles is graduate students. Hence to get a measure of how well the summarization system is performing, we asked 2 graduate students who have been working in the computational linguistics community to create gold standard summaries of a fixed length (8 sentences $\sim$ 200 words) for ten different randomly selected co-citations. The students were given guidelines to prepare summaries based on the design goals of the SciSumm system, but not any of its technical details. Thus, for 10 co-citations, we obtained two different gold standard summaries. For ROUGE-1 the average score between each pair of gold standard summaries was 0.518 (Min = 0.388, Max = 0.686). Similarly for ROUGE-2 the average score was 0.242 (Min = 0.119, Max=0.443). While these scores do not have a well-calibrated meaning to them, they give an indication of the complexity of the task. Since the annotators were creating extractive summaries which could justify the co-citation, they had to pay special attention to the section where the co-citation came from. One can consider this similar to the sense making process a reader might go through when using the citing paper as a lens through which to interpret the cited literature.

Note that while SciSumm provides users with an interactive interface that supports navigation between documents, the gold standard summaries are static. Thus, our evaluation is designed to measure the quality of the content selection when taking into consideration the citation context. This would also help us to evaluate the influence exerted by the citation context in the gold standard summaries. In future work, we will evaluate the usability of the SciSumm system using a task based evaluation.

In the absence of any other multi-document summarization systems in the domain of scientific article summarization, we used a widely used and freely available multi-document summarization system called MEAD (Radev, 2004) as our baseline. MEAD uses centroid based summarization to create informative clusters of topics. We use the default configuration of MEAD in which MEAD uses length, position and centroid for ranking each sentence. We did not use query focussed summarization with MEAD. We evaluate its performance with the same gold standard summaries we use to evaluate SciSumm. For generating a summary from our system we used sentences from the tiles which gets clustered in the top ranked cluster. When that entire cluster is exhausted we move on to the next highly ranked cluster. In this way we prepare a summary comprising of 8 sentences.

4.2 Results

For measuring performance of the two summarization systems (SciSumm and MEAD), we compute the ROUGE metric based on the 2 * 10 gold standard summaries that were manually created. ROUGE has been traditionally used to compute the performance based on the N-gram overlap (ROUGE-N) between the summaries generated by the system and the target gold summaries. For our evaluation we used two different versions of the ROUGE metric, namely ROUGE-1 and ROUGE-2, which correspond to measures of the unigram and bigram overlap respectively. We computed four metrics in order to measure SciSumm's performance, namely ROUGE-1 F-measure, ROUGE-1 Recall, ROUGE-2 F-measure, and ROUGE-2 Recall. To measure the statistical significance of this result, we carried out a Student T-Test, the results of which are presented in the results section. The t-test results displayed in Table 1 show that our systems performs significantly better than MEAD on three of the metrics ($p < .05$). On two additional metrics, SciSumm performs marginally better ($p < .1$).

This shows that using the query generated out of the co-citation is useful for content selection from cited papers. Intuitively, this makes sense as each researcher would have a unique perspective when reviewing scientific literature. Co-citations can be considered as micro-reviews which summarizes the thread unifying the research presented in each of the cited papers. This provides evidence that the co-citation context provides useful information for forming an effective query to focus the multi-document summary to reflect the perspective of the author of the citing paper.

Table 1: Average ROUGE results. * represents improvement significant at $p < .05$, † at $p < .01$.

Metric	MEAD	SciSumm
ROUGE-1 F-measure	0.3680	0.5123 †
ROUGE-1 Recall	0.4168	0.5018
ROUGE-1 Precision	0.3424	0.5349 †
ROUGE-2 F-measure	0.1598	0.3303 *
ROUGE-2 Recall	0.1786	0.3227 *
ROUGE-2 Precision	0.1481	0.3450 †

5 Conclusions and Future Work

In this work, we proposed the first unsupervised approach to the problem of multi-document summarization of scientific articles that we know of. In this approach, the document collection is a list of papers cited together within the same source article, otherwise known as a co-citation. The summary is presented in the form of topic labeled clusters, which provide easy navigation based on the user's topic of interest. Another contribution is a technical approach to query oriented multi-document summarization of scientific articles that has been evaluated in comparison with a competitive baseline that is not query oriented. Our evaluation shows that the SciSumm approach to content selection outperforms another multi-document summarization system for this summarization task.

Our long term goal is to expand the capabilities of SciSumm to generate literature surveys of larger document collections from less focused queries. This more challenging task would require more control over filtering and ranking in order to avoid generating summaries that lack focus. To this end, a future improvement that we plan to use a variant on MMR (Maximum Marginal Relevance) (Carbonell et al., 1998), which can be used to optimize the diversity of selected text tiles as well as the relevance based ordering of clusters in order to put a more diverse set of observations from the co-cited articles at the fingertips of users. A natural extension would also be to discover the nature of citations to generate improved summaries. Non-explicit citations (Qazvinian et al., 2010) which could be used to generate similar topic clusters.

Another important direction is to refine the interaction design through task-based user studies. As we collect more feedback from students and researchers through this process, we will use the insights gained to achieve a more robust and effective implementation. We also plan to leverage research in information visualization to enhance the usability of the system.

6 Acknowledgements

This work was supported by NSF EEC-064848 and NSF EEC-0935127.

References

Agrawal R. and Srikant R. 1994. Fast Algorithm for Mining Association Rules In *Proceedings of the 20th VLDB Conference* Santiago, Chile, 1994

Baxendale, P. 1958. Machine-made index for technical literature - an experiment. *IBM Journal of Research and Development*

Beil F., Ester M. and Xu X 2002. Frequent-Term based Text Clustering In *Proceedings of SIGKDD '02* Edmonton, Alberta, Canada

Carbonell J. and Goldstein J. 1998. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries In *Research and Development in Information Retrieval*, pages 335–336

Councill I. G. , Giles C. L. and Kan M. 2008. ParsCit: An open-source CRF reference string parsing package *INTERNATIONAL LANGUAGE RESOURCES AND EVALUATION European Language Resources Association*

Edmundson, H.P. 1969. New methods in automatic extracting. *Journal of ACM.*

Hearst M.A. 1997 TextTiling: Segmenting text into multi-paragraph subtopic passages *In proceedings of LREC 2004, Lisbon, Portugal, May 2004*

Joseph M. T. and Radev D. R. 2007. Citation analysis, centrality, and the ACL Anthology

Kupiec J. , Pedersen J. , Chen F. 1995. A training document summarizer. In *Proceedings SIGIR '95*, pages 68-73, New York, NY, USA. 28(1):114–133.

Luhn, H. P. 1958. *IBM Journal of Research Development.*

Mani I. , Bloedorn E. 1997. Multi-Document Summarization by graph search and matching In *AAAI/IAAI*, pages 622-628. [15, 16].

Nanba H. , Okumura M. 1999. Towards Multi-paper Summarization Using Reference Information In *Proceedings of IJCAI-99*, pages 926–931 .

Paice CD. 1990. Constructing Literature Abstracts by Computer: Techniques and Prospects *Information Processing and Management* Vol. 26, No.1, pp, 171-186, 1990

Qazvinian V. , Radev D.R 2008. Scientific Paper summarization using Citation Summary Networks In *Proceedings of the 22nd International Conference on Computational Linguistics*, pages 689–696 Manchester, August 2008

Radev D. R . , Jing H. and Budzikowska M. 2000. Centroid-based summarization of multiple documents: sentence extraction, utility based evaluation, and user studies In *NAACL-ANLP 2000 Workshop on Automatic summarization*, pages 21-30, Morristown, NJ, USA. [12, 16, 17].

Radev, Dragomir. 2004. *MEAD - a platform for multidocument multilingual text summarization.* In proceedings of LREC 2004, Lisbon, Portugal, May 2004.

Teufel S. , Moens M. 2002. Summarizing Scientific Articles - Experiments with Relevance and Rhetorical Status In *Journal of Computational Linguistics*, MIT Press.

Mohammad, Saif and Dorr, Bonnie and Egan, Melissa and Hassan, Ahmed and Muthukrishan, Pradeep and Qazvinian, Vahed and Radev, Dragomir and Zajic, David 2009. Using citations to generate surveys of scientific paradigms In *Proceedings of Human Language Technologies:The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*

Qazvinian, Vahed and Radev, Dragomir R. 2010. Identifying non-explicit citing sentences for citation-based summarization In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*

Hoang, Cong Duy Vu and Kan, Min-Yen 2010. Towards automated related work summarization In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*