



Literature review writing: how information is selected and transformed

Literature review
writing

303

Kokil Jaidka, Christopher S.G. Khoo and Jin-Cheon Na
*Wee Kim Wee School of Communication and Information, Nanyang
Technological University, Singapore, Singapore*

Abstract

Purpose – This paper aims to report a study of researchers' preferences in selecting information from cited papers to include in a literature review, and the kinds of transformations and editing applied to the selected information.

Design/methodology/approach – This is a part of a larger project to develop an automatic summarization method that emulates human literature review writing behaviour. Research questions were: how are literature reviews written – where do authors select information from, what types of information do they select and how do they transform it? What is the relationship between styles of literature review (integrative and descriptive) and each of these variables (source sections, types of information and types of transformation)? The authors analysed the literature review sections of 20 articles from the *Journal of the American Society for Information Science and Technology*, 2001-2008, to answer these questions. Referencing sentences were mapped to 279 source papers to determine the source sentences. The type of information selected, the sections of source papers where the information was taken from, and the types of editing changes made to include in the literature review were analyzed.

Findings – Integrative literature reviews contain more research result information and critique, and reference more information from the results and conclusion sections of the source papers. Descriptive literature reviews contain more research method information, and reference more information from the abstract and introduction sections. The most common kind of transformation is the high-level summary, though descriptive literature reviews have more cut-pasting, especially for information taken from the abstract. The types of editing – substitutions, insertions and deletions – applied to the source sentences are identified.

Practical implications – The results are useful in the teaching of literature review writing, and indicate ways for automatic summarization systems to emulate human literature review writing.

Originality/value – Though there have been several studies of abstracts and abstracting, there are few studies of literature reviews and literature review writing. Little is known about how writers select information from source papers, integrate it and present it in a literature review. This paper fills some of the gaps.

Keywords Literature review writing, Multi-document summarization, Information science, Information extraction, Information selection, Information, Information searches

Paper type Research paper

Introduction

A literature review is a summary of a set of related research papers. It selects information from the papers, and organizes and integrates it into a logical justification

The preliminary results of this study were earlier presented at the Asia-Pacific Conference on Library and Information Education and Practice 2011, Putra Jaya, Malaysia (Jaidka *et al.*, 2011). This paper provides more details on authors' information selection, as well as report new findings on the types of text transformations performed on the selected sentences.



for the author's research. Literature reviews are typically written by researchers who survey previous studies in order to identify research gaps 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:

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

Literature reviews are regularly found in books, journals and all manner of research writing, to cite, summarize and compare contemporary and significant research in a subject. However, little is known about how their authors select information, integrate it and present it. Several studies have documented how professional abstractors as well as authors write abstracts, what are the characteristics of such abstracts and how such abstracts should be constructed and evaluated (Cremmins, 1992; Endres-Niggemeyer *et al.*, 1995). Cremmins (1992) 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 help to explain how summaries of individual papers are written, but not how a number of papers are summarized into an integrated summary, or how they are compared and contrasted against each other. Our study aims to fill this gap; however, we used a content analysis approach to identify the choices authors made in constructing their literature review.

In this study we apply the results from our previous study (Khoo *et al.*, 2011), in which we conducted a linguistic analysis of literature review sections of information science journal articles, focusing on their discourse and rhetorical structure. We had identified two distinct styles of literature review writing, namely the integrative and descriptive literature reviews. Each type of literature review has a distinct profile of discourse elements and sentence-level rhetorical arguments. Descriptive literature reviews summarize individual papers/studies and 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, which is presented as a critical summary in the author's voice. We have extended the study to carry out a content analysis of information taken from the cited papers and map them to the source sentences in the cited papers. The aim is to compare the information selection and transformation strategies used in writing integrative versus descriptive literature reviews.

Research questions

Our study sought to answer the following questions:

- (1) How are literature reviews written?:
 - Where do authors select information from the cited papers – from the Abstract, Introduction, Method, Results or Conclusion sections?

- What types of information do they select – the research objectives, research methods or results information?
 - How do they transform that information – through paraphrasing, cut-pasting or higher-level summarizing?
 - What types of edits do they perform – insertions, removals, substitution of phrases, etc.?
 - Why do authors prefer one source of information over another (e.g. a sentence from the Abstract over a sentence from the Methodology section)?
- (2) What is the relationship between styles of literature review (integrative and descriptive) and each of these variables (source sections, types of information and types of transformation)?

We analyzed the literature review section of 20 articles from the *Journal of the American Society for Information Science and Technology*, 2001-2008, to answer these questions. We apply our findings to profile and compare different styles of literature review sections.

Literature review

Most studies of literature reviews are conducted by linguists who analyzed the texts to identify their rhetorical features and discourse structure (e.g. Swales, 1990; Kwan, 2006). Also, to our knowledge, no study has addressed the questions of what kind of information is selected from the source papers or from where. There are also few studies of how the source text is edited and transformed when it is incorporated into a summary or literature review. 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. Similarly, we analyzed the kinds of transformation performed on the source information when they are included in a literature review. In addition, we identified the types of information selected and their locations in the source document. Our hypothesis is that the transformations made to the source information depend on the type of information, its location in the source paper, and the style of writing – whether integrative or descriptive style of literature review writing.

Our analysis of referencing- and source-sentences bears some resemblance to citation analysis studies which explore the relationship between the citing and the cited paper (e.g. Teufel, 1999; Nanba and Kando, 2000; Cronin and Shaw, 2002). Chubin and Moitra (1975) explored why researchers cite other papers in the domain of Physics research. They found that citations fall in one of three broad categories: affirmative, wherein the writer supports or agrees with the cited study, negational, wherein the writer disproves or disagrees with previous work, and critical, which are used to criticize and raise questions about previous methods and results. Teufel (1999, p. 209) used a set of rhetorical relationships to categorize 17 types of relationships between the citing and the cited paper, for example, “previous context”, “negative adjective”, “positive adjective” and so on. These studies map citation to the source papers in order to identify the attitudes of the citer towards the cited, which is evidenced by the use of

cue phrases and other textual markers. These studies did not analyze the source sentences in the cited papers that the referenced information is taken from or their linguistic transformation, which are the foci of our study.

In our earlier study (Khoo *et al.*, 2011), we had conducted a linguistic analysis of the literature review sections of information science journal articles. We analyzed the texts at two levels of detail (Jaidka *et al.*, 2010):

- (1) Macro-level document structure – to identify the different sections of the literature, the types of information they contain and how they are organized hierarchically.
- (2) Rhetorical functions at the sentence level – to identify how each sentence is related to the overall purpose of the literature review.

We found that the literature reviews were structured as a set of topic sections, with each topic having a set of embedded study sections describing individual studies. Additionally, topic sections contain meta-summary sentences providing an overview of research in the field, and meta-critique sentences providing the reviewer’s critique or critical comparisons of previous research and the current study.

A list of rhetorical functions was derived from the sentence-level analysis (Jaidka *et al.*, n.d.). Rhetorical functions are used by authors to communicate intent by framing information in different ways. Certain rhetorical functions were found to be associated with certain macro-level discourse elements. For example, the dominant rhetorical function in meta-summary elements was introduce a topic.

Integrative and descriptive literature reviews were found to have different discourse and rhetorical profiles. Integrative reviews contained a much higher proportion of meta-summary statements than did descriptive reviews. They were also characterized by rhetorical functions such as indicate a relationship between concepts, find similarities between studies, and refer to historical development. On the other hand, descriptive literature reviews provide more information about individual studies and their methods and results. Accordingly, they were characterized by a higher presence of rhetorical functions such as to specify a research method, to describe an evaluation and to state the research results.

Research method

Twenty research articles were sampled from eight volumes of *The Journal of the American Society for Information Science and Technology (JASIST)* (2001-2008). *JASIST* is a leading journal in the field and carries high-quality research articles with substantive literature reviews. Two or three articles were haphazardly taken from each year (volume) of the electronic version of *JASIST*. The distribution is provided in Table I.

The literature review sections were extracted from the twenty journal articles. The papers were categorized into integrative or descriptive literature reviews by three

Table I.
Number of JASIST
articles selected from
each year/volume

Year	2001	2002	2003	2004	2005	2006	2007	2008	Total
Total	2	1	3	2	2	4	3	3	20

coders. It was found that the set comprised 11 descriptive and nine integrative literature reviews. The agreement was unanimous for 17 of the 20 literature reviews. The last three were decided by majority vote.

We analyzed the literature reviews line-by-line and retained all the sentences referencing previous work. These sentences were of three types:

- (1) Sentences which were explicitly referencing other work by including parenthesized citations, for example, "Lehtokangas and Airio (2002) conducted experiments in transitive translation".
- (2) Sentences which were implicitly referencing other work or adding onto the details of a cited study, for example "Researchers also are concerned with users' mental models of information seeking when using specific media (Marchionini, 1989)".
- (3) Sentences which provide broad generalizations without explicit cites. These included general descriptions of a topic or process which were not taken from any particular source paper. For example, "Citation analysis is a study of how citations are used by the citing authors".

The last category of sentences was filtered out because there was no identified source paper for the information. The remaining sentences were mapped to the source paper they cited, which was downloaded from bibliographic and full text databases and from the web. Our mapping was based on the premise that every referencing sentence is traceable to a single sentence or a group of sentences in the source paper. The source papers were retrieved from.

A total of 349 references were collected from the twenty literature review sections. Table II shows the number of source papers which were analyzed per literature review. Of the 349 original references, 292 were references to journal articles and conference papers, which we retained; 57 were references to other sources (e.g. books, technical reports, websites, professional articles, and literature survey articles). These sources were dropped because they do not conform to a research paper format and therefore could not be annotated with source section information. Of the remaining, 20 were not available online, usually when the paper dated back to the 1960s and 1970s, so these references were dropped from the analysis. Thus the final number of source papers analyzed was 272.

For every referencing sentence in a literature review, the referencing sentence in the source paper was located and the section (abstract, introduction, related work, methodology, conclusion, result, other or unknown) in which the source sentence occurred was noted. We refer to this as the source section.

We carried out four kinds of analysis on the referencing sentences and source sentences to answer our research questions:

- (1) Identifying the source sections from which the information is selected.
- (2) Identifying the types of transformations used to convert the source sentence to the referencing sentence. This was done by comparing the source sentence with the referencing sentence.
- (3) Identifying the types of information selected from the source papers.
- (4) Analysis of the reasons for preference of one source sentence over another, despite providing similar information. This was inferred by comparing source sentences against each other.

Literature review no.	No. of sources analyzed	Other sources ^a	Unavailable sources ^b	Total no. of sources
1	11	0	2	13
2	7	0	0	7
3	9	8	3	20
4	11	3	2	16
5	13	4	0	17
6	13	7	0	20
7	5	2	5	12
8	21	3	0	24
9	14	0	0	14
10	14	0	0	14
11	5	9	2	16
12	27	4	0	31
13	28	2	1	31
14	7	0	0	7
15	15	2	3	20
16	15	2	0	17
17	9	3	0	12
18	6	5	1	12
19	22	3	1	26
20	20	0	0	20
Total	272	57	20	349

Table II. Contents of the 20 sets of papers **Notes:** ^aOther sources: any source which was not a conference paper or journal article; ^bUnavailable sources: number of conference papers or journal articles which were not available online

The Appendix illustrates the above steps in our analysis: word-for-word similarities in the referencing and source sentence are underlined to show the extent of the resemblance. In the cases where there were other alternatives for the same information, we selected the closest match and considered the reasons why authors might have preferred one source sentence over another. These reasons are discussed in the next section.

Location of source sentences and source sections

Every referencing sentence was annotated with the name of the source sections from where the information was extracted, namely, the abstract, introduction, conclusion, results, method and related work (the literature review section of the source paper) sections. Other was used to represent non-typical source text such as headings, captions, titles, tables, etc. In some cases a source sentence could not be found, or the referencing information was a very short snippet that could not be pinpointed to a source sentence. In these cases, the source location was annotated as unknown (as illustrated in example 11 in the Appendix). This occurs when the citing author provides a high level summary of the source paper’s objectives, methods or findings. It also occurs when the citing author critiques or comments on the source paper.

Sometimes, the same sentence referenced information from more than one source section. In these cases, the sentence was “divided” into two aspects by annotating different segments with its sources. An illustration is provided as example 9 and 10 in the Appendix.

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

- (1) *Cut-paste*. Where source information is slightly modified into the cited information, through changes in tenses, parts of speech or sentence order. Some parts of the source information may have been dropped in the referencing sentence, such as introductory clauses, rhetorical devices, auxiliary clauses or adverbs. For example:

Referencing sentence: “Resnik, Oard and Levow proposed techniques for combining evidence from dictionary-based and corpus-based translation lexicons (Resnik *et al.*, 2001).”

Source sentence: “We present two techniques for combining evidence from dictionary-based and corpus-based translation lexicons.”

- (2) *Paraphrase*. Where source information is transformed through rewording, paraphrasing and replacement with synonyms into the cited information. The content remains the same as in the source sentence. For 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.”

- (3) *Summary*. Where the source information is semantically transformed 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. For 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.”

- (4) *Critical reference*. When the source information was transformed into a critical argument. The referencing sentence was a critique by the author, and could not be traced to a location in the source paper. For 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.”

In cases where duplicate information was present, the sentence requiring the least number of transformations into the cited information was selected. For example, simple edits such as deletion of a word (cut-pasting transformations) were preferred over rewording or substitution of words (paraphrasing transformations).

Identifying the types of information selected

The purpose was to identify the type of research information selected for inclusion in a literature review. We annotated every referencing sentence with one or more of the following information types:

- (1) *Research objective.* Referencing the purpose of the cited study (e.g. “Lehtokangas and Airio conducted experiments in transitive translation on several European languages (Lehtokangas and Airio, 2002)”).
- (2) *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”).
- (3) *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”).
- (4) *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”).

Each referencing sentence was mapped to its source sentences. Sometimes, a referencing sentence provided two kinds of information together. The research objective may be mentioned in combination with the research result or the research method. For example, the following sentence contains both research method and research objective information:

Using citations to the work of 25 library and information science faculty members as a case study, this paper examines the effects of using Scopus and Google Scholar (GS) on the citation counts and rankings of scholars as measured by WoS.

In these cases, the referencing sentence was “divided” into two text segments and analyzed separately. Example 12 and 13 in the Appendix illustrate how this was done.

Results

Profile of integrative and descriptive literature reviews

Tables III to VI present cross-tabulation tables which profile integrative and descriptive literature reviews. The numbers in italics indicate higher cell frequencies than expected. Pearson chi-square test of independence was performed to determine whether there were significant relationships between type of literature review (integrative versus descriptive) and the information type, source section and transformation type.

Table III profiles the integrative and descriptive literature reviews in terms of the types of information cited. The Pearson chi-square test did not find a significant relation between the style of literature review and type of information cited. However, a smaller cross-tabulation was conducted, focusing on the research method, and the research result and critique (combined). This time the relation was found significant at the 0.05 level. Descriptive literature reviews were found to have more research method information than expected; integrative literature reviews report more research results and provide more critique.

Table IV gives the cross-tabulation between style of literature review and type of source section selected. The relation was found to be significant at the 0.01 level.

Type of information	Type of literature review		Literature review writing
	Integrative	Descriptive	
<i>Research objective</i>			311
Count	62	62	
Expected count	60.7	63.3	
Percentage of column	27.1	25.9	
<i>Research method</i>			
Count	71	100	
Expected count	83.7	87.3	
Percentage of column	31.0	41.8	
<i>Research result</i>			
Count	67	54	
Expected count	59.2	61.8	
Percentage of column	29.3	22.6	
<i>Critique</i>			Table III. Profile of integrative and descriptive literature reviews: type of information
Count	29	23	
Expected count	25.4	26.6	
Percentage of column	12.7	9.6	

Note: Pearson chi-square p -value: 0.07

Descriptive literature reviews reference the Abstract section of the source papers more than expected (count: 64, expected count: 52.1). This accounted for 27 percent of the information – much higher than for the other sections of the source papers. Descriptive literature reviews also reference the introduction section more often than expected (count: 20, expected count: 14.8). This is expected, because the abstract and the introduction present the key details of a study in a concise manner, which makes them convenient sources of information. In contrast, integrative literature reviews reference the conclusion (count: 21, expected count: 16.6), results (count: 21, expected count: 16.1) and related work (count: 11, expected count: 7.3) sections much more than expected. Integrative literature reviews also reference other information sources, such as titles, headings and tables, more than expected (count: 20, expected count: 18.6). This may be because they report more research result information which is not all provided in the Abstract; instead, they cull information from tables and section headings.

Table V gives the cross-tabulation between style of literature review and type of transformation to the source sentence. The relation is significant at the 0.001 level. Overall, the most common kind of transformation is the high-level summary, accounting for over 50 percent of the source sentences. Comparing the two styles of literature review, we found that descriptive literature reviews have much more cut-paste than expected (observed count of 61 compared to expected count of 44.4). Perhaps the information selected from the abstract section (favored by descriptive literature reviews) are in a form that can be incorporated in the literature review with little editing. On the other hand, integrative literature reviews have more than expected number of the other three types of transformations, i.e. paraphrase, summary and critical reference.

Writers may be performing different transformations for different types of information. Table VI gives the cross-tabulation between type of transformation and type of information. It can be seen that, research objective and result information was

AP 65,3	Source section	Type of literature review	
		Integrative	Descriptive
312	<i>Abstract section</i>		
	Count	38	64
	Expected count	49.9	52.1
	Percentage of column	16.6	26.8
	<i>Introduction section</i>		
	Count	9	20
	Expected count	14.2	14.8
	Percentage of column	3.9	8.4
	<i>Conclusion section</i>		
	Count	21	13
	Expected count	16.6	17.4
	Percentage of column	9.2	5.4
	<i>Results section</i>		
	Count	21	12
	Expected count	16.1	16.9
	Percentage of column	9.2	5.0
	<i>Method section</i>		
	Count	28	27
	Expected count	26.9	28.1
	Percentage of column	12.2	11.3
	<i>Related work section</i>		
	Count	11	4
	Expected count	7.3	7.7
	Percentage of column	4.8	1.7
	<i>Headings and tables</i>		
	Count	20	18
	Expected count	18.6	19.4
	Percentage of column	8.7	7.5
	<i>Unknown</i>		
	Count	81	81
	Expected count	79.3	82.7
	Percentage of column	35.4	33.9
Note: Pearson chi-square <i>p</i> -value: 0.011			

Table IV.
Profile of integrative and
descriptive literature
reviews: type of source
section

more likely to be cut-pasted or paraphrased than expected, compared to research method information which is more likely to be summarized at a high level.

Table VII gives the same cross-tabulation for just the integrative reviews. It shows that research objectives are likely to be cut-pasted, research methods likely to be summarized at a high level, and research results likely to be paraphrased more often than expected.

Relation between type of transformation and source section

Writers may be transforming information differently, depending on which source section it belonged to. Table VIII shows the cross-tabulation between the source of information and the type of transformation. The relation was significant at the 0.05 level. For all the source sections except the abstract, about 50 percent or more of the information was found to have been summarized at a high level. In contrast,

Type of transformation	Type of literature review		Literature review writing
	Integrative	Descriptive	
<i>Cut-paste</i>			313
Count	26	61	
Expected count	42.6	44.4	
Percentage of column	11.4	25.5	
<i>Paraphrase</i>			
Count	42	40	
Expected count	40.1	41.9	
Percentage of column	18.3	16.7	
<i>Summary</i>			
Count	132	115	
Expected count	120.9	126.1	
Percentage of column	57.6	48.1	
<i>Critical reference</i>			
Count	29	23	
Expected count	25.4	26.6	
Percentage of column	12.7	9.6	
Note: Pearson chi-square <i>p</i> -value: 0.001			Table V. Profile of integrative and descriptive literature reviews: type of source section

Type of transformation	Type of information	
	Research objective and result	Research method
<i>Cut-paste and paraphrase</i>		
Count	109	60
Expected count	99.5	69.4
Percentage within row	64.4	35.6
<i>Summary</i>		
Count	136	111
Expected count	145.4	101.5
Percentage within row	55	45
Note: Pearson chi-square <i>p</i> -value: 0.05		

Table VI.
Cross-tabulation of type
of information and type
of transformation, for all
20 literature reviews

information taken from the abstract tended to be cut-pasted into the literature review, more often than expected (count: 42, expected count: 29.0). 41 percent of information from abstract sections was cut-pasted compared to 35 percent summarized. Taking a closer look at the abstract section, Table IX shows that descriptive literature reviews tended to apply cut-paste operations to source sentences from the abstract (count: 36, expected count: 30.9), whereas integrative literature reviews tended to paraphrase information (count: 19, expected count: 10.4).

Relation between type of information and source section

Table X shows the cross-tabulation between the source of information and the type of information. The relation was significant at the 0.001 level. Overall, the abstract section is the most popular source of information, accounting for 25 percent of the research objective, research method and research result information taken from the source

Table VII.
Cross-tabulation of type
of information and type
of transformation, for the
nine integrative literature
reviews only

Type of transformation	Research objective	Type of information Research method	Research result
<i>Cut-paste</i>			
Count	10	8	8
Expected count	8.1	9.2	8.7
Percentage of row	38.5	30.8	30.8
<i>Paraphrase</i>			
Count	11	9	22
Expected count	13.0	14.9	14.1
Percentage of row	26.2	21.4	52.4
<i>Summary</i>			
Count	41	54	37
Expected count	40.9	46.9	44.2
Percentage of row	31.1	40.9	28.0
Note: Pearson chi-square <i>p</i> -value: 0.03			

Table VIII.
Cross-tabulation between
source section and type of
transformation

Type of source section	Cut-paste	Type of transformation Paraphrase	Summary
<i>Headings and tables</i>			
Count	6	8	24
Expected count	10.7	9.4	17.9
Percentage of row	15.8	21.1	63.2
<i>Abstract section</i>			
Count	42	24	36
Expected count	29.0	26.0	47.0
Percentage of row	41.2	23.5	35.3
<i>Introduction section</i>			
Count	8	7	14
Expected count	8.2	7.2	13.7
Percentage of row	27.6	24.1	48.3
<i>Conclusion section</i>			
Count	9	11	14
Expected count	9.6	8.4	16.0
Percentage of row	26.5	32.4	41.2
<i>Results section</i>			
Count	5	9	19
Expected count	9.3	8.2	15.5
Percentage of row	15.2	27.3	57.6
<i>Method section</i>			
Count	12	13	30
Expected count	15.5	13.6	25.9
Percentage of row	21.8	23.6	54.5
Note: Pearson chi-square <i>p</i> -value: 0.04			

papers. We had anticipated that each type of information would be associated with particular source sections. This was supported by the data: research objective information tends to be taken from the Introduction section (count: 12, expected count: 8.6), research method information from the introduction (count: 15, expected count: 11.9) and method (count: 40, expected count: 22.6) sections, and research result information from the results (count: 22, expected count: 9.6) and conclusion (count: 22, expected count: 9.9) sections.

Unexpectedly, no particular source sentence (unknown) could be identified for a large proportion of the research objectives (35 percent) and research methods (28 percent) sentences. This information, with no identifiable source, was often summarized at a high-level. Sometimes, information from multiple source papers was synthesized to support the theme and context of the literature review (Hart, 1998), and to support the writer’s research.

Identifying the types of edits

We took our analysis of text transformations a step further by carrying out a finer analysis of the types of editing performed. Some of these edits are done to present information as reported speech, provide a citation or to follow a publication style (e.g. American Psychological Association, 2010) for a journal. Several studies of academic writing have elaborated on the features of academic writing, such as the “framing words” used to frame information (Pho, 2008), signal phrases used to alert the reader (Hyland, 2003), or the words describing the writer’s attitude and stance in academic writing (Schleppegrell, 2004). A few other studies have analyzed referencing sentences to identify the writer’s opinion of the cited work (Hyland, 2003). However, these studies have not been carried out from an editing but from a linguistic perspective.

The types of editing encountered are:

- (1) Types of substitutions:
 - Change of tense to past tense.
 - Substitution of personal pronouns with authors’ names.
 - Change of voice to passive voice in the case of research results; active voice in the case of research objectives and methods.

Type of transformation	Type of literature review	
	Integrative	Descriptive
<i>Cut-paste</i>		
Count	6	36
Expected count	11.0	30.9
Percentage of column	14.3	85.7
<i>Paraphrase</i>		
Count	19	11
Expected count	10.4	19.6
Percentage of column	63.3	36.7

Note: Pearson chi-square *p*-value: 1.6 × 10-5

Table IX.
Transformation of source sentences from the abstract in integrative versus descriptive reviews

AP
65,3

316

Type of source section	Research objective	Type of information Research method	Research result
<i>Abstract section</i>			
Count	31	40	31
Expected count	30.4	41.9	29.7
Percentage of column	25.0	23.4	26.1
<i>Introduction section</i>			
Count	12	15	2
Expected count	8.6	11.9	8.4
Percentage of column	9.7	8.8	1.7
<i>Conclusion section</i>			
Count	3	9	22
Expected count	10.1	14.0	9.9
Percentage of column	2.4	5.3	18.2
<i>Results section</i>			
Count	3	8	22
Expected count	9.8	13.6	9.6
Percentage of column	2.4	4.7	18.2
<i>Method section</i>			
Count	7	40	8
Expected count	16.4	22.6	16
Percentage of column	5.6	23.4	6.6
<i>Related work section</i>			
Count	4	4	7
Expected count	4.5	6.2	4.4
Percentage of column	3.2	2.3	5.8
<i>Headings and tables</i>			
Count	21	7	10
Expected count	11.3	15.6	11.1
Percentage of column	16.9	4.1	8.3
<i>Unknown</i>			
Count	43	48	19
Expected count	32.8	45.2	32.0
Percentage of column	34.7	28.1	15.7
Note: Pearson chi-square p -value: 8.1×10^{-15}			

Table X.
Cross-tabulation between
source section and type of
information

- Rearrangement of clauses.
 - Substitution of nouns with synonyms.
 - Substitution of dependent clauses with adjectives, e.g. substitute “articles in the field of webometrics” with “webometrics articles”.
 - Abstract nouns, or nouns referring to processes, are changed to verbs, e.g. substitute “comparison of” with “compared”.
 - Substitution of anaphoric references.
 - Substitution of singular nouns with plural nouns.
- (2) Types of insertions:
- Insertion of meta-frames:
 - for example, < authors >

- < authors > < verb >
 - More recently, < authors >
 - In these schemes,
 - Based on their surveys,
 - They found that,
 - Results were compared using
 - The study found that
 - Some authors have demonstrated
- Insertion of conjunctions between related sentences.
 - Insertion of elaborating clauses, e.g. “found in electronic dictionaries”.
- (3) Types of deletions:
- Deletion of modals, e.g. “tend to”, “might be”.
 - Deletion of introductory phrases, e.g. “In this paper”.
 - Deletion of affective words or qualitative judgments, e.g. “interesting finding...”.
 - Deletion of independent second clauses and dependent clauses, e.g. “The applicability of these measures is demonstrated and...”.

We grouped them into three categories:

- (1) *Substitutions*. Which pertains to how information in the source is substituted in the reference, and what are the kinds of substitutions made. Certain evident changes, such as to the tense and voice of source sentences, may have been done to conform to the academic writing genre. Rearrangement and paraphrasing may have been done either to highlight certain aspects of the information, or to make it more readable; trivial word-substitution may have been done to avoid plagiarism. Singular nouns may have been changed to plural to make a claim or description more general. For example:
- (2) “Comparison was facilitated using ____” is changed to “____ are compared using ____”.
- (3) “Strength” is substituted with “advantage”; “words” is replaced with “terms”.
- (4) *Insertions*. Which pertains to how the source has been appended with additional information in the reference, and what are the kinds of insertions made. Insertions typically comprise clauses which elaborate a method or the context of the research problem. Meta-frames of several types are often added to construct reported speech. Clauses are added to provide more details or context about nouns, such as sub-classes, numerical values, or full forms. For example:
- (5) “All the translation equivalents of a source query word are accepted in the target query” was transformed to “All the translation equivalents found in the electronic dictionaries of a source query word are accepted in the target query”.
- (6) *Deletions*. Which pertains to how the source had been shortened and what are the kinds of removals made. Deletions typically involve introductory phrases, dependent clauses, value judgments and modals. Independent second clauses are also often removed; in the source text, they usually provide elaborations to the main clause, such as purpose or definition. For example:

- (7) "Attempt to provide" is changed to "provided".
- (8) Independent second clauses such as "[. . .], a statistical method that measures the linear relationship between two sets of variables", are deleted.

Some of the types of editing seem obvious. We have listed all the types encountered in this study for the sake of completeness, in case they are useful to other researchers and teachers of academic writing.

Identifying reasons for source preference

In order to find out the likely reasons for selecting particular source sentences for summarization, we compared the best-match source with all likely candidates providing the same or similar information. We believe that the author may make his choice based on one whether one source sentence is the best at expressing information, or whether one source sentence is the best at supporting his justification or argument.

We expect that our content analysis will identify possible content-related reasons for source selection. By comparing candidate source sentences, we observed that the following types of source sentences are preferred:

- (1) Headings, titles and captions.
- (2) Candidates following an equation.
- (3) Bulleted candidates.
- (4) Short candidates, less clauses.
- (5) Candidates containing qualitative judgments.
- (6) Candidates emphasizing required information in its leading clause.
- (7) Candidate adjacent to another selected sentence.

Some candidates are more important because of their location, for example, the title or sub-headings in the source paper, or its adjacency to another selected sentence. On the other hand, some candidates provide extra information because of their format, for example, sentences in bulleted lists, or sentences describing an equation. In general, we noted that shorter candidates with fewer clauses were preferred in cases where a choice arose. This may be because they provide information in a concise manner. Specifically, sentences containing anaphora or co-references are avoided, as are those which comprise value judgments.

Summary of results

The results can be summarized as follows:

- (1) Integrative literature reviews have significantly more research result information and critique than descriptive literature reviews. Descriptive literature reviews have significantly more research method information than integrative literature reviews.
- (2) Integrative literature reviews reference more information from the results and conclusion sections than descriptive literature reviews. Descriptive literature reviews reference more information from the Abstract and Introduction sections than integrative literature reviews.

-
- (3) Integrative literature reviews have more critique, high-level summarizing and paraphrasing of source information than descriptive literature reviews. Descriptive literature reviews have more cut-pasting than integrative literature reviews.
 - (4) Either style of literature reviews frequently references the abstract section. Descriptive literature reviews are likely to cut-paste information from the abstract. Integrative literature reviews are likely to paraphrase information from the abstract.
 - (5) A large proportion of the research objectives (31 percent) and research methods (41 percent) sentences in either style of literature review are summarized at a high-level.

Discussion

This study has analyzed how authors select information and transform it to include in a literature review. In an earlier study, we had found 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.

Thus, from previous findings, we conjecture that an author begins a literature review with an overall strategy in mind, which includes the literature review style. Accordingly, he/she chooses the discourse structure and rhetorical arguments to implement the selected style. The contribution of the current study is to show that the author also selects and edits the information content based on the style of literature review. A writer may choose to write an integrative style of literature review to guide the reader along a critical survey of previous research. To support his argument, the author paraphrases information selected from the abstract and conclusion sections, and integrates information from the Results sections into a high-level overview of important findings.

On the other hand, a writer choosing to write a descriptive style of literature review may be more interested in providing synopses of previous research studies. This may be suitable where the readers are looking for a detailed background of research developments in an area, or where the details are important for the writer to challenge existing methods and suggest improvements. It is evident that in this case, the writer copy-pastes more information from the Abstract sections of the source papers, which are already available as summaries of the salient features of the cited paper.

Limitations

Though our study has yielded some significant results, it is based on a small sample of 20 articles from one journal. A larger scale study is suggested for future work, so that a comparison across journals can be made. One area of concern is the high proportion of cut-paste for source sentences from the Abstract of cited papers. There is a need to

investigate when, how and why authors would cut-paste, paraphrase or summarize source sentences. The results may be useful for studies in plagiarism.

Another area of concern is the larger percentage of sentences which are transformed as a high level summary or as a critique of previous work. This means that they require more synthesis, critique and higher levels of thinking by the writer, who situates the information in the context of his own research. Future work can investigate the semantic, reasoning and cognitive processes involved in producing high-level summaries of source information.

Conclusions and future research

Although we have attempted to delineate the distinguishing characteristics of descriptive versus integrative literature reviews, in reality, most literature reviews have both descriptive and integrative elements. For example, cited studies that are similar to the writer's work or provide a foundation for the writer's work may be described in detail. The same literature review may cite peripheral or distantly related studies as a high level summary. Likewise, the choice of style may depend on the type of research studies the author is reviewing. For experimental research, it may be more useful to construct a descriptive literature review which provides relevant details of previous studies. On the other hand, a theory-based paper may require more argumentation and interpretation in an integrative literature review. Therefore, students need to be taught both kinds of literature review writing, although integrative literature reviews appear to be considerably more difficult to construct.

We believe that our findings will find application in the teaching of literature review writing. Teachers can refer to our list of ways to select, edit and frame information selected from source papers. Future work can use our results to distil a set of patterns and best practices for the different aspects of literature review writing. Furthermore, the results of the study will also have implications for research in natural language processing: future work can build on our findings to develop automatic text summarization methods or automated tools to assist researchers.

References

- American Psychological Association (2010), *Publication Manual of the American Psychological Association*, American Psychological Association, Washington, DC.
- Chubin, D. and Moitra, S. (1975), "Content analysis of references: adjunct or alternative to citation counting?", *Social Studies of Science*, Vol. 5 No. 4, pp. 423-441.
- Cremmins, E. (1992), *The Art of Abstracting*, Information Resources Press, Arlington, VA.
- Cronin, B. and Shaw, D. (2002), "Identity-creators and image-makers: Using citation analysis and thick description to put authors in their place", *Scientometrics*, Vol. 54 No. 1, pp. 31-49.
- Endres-Niggemeyer, B., Maier, E. and Sigel, A. (1995), "How to implement a naturalistic model of abstracting: Four core working steps of an expert abstractor", *Information Processing and Management*, Vol. 31 No. 5, pp. 631-674.
- Hart, C. (1998), *Doing a Literature Review*, Sage, London.
- Hyland, K. (2003), "Self-citation and self-reference: credibility and promotion in academic publication", *Journal of the American Society for Information Science and Technology*, Vol. 54 No. 3, pp. 251-259.

- Jaidka, K., Khoo, C. and Na, J.-C. (2010), "Imitating human literature review writing: an approach to multi-document summarization", *Proceedings of the ICADL (International Conference on Asian Digital Libraries) in Surfer's Paradise, Australia, 2010*, Springer-Verlag, New York, NY, pp. 116-119.
- Jaidka, K., Khoo, C. and Na, J.-C. (2011), "Literature review writing: a study of information selection from cited papers", *Proceedings of the Asia-Pacific Conference On Library and Information Education and Practice 2011 (A-LIEP2011): Issues, Challenges and Opportunities in Pullman Putrajaya Lakeside, Malaysia*, pp. 467-477.
- Jaidka, K., Khoo, C. and Na, J.-C. (n.d.), "Building an argument: rhetorical functions in literature reviews of information science papers".
- Jing, H. and McKeown, K. (1999), "The decomposition of human-written summary sentences", in Gey, F. (Ed.), *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, 1999*, ACM, New York, NY, pp. 129-136.
- Khoo, C., Na, J.-C. and Jaidka, K. (2011), "Analysis of the macro-level discourse structure of literature reviews", *Online Information Review*, Vol. 35 No. 2, pp. 255-271.
- Kwan, B. (2006), "The schematic structure of literature reviews in doctoral theses of applied linguistics", *English for Specific Purposes*, Vol. 25 No. 1, pp. 30-55.
- Nanba, H. and Kando, N. (2000), "Classification of research papers using citation links and citation types", *Transactions of Information Processing Society of Japan*, Vol. 42 No. 11, pp. 2640-2649.
- Pho, P. (2008), "Research article abstracts in applied linguistics and educational technology: a study of linguistic realizations of rhetorical structure and authorial stance", *Discourse Studies*, Vol. 10 No. 2, pp. 231-250.
- Swales, J. (1990), *Genre Analysis: English in Academic and Research Settings*, Cambridge University Press, London.
- Schleppegrell, M. (2004), *The Language of Schooling: A Functional Linguistics Perspective*, Lawrence Erlbaum Associates, Hillsdale, NJ.
- Teufel, S. (1999), "Argumentative zoning: Information extraction from scientific text", PhD thesis, University of Edinburgh, Edinburgh.

Further reading

- Ansari, A., Essegaier, S. and Kohli, R. (2000), "Internet recommendation systems", *Journal of Marketing Research*, Vol. 37 No. 3, pp. 363-375.
- Baeza-Yates, R., Calderón-Benavides, L. and González-Caro, C. (2006), "The intention behind web queries", *Proceedings of the String Processing and Information Retrieval*, Spire, Glasgow, pp. 98-109.
- Balabanovic, M. and Shoham, Y. (1997), "Fab: content-based, collaborative recommendation", *Communications of the ACM*, Vol. 40 No. 3, pp. 66-72.
- Bar-Ilan, J. (2008), "Which h-index? A comparison of WoS, Scopus and Google Scholar", *Scientometrics*, Vol. 74 No. 2, pp. 257-271.
- Barzilay, R. and Elhadad, M. (1997), "Using lexical chains for text summarization", in Mani, I. and Maybury, M. (Eds) pp. 10-17, in *proceedings of the ACL Workshop on Intelligent Scalable Text Summarization in Madrid, Spain, 1997*, ACM.
- Bourner, T. (1996), "The research process: four steps to success", in Greenfield, T. (Ed.), *Research Methods: Guidance for Postgraduates*, Arnold, London, pp. 7-11.

-
- Desai, M. and Spink, A. (2005), "An algorithm to cluster documents based on relevance", *Information Processing & Management*, Vol. 41 No. 5, pp. 1035-1049.
- Edmundson, H. (1969), "New methods in automatic extracting", *Journal of the ACM*, Vol. 16 No. 2, pp. 264-285.
- Hargittai, E. and Hinnant, A. (2006), "Toward a social framework for information seeking", *New Directions in Human Information Behavior*, Springer, Dordrecht, pp. 55-70.
- Hinchliffe, L.J. (2003), "Having your say in a scholarly way", *Research Strategies*, Vol. 19 No. 3, pp. 163-164.
- Khan, K. and Locatis, C. (1998), "Searching through cyberspace: the effects of link display and link density on information retrieval from hypertext on the world wide web", *Journal of the American Society for Information Science*, Vol. 49 No. 2, pp. 176-182.
- Kupiec, J., Pedersen, J. and Chen, F. (1995), "A trainable document summarizer", in Fox, E., Ingwersen, P. and Fidel, R. (Eds) pp. 68-73, *proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Seattle, 1995, ACM.
- Lehtokangas, R. and Airio, E. (2002), "Translation via a pivot language challenges direct translation in CLIR", *Proceedings of the SIGIR 2002 Workshop: Cross-language Information Retrieval: A Research Roadmap*, August.
- Marchionini, G. (1989), "Information-seeking strategies of novices using a full-text electronic encyclopedia", *Journal of the American Society for Information Science*, Vol. 40 No. 1, pp. 54-66.
- Ou, S., Khoo, C. and Goh, D.H. (2007), "Design and development of a concept-based multi-document summarization system for research abstracts", *Journal of Information Science*, Vol. 34 No. 3, pp. 308-326.
- Paice, C. (1990), "Constructing literature abstracts by computer: techniques and prospects", *Information Processing and Management*, Vol. 26 No. 1, pp. 171-186.
- Resnik, P., Oard, D. and Levow, G. (2001), "Improved cross-language retrieval using backoff translation", *Proceedings of the First International Conference on Human Language Technology Research*, Association for Computational Linguistics, Stroudsburg, PA, pp. 1-3, March.
- Saito, H. and Miwa, K. (2001), "A cognitive study of information seeking processes in the WWW: the effects of searcher's knowledge and experience", *Proceedings of the Second International Conference on Web Information Systems Engineering, 2001*, Vol. 1, IEEE, New York, NY, pp. 321-327, December.
- Teufel, S. and Moens, M. (2002), "Summarizing scientific articles – experiments with relevance and rhetorical status", *Computational Linguistics*, Vol. 28 No. 4, pp. 409-445.

S.No.	Summary sentence	Source text	Type of content	Location of source information	Type of transformation
1	The authors concluded that researchers should consult Google Scholar in addition to Scopus or Web of Science, especially for relatively recent publications, but until Google Scholar provides a complete accounting of the material that it indexes and how often that index is updated, Google Scholar cannot be considered a true scholarly resource in the sense that Scopus and Web of Science are	Based on our preliminary examination and discovery of higher citation counts, we recommend that researchers should consult Google Scholar in addition to Web of Science or Scopus, especially for a relatively recent article, author or subject area. However, it is important for all researchers to note that until Google Scholar gives a full account of what material it is indexing and how often that index is updated, it cannot be considered a true scholarly resource in the sense that Web of Science and Scopus	Result	Conclusion section	Cut-pasted
2	More recently, Bar-Ilan (2008b) < abstract > compared the h scores (see below for a discussion of the development of the h-index) of a list of 40 highly-cited Israeli researchers based on citation counts from Google Scholar, Scopus, and Web of Science < / abstract >	This paper compares the h-indices of a list of highly-cited Israeli researchers based on citations counts retrieved from the Web of Science, Scopus and Google Scholar respectively	Method	Abstract section	Cut-pasted
3	Many studies have focused on what Hargittai and Hinant (2006) refer to as "convenience samples of undergraduate or graduate students at research university" or, in some cases, school children or university academics	As we discussed in the previous section, much research in information seeking is conducted on convenience samples of undergraduate or graduate students at research universities	Objective	Methodology section	Cut-pasted
4	Griffiths <i>et al.</i> (1986) found that searches that retrieved a single bottom-level cluster often returned only two or three documents	However, this (...) correspond to the retrieval of only two or three documents if a small bottom level cluster is identified as the best match for some query	Result	Methodology section	Paraphrased
5	We develop an ensemble of statistical methods, which we estimated using customer ratings on small, idiosyncratic subsets of products. We adopt a regression-based approach and model customer ratings as a function of product attributes, customer characteristics, and expert evaluations. The models we develop differ in how they account for unobserved sources of heterogeneity in customer preferences and product appeal structures.	Recent work by Ansari <i>et al.</i> (2000) has adopted the formal statistical approach to model the user ratings as a function of item attributes, user characteristics, and expert evaluations. Unobserved sources of heterogeneity in user preference and item appeal structures were accounted for using this approach	Method	Methodology section	Paraphrased
6	Desai and Spink (2005) devised an algorithm to cluster search engine results into three sets: relevant, partially relevant, and non-relevant, taking into account several document relevance criteria	In this paper, we present a clustering scheme that groups documents within relevant, partially relevant, and not relevant regions for a given search. A clustering algorithm accomplishes the task of clustering documents based on relevance	Objective	Abstract section	Paraphrased

(continued)

Table AI.
The steps in our analysis

Table AI.

S.No.	Summary sentence	Source text	Type of content	Location of source information	Type of transformation
7	The correlations were very low, and in most cases (except for Yahoo!), the correlation was negative	We see that most of the publicly available search engines taken in this study are much below the users expectations. (...) all but one of them is getting a negative value of Spearman Rank Order Correlation Coefficient averaged over all the queries	Result	Results section	High-level summary
8	They compared the effectiveness of clustering using the two types of similarity scores by means of a statistical test on the MK1 values computed for these two clustering schemes	This paper presents a pilot study of query-specific clustering that uses our novel document-context based similarity scores as compared with document similarity scores. Clustering is applied to the top 1,000 retrieved documents for a given query. Clustering effectiveness is evaluated based on the MK1 score for TREC-2, TREC-6 and TREC-7 test collections	Method	Abstract section	High-level summary
9	Perhaps the best that can be done with search engine logs is to use manual query classification to infer user intention, which some workers have recently used as a means to train automatic discrimination methods (Baeza-Yates <i>et al.</i> , 2006)	The identification of the user's intention or interest through queries that they submit to a search engine can be very useful to offer them more adequate results	Objective	Abstract section	High-level summary
10	Perhaps the best that can be done with search engine logs is to use manual query classification to infer user intention, which some workers have recently used as a means to train automatic discrimination methods (Baeza-Yates <i>et al.</i> , 2006)	After the manual classification of the queries was made, part of these labeled data was used like input to train an automatic classifier	Method	Results section	High-level summary
11		Efficiency measures related to browsing were also used, such as the number of pages viewed (Saito and Miwa, 2002) and the number of links traversed (Khan and Locatis, 1998)	Method	Unknown	High-level summary
12	To create a hybrid content-based, collaborative system, we maintain user profiles based on content analysis, and directly compare these profiles to determine similar users for collaborative recommendation	For example, Fab maintained user profiles based on content analysis of documents rated by users	Method	Methodology section	Cut-pasted
13	To create a hybrid content-based, collaborative system, we maintain user profiles based on content analysis, and directly compare these profiles to determine similar users for collaborative recommendation	It then compared user profiles to identify similar users to generate a collaborative recommendation (Balabanovic and Shoham, 1997)	Objective	Methodology section	Cut-pasted

About the authors

Kokil Jaidka is a PhD student at the Wee Kim Wee School of Communication and Information at Nanyang Technological University, Singapore. She is an engineering graduate from PEC University of Technology, India. She has previously worked on feature extraction and clustering algorithms on a short internship at the Department of Computer Science, Texas A&M University in 2007. For her doctoral dissertation at NTU, Kokil is exploring multi-document summarization techniques applicable to a research corpus. Kokil Jaidka is the corresponding author and can be contacted at: kokil@pmail.ntu.edu.sg

Dr Christopher S.G. Khoo is an Associate Professor in the Wee Kim Wee School of Communication and Information at Nanyang Technological University, Singapore. He teaches courses in knowledge organization, information architecture, data mining and web-based information systems. He obtained his PhD at Syracuse University and his MSc in Library and Information Science at the University of Illinois, Urbana-Champaign. He has also worked for several years as a science reference librarian, cataloger and online information searcher at the National University of Singapore Libraries. His main research interests are in text mining (information extraction and text summarization), medical decision support system, knowledge organization, and human categorization behavior.

Dr Jin-Cheon Na is an Associate Professor in the Wee Kim Wee School of Communication and Information at Nanyang Technological University (NTU), Singapore. Before he joined NTU, he was a senior researcher at the Agency for Defense Development, Korea. He obtained his PhD in the Department of Computer Science at Texas A&M University. He has published more than 40 papers mainly in the areas of digital libraries, hypertext and hypermedia, document engineering, and knowledge organization. He is a member of the Editorial Review Board of the *International Journal of Organizational and Collective Intelligence (IJOI)*. He was also a non-executive director of DideoNet Co Ltd, Korea.