Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. . 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi:10.7152/acro.v11i1.12774

Classification of Research Papers using Citation Links and Citation Types: Towards Automatic Review Article Generation

Hidetsugu NANBA(†), Noriko KANDO(††), Manabu OKUMURA(‡, †)

†School of Information Science,
Japan Advanced Institute of Science and Technology
nanba@jaist.ac.jp
††National Institute of Informatics
kando@nii.ac.jp
‡Precision and Intelligence Laboratory,
Tokyo Institute of Technology
oku@pi.titech.ac.jp

Abstract

We are investigating automatic generation of a review (or survey) article in a specific subject domain. In a research paper, there are passages where the author describes the essence of a cited paper and the differences between the current paper and the cited paper (we call them **citing areas**). These passages can be considered as a kind of summary of the cited paper from the current author's viewpoint. We can know the state of the art in a specific subject domain from the collection of citing areas. Further, if these citing areas are properly classified and organized, they can act as a kind of a review article. In our previous research, we proposed the automatic extraction of citing areas. Then, with the information in the citing areas, we automatically identified the types of citation relationships that indicate the reasons for citation (we call them **citation types**). Citation types offer a useful clue for organizing citing areas. In addition, to support writing a review article, it is necessary to take account of the contents of the papers together with the citation links and citation types. In this paper, we propose several methods for classifying papers automatically. We found that our proposed methods BCCT-C, the bibliographic coupling considering only type C citations, which pointed out the problems or gaps in related works, are more effective than others. We also implemented a prototype system to support writing a review article, which is based on our proposed method.

1 Introduction

The ultimate goal of the research reported in this paper is the automatic or computer-assisted generation of review articles. We discuss a prototype system called PRESRI that relies on citation relationships. PRESRI identifies citing areas in papers, identifies the type of the citing relationships, and uses this information for citation-based topical clustering of papers, and therefore the citing areas they contain. The clustered citing areas form a very rough review. The paper discusses the techniques used in detail.

ISSN: 2324-9773

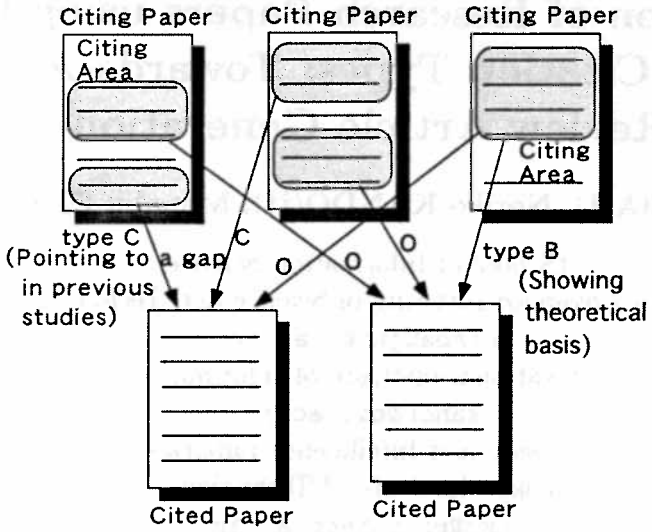

Figure 1: Citation relationships between papers

In our previous research, we considered citation relationships between papers for automatic generation of review articles [Nanba and Okumura, 1999]. In a research paper, there are passages where the author describes the essence of a cited paper and the differences between the current paper and the cited paper, as shown in Figure 1 (we call them **citing areas**). These passages can be considered as a kind of summary of the cited paper from the current author's viewpoint.

We can know the state of the art in a specific subject domain from citing areas collected from a set of papers in that subject domain. Further, if citing areas are properly classified and organized, they can act as a kind of a review article. In our previous research, we proposed the automatic extraction of citing areas. Then, with the information about the citing areas, we automatically identified the types of citation relationships that indicate the reasons for citation (we call them **citation types**). Citation types offer a useful clue for organizing citing areas. In addition, to support writing a review article, it is necessary to take account of the contents of the papers together with the citation links and citation types. In this paper, we propose several methods for classifying research papers in a database automatically, using citation links and citation types.

PRESRI provides two methods of document retrieval. One is retrieval by a query (using authors' names and/or terms in titles). The current version of PRESRI classifies papers based on cited papers they share in their bibliography (bibliographic coupling), taking into account the citation type.

Figure 2 shows PRESRI. The left window shows the citation relationships around 'CID:8001780' [Pereira and Sheabes, 1992]. Eight papers, which cite 'CID:8001780', are divided into three groups according to the reasons for citation. Further, they are classified by their topics. When an user of PRESRI clicks an icon 'CITING AREA' in the left window, the corresponding citing area is displayed in the right window. We think that the task of writing review articles consists of at least two sub-tasks: classifying multiple papers, and summarizing them. For these tasks, we pay attention to citation relationships between papers. From the results of several experiments, we found that

ISSN: 2324-9773

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. . 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi:10.7152/acro.v11i1.12774

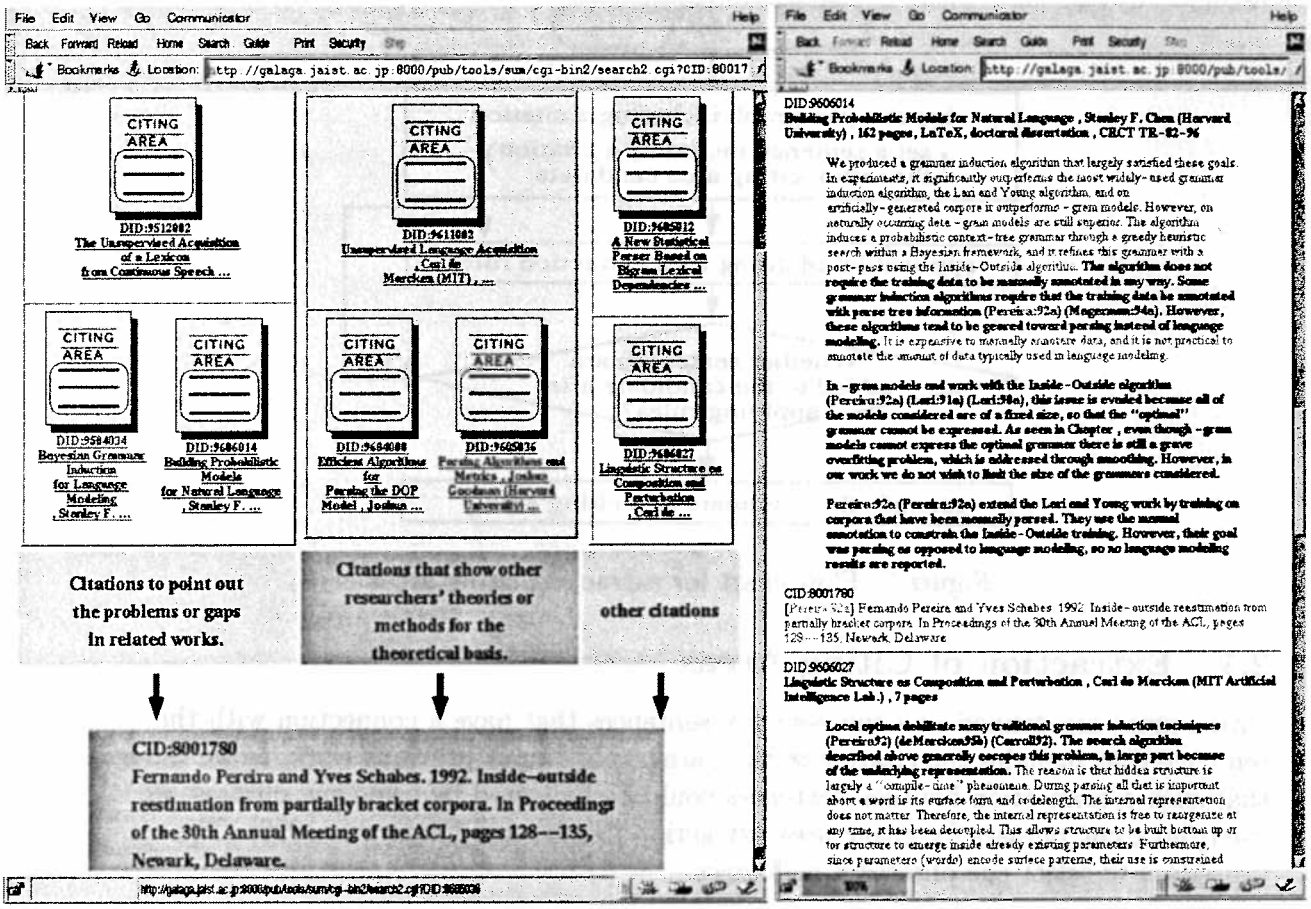

Figure 2: PRESRI: a system to classify research papers

knowing the reasons for citation is indispensable for classifying papers.
In the remainder of the paper we discuss the components of our system in detail. Section 2 deals with the extraction of citing areas and the determination of the citation type, Section 3 discusses methods for the classification of citing areas by subject, including experimental results and evaluation, and Section 4 shows briefly how these techniques are applied in PRESRI. Section 5 presents conclusions and an outlook on further research.

2 Citing Areas and Determination of Citation Types

In this section, we first explain our previous work, automatic extraction of citing areas, identification of citation types and classification of citing areas according to the citation types.

ISSN: 2324-9773

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi:10.7152/acro.v11i1.12774

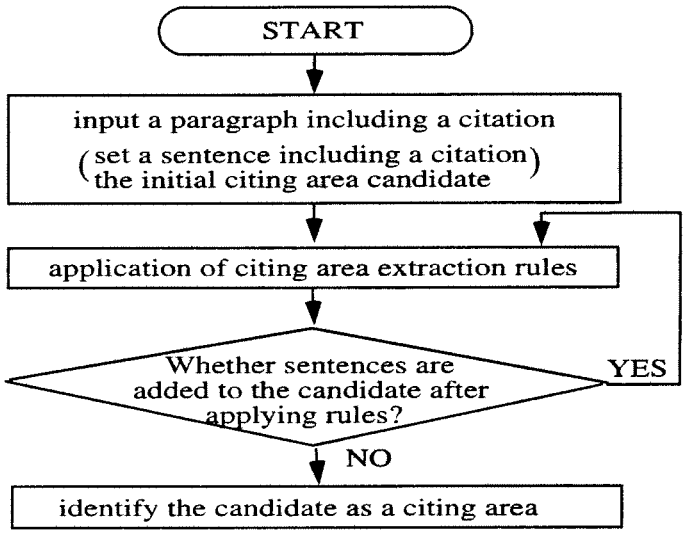

Figure 3: Flow chart for extracting citing areas

2.1 Extraction of Citing Areas

Citing areas are defined as a succession of sentences that have a connection with the sentence that includes the citation in the paragraph. In our previous work, as we thought that such a connection between sentences could be indicated by some cue phrases, we used those cue phrases for citing area extraction [Nanba and Okumura, 1999]. The procedure to select cue phrases is as follows:

1. Create the citing area corpus by hand,
2. Apply n-word gram analysis to this corpus,

As a result, 86 cue phrases were selected as shown in Table 4 in the appendix.
We then developed rules for extracting citing areas automatically using these 86 cue phrases. The flow chart for extracting citing areas is shown in Figure 3. Initially, a citing area candidate consists of one sentence that contains a citation. The citing area extraction rules are applied to the sentence just before and just after the citing area candidate determined so far (always staying within the selected paragraph). If either sentence includes any of the cue phrases from Table 4, it is added to the citing area candidate. This process continues until no more sentences are added.
We have conducted an experiment to evaluate the effectiveness of our method. For the experiment, we prepared 150 citing areas that were manually identified from paragraphs including citations. We used 100 for making rules and 50 for evaluation. We used the following equations for evaluation.

$$Recall = \frac{\left(\begin{array}{l} The\ number\ of\ sentences \\ correctly\ extracted\ by\ the\ rules \end{array} \right)}{\left(\begin{array}{l} The\ number\ of\ sentences \\ that\ should\ be\ extracted \end{array} \right)}$$

ISSN: 2324-9773

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. . 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi:10.7152/acro.v11i1.12774

$$Precision = \frac{\left(\begin{array}{l} The\ number\ of\ sentences \\ correctly\ extracted\ by\ the\ rules \end{array} \right)}{\left(\begin{array}{l} The\ number\ of\ sentences\ extracted \\ by\ the\ rules \end{array} \right)}$$

As a result, we obtained recall of 80% and precision of 76% in the data for evaluation [Nanba and Okumura, 1999].

2.2 Determination of Citation Types

For the classification of citing areas it is useful to distinguish citation types (reasons for citation). In our previous work, we proposed a method to identify citation types automatically [Nanba and Okumura, 1999]. In this section, we define citation types and discuss our method for determining citation types automatically. In Section 3 we discuss the application of citation types to the classification of citing areas.

2.2.1 Citation Types - reasons for citation -

Classification of reasons for citation

It is well known that there are several reasons for citation. To classify citing areas using citation relationships, we also have to take citation types into consideration. For example, Weinstock proposed the following 15 categories for the reasons of citations [Weinstock, 1971]:

(1) Paying homage to pioneers

(2) Giving credit for related work

(3) Identifying methodology, equipment, and so on.

(4) Providing background reading

(5) Correcting one's own work

(6) Correcting the work of others

(7) Criticizing previous work

(8) Substantiating claims

(9) Alerting to forthcoming work

(10) Providing leads to poorly disseminated, poorly indexed, or uncited work

(11) Authenticating data and classes of fact – physical constants, and so on.

(12) Identifying original publications in which an idea or concept was discussed

(13) Identifying original publications or other work describing an eponymous concept or term

ISSN: 2324-9773

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi: 10.7152/acro.v11i1.12774

(14) Disclaiming work or ideas of others (negative claims)

(15) Disputing priority claims of others (negative homage)

Instead of using Weinstock's categories, we classified the reasons for citation into the following three categories which are important for automatic generation of review articles. [Nanba and Okumura, 1999][1]:

- **type B** (corresponds to Weinstock's categories (3), (4) and (8))
  Citations that show other researchers' theories or methods for the theoretical basis.
- **type C** (corresponds to Weinstock's categories (14) and (15))
  Citations to point out the problems or gaps in related works.
- **type O**
  Citations other than types B and C.

To devise citation types, we also take account of the possibility of automatic identification by computer. In the following, we will explain how to identify citation types automatically.

2.2.2 Automatic identification of citation types

We manually created 160 rules for the automatic determination of citation types [Nanba and Okumura, 1999]. These rules are based on cue phrases, 84 cue phrases for type B (Table 5) and 76 for type C (Table 6) (Tables 5 and 6 are in the appendix). The rules for type C are applied first: Within a citing area, if a type C cue phrase appears in the sentence following the citation sentence, the citing area is assigned to type C and the process stops. If type C was not assigned, then the rules for type B are applied: If the citation sentence contains a type B cue phrase, the citing area is assigned to type B and the process stops. If neither type C nor type B are assigned, then the citing area is assigned to type O. The order of applying the rules was determined statistically using training data.
The results are shown in Table 1. The sum of bold numbers in Table 1 shows the number of citing areas with citation types correctly identified by the rules. Therefore, we obtain accuracies of 90(%) and 83(%) in the data for making rules and evaluation, respectively [Nanba and Okumura, 1999]. As the only information used for the identification of citation types is cue phrases in our system, when a citation seems somewhat like type C (or type B) but does not use the cue phrases, our system cannot identify the citation as type C (or type B). To improve the effectiveness of our rules, deep analysis using domain knowledge is inevitable.

[1]We devised three citation types, B, C and O. However, as anonymous reviewer pointed out, the following two types are also conceivable for automatic generation of review articles:
- **type S**: supporting other studies.
- **type P**: following or part of other studies.

The authors would like to express our gratitude to anonymous reviewers.

ISSN: 2324-9773

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. . 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi:10.7152/acro.v11i1.12774

Table 1: The accuracy of automatic identification of citation types using evaluation data

		citation type identified by rules			accuracy for each type(%)
		B	C	O	
citation correct type	B	**25**	2	5	78
	C	0	**12**	4	75
	O	5	1	**46**	89

the accuracy of citation type identification in evaluation data: 83(%)

3 Classification of Citing Areas by Subjects

3.1 Introduction

Topical Similarity

Each citing area in a paper is a component of the paper and closely related to the purpose and the method of the paper. Therefore, it is necessary to grasp the content of each citing area not only in its local context, but also in relation to the purpose and method of the entire paper. We therefore assumed that whole citing areas in a paper are topically the same. To classify citing areas from multiple documents by topical similarity, we therefore only need to classify the papers to which each citing area belongs.
Several techniques have been proposed for calculating topical similarity between papers. They all rely on identifying features two papers A and B have in common and measuring the degree of overlap. Approaches differ in the type of features chosen:

- **Word-based approach (words in common)**

  Measure the degree of overlap in the words that occur in the text (or some selection part of the text) of papers A and B. [Salton and McGill, 1983].

- **Citation-based approach (citations-to in common or citations-from in common)**

  It is well known that using citation analysis makes it possible to obtain topical collections of papers [Liu, 1993, Narin et al., 1994, White and McCain, 1989]. In these studies, two similar papers were found to cite many of the same papers (bibliographic coupling [Kessler, 1963]), or were cited from many other papers (co-citation analysis [Small, 1973]). We can measure the similarity by counting the number of couplings or co-citations.

  If the target papers for classification are old enough to be cited from many other papers, we can use co-citation analysis for these papers. In contrast, if the target papers are too new, we can apply bibliographic coupling. As our target papers, which we will describe later, are not old enough, we cannot adopt co-citation analysis.

ISSN: 2324-9773

Towards High Precision Classification

There are two essential points for automatic classification of citing areas. The first is that classification should be on-the-fly, to learn the outline of the subject domain in a short time, as our system is implemented as an interactive system. The second point is that classes should not include unrelated citing areas, even though citing areas in the class split into several classes. The users can grasp the outline of a research domain without reading whole citing areas in the domain. However, if many unrelated citing areas are included in a class, the users must read them, spending unnecessary time. We thus aim to implement high precision classification with reasonable computational cost.
Here, we point out the problems of these approaches as follows:

- **The problem of word-based approach (words in common):**
  It is quite time-consuming to search for words in common in full-length texts on-the-fly. To address this problem, we must reduce the amount of text that is searched.

- **The problem of citation-based approach (citations-to in common or citations-from in common):**
  Most previous research in citation analysis treats all citations equally, although there are actually several reasons for citations [Weinstock, 1971]. For more correct classification, it is necessary to take account of these reasons (citation types).

In the next section, we propose two classification methods that take account of these problems.

3.2 Methods of Classification

Our method 1 (words in common):

One way to reduce the calculation time of word-based approach is to use extracted text passages which represent the most salient contents of papers instead of the full-length text.
Kando proposed several rules using lexical cues to analyze the functional structure of technical papers. She extracted sentences that were assigned to particular categories (e.g. "Methods" and "Statement of Evidence") and used them for document and passage retrieval [Kando, 1997].
Similarly, we made use of the following two categories that are considered to represent salient contents of papers. We extracted sentences of these categories and used them for classification of papers.

- PURPOSE:

  The descriptions of the purpose of the research are closely related to the topic of the paper.

- METHODS:

  The descriptions of background theories or methods can also be considered as good indicators for topical classification.

ISSN: 2324-9773

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi:10.7152/acro.v11i1.12777

We used several lexical cues to extract sentences about 'PURPOSE' and 'METHODS'. For the extraction of 'PURPOSE', we used five cue phrases: 'our work', 'Our work', 'this paper', 'This paper', and 'purpose'. We assumed that the purpose of the research was written in sentences that include one of these five cue phrases and extracted them as the salient sentences of the paper. In the same way, we used 84 cue phrases for the extraction of 'METHODS'. These are the cue phrases used for identification of citation types (examples are shown in Table 5). We extracted sentences including these 84 cue phrases. We then measured the similarity by counting the number of words in common in the extracted sentences.

Our method 2 (BCCT: Bibliographic Coupling using Citation Types):

We refined bibliographic coupling by taking account of citation types. We measured similarity by counting the couplings of the same citation types.
The reason for using bibliographic coupling is that the database we used for our experiments (described below) included many very recent papers. In this case, co-citation analysis is unsuitable, because few papers in the database are cited from other papers.

3.3 Evaluation of Classification Methods : Methodology

In this section, we first explain our evaluation method, then show our proposed methods and other methods for classification of papers. We compare the effectiveness of each method using several measures, and discuss the results.

Document Collection

To evaluate the effectiveness of our methods, we have conducted some experiments. For the experiments, we used as the database 395 papers in TEX style source on computational linguistics from the E-Print archive [2], which we were kindly given permission to use for research purposes. This sample can be considered as papers on a specific subject domain collected indiscriminately from the WWW using an Internet search engine, or as a group of documents that have citation relationships in a bibliographic universe.

Automatic Identification of Citation Relationships

Because TEX has commands to write the bibliography such as "`\cite`" or "`\bibitem`", we could analyze such commands to obtain information of citation relationships between papers automatically. For the database, we can obtain the relationships with an accuracy of 95%.

[2] http://xxx.lanl.gov/cmp-lg/

ISSN: 2324-9773

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi:10.7152/acro.v11i1.12774

Relevance Judgments (Correct answers) and Queries

To evaluate the effectiveness of the automatic classification system, we prepared relevance judgments (correct answers) by classifying the 395 papers. We classified them into 58 categories manually. Each paper belonged to only one category. Three hundred and fifty papers belonged to ten categories, which are typical research fields on computational linguistics, while the others were not related to each other[3].
For the experiments, we first selected one paper from the 395 papers and conducted 395 searches. We considered the target paper as a query and attempted to collect all other papers in the same category. Our system then inspected papers from the database and returned ranked papers for each query.

Search Engine

We implemented the search engine based on a vector space model. Our system first extracts all nouns from passages using Brill's part-of-speech tagger [Brill, 1994]. Then the system calculates the similarity by cosine distance using the extracted nouns.

Alternatives

We conducted experiments using the following eight methods.

- **'FULL', 'TITLE', 'ABST':** using words in the full-length text, title and abstract.
- **'METHOD', 'PURPOSE' (our methods):** using words in sentences extracted using cue phrases.
- **'NBC':** using (normal) bibliographic coupling without considering citation types.
- **'BCCT-C', 'BCCT-BCO' (our methods):** using BCCT considering type C citations (BCCT-C) and all types (BCCT-BCO). Even if two papers commonly cite a paper, the coupling is not counted if their citation types do not accord.

Evaluation Measures

We compared the effectiveness of the above methods by the measures shown below;

- precisions at top-ranked documents
- fallout
- calculation time

Eleven points Recall-Precision is the most typical evaluation measure in the IR community. This measure is a good indicator to show the total balance of search effectiveness of the engines. However, for the evaluation of our system, 11 points R/P is not a sensitive measure for two reasons. One is that our system aims at high precision.

[3]We consider each of these papers as a category that includs only one paper.

ISSN: 2324-9773

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi:10.7152/acro.v11i1.12771

The other is that the numbers of papers collected by the eight methods are remarkably different.
The scores of precision and fallout are given by the following equations.

$$P(Precision) = \frac{\left(\begin{array}{l}The\ number\ of\ correct\ papers\\ collected\ by\ the\ search\ engine\end{array}\right)}{\left(\begin{array}{l}The\ number\ of\ papers\ collected\\ by\ the\ search\ engine\end{array}\right)}$$

$$F(Fallout) = \frac{\left(\begin{array}{l}The\ number\ of\ incorrect\ papers\\ collected\ by\ the\ search\ engine\end{array}\right)}{\left(\begin{array}{l}The\ number\ of\ papers\ that\\ belong\ to\ categories\ other\\ than\ a\ query's\ category\end{array}\right)}$$

For the calculation of precision and fallout, we made use of 'trec_eval' [trec_eval], which is an evaluation tool developed for the Text Retrieval Conference (TREC). If we give a ranked document list given by a search engine and a set of correct documents to trec_eval, we can obtain an eleven points of precision score, where recall scores are 0 to 1 at 0.1 intervals, respectively. Here, if we give a set of incorrect documents instead of a set of correct documents with a ranked document list to trec_eval, we can determine eleven points of '1-precision' score, where fallout scores are 0 to 1 at 0.1 intervals, respectively.

3.4 Evaluation of Classification Methods : Results

3.4.1 Precisions for Top-ranked Documents

In our research, precision is more important than recall; we have already explained the reason for this in section 2. If one class includes many unrelated papers, we must read them, spending unnecessary time, though we can grasp the outline of the subject domain without reading whole papers in the domain.
We show the results of averages of precisions for top-ranked documents over all 395 searches in Figure 4. Our method 'BCCT-C' was more effective than the others, especially for higher rankings. We think that type C is more important than the other citation types. If there are many couplings of type C between papers, it is considered that such papers have similar research motivations. In contrast, 'BCCT-BCO' was worse than 'NBC'. From the results of our experiments, we feel that type B seems not to work well, or works poorly rather frequently.
We compared the averages of the precisions for top-ranked documents of 'METHOD (our method)', 'TITLE' and 'ABST' at the 5th, 10th, 15th and 20th documents (in Table 2). In Table 2, we see that 'METHOD' is better than 'ABST' in each ranking, especially at the 5th document (9.37%). 'TITLE' and 'METHOD' gave almost the same effectiveness. Titles generally include many good content words of texts, so it is considered that this result is valid. However, if the titles are short, 'METHOD' seems to be more effective. One reason for the poor effectiveness of 'PURPOSE' was that the number of extracted sentences was very small (sometimes no sentences were extracted). So, it is necessary to

ISSN: 2324-9773

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi: 10.7152/acro.v11i1.12774

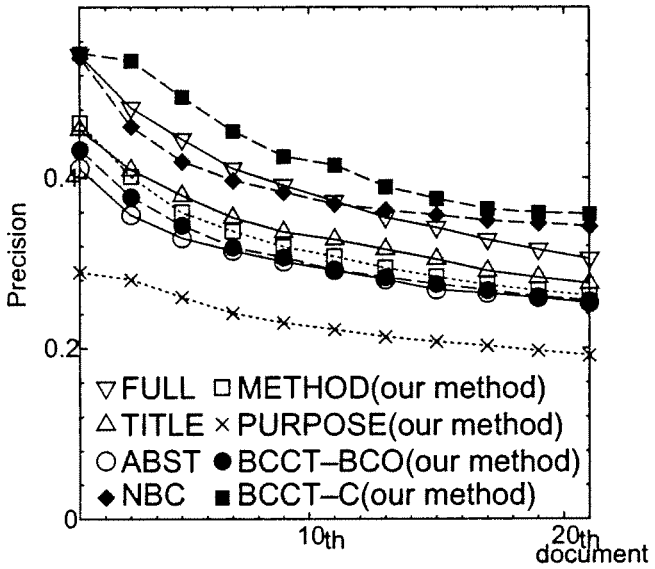

Figure 4: Precision for each ranking

increase the number of cue phrases for extracting salient sentences of 'PURPOSE'. The words in 'PURPOSE' sometimes represented the contents well, but in most cases they were too abstract, using broader terms, or details of papers were described there. Similar results were reported by Kando [Kando, 1997]. She found that searches using the categories of 'Method and Validity' and 'Evidence' are more effective than using the category of 'Research Topic', though she used manually category assigned documents and each paper in the database contained one or more sentences for "Research Topic".

Table 2: Precisions for top-ranked documents

ranking	ABST	METHOD (our method) (METHOD/ABST)	TITLE (TITLE/METHOD)
5th	0.36	0.39 (**+9.4 %**)	0.41 (+5.1 %)
10th	0.33	0.34 (+4.8 %)	0.36 (+6.2 %)
15th	0.29	0.31 (+6.0 %)	0.33 (+5.0 %)
20th	0.28	0.29 (+3.3 %)	0.30 (+4.5 %)

ISSN: 2324-9773

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. : 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi:10.7152/acro.v11i1.12774

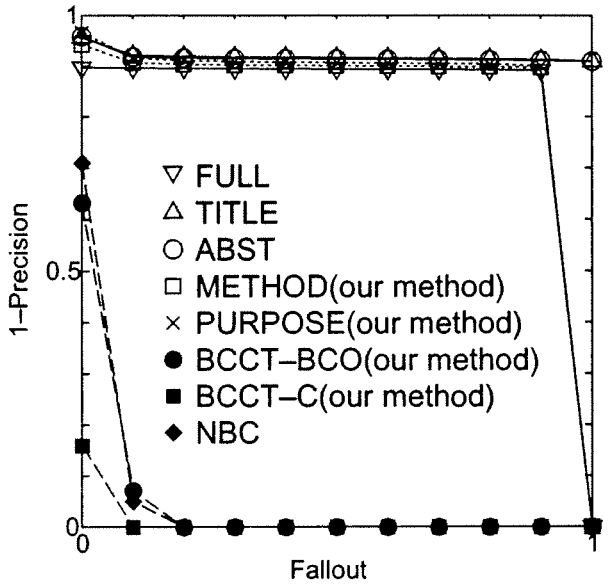

Figure 5: Evaluation by fallout and precision

3.4.2 Fallout

We also compare the eight methods by fallout. Fallout is a criterion for measuring the error of search engines. If the fallout score is small, the number of errors by the search engine is small and users need not read many unrelated papers.
We show the result in Figure 5. The three methods based on bibliographic coupling are more effective than the five word-based approach. In particular, 'BCCT-C' (our method) is better than the others.

3.4.3 Calculation time

Finally, we compare the eight methods by calculation time. The prototype classification system is implemented by a Perl script. We measure the time spent for calculating topical similarity per query on a 550 MHz Pentium III. This does not include the time spent for part-of-speech tagging and for extracting salient sentences ('METHOD' and 'PURPOSE').
Of the eight methods, 'TITLE' is fastest and 'FULL' is slowest, though the effectiveness of 'FULL' in top-ranked document is good. Calculation time of all methods except 'FULL' is within one hour. Of our four proposed methods, 'BCCT-C' and 'PURPOSE' are particularly fast.

3.4.4 Discussion

Generally, the methods based on citation relationships worked better than word-based approaches. However, fewer papers were collected by methods based on citation relationships than by word-based approaches. We conclude that 'BCCT-C' is effective if we want to know the outline (survey) of a specific subject domain in a short time. We

ISSN: 2324-9773

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi:10.7152/acro.v11i1.12774

Table 3: Calculation time of eight methods (per query)

method	time (second)
FULL	232
TITLE	0.25
ABST	1.2
METHOD (our method)	8.1
PURPOSE (our method)	0.77
BCCT-BCO (our method)	14
BCCT-C (our method)	1.3
NBC	14

further conclude that our proposed method 'BCCT-C' should be effective for the final goal of our research, automatic generation of review articles.

4 PRESRI: a Prototype of the Support System for Writing a Review Article

We implemented our method 'BCCT-C' by expanding PRESRI[4], which was developed as a system to support writing a review article in our previous work.
Figure 2 shows the display of the support system. The left window shows the citation relationships around 'CID:8001780' [Pereira and Sheabes, 1992]. In the database, this paper is cited by eight other papers, three by type C ('DID:9512002', 'DID:9504034', 'DID:9606014'), three by type B ('DID:9611002', 'DID:9604008', 'DID:9605036') and the remaining two by type O ('DID:9605012', 'DID:9606027').
Furthermore, 'DID:9504034' [Stanley, 1995] and 'DID:9606014' [Stanley, 1996] are grouped together, because these two papers have in common a type C citation to a paper other than 'CID:8001780' (our method 'BCCT-C'). Both papers are about probabilistic language modeling and by the same author, and the topic is different from that of the third paper ('DID:9512002') [Marcken] in the type C group. We can validate the effectiveness of our method 'BCCT-C' from the output of PRESRI.
By clicking icons of citing areas, we can see citing areas in the papers. The right window displays some paragraphs including the citation of 'CID:8001780'. In these paragraphs, boldfaced sentences are citing areas, where the authors of 'DID:9606014' and 'DID:9606027' describe 'CID:8001780'. In this way, by displaying several classified citing areas, the system can assist our overview of similarity and differences between papers. Therefore, we think it is useful for writing a review article.
For the implementation of the current version of PRESRI, we partially referred to the source code and the interface of ResearchIndex [Lawrence et al., 1999]. ResearchIndex is a similar system to PRESRI. It is constructed by collecting research papers from the

[4]http://galaga.jaist.ac.jp:8000/pub/tools/sum

ISSN: 2324-9773

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi:10.7152/acro.v11i1.12774

World Wide Web in Postscript and PDF formats. ResearchIndex provides several methods for retrieving papers including bibliographic coupling.

5 Conclusion and Future Research

The purpose of our research is to implement a system to classify research papers with high precision, even though the recall may not be high. For this task, we proposed two methods of classification. The results of our experiments showed that our method based on bibliographic coupling is more effective than others, especially at higher rankings. We are now working towards automatic generation of review articles. We think that the method proposed in this paper will be useful for our future work.
In this paper, we proposed three citation types. But, as anonymous reviewer suggested, type S (supporting other studies) and type P (following or part of other studies) are important for automatic generation of review articles. We will examine whether these fine citation types could contribute to improve the effectiveness of classification.
In this paper, we used 395 papers on computational linguistics. We will next investigate the generality of our methods using larger databases. We will also use databases from other subject domains to investigate whether our methods are domain-independent.
We evaluated our methods by comparing the system outputs with papers classified manually. We also plan to evaluate our system by letting other users apply it.

Acknowledgments

The authors would like to express our gratitude to Dr. Dagobert Soergel of University of Maryland and anonymous reviewers for their suggestions to improve our paper.

References

[Brill, 1994] Brill, E. *Some advances in rule-based part of speech tagging.* Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94), pp. 722–727, 1994.

[Kando, 1997] Kando, N. *Text-level Structure of Research Articles and Its Implication for Text-based Information Processing Systems.* In Proceedings of the 19th Annual BCS-IRSG Colloquium on IR Research, pp.68–81, 1997.

[Kessler, 1963] Kessler, M. M. *Bibliographic Coupling between Scientific Papers.* American Documentation, Vol.14, No.1, pp. 10–25, 1963.

[Lawrence et al., 1999] Lawrence, S., Giles, L. and Bollacker, K. *Digital Libraries and Autonomous Citation Indexing.* IEEE Computer, Vol. 32, No. 6, pp. 67–71, 1999.

[Liu, 1993] Liu, M. *Progress In Documentation: The Complexities of Citation Practice: A Review of Citation Studies.* Journal of Documentation, Vol.49, No.4, pp. 370–409, 1993.

ISSN: 2324-9773

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi:10.7152/acro.v11i1.12774

[Nanba and Okumura, 1999] Nanba, H. and Okumura, M. *Towards Multi-paper Summarization Using Reference Information.* Proceedings of the 16th International Joint Conferences on Artificial Intelligence(IJCAI-99), pp. 926–931, 1999.

[Narin et al., 1994] Narin, F., Olivastro, D. and Stevens, K. A. *Bibliometrics/Theory, Practice and Problems.* Evaluation Review, Vol.18, No.1, pp. 65–76, 1994.

[Salton and McGill, 1983] Salton, G. and McGill, M. J. *Introduction to Modern Information Retrieval.* New York, McGraw-Hill, 448p, 1983.

[Small, 1973] Small, H. *Co-citation in the Scientific Literature: A New Measure of the Relationship between Two Documents.* Journal of the American Society for Information Science, Vol.24, pp. 265–269, 1973.

[trec_eval] trec_eval. "ftp://ftp.cs.cornell.edu/pub/smart", 1992.

[Weinstock, 1971] Weinstock, N. *Citation indexes, in Kent A. (Ed.).* Encyclopedia of Library and Information Science, New York: Marcel Dekker, Vol.5, pages 16–41, 1971.

[White and McCain, 1989] White, H. D. and McCain, K. W. *Bibliometrics.* Annual Review of Information Science and Technology(ARIST), Vol.24, pages 119–186, 1989.

Papers Used in the Explanation of PRESRI

[Marcken] Carl de Marcken. *The Unsupervised Acquisition of a Lexicon from Continuous Speech.* MIT AI Memo No. 1558/CBCL Memo No. 129.

[Pereira and Sheabes, 1992] Pereira, F. and Schabes, Y. *Inside-outside re-estimation from partially bracketed corpora.* Proceedings of the 30th Annual Meeting of the ACL, pages 128–135, 1992.

[Stanley, 1995] Stanley, F. C. *Bayesian Grammar Induction for Language Modeling.* Proceedings of the 33rd Annual Meeting of the ACL, pages 228–235, 1995.

[Stanley, 1996] Stanley, F. C. *Building Probabilistic Models for Natural Language.* Doctoral dissertation in Harvard University, CRCT TR-02-96, 1996.

ISSN: 2324-9773

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. 11th ASIS SIG/CR Classification Research Workshop, 117-134. doi:10.7152/acro.v11i1.12774

Appendix

Table 4: Cue phrases for citing area extraction

types	cue phrases
(1) anaphor	For this, For these, On this, On these, In this, in this in these, In these, This, These, Therefore
(2) negative expression	yet, less, but, in spite of, unlike, rarely in contrast, although, Still, Nevertheless, instead, despite, irrelevant, has not been, not attempt to not possible to, this is not, but is not, less, has not, have not
(3) 1st person pronoun	I, in our example, our analysis was, our analysis of by using our, in our work, our analysis is, to our concept, our analysis, our work our example, using our, we
(4) 3rd person pronoun	they, their, them, he, his, him, she her, hers
(5) adverb	And, Furthermore, Because, Again, Additionally, Such, In such, So
(6) other	difference between, defect, drawback, impossible, Using, we incorporate, in the implementation, is implemented, first, second, theory, theoretical, origin, based, base, basis, adopt, apply, applied, foundation, fundamental, radical, element, underlie, underlay, underlain, underlying, In particular, follow

ISSN: 2324-9773

Table 5: Cue phrases for determining type B(refers to theory or methods)

based mainly on	based on ... in	is based on	are based on
, based on	this ... based on	employ	invoke
Based on	the basic	underlie	underlain
assume	widely used	has used	used by
use ... to	can use	used as a	by using
Using the	is checked	we use	We will use
we used	result	make use of	made use of
Making use of	advantage of	we describe	is described in
are described in	accord with	accorded with	correspond to
mentioned	benefit	beneficial	we introduce
And	We argue	In such	we present
support	is given in	are given in	offer
For ... reason	we ... influence	assume	is needed to
are needed to	been given	a given	given the
follows	following	we believe	implementation
we investigate	to consider	which can be	the possible
available for	apply	applied to	application to
adopt	We adopted	extend the	we extended
we consider	extended to	expands	expanded
refer ... to	referred to	This ... importance	This ... important
implement			

Table 6: Cue phrases for determining type C (refers to problems or gaps in related works)

although the	, although	Though,	however, ... our
however, ... they	recently ... however	, however	however, ... not
However,	however, the	however, that	the only
But	but a	but the	but it
but is	but are	but rather	but no
but they	but their	but he	but his
but she	but her	but it	but instead
Instead,	In spite of	does not	did not
that is not	not be	it is not	this is not
was not	were not	it does not	may not
might not	will not	would not	wouldn't
should not	can not be	could not	(citation) ... can not
need not	not always	not have	have not
is too	has not	little influence	that do not
they do not	he does not	she does not	not require
not provide	not cover	not in effect	more efficient than ... (citation)
not enough	less studied	difference from	different from
more difficult	a difficult	difference between	

ISSN: 2324-9773