



Automatic Title Generation in Scientific Articles for Authorship Assistance: A Summarization Approach

Jan Wira Gotama Putra & Masayu Leylia Khodra*

Department of Computer Science, School of Electrical Engineering & Informatics,
Bandung Institute of Technology, Jalan Ganesa No.10, Bandung 40132, Indonesia

*E-mail: masayu@stei.itb.ac.id

Abstract. This paper presents a study on automatic title generation for scientific articles considering sentence information types known as rhetorical categories. A title can be seen as a high-compression summary of a document. A rhetorical category is an information type conveyed by the author of a text for each textual unit, for example: background, method, or result of the research. The experiment in this study focused on extracting the research purpose and research method information for inclusion in a computer-generated title. Sentences are classified into rhetorical categories, after which these sentences are filtered using three methods. Three title candidates whose contents reflect the filtered sentences are then generated using a template-based or an adaptive K-nearest neighbor approach. The experiment was conducted using two different dataset domains: computational linguistics and chemistry. Our study obtained a 0.109-0.255 F1-measure score on average for computer-generated titles compared to original titles. In a human evaluation the automatically generated titles were deemed ‘relatively acceptable’ in the computational linguistics domain and ‘not acceptable’ in the chemistry domain. It can be concluded that rhetorical categories have unexplored potential to improve the performance of summarization tasks in general.

Keywords: *adaptive K-nearest neighbor(AKNN); chemistry domain; computational linguistics domain; rhetorical categories; scientific article; summarization; title generation.*

1 Introduction

The literature review is a key research activity, where researchers evaluate publications based on their relevance to the research topic. As there are many scientific articles available, the title of a scientific article is important in two ways. Firstly, researchers judge the relevance of an article promptly by its title instead of reading the whole document [1-2]. Secondly, the quality of the article’s title affects the number of prospective readers, hence affecting the number of citations [3-5]. For these reasons, writing a good title is crucial for researchers; yet, some spend only little time on it [2]. This results in the

Received December 9th, 2016, 1st Revision July 17th, 2017, 2nd Revision August 15th, 2017, Accepted for publication August, 31st, 2017.

Copyright © 2017 Published by ITB Journal Publisher, ISSN: 2337-5787, DOI: 10.5614/itbj.ict.res.appl.2017.11.3.3

conception of uninformative titles that do not reflect the overall content of the scientific articles.

Title generation can be approached from a text summarization perspective, where it is considered as compressing a scientific article to reflect its content [1,3,6-9]. The key challenge of title generation is the sparseness of the information. Given a text containing many terms, a short and concise summary must be produced, conveying the overall information of the text using only a few terms. Hence, this task cannot be considered easy.

The title of scientific articles generally reveals the purpose and method of the research, e.g. '*Scientific Paper Title Validity Checker Utilizing Vector Space Model and Topics Model*'. Hence, detecting the information type of each textual unit is important to produce a good title. Rhetorical categories denote the information type/communicative purpose of textual units as conveyed by the author to the reader, e.g. research background, proposed method, or experimental result [9]. Rhetorical categories can be used to evaluate the importance of each textual unit during the summarization process, which filters out the less useful information to get good coverage and saliency of the produced summary [10], hence decreasing the number of irrelevant textual units to be considered for the final summary (overcoming sparseness).

Automatic text summarization that considers rhetorical sentence categories has been done by Contractor, *et al.* in [11] to generate paper abstracts. However, to the best of our knowledge, the present study is the first to incorporate information type in automatic title generation. While automatic title generation cannot replace the author's expertise in conceiving a title, it is helpful by suggesting a title. The present study wants to support novice authors in creating a title for their scientific articles and make sure they do not miss important information that should be present in the title.

In this paper, a study is presented on automatic title generation for scientific articles considering rhetorical categories, i.e. information types of textual units. Sentences, i.e. the textual units that are analyzed, are classified into one of three rhetorical categories: AIM (research goal), OWN_MTHD (research method), and NR (not relevant) [12]. The proposed system generates several title candidates based on processing a paper's abstract. The abstract is used because it is relatively short, has sufficient information to represent the research idea, and can be easily obtained [13]. The final automatic summary, in the form of a title, is aimed to be as close to a human-written title (gold standard) as possible. Our contribution is the experiment on incorporating rhetorical sentence categories for automatic scientific article title generation.

The rest of this paper is organized as follows: Section 2 contains related works on rhetorical sentence classification and automatic title generation. Section 3 provides an explanation of the proposed method. Section 4 presents the title generation experiment and its result. Finally, Section 5 contains the conclusion of the paper.

2 Related Works

2.1 Rhetorical Sentence Classification

Previous researchers have done sentence classification as supervised learning task [12,14-17]. In this study, of particular interest was rhetorical sentence classification where sentences are classified according to their information type/communicative purpose to judge their importance [12,18].

Sentences with different rhetorical categories may be treated differently during summarization, especially during the information selection process [15]. Fifteen different rhetorical categories have been applied to computational linguistics and the chemistry domain in [10]. Seaghdha and Teufel [19] argued that words and linguistic forms in scientific writing are not unique to the research topic, while the writing structure can have different inter-domains but a similar intra-domain. This means that the text structure of scientific articles from the same domain can be different with that from another domain and hence capturing the writing patterns of each domain is useful for building a domain-specific classifier [12].

Table 1 Rhetorical categories [12].

Category	Description
AIM	Statement of the specific research goal or hypothesis of the paper
OWN_MTHD	New knowledge claim or proposed method
NR	Other information that does not belong to AIM or OWN_MTHD

Fifteen different rhetorical categories were heuristically tailored into three categories for the purpose of title generation [12], as shown in Table 1. This annotation scheme was used to build sentence classifiers separately for computational linguistics and chemistry scientific articles. As mentioned in [19], writing patterns exist in scientific articles, so rhetorical classification was approached as a sequence-labeling task, employing the C4.5 decision-tree learning algorithm [12]. The classification models had an F1-measure score of around 0.70-0.79 and tended to be overfitted to the most common writing patterns in the training set [12]. In the current work, the model proposed in [12] was used to label sentences.

2.2 Automatic Title Generation

Several existing title generation studies followed the pipeline summarization approach, such as [6,7,20]. They addressed the title generation problem as a summarization task following general summarization processes: preprocessing, information selection, and summary generation.

Other researches were focused on extracting important terms to be included in the title (the title as a sequence of terms) [6-7]. On the other hand, Chen and Lee [20] proposed to adapt the title of another article by using the terms from a given input document. These studies judged the importance of textual units based on their statistical properties without taking into account the information type/communicative purpose. While these methods can generate titles of good quality to some extent, we argue that incorporating the information type is indispensable considering the purpose of title writing.

3 Proposed Method

In this study, title generation is addressed as a summarization task. Our proposed system architecture consists of three modules executing the primary pipeline summarization processes: pre-processing, information selection, and title (summary) generation. The proposed architecture is shown in Figure 1 **Error! Reference source not found.**. The following subsections describe each module.

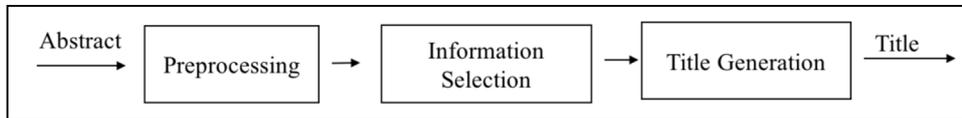


Figure 1 Proposed architecture.

3.1 Pre-processing

The abstract is used as input because it is sufficient to represent the research idea in a short manner [12]. In the pre-processing module, several steps are involved: sentence splitting, tokenization, POS tagging, and stop word removal. Stanford CoreNLP is used for pre-processing [21].

3.2 Information Selection

The rhetorical categories as shown in Table 1 (AIM, OWN_MTHD, and NR) were used to evaluate the importance of each sentence during the information selection process. The AIM and OWN_MTHD categories are considered relevant, while the NR rhetorical category is not. The focus is on extracting the

research goal and research method information and then squashing them to form a title (summary), following the idea of [12]. A manually annotated example for Figure 2 can be seen in Table 2.

Analysis of Japanese Compound Nouns by Direct Text Scanning	
s_1	[This paper aims to analyze word dependency structure in compound nouns appearing in Japanese newspaper articles].
s_2	[The analysis is a difficult problem because such compound nouns can be quite long, have no word boundaries between contained nouns, and often contain unregistered words such as abbreviations].
s_3	[The non-segmentation property and unregistered words cause initial segmentation errors which result in erroneous analysis].
s_4	[This paper presents a corpus-based approach which scans a corpus with a set of pattern matchers and gathers co-occurrence examples to analyze compound nouns].
s_5	[It employs boot-strapping search to cope with unregistered words: if an unregistered word is found in the process of searching the examples, it is recorded and invokes additional searches to gather the examples containing it].
s_6	[This makes it possible to correct initial over-segmentation errors, and leads to higher accuracy].
s_7	[The accuracy of the method is evaluated using the compound nouns of length 5, 6, 7, and 8].
s_8	[A baseline is also introduced and compared].

Figure 2 An example of a paper's abstract.

Table 2 Annotated abstract of Figure 2.

Original Title	Analysis of Japanese Compound Nouns by Direct Text Scanning
S1: AIM	This paper aims to analyze word dependency structure in compound nouns appearing in Japanese newspaper articles.
S2: NR	The analysis is a difficult problem because such compound nouns can be quite long, have no word boundaries between contained nouns, and often contain unregistered words such as abbreviations.
S3: NR	The non-segmentation property and unregistered words cause initial segmentation errors which result in erroneous analysis.
S4: OWN_MTHD	This paper presents a corpus-based approach which scans a corpus with a set of pattern matchers and gathers co-occurrence examples to analyze compound nouns.
S5: OWN_MTHD	It employs boot-strapping search to cope with unregistered words: if an unregistered word is found in the process of searching the examples, it is recorded and invokes additional searches to gather the examples containing it.
S6: NR	This makes it possible to correct initial over-segmentation errors, and leads to higher accuracy.
S7: OWN_MTHD	The accuracy of the method is evaluated using the compound nouns of length 5, 6, 7, and 8.
S8: OWN_MTHD	A baseline is also introduced and compared.

The proposed method utilizes the C4.5 (also known as J48) sentence classification model produced in [12]. After automatic classification, sentences are filtered using one of the three following configurations:

1. *Delete the non-relevant.* This configuration omits NR sentences to satisfy the heuristic that a title contains AIM and OWN_MTHD information.

2. *Retain the relevant.* This configuration omits NR sentences and only leaves one OWN_MTHD sentence, which is the most relevant to the first AIM sentence. Relevance is measured using the total number of overlapping terms. If there is no AIM sentence at all, then the first OWN_MTHD sentence appearing in the abstract is extracted. The heuristic rationale for this configuration is that the sentence providing the most general information regarding the method of research is assumed to be the first OWN_MTHD sentence in the abstract.
3. *Retain all.* This configuration does not do anything, it keeps all sentences intact.

3.3 Title Generation

After filtering the sentences using one of the three configurations mentioned in the previous subsection several sentences are left over. Terms appearing in the computer-generated title are taken from these sentences. A template-based [22] and an adaptive K-nearest neighbor (AKNN) approach were used [20].

3.3.1 Template-based Approach

The template-based approach generates titles using a number of predefined templates. POS tagging task was performed on titles of papers from our previous dataset [12] to create 50 clusters of title patterns based on POS tag patterns. The resulting patterns were generalized by manually merging the clusters into two clusters and producing two title templates in the form of a regex as follows (the regex element is a POS tag).

```

Template 0 (T0) = DT? (JJ+)?Noun+ (VBG|VBN|TO|IN) DT?(JJ+)? Noun+
Template 2 (T1) = (VBG|VBN)? DT?(JJ+)? Noun+ IN Noun+
*Noun          = (NN|NNP|NNPS|NNS)

```

These templates are expected to realize a title in the following forms:

1. <research task><utilization phrase><method phrase>, or
2. <utilization phrase>of <method phrase>in<research task>

Each term is weighed using the TF method and an N-gram (bigram) model is created based on the filtered sentences. Phrases are created, which are the longest sequences of terms with the same POS tag based on the bigram model. The proposed system then generates a title based on the algorithm in Figure 3. In the template-based approach, the length is limited to 10 terms for ensuring that the generated title is not too long, following a heuristic for good titles [1,4,5,11]. There are cases in which the proposed system cannot generate a title due to these constraints (length and pattern).

```

Foreach  $t[i]$  in template do begin
1. Choose phrase  $f[x]$  which satisfies following constraints:
  a. Has the highest TF summation from its consisting terms with respect to the
     POS tag  $t[i]$ .
  b. The probability of occurrence of the first term in  $f[x]$  given the last term of  $f[x-1]$ 
     should be more than 0 (based on bigram).
2. If there is no suitable  $f[x]$  candidate, then backtrack to replace  $f[x-1]$  with the next
   highest sum of TF values with respect to POS tag  $t[i-1]$  and  $f[x-2]$ .
3. If the first and second rule cannot be satisfied then terminate.
4. Increment  $i$  by 1.

```

Figure 3 Pseudo code template-based title generation.

At first glance, this algorithm may look similar to complete search, i.e. finding all possible combinations to satisfy constraints so that the worst case gives a complexity measurement in factorial time. However, our algorithm is bounded by the template's length being equal to 7 (T0). The algorithm involves two main operations for each iteration step: (1) finding the phrase with the highest TF summation value with respect to the POS tag; (2) checking whether the phrase satisfies the constraint with the previous phrase using bigram lookup. In the worst case every phrase consists of only one term, so there are n possible phrases for each POS tag element in the template. Thus, the complexity becomes $O(n^7)$.

In reality, a phrase usually consists of two terms on average, an abstract contains a variety of terms, backtracks rarely occur, and one term can only be succeeded by particular terms. Even if backtrack occurs many times, most probably the program will terminate because it cannot satisfy the bigram probability constraint between phrases. By this rationale, the expectation of average running complexity equals $O(7n) = O(n)$ at a cost of bigram lookup for seven times. In short, the algorithm can be regarded as a greedy approach with backtracking permitted, bounded by the length of the template.

3.3.2 Adaptive K-nearest Neighbor Approach (AKNN)

For AKNN, Chen and Lee's approach [20] was adapted. The most similar abstract in the AKNN corpus with respect to the input abstract is selected, where its similarity is measured as the summation of TF weight multiplication of overlapping terms between abstracts. The title of the most similar instance in the corpus is selected as the template. The template's nouns (NN, NNP, NNS, NNPS) and verbs (VB, VBD, VBG) are adapted by taking phrases from sentences left by information filtering. Phrases with the highest TF summation value of their terms are prioritized. In this approach, no title length constraint is present.

This algorithm has complexity $O(l^2m)$ to find the nearest neighbors, where l equals the length of the longest sentence in the input abstract and m equals the number of instances in the AKNN corpus. The adaptation process is fast as only phrases with the highest TF summation values are adapted, resulting in $O(Xn)$ complexity if every phrase only consists of one word, where n is the length of the abstract and X is the longest title length in the corpus, which we can discard as a constant. Therefore, the complexity of AKNN is $O(l^2m) + O(n) = O(l^2m + n)$, which is expected to be slower than the template-based approach.

4 Experiment

4.1 Experimental Setting

Our dataset was sourced from two domains: computational linguistics (CL) and chemistry (GaN). 250 randomly selected, rhetorically un-annotated LREC 2014 papers and 250 un-annotated GaN papers [13] were used as our test set to evaluate the proposed title generation system's performance.

The C4.5 model from [12] was used as the sentence classifier to categorize each sentence into its rhetorical category. Also, previous training data [12] were used as the AKNN corpus. To the best of our knowledge, there is no similar study to automatically generate titles of scientific articles, so there was no competitive method to compare with. As mentioned in Section 3, three configurations to filter classified sentences are used: *delete the non-relevant*, *retain the relevant*, or *retain all*. To ensure that the computer-generated titles are as close as possible to human-written titles, the computer-generated titles were compared with the original titles using F1-measure as performance measure, which was computed as in Eq. (1).

Precision

$$= \frac{\# \text{ of overlapping terms between computer generated and original title}}{\text{length of computer generated title}}$$

Recall

$$= \frac{\# \text{ of overlapping terms between computer generated and original title}}{\text{length of original title}}$$

$$F1 - \text{measure} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (1)$$

Using F1-measure instead of BLEU as performance measure means that we ignore the order of the terms. Because an AKNN-based approach is used, taking into account the term ordering will result in a relatively low BLEU score, thus providing uninformative analysis. Instead, the focus is on extracting the phrases that should be present in the title. For this reason, F1-measure is more informative for the present research to see whether phrases about the stated research goal or research method of the paper appear in the computer-generated title (analogous with considering ROUGE-1).

Also, the generated titles were evaluated by a survey among fourth-year undergraduate students (questionnaire). A three-set questionnaire was made for each domain (10 questions were picked using stratified sampling for each set) to get a total of 30 questions for each domain.

Each question consisted of an abstract and computer-generated title pair (T0, T1, and AKNN), where respondents had to choose one out of three ordinal scales to judge the quality of each generated title: 1 ('not relevant to abstract/unreadable'), 2 ('not sure'), or 3 ('relevant to abstract/readable'). The assumption is that the original human-written title must be readable (score = 3) and relevant (score = 3), therefore there was no need to evaluate the original title in the questionnaire.

4.2 Result

Details of the proposed system's performance can be seen in Table 3. In general, the proposed method performed better on the CL dataset than on the GaN dataset as the average value of the template-based approach for the GaN domain was around 0.109-0.192 while yielding more than 0.200 for the CL domain.

The samples used were investigated before including them in the questionnaire and it was discovered that the CL dataset tended to contain many repetitive terms in the abstract. On the other hand, repetition of terms happened relatively less in the GaN dataset, even for terms appearing in human-written titles. This suggests that the TF weight has great influence on the selected terms appearing in the title since phrases with TF weight summation included were prioritized. This resulted in a more noticeable effect of using rhetorical categories in the GaN domain than in the CL domain.

We intend to use the full paper text instead of only the abstract in a future study. As an abstract is short, it is reasonable that the TF weight of the terms has a great influence. However, a greater sparseness problem will arise. Further

investigation also needs to be done to check empirically which information is conveyed by titles of papers to refine the heuristic for each domain of the dataset. Although rhetorical categories were used, the current phrase selection method using TF summation of its consisting terms makes a non-noticeable difference in the CL domain compared to without using the rhetorical categories. The phrase generation and selection method needs to be refined to capture the ‘real’ important terms from the filtered sentences.

Table 3 Title generation experiment.

Configu- ration	Value	CL Domain			GaN Domain		
		T0	T1	AKNN	T0	T1	AKNN
Delete non- relevant	Avg. F1- measure	0.209	0.203	0.243	0.124	0.169	0.245
	Max F1- measure	0.757	0.690	0.627	0.625	0.652	0.668
	Realization percentage	0.792	0.952	1.000	0.848	0.988	1.000
Retain the relevant	Avg. F1- measure	0.212	0.202	0.243	0.131	0.192	0.251
	Max F1- measure	0.757	0.727	0.667	0.625	0.533	0.694
	Realization percentage	0.784	0.952	1.000	0.824	0.992	1.000
Retain all	Avg. F1- measure	0.215	0.205	0.255	0.109	0.149	0.231
	Max F1- measure	0.833	0.769	0.625	0.625	0.706	0.668
	Realization percentage	0.852	0.996	1.000	0.880	0.976	1.000

In general, the AKNN method performed best while the manual title templates were not consistent across different domains (T0 was better for the CL dataset while T1 was better for the GaN dataset). The filtering configuration also differs across the dataset, as a consequence of which it is better to have less OWN_MTHD sentences in the GaN dataset since *retain the relevant* configuration performed best.

A survey was held among 17, 15, 15 fourth-year undergraduate computer science students for each questionnaire set on the CL dataset and 11, 4, 4 chemistry students for each questionnaire set on the GaN dataset. Table 4 shows the questionnaires result, which suggests that our proposed method in general generated better quality of computer-generated titles in the CL domain than in the GaN domain. This is consistent with the performance values in Table 3. It is

reasonable that the AKNN approach receives the highest readability judgment score since it is relatively more syntactically well-formed being directly adapted from human-written titles.

Table 4 Questionnaires average score (ordinal scales).

Parameter/Configuration	CL Domain			GaN Domain		
	T0	T1	AKNN	T0	T1	AKNN
Average relevance (1-3)	2.070	1.699	2.237	1.792	1.595	2.205
Average readability (1-3)	2.302	1.964	2.405	1.600	1.448	1.684

Table 5 shows the Cronbach-alpha measurement for reliability, in this case depicting the questionnaire's consistency/whole agreement. The value ranged from 0-1. A questionnaire result is acceptably consistent if the reliability value is more than 0.70. The result in Table 5 shows relatively satisfying reliability.

Table 5 Average questionnaires whole agreement.

Parameter/Configuration	CS Domain			GaN Domain		
	CL	T1	AKNN	T0	T1	AKNN
Relevance	0.559	0.800	0.781	0.825	0.799	0.938
Readability	0.595	0.697	0.671	0.865	0.752	0.952

Abstract: We construct a large corpus of Japanese predicate phrases for synonym-antonym relations. The corpus consists of 7,278 pairs of predicates such as "receive-permission (ACC)" vs. "obtain-permission (ACC)", in which each predicate pair is accompanied by a noun phrase and case information. The relations are categorized as synonyms, entailment, antonyms, or unrelated. Antonyms are further categorized into three different classes depending on their aspect of oppositeness. Using the data as a training corpus, we conduct the supervised binary classification of synonymous predicates based on linguistically-motivated features. Combining features that are characteristic of synonymous predicates with those that are characteristic of antonymous predicates, we succeed in automatically identifying synonymous predicates at the high F-score of 0.92, a 0.4 improvement over the baseline method of using the Japanese WordNet. The results of an experiment confirm that the quality of the corpus is high enough to achieve automatic classification. To the best of our knowledge, this is the first and the largest publicly available corpus of Japanese predicate phrases for synonym-antonym relations.

Machine Generated Title:

1. [T0] A Japanese predicate phrases to a synonym-antonym relations
2. [T1] Accompanied an available corpus in predicate pair
3. [AKNN] Corpus for Japanese Predicate Phrases
4. [Original Title] Constructing a Corpus of Japanese Predicate Phrases for Synonym/Antonym Relations

Figure 4 Example of computer-generated titles.

It was found that the generated titles with a lower F1-measure score did not always get a lower questionnaire score. This suggests two things: (1) employing rhetorical sentence categories does have potential in improving the quality of computer-generated titles when the key terms can be captured effectively, despite the produced titles being relatively dissimilar to the original ones, and (2) all terms appearing in the title of a paper were not directly taken from the paper's body as they were. To provide an illustration of the quality of the computer-generated titles, abstract-generated titles are provided in Figure 4.

We suggest introducing a post-processing step to refine the morphology of the terms in the computer-generated titles. This is useful to satisfy the terms of feature agreement (analogous to augmented grammar) to make the title more readable (human-like). Another strategy is to use an abstractive summarization approach.

5 Conclusion

In this paper, a novel scientific article title generation method was introduced that considers the author's intentions (represented as the communicative purposes/information types of sentences). The key challenge to title generation is sparseness. Given a document with many optional terms, a short and concise title must be produced. The important terms denoting the research goal and research method from the scientific article's abstract were considered to appear in the artificially generated title. Two title generation approaches were used: a template-based and an adaptive K-nearest neighbor approach.

Computer-generated titles were compared with the original human-written titles and the experimental result showed that the computer-generated titles obtained a 0.109-0.255 F1-measure on average when compared to the original human-written titles. Generally, the adaptive K-nearest neighbor based title generation approach produced the best result in the experiment, both regarding F1-measure, human judgment as well as scalability. Human judgment obtained through a survey showed that the computer-generated titles were somewhat acceptable in the computational linguistics domain while low in quality for the chemistry domain.

Articles' titles from different domains can have different communicative purposes. We are aware that this study tends to over-generalize this part. There is still work left in the future to empirically check the underlying heuristic assumption in this study (the title contains research goal and method information). As our study only used the abstracts of scientific articles, it is reasonable that the TF weight of the terms greatly affects the generated phrase and title. While it is true that the abstract reflects the article in a short manner,

authors may have written the abstract in a hurry. Therefore, we suggest using the full paper text and introducing more refined weighing, phrase generation, and selection to better capture the salient phrases in sentences of particular categories. However, this could increase sparseness problem. Convolutional neural network is probably a good solution considering the nature of the task. Another work left in the future is to analyze the execution time of our method towards discussing its efficiency with many data. This is important in order to automate the title generation process.

Rhetorical sentence categories are considered to have potential to improve the quality of automatic title generation when used effectively. For future study, we propose to consider rhetorical sentence classification as a multi-label classification (one sentence has several communicative purposes). This would be useful in the case of compound/complex sentences. Another possible way is to annotate rhetorical categories at the level of clauses or phrases instead of sentences to capture the salient phrases as well as overcoming the sparseness problem.

Acknowledgements

We would like to give the greatest thankfulness to our respondents for their time helping us.

References

- [1] Jamali, H.R. & Nikzad, M., *Article Title Type and Its Relation with the Number of Downloads and Citations*, *Scientometrics*, **88** (2), pp. 653-661, 2011.
- [2] Xu, H., Martin, E. & Mihidadia, A., *Extractive Summarization Based on Keyword Profile and Language Model*, In Proceedings of North American Chapter of the ACL - Human Language Technologies (HLT), pp. 123-132, 2015.
- [3] Pavia, C.E., da Silveira Nogueira Lima, J.P. & Paiva, B.S.R., *Articles with Short Titles Describing the Results are Cited More Often*. *CLINICS*, **65**(6), pp. 509-513, 2012.
- [4] Tolga, A., *Selection of Authors, Titles and Writing a Manuscript Abstract*, *Turkish Journal of Urology*, **39**(1), pp. 5-7, 2013.
- [5] Letchford, A., Moat, H.S. & Preis, T., *The Advantage of Short Paper Titles*, Royal Society Open Science, 2015.
- [6] Jin, R. & Hauptmann, A.G., *Automatic Title Generation for Spoken Broadcast News*, In Proceedings of North American Chapter of the ACL - Human Language Technologies (HLT), pp.1-3, 2001.
- [7] Kong, S-Y., Wang, C-C., Kuo, K-C. & Lee, L-S., *Automatic Title Generation for Chinese Spoken Documents with A Delicate Scored*

- Viterbi Algorithm*, In Spoken Language Technology (SLT) Workshop, pp. 165-168, 2008.
- [8] Colmenares, C.A., Litvak, M., Matrach, A. & Silvestry, F., *HEADS: Headline Generation as Sequence Prediction Using an Abstract Feature-Rich Space*, In Proceedings of North American Chapter of the ACL – Human Language Technologies (HLT), pp. 133-142, 2015.
- [9] Teufel, S., *Argumentative Zoning: Information Extraction from Scientific Text*, PhD Thesis, Edinburgh: University of Edinburgh, 1999.
- [10] Teufel, S., Siddharta, A. & Batchelor, C., *Towards Discipline-Independent Argumentative Zoning: Evidence From Chemistry and Computational Linguistics*, In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.1493-1502, 2009.
- [11] Contractor, D., Fan, G.Y. & Koheren, A., *Using Argumentative Zones for Extractive Summarization of Scientific Articles*, In Proceedings of Computational Linguistics (COLING), pp. 663-678, 2012.
- [12] Putra, J.W.G. & Khodra, M.L., *Rhetorical Sentence Classification for Automatic Title Generation in Scientific Article*. In Journal of TELKOMNIKA, **15**(2), pp. 656-664, 2017.
- [13] Putra, J.W.G. & Fujita, K., *Scientific Paper Title Validity Checker Utilizing Vector Space Model and Topics Model*, In Proceedings of Konferensi Nasional Informatika (KNIF), pp. 69-74, 2015.
- [14] Kupiec, J., Pedersen, J. & Chen, F., *A Trainable Document Summarizer*, In Proceedings of Special Interest Group in Information Retrieval, pp. 68-73, 1995.
- [15] Teufel, S. & Moens, M., *Summarizing Scientific Articles – Experiments with Relevance and Rhetorical Status*. Journal of Computational Linguistics, **28**, 4, pp. 409-445, 2002.
- [16] Wong, K-F., Wu, M. & Li, J.W., *Extractive Summarization Using Supervised and Semi-Supervised Learning*, In Proceedings of Computational Linguistics, pp. 985-992, 2008.
- [17] Widyantoro, D.H., Khodra, M.L., Riyanto, B. & Aziz, E.A., *A Multiclass-Based Classification Strategy for Rhetorical Sentence Categorization from Scientific Papers*, Journal of ICT Research and Applications, **7**(3), pp. 235-249, 2013.
- [18] Teufel, S. & Moens, M., *Discourse-Level Argumentation in Scientific Articles: Human Automatic Annotation*, In Towards Standards and Tools for Discourse Tagging – ACL 1999 Workshop, 1999.
- [19] Seaghdha, D.O. & Teufel, S., *Unsupervised Learning of Rhetorical Structure with Un-Topic Models*, In Proceedings of Computational Linguistics (COLING), pp.2-13, 2014.
- [20] Chen, S-C. & Lee, L-S., *Automatic Title Generation for Chinese Spoken Documents Using an Adaptive K-Nearest-Neighbor Approach*, In

- Proceedings European Conference of Speech Communication and Technology, pp. 2813-2816, 2003.
- [21] Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J. & McClosky, D., *The Stanford CoreNLP Natural Language Processing Toolkit*, In Proceedings of the 52nd Annual Meeting of the Association of Computational Linguistics, 2014.
- [22] Clark, A., Fox, C. & Lappin, S., *The Handbook of Computational Linguistics and Natural Language Processing*, John Wiley & Sons, Singapore, 2010.