TRANSFORMING RHETORICAL DOCUMENT PROFILE INTO TAILORED SUMMARY OF SCIENTIFIC PAPER

Maayu Leylia Khodra, Mohammad Dimas, Dwi Hendratmo Widyantoro, E. Aminudin Aziz, Bambang Riyanto Trilaksono

Abstract

Since abstract of scientific paper is author biased, readers’ required information may not be included in the abstract. Tailored summary may help them to get a summary based on their information needs. This research is the first one that implements tailored summary system for scientific paper. Tailored summary applies information extraction that transforms a scientific paper into Rhetorical Document Profile, a structured representation of paper content based on rhetorical scheme of fifteen slots. This research adapted building plan that used rhetorical scheme of seven slots. We also implement tailored summary system. After generating initial summary, surface repair is conducted to improve summary readability. Each sentence in initial summary is combined with template phrase based on syntax-tree combination method. There are five groups of template phrases provided in surface repair. We construct evaluation standards by asking five human raters. The best method for sentence selection subsystem that uses Maximal Marginal Importance-Multi Sentence is employing TF.IDF weighting system with precision/recall of 0.61. The surface repair subsystem has acceptance of 0.91.

Key words: Tailored Summary, Rhetorical Document Profile, Building Plan, Maximal Marginal Importance – Multi Sentence, Surface Repair.
INTRODUCTION

Abstract of scientific paper is often the first part to be read because it is brief and concise. However, it is author-biased [1]. The reader cannot find some information because author considered that information as unimportant to be included in the abstract. In assisting the reader to get his needed information from scientific paper, our research aims to investigate automatic summarization on scientific paper to produce tailored summary, which is a summary that takes into account user information needs.

Summarizing paper is the process of extracting important information from a scientific paper and transform it into a shorter text (summary). Automatic summarization is applied computational linguistics to develop intelligent system by acting humanly for natural language processing.

A paper summarization system generally produced generic summary. It is not much different from scientific abstract that is written by the authors. It means that generic summary cannot help user more than abstract. Our research will produce tailored summary based on user task and background knowledge.

Although scientific paper is unstructured document, it has common structures, namely sections and paragraphs as explicit structure, and implicit structure of problem solving in the form of rhetorical information in each sentence. Rhetoric is the intention information to be conveyed to the reader by the author of the paper. Therefore, information in scientific paper can be structured in the form of Rhetorical Document Profile (RDP) [2] based on rhetorical structure.

RDP and tailored summary were proposed as extended concepts of argumentative zoning [2]. No previous study has implemented both concepts, and our study is the first one to implement them. In addition, our study also contributes in adapting the building plan (summary of the composition pattern) for the fifteen rhetorical categories.

There are two main stages to produce a tailored summary of scientific paper: (1) transforming scientific paper into RDP, and (2) transforming RDP to a tailored summary according to user information needs. The first phase of generating RDP is processed by taking all the sentences in the abstract and the main sentence of each paragraph in the other section [3], and then classifying the rhetorical category of each sentence [4]. The second phase uses the user needs to determine the building plan, takes the sentences of RDP, and surface repair. Our focus is to describe series of processes carried out in the second stage and some sample of various summary generated from a scientific paper.

The rest of the paper is organized as follows. The next section provides related work on automatic paper summarization and RDP. Section 3 discusses summarization using RDP, and section 4 describes the evaluation. In the last section, we discuss the conclusions and further research to be conducted.

AUTOMATIC SUMMARIZATION OF SCIENTIFIC PAPER

Automatic summarization has been studied since 1958 for producing scientific abstract in order to facilitate the reader in identifying the topic of scientific paper quickly and accurately [5]. Generating abstract automatically helps the author in making abstract or completing papers without abstract. However, summarizing scientific papers has less attention nowadays because the researchers focus to develop newspaper summarization.

Since summarization methods depends on the genre of the document, methods for news genre can not be applied directly for scientific articles. It is different in focus of identification, position of important information, context, and degree of compression [6].

Existing summarization system usually produced generic summary for all users. Some systems generates different summaries for different topic, and are known as topic-focused summarization [7]. Jiaing [8] identified topics in corpus and created summary for each topic. Moreover, user-focused summarization consideres user information needs [9][10][11] to generate tailored summary [2][12]. Personalized summarization adds personalization by keeping parameter values of user information needs in user profiles [13][14]. User does not need to input the parameter values if he want to use the system again [15].

Summarizing scientific paper generally uses extractive approach. This approach selects
important sentences and arrange the extraction into summary. Advantages of extractive approach is grammatical summary that is still easy to read by humans. The disadvantage is low coherence between the sentences. The main components of extractive summarization is important identification and summary generator. The first component will assess each sentence in the document and select some sentences with the highest score. The second component will put together a collection of sentences that are considered essential to form a summary.

Filho [16] modified GistSumm, newspaper summarizer, by adding structure detector and additional constraint that there is minimum one summary sentence representing each section. He compared this modified system and the original system, and concluded that performance of the modified system is better than the original.

Kupiec [17] developed trainable summarizer by using assumption that every sentence can be binary classified based on its relevance. This approach is popular and it used machine learning.

Qazvinian [18] summarized scientific paper based on clustering of all citations, and produced summary in the form of citation network. The summary is extraction of centroids of clusters. Agarwal [19] also developed summarization system of scientific paper by using clustering of co-citations based on user query.

**RHETORICAL DOCUMENT PROFILE**

*Rhetorical Document Profile* (RDP) is representation of extracted information of scientific paper. RDP is filled by argumentative zoning (AZ) [2] or rhetorical classification of topic sentences of paragraphs [20]. Summary of scientific paper is produced by using the information of filled RDP.

Teufel [21] defined fifteen rhetorical categories that state the intention information to be conveyed by the author of the paper. This scheme is also known as AZ-II [21] and it is improvement of AZ scheme that has seven categories [2]. Scheme AZ-II (see Table 1) is more informative because AZ-II can identify problem solving structure better than AZ. This paper uses scheme AZ-II.

In our previous research, filled RDP has been generated as shown by Figure 1. All sentences in each slot are extracts of a scientific paper.

**Table 1. Rhetorical Categories of Scientific Paper [21].**

<table>
<thead>
<tr>
<th>Kategori</th>
<th>Deskripsi</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIM</td>
<td>Statement of specific research goal, or hypothesis of current paper</td>
</tr>
<tr>
<td>NOV_ADV</td>
<td>Novelty or advantage of own approach</td>
</tr>
<tr>
<td>CO_GRO</td>
<td>No knowledge claim is raised (or knowledge claim not significant for the paper)</td>
</tr>
<tr>
<td>OTHR</td>
<td>Significant knowledge claim held by somebody else. Neutral description</td>
</tr>
<tr>
<td>PREV_OWN</td>
<td>Significant knowledge claim held by authors in a previous paper. Neutral description</td>
</tr>
<tr>
<td>OWN_MTHD</td>
<td>New Knowledge claim, own work: methods</td>
</tr>
<tr>
<td>OWN_FAIL</td>
<td>A solution/method/experiment in the paper that did not work</td>
</tr>
<tr>
<td>OWN_RES</td>
<td>Measurable/objective outcome of own work</td>
</tr>
<tr>
<td>OWN_CONC</td>
<td>Findings, conclusions (non-measurable) of own work</td>
</tr>
<tr>
<td>CODI</td>
<td>Comparison, contrast, difference to other solution (neutral)</td>
</tr>
<tr>
<td>GAP_WEAK</td>
<td>Lack of solution in field, problem with other solutions</td>
</tr>
<tr>
<td>ANTISUPP</td>
<td>Clash with somebody else’s results or theory; superiority of own work</td>
</tr>
<tr>
<td>SUPPORT</td>
<td>Other work supports current work or is supported by current work</td>
</tr>
<tr>
<td>USE</td>
<td>Other work is used in own work</td>
</tr>
<tr>
<td>FUT</td>
<td>Statements/suggestions about future work (own or general)</td>
</tr>
</tbody>
</table>
In this paper, we propose a learning-based approach to combine various sentence features.

Traditionally, features for summarization were studied separately.

We investigate the effectiveness of different sentence features with supervised learning to decide which sentences are important for summarization.

Recently, semi-structure events (<REF>; <REF>; <REF>) have been investigated by many researchers as they balanced document representation with words and structures.

An automatic evaluation package, ie, ROUGE (<REF>) is employed to evaluate the summarization performance.

Table 2. Building Plan for Short Summary.

<table>
<thead>
<tr>
<th>Task \ User</th>
<th>Length of summary = short</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>2 Aim</td>
</tr>
<tr>
<td></td>
<td>Gap_Weak + 2 AIM</td>
</tr>
<tr>
<td></td>
<td>1 Co_Gro + 1</td>
</tr>
<tr>
<td></td>
<td>1 Co_Gro + 1</td>
</tr>
<tr>
<td>Contrastive</td>
<td>2 AIM +1-2 Codi +1</td>
</tr>
<tr>
<td></td>
<td>Gap_Weak + 2 AIM +1-2 Codi</td>
</tr>
<tr>
<td></td>
<td>1 Co_Gro + 1</td>
</tr>
<tr>
<td>Ancestry</td>
<td>2 AIM +1-2 Use +1 Fut</td>
</tr>
<tr>
<td></td>
<td>Gap_Weak + 2 AIM +1-2 Use +1 Fut</td>
</tr>
</tbody>
</table>

Table 3. Building plan for long summary

<table>
<thead>
<tr>
<th>Task \ User</th>
<th>Length of summary = long</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>2-3 AIM +1-2 Own_Conc</td>
</tr>
<tr>
<td></td>
<td>Gap_Weak + 2 AIM +1 Own_Conc</td>
</tr>
<tr>
<td></td>
<td>1 Co_Gro +1</td>
</tr>
<tr>
<td></td>
<td>1 Co_Gro +1</td>
</tr>
<tr>
<td>Contrastive</td>
<td>2-3 AIM +1-2 Codi +1</td>
</tr>
<tr>
<td></td>
<td>Gap_Weak + 2 AIM +1 Use +1 Fut</td>
</tr>
<tr>
<td></td>
<td>1 Co_Gro +1</td>
</tr>
<tr>
<td></td>
<td>1 Co_Gro +1</td>
</tr>
<tr>
<td>Ancestry</td>
<td>2-3 AIM +1-2 Use +1 Fut</td>
</tr>
<tr>
<td></td>
<td>Gap_Weak + 2 AIM +1 Use +1 Fut</td>
</tr>
</tbody>
</table>

SUMMARIZING SCIENTIFIC PAPER
BYP PROCESSING RDP

If filled RDP has been generated as shown in Figure 1, the next step is processing RDP to produce tailored summary based on user information needs. This needs can be represented by three variables: task (general, ancestry, contrastive), background knowledge (informed, uninformed), and summary length (short, longer) [2]. Consequently, there are twelve user types based on user information needs.

For each user type, building plan is defined [2]. It contains summary composition based on RDP. For example, as shown in Table 2, summary composition for user type <general task, informed user, short summary> is 2 AIM sentences, but summary composition for user type <general task, uninformed user, short summary> is 2 CO_GRO sentences and 2 AIM sentences.

Since extractive summary is composed by important extracts that is loss of context, one of the problems is low coherence of summary. In this paper, improvement of summary coherence...
is conducted by surface repair that makes summary is easier to be read.

Figure 2 shows processes of second stage that is transforming filled RDP into tailored summary. It consists of two processes: generating initial summary and surface repair.

**GENERATING INITIAL SUMMARY ACCORDING TO BUILDING PLAN**

Before generating initial summary, we have to define building plan first. Teufel [2] has defined building plan, but it was built for AZ with 7 rhetoric classes. That’s why adaption of building plan for AZ-II with 15 rhetoric classes needs to be done.

Modification of building plan will be done by following AZ to AZ-II conversion scheme [21], for example background problem and background aim those are exist on AZ will be converted into CO_GRO. Besides the modification of existing building plan, there is also necessity to define user’s needs combination that is not exist yet, it is summary with requirement long for task contrastive and ancestry. Result of this modification and improvement are shown in Table 2 and Table 3, it shows all combination of needed information from three defined variables.

Based on the user information needs, the matched building plan will be selected. If the required sentence in building plan is less than the number of sentences in the slot, sentence selection must be conducted. This research used Maximal Marginal Importance – Multi Sentence (MMI-MS) [22]. This method is a selection method that will choose some most important sentences and have least similarity among the other sentences. In the following example, K2 is a surface repair from K1.

K1: This paper describes recent work on developing an integrated heuristic scheme for selecting the parse that is deemed `best` from such a collection.

K2: This paper’s goal is to describes recent work on developing an integrated heuristic scheme for selecting the parse that is deemed `'best`' from such a collection.

**SURFACE REPAIR**

Surface repair is used to enhance the readability of initial summary and also to make sentences in the summary more related each other. This step will increase the coherency between sentences in the final summary.

This research used syntax tree based combination method [24]. After preparing some template phrases, sentence combining will be conducted based on the collected phrases [2]. Five groups of template phrases will be explained later.

**Group I: “This paper’s goal is to”, “This paper’s topic is to”**

This phrase group is only able to be combined for sentence which type is AIM and the sentence is the first sentence in summary paragraph. This phrase is not able to be used for passive sentence, AIM sentence with “we/i” as subject, and sentence that already has word “to”. Those restrictions are there in order to prevent changing the meaning and grammatical errors on the result of surface repair. For example, K2 is a surface repair from K1.

K1: This paper describes recent work on developing an integrated heuristic scheme for selecting the parse that is deemed `best` from such a collection.

K2: This paper’s goal is to describes recent work on developing an integrated heuristic scheme for selecting the parse that is deemed `'best`' from such a collection.

**Group II: “This paper’s specific goal is to”, “Another goal is to”, “The goal is to”, “This approach”**

This phrase groups is only able to be used when the combined sentence’s type is aim and not the first sentence of the summary. Similar with the previous group, phrase can’t be used for passive sentence, sentence that contains word “to”, and sentence with subject “we/i”. If K1 on the previous sentence’s type is AIM and not the first sentence, then K3 is the surface repair for K1.

K3: This approach describes recent work on developing an integrated heuristic scheme for selecting the parse that is deemed `'best`' from such a collection.

**Group III: “More specifically”**

Template phrase ”more specifically” will be used when the combined sentence is not first sentence and preceded with sentence which type is AIM too. Because this phrase is a conjunction between two AIM sentences, so the form of the sentence is free. In the following example, K5 combined with the template phrase and become clause of K4 as shown in K6.

K4-K5: I describe a compiler and development environment for feature-augmented two-level morphology rules
Equation (4) modifies slightly the authors, there will be 12 summary combinations that can be generated from Table 2 and Table 3. Based on the building plan that is shown in the beginning of sentence. The objective is to obtain sentence’s POS (part of speech) which is used for sentence’s continuation characteristic, it is identified by the type of sentence that is “use”. The combination method is different with the other groups, it is by replacing “i” or “we” from the original sentence directly. In the previous example, K6 is result of surface repair for K4.

**Group IV: “The authors”**

Template phrase “the authors” is used as word substitute in summary’s sentences those refer to the writer, for instance we or i. For this phrase the combination method is different with the other groups, it is by replacing “i” or “we” from the original sentence directly. In the previous example, K6 is result of surface repair for K4.

**Group V: “It uses”**

Template phrase “it uses” is used for sentence that has continuation characteristic, it is identified by the type of sentence that is “use”. For the combination of this phrase, the method is a little different with the other phrase combination method. In this combination, the method will look for NP (noun phrase) from the sentence’s POS (part of speech) which belongs to the first VP (verb phrase) of the sentence. The objective is to obtain sentence’s object. After that POS is found, words those are appeared before VP and NP will be removed. And the last step is to add phrase “it uses” in the beginning of sentence.

**K7**: This is basically the entropy used in Quinlan, 1986.

**K8**: It uses the entropy used in Quinlan, 1986.

**Example of Tailored Summary**

Based on the building plan that is shown in Table 2 and Table 3, there will be 12 summary combinations that can be generated from a scientific paper. The sentence’s order in the summary follows the sentence’s order in the original source.

---

<?xml version="1.0" ?>

- <root> <slot> <slotName>aim</slotName> </slot> <sentence> In this paper we presented a new model that implements the similarity-based approach to provide estimates for the conditional probabilities of unseen word cooccurrences. </sentence> <sentence> In this work we propose a method for estimating the probability of such previously unseen word combinations using available information on “most similar” words. </sentence> <sentence> We focus here on a particular kind of configuration, word cooccurrence. </sentence> </sentences> </slot> <slot> <slotName>co_gro</slotName> </slot> <sentence> For example, a speech recognizer may need to determine which of the two word combinations “eat a peach” and “eat a beach” is more likely. </sentence> <sentence> In many applications of natural language processing it is necessary to determine the likelihood of a given word combination. </sentence> <sentence> The MLE for the probability of a bigram (w1,w2) is simply: where is the frequency of (w1,w2) in the training corpus and N is the total number of bigrams. </sentence> <sentence> However, the nature of language is such that many word combinations are infrequent and do not occur in a given corpus. </sentence> <sentence> Because of data sparseness, we cannot reliably use a maximum likelihood estimator (MLE) for bigram probabilities. </sentence> <sentence> However, this estimates the probability of any unseen bigram to be zero, which is clearly undesirable. </sentence> <sentence> Arc scores in those lattices are sums of an acoustic score (negative log likelihood) and a language-model score, in this case the negative log probability provided by the baseline bigram model. </sentence> <sentence> Arc scores in those lattices are sums of an acoustic score (negative log likelihood) and a language-model score, in this case the negative log probability provided by the baseline bigram model. </sentence> <sentence> We describe a probabilistic word association model based on distributional word similarity, and apply it to improving probability estimates for unseen word bigrams in a variant of Katz’s back-off model. </sentence> <sentence> Following Pereira, Tibby, Lee, 1993 , we measure word similarity by the relative entropy, or Kullback-Leibler (KL) distance, between the corresponding conditional distributions. </sentence> <sentence> We evaluated our method by comparing its perplexity and effect on speech-recognition accuracy with the baseline bigram back-off model developed by MIT Lincoln Laboratories for the Wall Street Journal (WSJ) text and dictation corpora provided by ARPA’s HLT program Paul, 1991. </sentence> <sentence> For perplexity evaluation, we tuned the similarity model parameters by minimizing perplexity on an additional sample of 57.5 thousand words of WSJ text, drawn from the ARPA HLT development test set. </sentence> </root>
In this work the authors propose a method for estimating the probability of such previously unseen word combinations using available information on most similar words, more specifically in this paper the authors presented a new model that implements the similarity-based approach to provide estimates for the conditional probabilities of unseen word cooccurrences.

Figure 4. Example for Short Summary for User with General Task and Informed User Background.

In this work the authors propose a method for estimating the probability of such previously unseen word combinations using available information on most similar words, more specifically the authors focus here on a particular kind of configuration, word cooccurrence. Arc scores in these lattices are sums of an acoustic score (negative log likelihood) and a language-model score, in this case the negative log probability provided by the baseline bigram model. In this paper the authors presented a new model that implements the similarity-based approach to provide estimates for the conditional probabilities of unseen word cooccurrences.

Figure 5. Example for Long Summary for User With General Task and Informed User Background.

In this work the authors propose a method for estimating the probability of such previously unseen word combinations using available information on most similar words. It uses their method by comparing its perplexity and effect on speech-recognition accuracy with the baseline bigram back-off model developed by MIT Lincoln Laboratories for the Wall Street Journal -LRB- WSJ -RRB- text and dictation corpora provided by ARPA’s HLT program Paul, 1991. It uses the similarity model parameters by minimizing perplexity on an additional sample of 57.5 thousand words of WSJ text, drawn from the ARPA HLT development test set. In this paper the authors presented a new model that implements the similarity-based approach to provide estimates for the conditional probabilities of unseen word cooccurrences.

Figure 6. Example for short summary for ancestry user task and informed user background.

For example, Figure 4 until Figure 7 show 4 summaries, with various length parameter (Figure 4 and Figure 5), task parameter (Figure 4 and Figure 6), and also user’s background knowledge (Figure 6 dan Figure 7). The summaries are composed by sentences that are obtained from Aim, Own_Mtd, and Use in XML form (Figure 3).

Figure 4 shows short summary with general user task and informed user background, besides that Figure 5 shows longer version of summary for the same user. Based on building plan, short summary will be consisted of 2 Aim sentences those are connected with template phrase “more specifically”. Long summary will be consisted 2-3 Aim sentences and 1 OWN_MTHD Sentence. In Figure 5, third sentence’s type is OWN_MTHD.

Difference between user tasks in Figure 4 and Figure 6 for short summary and informed user background is 1-2 additional sentences with use type. In Figure 6, the second and third sentence’s type is USE. In Figure 4, there is a merger process between two AIM sentences with template phrase, but this does not happen in the summary in Figure 6 because the AIM sentences are the first and the fourth sentence.

Figure 7. Example for Short Summary with Ancestry User Task and Uninformed User Background.

However, the nature of language is such that many word combinations are infrequent and do not occur in a given corpus. In this work the authors propose a method for estimating the probability of such previously unseen word combinations using available information on most similar words. The MLE for the probability of a bigram (w1, w2) is simply: where is the frequency of (w1, w2) in the training corpus and is the total number of bigrams. It uses their method by comparing its perplexity and effect on speech-recognition accuracy with the baseline bigram back-off model developed by MIT Lincoln Laboratories for the Wall Street Journal -LRB- WSJ -RRB- text and dictation corpora provided by ARPA’s HLT program Paul, 1991. It uses the similarity model parameters by minimizing perplexity on an additional sample of 57.5 thousand words of WSJ text, drawn from the ARPA HLT development test set. In this paper the authors presented a new model that implements the similarity-based approach to provide estimates for the conditional probabilities of unseen word cooccurrences.

Instruction:
You’ll be given some sentences those are already grouped. Please choose some number of sentences based on the instruction for each group by giving circle on the number besides the sentence. Choose sentence that is most representative or the most different compared to the other sentences.
Paper’s Title: AN INTEGRATED HEURISTIC SCHEME FOR PARTIAL PARSE EVALUATION

Group of sentences that contain the purpose of written paper:

1. This paper describes recent work on developing an integrated heuristic scheme for selecting the parse that is deemed ‘‘best’’ from such a collection.

2. Preliminary results from experiments conducted on parsing speech recognized spontaneous speech are also reported.

3. First, we wanted to compare the parsing capability of the GLR* parser with that of the original GLR parser.

Table of Sentences

<table>
<thead>
<tr>
<th>Sentences (choose 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. This paper describes recent work on developing an integrated heuristic scheme for selecting the parse that is deemed ‘‘best’’ from such a collection.</td>
</tr>
<tr>
<td>2. Preliminary results from experiments conducted on parsing speech recognized spontaneous speech are also reported.</td>
</tr>
<tr>
<td>3. First, we wanted to compare the parsing capability of the GLR* parser with that of the original GLR parser.</td>
</tr>
</tbody>
</table>

Figure 8. Example of Grouped Sentences in The Feedback form for Evaluating Sentence Selection Process.

Figure 9. Example of Paired Sentence in Feedback form for Evaluating Result of Surface Repair.

Figure 7 shows summary for user with informed background that is different with Figure 6. Building plan for this summary is 1 CO_GRO sentence, 1 GAP_WEAK sentence, 2 AIM sentences, and 1-2 USE sentence(s). Sentence’s rhetorical category in summary as shown in Figure 7 are GAP_WEAK, AIM, CO_GRO, USE, and AIM.

RESULT AND DISCUSSION

Evaluation was conducted in each subsystem. There are evaluation of initial summary step and evaluation of surface repair step. There were five respondents which were fourth year students of Informatics Engineering. Each respondent was asked to fill feedback form related to sentence selection and surface repair.

Evaluation in Sentence Selection

In order to evaluate sentence selection, it is required to have a standard sentence selection that is done by human. This standard will be compared to the result of sentence selection that is done by the system.

In sentence selection process for generating initial summary, system chooses some sentences in certain slot. Number of selected sentences depends on building plan.

All respondents were given feedback form that contains 18 groups of sentences which total was 51 sentences from one paper. Figure 8 shows example of sentence group within the feedback form. Sentences in feedback form were grouped based on its slot in RDP. The result was 18 sentences from human respondent.

Measurement of the evaluation uses precision and recall, Equation (1) and (2) is counted based on the number of sentences those are chosen by both respondent and system. A sentence will be included to the evaluation measurement if that sentence is chosen by at least 3 respondents from total 5 respondents. If a sentence is chosen by both system and respondent, then that sentence will be counted as true positive (TP). If a sentence is chosen by system but not chosen by respondent, then that sentence will be counted as false positive (FP). And the other way, sentence will be counted as false negative (FN).

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (1)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (2)
\]

The weighting process for counting MIMs uses three type combinations. The first combination is idf, tf, normalization, and IGR. The second combination is tf, and idf. And the third combination is tf, idf, and normalization.

For the first and third weighting combination, it is given value 0.56 for recall and precision of system. And for second weighting combination, the value of system’s recall and precision is 0.61 which is also the best value among other combinations. Precision and recall have the same value because the value of FP and FN are same. Complete result of selection is shown in Table 4.
Table 4. Result of sentence selection compared to sentence selection standard that is done by human respondent.

<table>
<thead>
<tr>
<th>Combination</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tf, idf, normalization, IGR</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>Tf, idf</td>
<td>11</td>
<td>7</td>
<td>7</td>
<td>0.61</td>
<td>0.61</td>
</tr>
<tr>
<td>Tf, idf, normalization</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>0.56</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Evaluation in Surface Repair

Just like the evaluation for sentence selection, evaluation for surface repair was done based on standard that was the result of feedback form to human respondent. In that feedback form, there are some sentences that were taken from the system’s summary. For one RDP, there will be some variation of summary based on parameters for customizable user’s needs. Basically, this evaluation is done to evaluate the surface repair in the final summary. Therefore, evaluation is not done on each template in surface repair, but it is done on each sentence that is being surface repaired in the final summary. Respondent answer for each sentence is a yes or no, yes if respondent accepts and no if respondent doesn’t accept. Figure 9 shows example of paired sentence in feedback form.

From 15 pair of sentences those are given to respondents, 14 pairs are accepted so the acceptance value is 0.91. Sentence is assumed as accepted if that sentence is accepted by at least 3 out of 5 respondents. Acceptance value is ratio between accepted sentences and total sentences. Sentence can’t be accepted by respondent basically caused by the changing of meaning, or resulting grammatical error but does not give meaning change.

CONCLUSION

This paper has discussed the phase to transform RDP to tailored summary which is the second step of summary generation after generating RDP. This research is aim to implement tailored summary that has already proposed by Teufel [2].

There are two main processes, generating initial summary and surface repair. Based on the information needs, initial summary is created by selecting sentence from RDP based on the slot listed in building plan. Surface repair combines sentence in initial summary with template phrase based on syntax tree combination method.

Further works will integrate each subsystem that is already developed into one integrated system for summarizing scientific paper. Besides that, this research is the initial research towards research in summarizing multiple scientific paper.

REFERENCES


