

# Deconstructing Human Literature Reviews – A Framework for Multi-Document Summarization

Kokil Jaidka, Christopher S.G. Khoo, Jin-Cheon Na  
Division of Information Studies

Wee Kim Wee School of Communication and Information  
Nanyang Technological University, Singapore

[kokil, chriskhoo]@pmail.ntu.edu.sg, tjcna@ntu.edu.sg

## Abstract

This study is conducted in the area of multi-document summarization, and develops a literature review framework based on a deconstruction of human-written literature review sections in information science research papers. The first part of the study presents the results of a multi-level discourse analysis to investigate their discourse and content characteristics. These findings were incorporated into a framework for literature reviews, focusing on their macro-level document structure and the sentence-level templates, as well as the information summarization strategies. The second part of this study discusses insights from this analysis, and how the framework can be adapted to automatic summaries resembling human written literature reviews. Summaries generated from a partial implementation are evaluated against human written summaries and assessors' comments are discussed to formulate recommendations for future work.

## 1 Introduction

This project proposes a framework for literature reviews, which has applications in automatic summarization of scientific papers. A literature review is the traditional multi-document summary of research papers which is constructed by a researcher to survey previous findings and its structure follows certain linguistic rules. Several studies have identified that literature reviews are used to achieve distinct rhetorical purposes (Hart, 1998; Bourner, 1996; Boot & Beile, 2005; Jonsson, 2006; Massey, 2006; Torraco, 2005; Hinchliffe, 2003; Bruce, 1994), such as to:

- Compare and contrast previous research.
- Identify gaps in the literature

- Identify new research questions
- Define the proposed research contributions
- Build the justification for the current work
- Situate the work in the research literature
- Reinterpret and critique previous results

These rhetorical characteristics of literature reviews make it a challenging research problem in automatic multi-document summarization – not only should the summarizer identify salient information, but it should also synthesize the summary in a way that achieves certain argumentative purposes. The problem of summarization in context was first identified by Sparck Jones and Endres-Niggemeyer (1995) and subsequently in Sparck Jones' follow-up article (2007), wherein they questioned the usefulness of state-of-the-art summarization methods in addressing users' information needs. As articulated by Sparck Jones (2007) and echoed by Nenkova and McKeown (2011), summarization needs to be viewed as a part of the larger discourse (academic writing) it belongs to, tailored to the purpose (literature review) of summarization, the reader (in this case, a researcher) and the genre being summarized (research papers). Motivated by this research gap, we outline the aims of our analyses:

- To identify how to emulate the purpose of literature reviews, we conducted a discourse analysis to identify the macro-level structure and the sentence-level linguistic expressions embedded in literature review sections.
- To identify the relationship between research paper and literature review, we conducted an information analysis to identify rules for selecting and

transforming information from research papers.

The focus of the paper is to draw insights from the framework to propose strategies for automatic literature review generation. An automatic summary fashioned as a literature review can function as a tool to help literature review writers by pointing out ways in which information in the source papers can be compared and integrated. For information searchers, it can provide a customisable overview of a set of retrieval results that is more readable and more logical than a list of salient sentences.

## 2 Previous Work

This paper investigates the human summarization process through an extensive discourse analysis. Human summarization is a process comprising document exploration to investigate the document macrostructure, relevance assessment by constructing a mental representation and summary production by selecting and transforming text from the source(s) (Endres-Niggemeyer, Maier, and Sigel, 1995). The underlying principle is the theory of human synthesis of information, by Van Dijk and Kintsch (1983).

This study proposes a linguistically motivated framework for summarization. In previous work, a summarization framework developed by Marcu (2000) compressed information from general texts by identifying rhetorical relationships between clauses and sentences, and extracting sentence nuclei. Shiyan, Khoo & Goh (2008) summarized social science dissertation abstracts by referencing a social science taxonomy to identify important information and a specially constructed knowledge bank to identify important inter-relationships. In earlier work, a summarization framework designed by Teufel and Moens (2002) identified 7 categories of scientific arguments and extracted single-document summaries from chemistry and computational linguistics papers (Teufel, Siddharthan & Batchelor, 2009) based on user's queries. However, it required large corpora of manually annotated papers to be applied to any field, and it generated only single-document summaries.

Some other scientific summarization systems aim to model information relationships accurately without concerning themselves with summary

structure. Centrifuser, a framework for summarizing medical literature (Elhadad, Kan, Klavans and McKeown, 2005) produced a multi-document, query-focused indicative summary highlighting the similarities and differences between source documents. The topic tree for the final summary was constructed offline by clustering a large number of documents, thus it was not suitable for real-time user queries. In a related recent approach, Hoang and Kan (2010) presented preliminary results from automatically generating related work sections for a target paper by taking a hierarchical topic tree as an input; however, the requirement of a pre-conceived topic tree limits the scalability of this system. To sum up, these scientific summarization systems are typically delimited by their scalability and generalizability for multiple documents and domains.

Newer approaches in scientific paper summarization rely on preselected information cited in other papers to judge whether information is influential or not, and generate a multi-document summary of a topic (Nanba, Kando & Okumura, 2011) or a single document summary for a paper using its relevant cited information (Qazvinian & Radev, 2008). A system for generating literature surveys through citations was proposed by Mohammad et al. (2009) which applied superficial analysis of research paper citation sentences to suggest model sentences; the present study describes parallel efforts to refine a summarization framework after extensive discourse analysis. We consider providing not just a synopsis of information, but also integrating the synopsis with the contextual and rhetorical features which make a human written literature review a coherent, cohesive and useful reference. Our study thus addresses a different, and more challenging, set of objectives than the citation-based summarizers of recent work.

## 3 Developing the Literature Review Framework

Following the first research aim, we carried out an analysis of the discourse structure of a sample of 30 literature review sections in research papers haphazardly selected from the Journal of the American Society for Information Science and Technology between the years 2000-2008, 2 or 3

from each year. On average, a literature review section was 1146 words in length and it cited 17 studies. The texts were analyzed at 3 levels of detail:

- Macro-level document structure: to identify the different sections of the literature, the types of information they contain and how they are organized hierarchically.
- Sentence-level rhetorical structure: to identify how sentences are framed according to the overall purpose of the literature review.
- Summarization strategies: to identify how information was selected and synthesized for the literature review.

Preliminary findings of these discourse analyses have been discussed in previous work by the authors, notably, in a discussion of the features of the macro-structure of information science literature reviews (Khoo, Na & Jaidka, 2011), rhetorical functions found in literature reviews (Jaidka, Khoo & Na, 2010) and associations between sections in source papers and their citing sentences in literature reviews (Jaidka, Khoo & Na, 2013). The current study applies the discourse characteristics thus identified to develop and test a literature review framework for multi-document summarization.

### 3.1 Designing Document Structure Templates

As noted in academic writing textbooks (Hart, 1998), literature reviews are structured as a hierarchy of topics and each “paragraph” fulfills certain functions. To identify these macro-structures and their functions, we conducted this discourse analysis, proceeding with the assumption that a literature is structured as a set of topic elements, with each topic having a set of embedded study elements (i.e. descriptions of research studies relevant to the topic). An exploratory study was conducted to identify the structures within these topics and their hierarchical relationships. Two Research Assistants holding graduate degrees annotated every sentence with one or more of the following tags:

- title tag: to provide a statement of the topic theme or study objective

- description tag: to encapsulate the details of the topic or study
- meta-summary tag: to provide the writers’ comments as an overview summary of the research in the field
- meta-critique tag: to contain the writers’ critique or interpretation of cited studies, critical comparison of research or justification for the current study
- current-study tag: to refers to and compare with the current work being described in the paper.
- method and result tags: to provide a description of the research methods and research results reported in the cited papers.

In this coding scheme, the meta-summary and meta-critique tags provide the writers’ comments, citing one or more studies together. The rest of the elements comprise descriptive text about individual studies. The average inter-coder reliability score (Cohen’s Kappa) obtained was high at 0.76. Disagreements between the coders were resolved through discussion until a mutually agreeable solution was reached. The analysis identified different types of literature reviews as well as different structures. In our literature review framework, these findings suggested rules for generating different types of literature reviews:

- Integrative literature reviews should comprise a large proportion of meta-summary and meta-critique elements. This is because they discuss and critique ideas from a number of studies in a high-level summary.
- Descriptive literature reviews should report the results of individual studies in detail, outlining their methodology and recommendations. This is because they were found to comprise significantly more study elements.
- Integrative literature reviews should be organized as a hierarchical structure with embedded topics. Comparatively, descriptive literature reviews should be organized as a flat structure, with many more topic elements per text but less embedded topics. This is because

integrative literature reviews were found to comprise an average of 2.5 embedded topics, and descriptive literature reviews had an average of 1.4 embedded sub-topics.

These rules have been applied in designing several integrative and descriptive literature review templates. Fig 1 illustrates one of the template integrative literature reviews we designed. It comprises a level 1 starting topic which acts as the overall topic of the literature review. The topic has other sub-topic elements within it, each of which begins with a meta-summary element which introduces it, followed by study elements to illustrate it. The topic elements determine the logical organization of the literature review; meta-summary are incorporated into the structure because they provide research overviews and highlight the similarities across related papers. The study elements highlight the unique features for individual papers. These templates will be instantiated in the automatic literature review generation process.

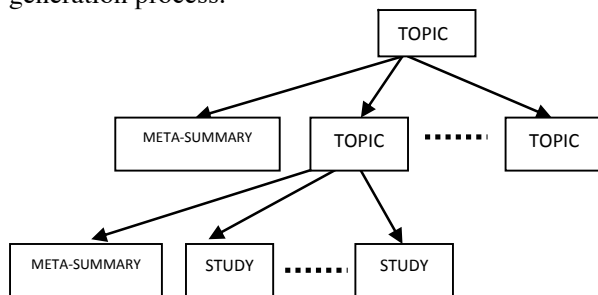


Figure 1. A template document structure in the literature review framework

### 3.2 Designing Sentence Templates

Previous studies of literature reviews (Bunton, 2002; Kwan, 2006) have highlighted the broad rhetorical “moves” which organize the text, but none have attempted to identify their linguistic structure or specific functions. In the clause-level analysis, we annotated linguistic expressions framing research descriptions, defined as discourse markers by Hyland (2004). Although discourse markers include generic logical connectives such as “so”, “therefore” and “because”, we followed Teufel’s criteria (Teufel, 1999 pp. 76) to focus on only those discourse markers which are used in scientific discourse to perform one of the functions listed below:

- Describe a topic: Present a broad overview of research (e.g., “Previous research focused on”) or its context (e.g., “Research in the area of”)
- Describe a study: Cite an author (e.g., “In a study by”) or describe research processes (e.g., “X identified...”, “Y has conducted an experiment to...”)
- Compare studies: Highlight similarities or differences in research (e.g., “Several studies have applied”).
- Provide additional information: Frame examples or enumerate research studies (e.g., “For example”, “A list includes”).

It was found that a total of 110 expressions were used in 1298 variations to frame different types of information in different ways and achieve different rhetorical functions. We have applied these findings in the literature review framework to develop sentence templates for text generation, and to formulate rules for selecting templates which are significantly associated with the type of literature review and discourse element to be populated:

- In integrative literature reviews: apply regular expressions which describe research objectives in the description elements. In the meta-summary elements in integrative literature reviews, apply expressions which “state the common aims”.
- In descriptive literature reviews: choose expressions which “state the research method” and “state the common approaches” in the description and meta-summary elements.

Regular expressions are applied for text-to-text generation, serving as a means to extract information from source papers as well as to map them into appropriate sentence templates. Those applied to extract and instantiate research objective sentences within topics, studies and comparisons are illustrated in Table 1.

### 3.3 Designing Information Selection and Summarization Strategies

In accordance with the second research aim, we conducted a content analysis to identify the relationship between the source papers and the final literature review. Similar work describing text editing strategies has been done by Jing and

McKeown (1999); however, in this analysis we extend their objectives to additionally identify:

- The source sections of the paper from where information was selected (i.e., Abstract Introduction, Methodology, Results or Conclusion).
- The types of transformations used to convert the source sentence to the referencing sentence (i.e., copy-paste, paraphrase, or higher-level summary).
- Identifying the types of information selected from the source papers (i.e., objective, methodology, results and critical summary).
- Analysis of the reasons for preference of one source sentence over another, despite providing similar information. This was inferred by comparing candidate source sentences against each other.

The corpus for analysis was constructed by analyzing the 20 literature reviews line-by-line and retaining all the sentences referencing previous work, either explicitly (e.g., “X and Y (1998) conducted experiments in transitive translation”) or implicitly by adding onto the details of a cited study (e.g., “Studies have also focused on users' mental models of information seeking (X, 1989)”.

A total of 349 references were collected from the twenty literature review sections. Sentence

providing definitions, or citing sources other than research papers, were further discarded because they lay outside the scope of our analysis. The findings, revealed that more than a quarter of all selected information was from the Abstract of the source paper. The information selected by the reviewer is copy-pasted more often in descriptive as compared to integrative literature reviews. Some of these findings have been applied to suggest strategies for information selection and summarization in the literature review framework:

- For research objective information: choose sentences from the Abstract and Introduction of source papers; copy-paste it into descriptive literature reviews, but paraphrase it in integrative literature reviews.
- In descriptive literature reviews: provide detailed method information, copy-pasted from the Introduction or Method of source papers.
- In integrative literature reviews: provide detailed result information, summarized at a higher level from the Results and Conclusions.

When more than one sentence provides the same factual information, the sentence selection criteria listed in Table 2 should be followed to choose the more concise alternative.

Function	Type of Information Required	Regular Expression which map into Sentence Templates
Describe a topic	Introduce a topic through its research aspects	(Researchers   Research) (have   has) (in   are concerned with   have addressed   proposed   observed   investigated   focused on)
	Introduce a topic through its literature review	The (literature review   prior work) (covered   dealt with   looked at   focused on )?
	Introduce area of research	research   studies   findings) in the (field   area   domain   context) of
Describe a study	State the study objective	(the study   we   who) (conducted   explored   proposed   pursued   described   attempted to   represented   analyzed   examined   investigated   deals with   seeks to discover)
	State the study motivation	(The   Their) underlying research (question   objective) (was   is)
	State the study hypothesis	(They) (argue   opine   hold   debate   believe) that
Compare studies	State the common aim of studies	The (common)? (issue   motivation   aim   principle) (for   behind) (many   most   some   these   such   existing) studies (Many   Most   These   Some   Such   Existing   Various)? (studies   work) have (explored   focused on)

Table 1. Regular expressions obtained from clause-level analysis

Type of Criteria	Order of Priority
Lexical	• “This article/paper...”
	• “The aim/goal/objective is...”
	• “We present/ describe...”
	• “Recent research into...”
	• Sentences with how/what/why questions
Syntactic	• Sentence having the main topic in its main clause
	• The sentence with fewer clauses
	• The sentence with no back-referencing
Surface	• Sentence from the first paragraphs of a section
	• The title of the source paper
	• The sentence which is the shortest

Table 2. Criteria for selecting sentences

#### 4 Evaluation

To evaluate the framework, the objective was to compare its “human-ness” represented by its **Comprehensibility, Readability and Usefulness** against human-written literature reviews and machine-generated sentence extracts. For this purpose, the framework was partially adapted in a summarization method focusing on comparing research objective information extracted from Abstracts and Introduction sections, and presenting a topical overview resembling a three-level literature review. The output generated is similar to the summaries generated by Centrifuser (Elhadad et al., 2005) – sentences are extracted to provide a synopsis of similarities and unique features of studies are highlighted for individual papers; however our prototype does so without rely on external domain knowledge. The method was implemented in Java on the Eclipse IDE, and it comprised three stages:

- Text pre-processing: to extract sentences from the Abstract and Introduction of the input source papers. Here the text is segmented, tokenized, parsed, stop-words are filtered and n-grams of noun phrases are created to represent concepts in the source papers.
- Information selection and integration: to identify similarities and differences across the research objective sentences of source papers. It selects important concepts based on the document frequency of lexical concept chains (Barzilay and McKeown, 2005), and applies the research objective sentence selection rules developed in the

framework to select important information for summarization.

- Text presentation: to produce text that has the characteristics of the literature review. It applies the document structure described in the framework, to organize the literature review, and sentence templates particular to research objective information in integrative literature reviews (the ones listed in Table 1).

The resultant summaries resemble a human written literature review because they are laid out as a topic tree and present a comparative overview of similarities and unique features. However, some grammatical errors can be spotted, which would need a post-processing module to remove.

30 sets of information science source papers were prepared by sampling topics from 30 literature reviews from 2000-2008 issues of JASIST, Journal of Documentation and Journal of Information Science and downloading the papers they cited. Only 3-10 source papers were downloaded for every sampled topic; this was so that the task could be manageable for the researchers constructing the human summaries. An excerpt system summary is provided in Table 3.

For each input set of related research papers, three types of summaries were generated, each with a different kind of method – framework-based structure (by our method), sentence-extraction structure (by the baseline, MEAD) and a human-written summary by a researcher:

- MEAD: The MEAD summarization system (Radev, Jing, Stys, & Tam, 2004) was the baseline; it followed a sentence-

extraction approach to generate multi-document extracts of information (generally news articles).

- System: Our system based on the framework, and focusing on the similarities and differences between research objectives at the lexical and syntactic level.
- Human: Five researchers from the School of Humanities and Social Sciences of our university summarized the research objective sentences from set of source papers in the context of a given (main) topic.

---

This literature review presents research in relevance published by Barry (1994), Harter (1992), Tang and Solomon (1998), Vakkari and Hakala (2000) and Wang and Soergel (1998). Studies by Barry (1994) and Tang et al. (1998) focus on retrieval mechanism. Researchers in relevance have also considered users (Harter, 1992; Vakkari et al., 2000; Wang et al., 1998). The study by Vakkari et al. (2000) demonstrates that it is productive to study relevance as a task and process-oriented user construct. Studies by Wang et al. (1998) and Tang et al. (1998) focus on dynamic models. The study by Tang et al. (1998) is a step in the empirical exploration of the evolutionary nature of relevance judgments.

---

Table 3: Excerpt from a system summary

In the human summaries, the coders selected an average of 3 sub-topics and 8 unique sub-topics in their summaries. Human summaries also had the highest compression rate of 18%, as compared to a compression rate of 25% by MEAD and our System. An inter-coder agreement was conducted over 10 summaries by taking the summaries done by one of the post-graduate researchers as reference and comparing each pair of summaries, considering each of the “similarities” or “differences” as a “common” or “unique” sub-topic. Comparisons revealed that the coders usually had the same idea of what constituted an important “similarity” or common sub-topic (percent agreement= 70%) though they often chose

different “differences” or unique sub-topics in their summaries (percent agreement= 56%).

Content evaluation of the 30 sets of summaries by the ROUGE-1 metric (Lin & Hovy, 2003) revealed that system summaries had a higher but not significantly different effectiveness or f-measure of 0.38 as compared to the baseline (0.33). We developed our own version of ROUGE to measure information overlap by comparing the information concepts extracted from summaries. It was different from the standard ROUGE-1 in three ways: it filtered out “research stopwords” such as “method”, “experiment” and “study”, which didn’t represent research information; it aggregated words which shared the same lemma; and it also conflated co-occurring adjacent words into the same information concepts. Consequently, we obtained real scores of effectiveness in terms of higher f-measure scores for both the system and the baseline. The system’s f-measure (0.57) was a significant improvement over the baseline (0.50) at the 0.01 level. The results are provided in Table 4.

For the quality evaluation, 90 questionnaires were prepared from the 30 sets of summaries, using permutations of presentation orders to account for carry-over effects during assessment. To recruit assessors, a call for participation in the evaluation was broadcast over the internet, through postings in discussion boards, personal emails and library sciences mailing lists. The invitation was also personally extended to authors of other publications in JASIST, JDoc and JIS. The invitation for participation was restricted to only Library and Information Science and Computer Science researchers and PhD students who had passed their qualifying exam. It was anticipated that such assessors would be more familiar with the topics in the summary, and would be able to make meaningful comments about the summaries and their characteristics, such as lack of evident comparisons and generalizations, or incorrect comparisons and generalizations among unlike information. There were a total number of 35 assessors with a mean research experience of 6 years, who provided 67 responses, by filling out 1 or 2 each, over a period of two months. The assessors were from reputable international universities in different countries. The highest degrees held by the assessors varied from Bachelors (for PhD students who had passed their qualifying exam) to PhD. They scored the



summaries on their Comprehensibility, Readability and Usefulness and also provided qualitative comments to the following questions:

- What did you like about this summary?
- What did you find confusing about this summary?
- How is this summary, a good/bad literature review?

The quantitative results in Table 5 show that the System summary was significantly more readable and more useful than the baseline at the 0.05 level. The qualitative results (provided in Table 6) are equally interesting and show that researchers with different number of years of research liked or disliked different things about the System summary. Researchers with 0-4 years of experience did not have any specific preference of one type of summary over another. Researchers with 5-8 years of experience were more conscious of grammatical errors and repetition mistakes in the system summary. Researchers with 9-12 years of experience ignored the grammatical errors in Human summaries and System and instead criticized their lack of detail. Researchers with 13 years or experience or more were sensitive to the overall “context” and “flow” of the summary. Most of the assessors were able to identify the main topic and its related sub-topics; however, they experienced the System as being more disjointed, lacking “focus” as compared to the Human summaries. On the whole, researchers were satisfied with the overview provided as well as the hierarchical organization. It would be interesting to see whether these findings and differences would be replicated in a larger study.

Measures	System	MEAD
Recall	0.70	0.63
Precision	0.49	0.44
F-measure	0.57	0.50

Table 4. Results from the content evaluation (N=30)

	MEAD	System	Human
Comprehensibility	5.6	5.6	6.2
Readability	4.9	5.3	5.6
Usefulness	5.7	6.4	6.3

Table 5. Results from the quality evaluation (N=67)

## 5 Conclusion and Future Work

This study has analyzed how authors select information, transform it and organize it in a definite discourse structure as a literature review. Our findings identified two styles of literature reviews – the integrative and descriptive literature reviews, with different profiles of discourse elements and rhetorical expressions. Integrative literature reviews present information from several studies in a condensed form as a critical summary, possibly complemented with a comparison, evaluation or comment on the research gap. The focus is on highlighting relationships amongst concepts or comparing studies against each other. Descriptive reviews present experimental detail about previous studies, such as the approach followed, their results and evaluation. The focus is on providing important details of previous studies in a concise form.

From these findings, we conjecture that authors begin a literature review with an overall strategy in mind. They select and edit the information content based on the style of literature review. They may choose to write an integrative style of literature review to guide the reader along a critical survey of previous research. To support their argument, they paraphrase information selected from the Abstract and Conclusion sections, and integrate information from the Results sections into a high-level overview of important findings. Accordingly, they choose the discourse structure and linguistic expressions to frame their argument.

Our framework has since been validated on a larger sample size of 90 articles selected from 3 top journals in information science. It is recommended for application in a complete automatic literature review generation system, wherein a user would be able to control the style of literature review, the level of detail and analysis required, as well as the structure of the layout and the number of topics. At the information selection stage, it would be able to apply different information selection and transformation strategies to generate different parts of a literature review. At the text generation stage, it would be able to introduce a topic and describe its context and core concepts, describe a study and its objectives, methods and findings, delineate a research gap and identify the common and different features among studies, and illustrate its argument with examples.



	Year 0-4	Year 5-8	Year 9-12	Year 13+
Comprehensibility	<ul style="list-style-type: none"> <li>- It gives a good overview on the topic and points</li> </ul>	<ul style="list-style-type: none"> <li>- I liked the structure.</li> <li>- It summarizes the research and connects the authors to the topic by the use of "these authors."</li> <li>- It's not too short nor too long.</li> </ul>	<ul style="list-style-type: none"> <li>- Easy to read and understand.</li> <li>- It is better review than the others because it tries to tie the literature together in some fashion.</li> </ul>	<ul style="list-style-type: none"> <li>- There seemed to be no reason for the ordering of the sentences about the different research papers</li> <li>- Each individual statement in the summary seems relevant (of some objective value) by itself, but all together lacks uniformity in subject.</li> <li>- However it does seem to get the core issues.</li> </ul>
Readability	<ul style="list-style-type: none"> <li>- Continuity</li> <li>- Yet, the linking of sentence could be better.</li> <li>- Too many repetitions, but gives some information</li> </ul>	<ul style="list-style-type: none"> <li>- This summary is neither readable nor informative.</li> <li>- The same studies are cited several times</li> <li>- It kept repeating all the studies.</li> <li>- It felt very disjointed, maybe because of all the small paragraphs.</li> <li>- Badly written, hard to read.</li> </ul>	<ul style="list-style-type: none"> <li>- It flows well</li> <li>- Has some sentences seemingly unrelated to neighboring sentences</li> </ul>	<ul style="list-style-type: none"> <li>- Generally easy to read.</li> <li>- There are a few mistakes in grammar, which is distracting.</li> <li>- Very readable.</li> <li>- Like: seems to have a bit of flow.</li> </ul>
Usefulness	<ul style="list-style-type: none"> <li>- This summary seems quite good</li> <li>- I feel I got an overview over the research in the area.</li> <li>- The summary covered a good deal of literature</li> <li>- The overview is nice but still really flat.</li> </ul>	<ul style="list-style-type: none"> <li>- This is the best summary of the sample.</li> <li>- Comprehensively covers the text</li> <li>- The summary provides information about groups of studies researching certain topics</li> <li>- This summary provides an overview of research in web search with more informative details</li> </ul>	<ul style="list-style-type: none"> <li>- Comparison between studies is helpful.</li> <li>- More info required about study, including methods, findings.</li> <li>- It would be pretty useful for lit review.</li> <li>- While comparisons of different papers are well done, it would also be useful to have more description of each study.</li> </ul>	<ul style="list-style-type: none"> <li>- Should give an indication of these trends in order to help the reader contextualize the research field.</li> <li>- There is an attempt at relating studies to each other so that one gets an overview of the research area.</li> </ul>

Table 6. Comments on System by assessors with different years of research experience

## References

- Barzilay, R., & McKeown, K. R. (2005). Sentence fusion for multidocument news summarization. *Computational Linguistics*, 31(3), 297-328.
- Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational researcher*, 34(6), 3-15.
- Bourner, T. (1996). The research process: four steps to success. *Research methods: guidance for postgraduates*, Arnold, London, 7-11.

- Bruce, C. S. (1994). Research students' early experiences of the dissertation literature review. *Studies in Higher Education*, 19(2), 217-229.
- Bunton, D. (2002) Generic moves in Ph.D Introduction chapters. In J. Flowerdew (Ed.), *Academic Discourse*. London: Longman.
- Cooper, H. M. (1988). The structure of knowledge synthesis. *Knowledge in Society*, 1, 104-126.
- Hoang, C., & Kan, M.Y. 2010. Towards automated related work summarization. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING'10): Posters* (pp. 427-435).
- DUC. (2002). The Document Understanding Conference. Retrieved Oct 2010, from <http://duc.nist.gov>.
- Elhadad, N., Kan, M. Y., Klavans, J. L., & McKeown, K. R. (2005). Customization in a unified framework for summarizing medical literature. *Artificial Intelligence in Medicine*, 33(2), 179.
- Endres-Niggemeyer, B., Maier, E., & Sigel, A. (1995). How to implement a naturalistic model of abstracting: four core working steps of an expert abstractor. *Information Processing & Management*, 31(5), 631-674.
- Guo, Q., & Li, C. (2007, August). The Research on the Application of Text Clustering and Natural Language Understanding in Automatic Abstracting. In *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on* (Vol. 4, pp. 92-96). IEEE.
- Hart, C. (1998). *Doing a literature review*. London: Sage.
- Hinchliffe, L. (2003). Having your say in a scholarly way. *Research Strategies*, 19, 163-164.
- Hyland, K. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, 13(2), 133-151.
- Jing, H., & McKeown, K. R. (1999). The decomposition of human-written summary sentences. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 129-136). ACM.
- Jaidka, K., Khoo, C., and Na, J.-C. (2010). Imitating Human Literature Review Writing: An Approach to Multi-Document Summarization. In *Proceedings of the International Conference on Asian Digital Libraries (ICADL)* (pp. 116-119). Australia: Springer-Verlag.
- Jaidka, K., Khoo, C., & Na, J. C. (2013). Literature Review Writing: How Information is Selected and Transformed. *Aslib Proceedings*, 65(3), 303-325.
- Khoo, C., Na, J. C., & Jaidka, K. (2011). Analysis of the macro-level discourse structure of literature reviews. *Online Information Review*, 35(2), 255-271.
- Kwan, B. S. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes*, 25(1), 30-55.
- Lin, C. Y., & Hovy, E. (2003, May). Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1* (pp. 71-78). Association for Computational Linguistics.
- Marcu, D. (1997, July). From discourse structures to text summaries. In *Proceedings of the ACL* (Vol. 97, pp. 82-88).
- Nenkova, A., & McKeown, K. (2011). *Automatic summarization*. Now Publishers Inc.
- Nanba, H., Kando, N., & Okumura, M. (2011). Classification of research papers using citation links and citation types: Towards automatic review article generation. *Advances in Classification Research Online*, 11(1), 117-134.
- Ou, S., Khoo, C. S. G., & Goh, D. H. (2008). Design and development of a concept-based multi-document summarization system for research abstracts. *Journal of information science*, 34(3), 308-326.
- Radev, D. R., Jing, H., Styś, M., & Tam, D. (2004). Centroid-based summarization of multiple documents. *Information Processing & Management*, 40(6), 919-938.
- Saggion, H., & Lapalme, G. (2002). Generating indicative-informative summaries with sumum. *Computational linguistics*, 28(4), 497-526.
- Mohammad, S., Dorr, B., Egan, M., Ahmed, H., Muthukrishnan, P., Qazvinian, V., Radev, D., Zajic, D. (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* (pp. 584-592). Association for Computational Linguistics.
- Sparck Jones, K., & Endres-Niggemeyer, B. (1995). Automatic summarizing. *Information Processing & Management*, 31(5), 625-630.

- Sparck Jones, K. (2007). Automatic summarising: The state of the art. *Information Processing & Management*, 43(6), 1449-1481.
- Teufel, S. (1999). Argumentative zoning: Information extraction from scientific text (Doctoral dissertation, University of Edinburgh).
- Teufel, S., & Moens, M. (2002). Summarizing scientific articles: experiments with relevance and rhetorical status. *Computational linguistics*, 28(4), 409-445.
- Teufel, S., Siddharthan, A., & Batchelor, C. (2009, August). Towards discipline-independent argumentative zoning: Evidence from chemistry and computational linguistics. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3* (pp. 1493-1502). Association for Computational Linguistics.
- Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review*, 4(3), 356-367.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.