Literature Review Writing: A Study of Information Selection from Cited Papers

KOKIL JAIDKA, kokil@pmail.ntu.edu.sg
CHRISTOPHER KHOO, chriskhoo@pmail.ntu.edu.sg
JIN-CHEON NA, tjcna@e.ntu.edu.sg
Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore

ABSTRACT

This paper reports the results of a small study of how researchers select and edit research information from cited papers to include in a literature review. This is part of a bigger content analysis and linguistic analysis of literature reviews. This study aims to answer the following questions: *where* do authors select information from the cited papers (e.g., Abstract, Introduction, Conclusion section, etc.)? *What* types of information do they select (e.g., research objectives, results, etc.), and *How* do they transform that information (e.g., paraphrasing, cut-pasting, etc.)? In order to answer these questions, we analyzed the literature review section of 20 articles from the Journal of the American Society for Information Science & Technology, 2001-2008, to answer these questions. Referencing sentences were mapped to source papers to determine their origin. Other features of the source information were also annotated, such as the type of information selected and the types of editing changes made to it before including into the literature review. Preliminary results indicate that authors prefer to select information from the Abstract, Introduction and Conclusion sections of the cited papers. This information is transformed through cut-paste, paraphrase or higher-level semantic transformations to describe the research objective, methodology and results of the referenced study. The choices made in selecting and transforming the source information appeared to be related to the two styles of literature review finally constructed – integrative and descriptive literature reviews.

Keywords: Literature reviews; Multi-document summarization; Information science; Information extraction; Information selection.

INTRODUCTION

This study is part of a larger project to develop an automatic literature-review generation system that summarizes a set of related research papers into a literature review (Jaidka, Khoo & Na, 2010). Our approach is to emulate human literature-review writing behavior, right from the selection of information from the source papers, integrating the information into a logical argument and presenting the information using appropriate rhetorical devises, discourse organization and linguistic expression. We seek to understand and model human literature review writing behaviour by carrying out content and linguistic analyses of literature reviews included in journal articles published in major information science journals.

This paper reports the findings of an initial study of the authors' information selection strategy when crafting a literature review. Specifically, we seek to answer the following questions:

- *Where* do authors select information from the cited papers (e.g., Abstract, Introduction, Conclusion section, etc.)?
- *What* types of information do they select (e.g., research objectives, results, etc.)?
- *How* do they transform that information (e.g., paraphrasing, cut-pasting, etc.)?

We analyzed the literature review section of 20 articles from the Journal of the American Society for Information Science & Technology, 2001-2008, to answer these questions.

PREVIOUS STUDIES

A literature review can be considered a multi-document summary of related research papers, integrating information from the source papers and presenting them as a logical justification for the author's research.

A literature review is our natural choice in format for modeling a multi-document research summary because it is the way scientific research information has traditionally been summarized and reviewed. They are typically written by researchers who survey previous

studies in order to identify their shortcomings and to place their own work in the context of previous findings. Hart (1998, p. 27) listed the functions that a literature review can serve:

- To distinguish what has been done from what needs to be done
- To identify important variables relevant to the topic
- To synthesize earlier results and ideas, and gain a new perspective
- To rationalize the significance of the problem
- To identify the main methodologies and research techniques that have been used
- To place the research in context with state-of-art developments, and so on.

Little is known about how literature review authors select information, integrate it and present it. Several studies have documented how professional abstractors as well as authors write abstracts—and the characteristics of such abstracts – and have formulated standards for constructing and evaluating abstracts (Cremmins, 1982; Endres-Niggemeyer, Maier, & Sigel, 1995). Cremmins (1982) introduced an analytical reading model for abstract writing comprising *retrieval reading, creative reading* and *critical reading.* Similarly, Endres-Niggemeyer et al. (1995) identified the abstracting strategies of experts as *document exploration, relevance assessment* and *summary production.* These studies do not cover the questions of *what kind of* information is selected, or *why* a particular source is regularly preferred over others. In our research, we are less interested in the abstracting or literature review *process* and more interested in its *inferred characteristics,* since our aim is to emulate the *output* of these processes*.* Besides, these studies of abstract writing may not apply to literature review writing because the latter are intended as a structured argument justifying the current research in the context of previous research. While abstracts are summaries of individual studies, literature reviews compare and contrast numerous studies to suit their purpose.

There have also been few studies of how the source text are transformed and edited when they are incorporated in a summary. One study in the context of news summarization was carried out by Jing and McKeown (1999) who identified the kinds of transformations which are manually performed on source sentences when they are being included into a news summary, namely, sentence reduction, sentence combination, syntactic transformation, paraphrasing, generalization/specification and reordering. They trained a Hidden Markov Model classifier to detect cut-paste transformations by comparing with the source sentences. Although their study details syntactic and semantic source transformations, it was conducted at the sentence level and focused on cut-paste strategies. In our study, we take this analysis a step further by relating the transformations performed to the types of information and the section of the source paper from which the information is selected.

It may be argued that our analysis of referenced- and source-sentences bears resemblance to citation analysis studies which explore the relationship between the citing and the cited paper (Teufel, 1999; Nanba & Kando, 2000; Cronin & Shaw, 2002). Teufel (1999) used a set of rhetorical relationships to categorize the relationship between the citing and the cited paper. In older research, Chubin and Moitra (1975) extended previous research in citation behavior analysis to explore why researchers cite other papers. These studies typically use cue phrases and other textual markers to gauge the attitude of the citer towards the cited, and frame their categories based on this surmise. While these studies provide useful insights into citation behavior, they do not answer any of our research questions about the preferential relationship between a reference sentence and different sections of a source paper, or the preferential relationship between the type of information selected in a source sentence and the corresponding editing changes performed on it.

In previous work, we have analyzed the macro-level and rhetorical-level discourse structure of literature review sections of information science papers to identify patterns in the discourse organization of literature reviews. We identified two distinct styles of literature review writing, namely the *integrative* and *descriptive literature reviews*. Each type of literature review is written with a distinct profile of discourse elements and rhetorical arguments used (Khoo et al., in press). *Descriptive literature reviews* summarize individual papers/studies and have more individual study-related discourse elements which provide details of the research methods and results of cited studies. *Integrative literature reviews* summarize the ideas and results from the cited studies at a higher level; they have more critique-related discourse elements which build a critical summary of topics or illustrate the author's argument. Building on this finding, we are

comparing the information selection strategies used for integrative versus descriptive literature reviews, to determine if different strategies are used.

RESEARCH METHOD

Twenty research articles were sampled from eight volumes of the Journal of the American Society for Information Science & Technology (JASIST) (2001-2008), two or three articles from each year. JASIST was selected as the corpus of the study because JASIST is a leading journal in the field and carries high-quality research articles with substantive literature reviews. This study will later be extended to other information science journals to carry out comparisons between journals.

The literature review sections from these twenty articles were extracted and annotated to identify the document macro-structure (Khoo et al., in press). During this analysis, it was observed that some literature reviews were *descriptive* and summarized individual papers/studies. Other literature reviews were *integrative* because the authors synthesized a summary of trends and milestones in the research area, based on the information in different cited papers. The literature reviews were then categorized according to their writing style, and it was found that the set comprised of 11 descriptive and 9 integrative literature reviews. In the following sections, we provide the procedure and findings of a study of the information selection strategies in these literature reviews.

We analyzed the literature reviews line-by-line and extracted every sentence which referenced the work of another study. These included sentences which did not carry an explicit reference but continued a description by a previous sentence that contained the explicit cite. Of the leftover sentences which did not reference another study, most of them were general descriptions of a topic or process, and not specific to one research study. Of the selected sentences, we eliminated those which referenced information from sources which were not research papers. These sources include books, technical reports, websites, professional articles, and literature survey articles. This was done because journal source papers were easier to obtain and analyse. The cited research papers were retrieved from bibliographic and fulltext databases and from the Web. Occasionally, the online version of a research paper would not available, for example if it was a very old paper dating back to the 1960s or 1970s. In these cases, the referencing sentence would not be analysed any further.

Our analysis was based on the premise that every referencing sentence contains some information taken from the source paper. This information would be about a particular aspect of the cited study, such as its study objective or research methods. The reference would be traceable to a single sentence or a group of sentences in the source paper.

For each referencing sentence, the source sentence in the cited research paper was located and the paper section (e.g., abstract, introduction, etc.) in which the source sentence occurred was noted. We refer to this as the *source section*. If there was more than one possible candidate source sentence in the cited paper, the sentence was selected if it required minimal transformation to be changed into the referencing sentence. For example, simple edits such as deletion of a word (cut-pasting transformations) were preferred over rewording or substitution of words (paraphrasing transformations).

We carried out three kinds of analysis on the referencing sentences and source sentences:

- Analysis of the referencing sentences to identify the *types of information* selected from the source papers.
- Analysis of the source papers to identify the *types of source sections* from which the information is selected.

Comparison of the referencing sentence and the associated source sentence to identify the types of transformations used to convert the source sentence to the referencing sentence.

Types of Information

The purpose was to identify the type of research information selected for inclusion in a literature review. We have adapted the categories of information from our previous study which classified sentences using a decision tree induction algorithm (Ou, Khoo & Goh, 2006). We labelled the

information type and conceptual structure of every referencing sentence. We referred to this stage as *information analysis* and conducted it at the *semantic* and the *conceptual* level as described below.

We coded the referencing sentence with the following information types:

- ***Research Objective*** – referencing the purpose of the cited study (e.g., *"Lehtokangas and Airio conducted experiments in transitive translation on several European languages (Lehtokangas & Airio, 2002)."*)
- ***Research Method*** – referencing the procedure followed in the cited study (e.g., *"They tagged the source query terms with part-of-speech tags and find all the term translations with matching part-of-speech."*)
- ***Research Result*** – reporting the finding or conclusion of the cited study (e.g., "*Their data showed a significant difference in the mean citation rates between all pairs of resources except between Google Scholar and Scopus for condensed-matter physics in 2003."*)
- ***Critique*** – providing the author's critique of the cited study (e.g., *"This evaluation did not use recall and precision measurement to indicate the evaluated system's performance either."*).

For each information type, we constructed a list of related concepts and indicative keywords and phrases to help us to code the referencing sentences consistently. For example, the indicative words/phrases for research method include: *model, output, perform, conduct, estimate, construct, calibrate, control, compute, measure, technique* and so on.

Types of Source Sections

Information in every referencing sentence was coded with the source sections from where it was extracted:

- Abstract
- Introduction section
- Conclusion section
- Results section
- Method section
- Related Work section (the literature review section of the source paper)
- Other
- Unknown.

Other was used to represent non-typical source text such as Headings, Captions, Titles, Tables etc. In case a source sentence information could not be found, the source location was annotated as *Unknown*. This occurs when the citing author provided a high level summary of the source paper's objectives, methods or findings in such a way that no one sentence can be identified to contain the information. It also occurs when the citing author critiques or comments on the source paper.

Types of Transformation

We annotated the type of transformation performed for generating every referencing sentence:

- ***Cut-paste*** – which involved little or no changes made to the source information. Minor modifications like change of tenses and parts of speech or reordering sentences were allowed. Some parts of the source information may be dropped, like introductory clauses, rhetorical devices, auxiliary clauses or adverbs.
  Here is an example:
  Referencing Senctence: *"Resnik, Oard and Levow proposed techniques for combining evidence from dictionary-based and corpus-based translation lexicons (Resnik, Oard, & Levow, [2001])."*
  Source Sentence: *"We present two techniques for combining evidence from dictionary-based and corpus-based translation lexicons."*

- ***Paraphrase*** – which involved significant lexical transformations like rewording, paraphrasing and using synonyms to convey the source information. The content remains the same as in the source sentence.
  Example:
  Referencing Sentence: "*They tagged the source query terms with part-of-speech tags and find all the term translations with matching part-of-speech."*
  Source Sentence: *"Source language (Spanish) queries are first tagged using a part-of-speech (POS) tagger. Each Spanish source term is replaced by all possible target language (English) translations for the term's POS."*
- ***Summary*** – which involves semantic transformation of the source information in order to provide a higher-level gist of its information. This involves significant modifications to the source form, which may not be isolated to a single sentence but may be summarized from the information in many sentences from different locations in the text.
  Example:
  Referencing sentence: *"Their experiments have shown that their schemes can accomplish truthful predictions while preserving individual user's privacy."*
  Source sentence: *"Our solution makes it possible for servers to collect private data from users for collaborative filtering purposes without compromising users' privacy requirements. Our experiments have shown that our solution can achieve accurate prediction compared to the prediction based on the original data."*
- ***Critical reference –*** which involves transforming the information in the paper into a critical argument. The reference is embedded in the author's critique of its approach or results and cannot be traced to a location in the source paper.
  Example:
  Reference sentence: *"Therefore, the effect of personal subscriptions when measuring institutional user statistics may be problematic, having the effect of under-representing the use of popular browsing journals."*

RESULTS

Profile of Integrative and Descriptive Literature Reviews

Table 1 presents cross-tabulation tables that compare integrative and descriptive literature reviews in terms of the types of information cited, source sections from which the information was extracted, and types of transformations performed on the source sentences. The numbers in bold indicate higher cell frequencies than expected. Pearson Chi-Square test of independence was performed to find whether there were significant relationships between type of literature review (integrative versus descriptive) and information type, source section and transformation type.

Table 1 suggests that descriptive literature reviews reference a higher proportion of *research method* information than expected (41%, compared to 31% for integrative literature reviews). Integrative literature reviews reference a higher proportion of *research results* and has more critiques. However, the Pearson Chi-Square test indicates no significant relation between literature review type and information type. Overall (combining both integrative and descriptive literature reviews), a relatively high proportion of referencing sentences carry information on *research method* (36%), compared to 26% for *research objective* and 28% for *research result*. This is unexpected, and could be because the research method of a study requires more sentences to describe its details.

The second part of Table 1 shows the relation between type of literature review and type of source section. They are significantly associated at the 0.001 level (α= 5.299E-5). Integrative literature reviews reference more information from the *Conclusion*, *Results* and *Related work* sections of the source papers. Descriptive literature reviews reference more information from the *Abstract* and Introduction sections. The *Abstract* section accounts for 31% of the referencing sentences, compared to 15% for integrative literature reviews.

The third section of Table 1 shows a significant relation (α=0.0017) between type of literature review and type of transformation to the source sentence. Descriptive literature reviews use more *cut-paste* than expected (26% compared to 12% for integrative literature

reviews). Overall *summary* accounts for the majority (53%) of the referencing sentences, with *cut-paste* 19% and *paraphrase* 17%.

Relation between Type of Information and Source

It should be noted that in all instances, the *critique* information type involved a *critical reference* transformation and originated from an *Unknown* source; no other significant associations could be ascertained. Therefore, it was filtered out in our cross-tabulations.

Table 2 shows the cross-tabulation between the source of information and the type of information. The relation was significant at the 0.001 level. Overall, 26% of the *research objective*, *research method* and *research result* information is taken from the source *Abstract*. The *research objective* information is sometimes taken from the *Abstract* (27%) and *Introduction* (8%) sections. But a large majority of the referencing sentences (49%) summarize the research objective at a high level, and no particular source sentence could be identified.

The *research method* information is taken from the *Method* section of the source paper (22%), in addition to the *Abstract* (26%). Some 30% are summarized at a high level with no identified source sentence. Information on the *research result* is taken from the *Conclusion* (19%), the *Results* section (19%), and the *Abstract* (26%).

Relation between Source of Information and Type of Transformation

Table 3 shows the cross-tabulation between the source of information and the type of transformation. As indicated earlier, overall, *summary* accounts for the majority (53%) of the transformations to the source sentences. There is a significant relation between *Type of transformation* and *Source of information.* Not surprisingly, source sentences from the *Abstract* are *cut-paste* more often than expected (43%). The proportion of paraphrasing is also higher than expected. Source sentences from the *Conclusion* section are also *paraphrased* more often than expected. Source sentences from the *Methodology* section are *cut-paste* and *paraphrased* more often than expected. Source sentences from the *Results* section are usually paraphrased.

Examining the interaction with Type of literature review, we found that in integrative literature reviews, source sentences from the *Abstract* tend to be *paraphrased* whereas in descriptive literature reviews, they tend to be *cut-pasted* (see Table 4). The *research method* information is taken from the *Method* section of the source paper (22%), in addition to the *Abstract* (26%). Some 30% are summarized at a high level with no identified source sentence. Information on the *research result* is taken from the *Conclusion* (19%), the *Results* section (19%), and the *Abstract* (26%).

Table 1: Profile of Integrative & Descriptive Literature Reviews

Type of Information		Type of Literature Review	
		Integrative	Descriptive
Research Objective	Count	55	55
	Expected Count	53.2	56.8
	% within lit-review	26.3%	24.7%
Research Method	Count	64	**92**
	Expected Count	75.5	**80.5**
	% within lit-review	30.6%	**41.3%**
Research Result	Count	**64**	56
	Expected Count	**58.1**	61.9
	% within lit-review	**30.6%**	25.1%
Critique	Count	**26**	20
	Expected Count	**22.3**	23.7
	% within lit-review	**12.4%**	9.0%
Pearson Chi-Square Asymp. Sig: 0.117			
Type of Source Section			
Abstract Section	Count	32	**70**
	Expected Count	49.3	**52.7**
	% within lit-review	15.3%	31.4%
Introduction Section	Count	6	**17**
	Expected Count	11.1	**11.9**
	% within lit-review	2.9%	7.6%
Conclusion Section	Count	**26**	11
	Expected Count	**17.9**	19.1
	% within lit-review	12.4%	4.9%
Results Section	Count	**22**	12
	Expected Count	**16.4**	17.6
	% within lit-review	10.5%	5.4%
Method Section	Count	25	23
	Expected Count	23.2	24.8
	% within lit-review	12.0%	10.3%
Related Work Section	Count	**11**	4
	Expected Count	**7.3**	7.7
	% within lit-review	5.3%	1.8%
Other	Count	9	9
	Expected Count	8.7	9.3
	% within lit-review	4.3%	4.0%
Unknown	Count	78	77
	Expected Count	75.0	80.0
	% within lit-review	37.3%	34.5%
Pearson Chi-Square Asymp. Sig: 5.29968941765105E-5			

Table 1: Profile of Integrative & Descriptive Literature Reviews (cont.)

Type of Transformation		Type of Literature Review	
		Integrative	Descriptive
Cut-Paste	Count	24	**58**
	Expected Count	39.7	**42.3**
	% within lit-review	11.5%	26.0%
Para-phrase	Count	**39**	35
	Expected Count	**35.8**	38.2
	% within lit-review	**18.7%**	15.7%
Summary	Count	**120**	110
	Expected Count	**111.3**	118.7
	% within lit-review	57.4%	49.3%
Critical Reference	Count	**26**	20
	Expected Count	**22.3**	23.7
	% within lit-review	12.4%	9.0%
Pearson Chi-Square Asymp. Sig: 0.0017386153510299597			

Table 2: Sources & Types of Information

Type of Source Section		Objective	Method	Result
Abstract Section	Count	30	41	31
	Expected Count	29.1	41.2	31.7
	% in lit-review	27.3%	26.3%	25.8%
Introduction Section	Count	**9**	**12**	2
	Expected Count	**6.6**	**9.3**	7.2
	% in lit-review	**8.2%**	**7.7%**	1.7%
Conclusion Section	Count	5	9	**23**
	Expected Count	10.5	15.0	**11.5**
	% in lit-review	4.5%	5.8%	**19.2%**
Results Section	Count	3	8	**23**
	Expected Count	9.7	13.7	**10.6**
	% in lit-review	2.7%	5.1%	**19.2%**
Method Section	Count	7	**34**	7
	Expected Count	13.7	**19.4**	14.9
	% in lit-review	6.4%	**21.8%**	5.8%
Related Work Section	Count	4	4	7
	Expected Count	4.3	6.1	4.7
	% in lit-review	3.6%	2.6%	5.8%
Other	Count	**9**	2	7
	Expected Count	**5.1**	7.3	5.6
	% in lit-review	**8.2%**	1.3%	5.8%
Unknown	Count	43	46	20
	Expected Count	31.1	44.1	33.9
	% in lit-review	39.1%	29.5%	16.7%
Pearson Chi-Square Asymp. Sig: 1.3813510414805267E-11				

Table 3: Cross-tabulation for Type of Source & Type of Transformation

Type of Source Section		Type of Transformation		
		Cut-paste	**Para-phrase**	**Summary**
Abstract Section	Count	**44**	**24**	34
	Expected Count	**21.7**	**19.6**	60.8
	% in source	**43.1%**	**23.5%**	33.3%
Introduction Section	Count	4	7	12
	Expected Count	4.9	4.4	13.7
	% in source	17.4%	30.4%	52.2%
Conclusion Section	Count	9	**11**	17
	Expected Count	7.9	**7.1**	22.0
	% in source	24.3%	**29.7%**	45.9%
Results Section	Count	4	**11**	19
	Expected Count	7.2	**6.5**	20.3
	% in source	11.8%	**32.4%**	55.9%
Method Section	Count	12	**12**	24
	Expected Count	10.2	**9.2**	28.6
	% in source	25.0%	**25.0%**	50.0%
Related Work Section	Count	6	6	3
	Expected Count	3.2	2.9	8.9
	% in source	40.0%	40.0%	20.0%
Other	Count	3	3	**12**
	Expected Count	3.8	3.5	**10.7**
	% in source	16.7%	16.7%	**66.7%**
Unknown	Count	0	0	**109**
	Expected Count	23.2	20.9	**64.9**
	% in source	.0%	.0%	**100.0%**
Pearson Chi-Square Asymp. Sig: 5.976037016183889E-6				

Table 4: Transformation of Source Sentences from Abstract in Integrative versus Descriptive Reviews

Type of Transformation		Type of Literature Review	
		Integrative	**Descriptive**
Cut-paste	Count	5	**39**
	Expected Count	10.4	**33.6**
	% of column	20.8%	67.2%
Paraphrase	Count	**11**	13
	Expected Count	**5.6**	18.4
	% of column	28.3%	**37.1%**

Table 5: Types of Information & Types of Transformation

Type of Information		Type of Transformation		
		Cut-paste	**Paraphrase**	**Summary**
Research Objective	Count	23	20	67
	Expected Count	23.4	21.1	65.5
	% within type	20.9%	18.2%	60.9%
Research Method	Count	30	26	**100**
	Expected Count	33.1	29.9	**93.0**
	% within type	19.2%	16.7%	**64.1%**
Research Result	Count	**29**	**28**	63
	Expected Count	**25.5**	**23.0**	71.5
	% within type	**24.2%**	**23.3%**	52.5%
Pearson Chi-Square Asymp. Sig: 0.404				

Relation between Source of Information and Type of Transformation

Table 3 shows the cross-tabulation between the source of information and the type of transformation. As indicated earlier, overall, *summary* accounts for the majority (53%) of the transformations to the source sentences. There is a significant relation between *Type of transformation* and *Source of information.* Not surprisingly, source sentences from the *Abstract* are *cut-paste* more often than expected (43%). The proportion of paraphrasing is also higher than expected. Source sentences from the *Conclusion* section are also *paraphrased* more often than expected. Source sentences from the *Methodology* section are *cut-paste* and *paraphrased* more often than expected. Source sentences from the *Results* section are usually paraphrased. We also found that in integrative literature reviews, source sentences from the *Abstract* tend to be *paraphrased* whereas in descriptive literature reviews, they tend to be *cut-pasted* (see Table 4).

Relation between Type of Transformation and Type of Information

Table 5 shows the cross-tabulation between the type of transformation and the type of information referenced. When the *research objective* is referenced, the majority (60%) of transformations are of the *summary* type. Similarly, for the referencing the *research method*, almost two-third of the transformations (64%) applied are of the *summary* type and occur more often than the expected count. The *research result* shows more counts than expected of *cut-paste* (24%) and *paraphrase* (23%) transformation and also shows a high frequency of *summary* (52%). However, the Pearson Chi-Square indicates no significant relation between the transformation type and the information type.

DISCUSSION AND CONCLUSION

We analyzed 20 literature reviews taken from journal articles published in JASIST for the types of information that are extracted from the cited (source) papers, the sections in the source papers that are referenced (source sections), and the types of transformation performed on the source sentences. The results can be summarized as follows:

- Descriptive literature reviews referenced a higher proportion of methodology information than expected, whereas integrative literature reviews referenced a higher proportion of research results and had more critiques. However, the Pearson Chi-square test found no significant relation. The relationship may become significant with a larger sample size.
- The highest proportion of referencing sentences carried information about the research method of a cited study. We had expected the research findings to be the most often referenced. It may be because a research method of a study requires more sentences to describe its details.
- Of the sources of information, the Abstract section is referenced more often than others. In integrative literature reviews, source sentences from the Abstract tend to be paraphrased whereas in descriptive literature reviews, they tend to be cut-pasted.
- Integrative literature reviews reference more information from the *Conclusion*, *Results* and *Related work* sections, whereas descriptive literature reviews reference more information from the Abstract and Introduction sections.
- There is more cut-paste in descriptive literature reviews than integrative literature reviews.
- A large proportion of the research objective and research method information are summarized at a high level, with no specific source sentences.
- Source sentences from the Results section tend to be paraphrased, rather than cut-pasted.

FUTURE WORK

Choice of Source

In future work, we will be exploring the reasons why an author chooses or prefers one source location over another. We conjectured at the likely reason for choosing the particular source

sentence from all likely candidates holding the same type of information. We deduced that the author's reasons could be two-fold:

- ***Argument-related*** – The source sentence may be the best supporter of the author's justification or argument.
- ***Content-related*** – The source sentence may be the best candidate out of all the likely candidates.

Type of Editing

In the context of our auto-summarization system, we will be making a note of the type of edits being made in each kind of transformation. Our aim will be to emulate a similar effect in our automatically generated literature reviews. The information we are collecting is of three types:

- ***Type of substitution*** – How information in the source had been substituted in the reference, and what were the kinds of substitutions made. The substitutions made were generally at the word level, where like-meaning verbs would be substituted, and at the noun level, where pronouns would be substituted with the authors' names.
- ***Type of insertion*** – How the source had been appended with additional information in the reference, and what were the kinds of insertions made. Insertions typically involved elaboration through integrative clauses related to the context of the source.
- ***Type of removal*** – How the source had been shortened to remove any information, and what were the kinds of removals made. Removals typically involved elaboration, dependent clauses and value judgments.

REFERENCES

Bowen, D. E. (1986). Managing customers as human resources in service organizations. *Human Resource Management*, 25(3), 370-383.

Azzam, S., Humphreys, K., & Gaizauskas, R. (1999). Using coreference chains for text summarization. In A. Bagga et al. (Eds.), *Proceedings of the ACL 1999 Workshop on Coreference and its Applications* (pp. 77-84). Maryland: ACM.

Barzilay, R., & Elhadad, M. (1997). Using lexical chains for text summarization. In U. Hahn et al. (Eds.), *Proceedings of the ACL 1997 Workshop on Intelligent Scalable Text Summarization* (pp.10-17). ACM.

Bourner, T. (1996). The research process: four steps to success. In T. Greenfield (Ed.), *Research Methods: Guidance for Postgraduates*. London: Arnold.

Chubin, D., & Moitra, S. (1975). Content Analysis of References: Adjunct or Alternative to Citation Counting?. *Social Studies of Science*, 5(4), 423-441.

Cremmins, E. (1992). *The art of abstracting*. Arlington, VA: Information Resources Press.

Cronin, B., & Shaw, D. (2002). Identity-creators and image-makers: Using citation analysis and thick description to put authors in their place. *Scientometrics,* 54(1), 31-49.

Edmundson, H. (1969). New methods in automatic extracting. *Journal of the ACM*,16(2), 264-285.

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.

Hart, C. (1998). *Doing a literature review*. London: Sage.

Hinchliffe, L.J. (2003). Having your say in a scholarly way. *Research Strategies,* 19, 163– 164.

Jaidka, K., Khoo, C., Na, J.-C. (2010). Imitating Human Literature Review Writing: An Approach to Multi-Document Summarization. In *Proceedings of the ICADL (International Conference on Asian Digital Libraries)* (pp. 116-119). Australia: Springer-Verlag.

Jing, H., & McKeown, K. R. (1999). The decomposition of human-written summary sentences. In F. Gey (Ed.), *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 129-136). Berkeley, California, United States: ACM.

Khoo, C., Na, J.-C., Jaidka, K. In press. Analysis of the Macro-Level Discourse Structure of Literature Reviews. *Online Information Review*.

Kupiec, J., Pedersen, J., & Chen, F. (1995). A trainable document summarizer. In E. Fox, P. Ingwersen and R. Fidel (Eds.), *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 68-73). New York: ACM.

Nanba, H., & Kando, N. (2000). Classification of Research Papers Using Citation Links and Citation Types. *Transactions of Information Processing Society of Japan,* 42(11), 2640-2649.

Paice, C. (1990). Constructing literature abstracts by computer: techniques and prospects. *Information Processing and Management* 26(1), 171-186.

Teufel, S. (1999). *Argumentative zoning: Information extraction from scientific text*. (Ph.D. Thesis). Edinburgh: School of Cognitive Science, University of Edinburgh.