

## Rhetorical Sentence Classification for Automatic Title Generation in Scientific Article

Jan Wira Gotama Putra<sup>\*1</sup>, Masayu Leylia Khodra<sup>2</sup>

Department of Informatics/Computer Science, School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Jalan Ganesha No. 10 Bandung 40132, Jawa Barat, Indonesia (+6222) 2502260  
Corresponding author, e-mail: wiragotama@gmail.com<sup>\*1</sup>, masayu@stei.itb.ac.id<sup>2</sup>

### Abstract

*In this paper, we proposed a work on rhetorical corpus construction and sentence classification model experiment that specifically could be incorporated in automatic paper title generation task for scientific article. Rhetorical classification is treated as sequence labeling. Rhetorical sentence classification model is useful in task which considers document's discourse structure. We performed experiments using two domains of datasets: computer science (CS dataset), and chemistry (GaN dataset). We evaluated the models using 10-fold-cross validation (0.70-0.79 weighted average F-measure) as well as on-the-run (0.30-0.36 error rate at best). We argued that our models performed best when handled using SMOTE filter for imbalanced data.*

**Keywords:** rhetorical corpus construction, rhetorical classification, automatic title generation, scientific article

Copyright © 2017 Universitas Ahmad Dahlan. All rights reserved.

### 1. Introduction

Motivated by flourishing availability of scientific articles over internet, keeping updated on relevant articles is important. Since article title represents a paper in a brief manner [1], a title is fundamental guide to quickly determine whether a paper is relevant to reader needs in literature review. Although its title is substantial for a paper [2], writers usually spend a bit of time to construct their article title despite its importance. It makes the title sometimes does not reflect what the research is about. Some recent studies showed that quality of article title influences its number of citation [3-5]. Therefore, it is very important to write a good title.

Automatic title generation is developed to help authors in considering good title. There are several automatic title generation researches have been conducted for indexing purposes [6-7]. Title generation could be considered as summarization task as a very short summary of a document [1], [8-10]. Some researches of automatic title generation have been conducted utilizing extractive summarization approach for news article [9-11]. The majority of these previous works were focused on how to extract relevant words in article for generating title, without considering discourse structure of document. Discourse actually provides information type that represents communication goal conveyed to the reader by an author of the paper [12-13]. Information type of sentence in document is argued could improve the performance of automatic title generation, as title of a document presents certain type of information [12-13]. This research considers discourse structure in form of information type conveyed by sentences. As we want to generate title as similar to human generated title, the problem is to differentiate useful and unuseful information to get nice coverage and saliency [13-15]. We can model title communication goals more accurately by taking discourse of text into consideration. Information type of sentence (known as rhetorical categories) in document is argued could be incorporated in information selection to judge the importance of sentence. Rhetorical categories, for example research aim or method, indicated the discourse structure of a document. According to our observation, title usually contains research aim and method [2-5], [13]. As an example, "Scientific Paper Title Validity Checker Utilizing Vector Space Model and Topics Model". "Validity Checker" phrase is the purpose of the research. Meanwhile "Vector Space Model and Topics Model" phrase is the method. Therefore, useful information in this case are sentences which yield information of research purpose and method. This paper is focused on work of

rhetorical sentence categorization [2-5], [13]. It can be specifically incorporated in scientific article title generation task.

Some previous similar researches had been done on sentence classification and information structure identification on abstract of scientific article [16-21]. They defined various rhetorical categories on various domains, as well as feature selection. Teufel and Moens [18] worked on rhetorical sentence classification of abstract section which yielded 0.6861 in precision and recall. Widyantoro et al [20] employed rhetorical classification using 15 categories of sentence from [13] for full paper rhetorical classification. Despite the high performance in precision, performance of their models suffered in recall [20]. We address the issue by modifying the annotation scheme, specific for title generation task. Each sentence is classified into one of three rhetorical categories, i.e. AIM (research purpose), OWN\_MTHD (research method), and NR (not relevant). Based on previous research, we assumed that abstract part of article is sufficient as input for title generation task, as it depicts the most important things in the paper [22, 23]. Therefore, we focused on extracting AIM and OWN\_MTHD information in abstract part of scientific article to satisfy title communication goals.

The main contributions of our research work are two-fold: (1) developing a corpus for automatic title generation task based on several research paper collections that has been annotated into one of the three categories, and (2) providing model for rhetorical sentence classification that could be incorporated in automatic title generation task.

The rest of the paper is organized as follows. Section 2 provides an overview of related work on annotation scheme for paper sentence rhetorical categories. The proposed method is then explained in section 3. Experiments results and analysis are presented in Section 4, followed by concluding remarks on Section 5.

## 2. Related Work

An information category that is conveyed to readers is known as rhetorical categories. These categories are usually applied to sentence unit for extracting particular sentence or argumentative zoning [12-13]. Teufel et al [13] proposed 15 different rhetorical categories as shown by Table 1. This annotation scheme has been applied to computational linguistics and chemistry domains. Séaghdha and Teufel [23] argued that words and linguistics forms in paper writing is not specific to research topic. In general, document writing structure could be different across-domains, but similar intra-domain.

Table 1. Full Annotation Scheme [13]

Category	Description
AIM	Statement of specific research goal, or hypothesis of current paper
NOV_ADV	Novelty or advantage of current paper
CO_GRO	No knowledge claim is raised (or knowledge claim not significant for the paper)
OTHR	Knowledge claim (significant for paper) held by somebody else. Neutral description
PREV_OWN	Knowledge claim (significant) held by authors in a previous paper. Neutral description
OWN_MTHD	New knowledge claim, own work: methods
OWN_FAIL	A solution/method/experiment in the paper that did not work
OWN_RES	Measureable/objective outcome of own work
OWN_CONC	Findings, conclusions (non-measureable) of work
CODI	Comparison, contrast, difference to other solution (neutral)
GAP_WEAK	Lack of solution in field, problem with other solutions
ANTI_SUPP	Clash with somebody else's results or theory, superiority of own work
SUPPORT	Other work supports current work or is supported by current work
USE	Other work is used in own work
FUT	Statements/suggestions about future work (own or general)

Abstract of a paper is a short description of the research, which differs from the full papers. An article abstract usually consists of several information types: background, previous work, research aim/problem, proposed method, experimental result, and comparison to other works. An example of writing style is shown in Figure 1. On the other hand, full paper contains more information type as researched by Teufel et al [13]. Since abstract contains research aim/problem and proposed method, we argue that abstract is sufficient for title generation task.

The Problem of \_\_\_ has received a lot of attention because of its relevance to \_\_\_\_\_. CITATION proposed an approach based on the method of \_\_\_\_\_. In this paper we present a method for \_\_\_ that has the following advantages over prior work: \_\_\_\_\_. We demonstrate the empirical effectiveness of our method by reporting experiments on \_\_\_ data, where it outperforms the approach of CITATION by \_\_\_%.

Figure 1. Abstract Writing Style in NLP Domain [23]

Widyantoro et al [20] worked on rhetorical sentence classification using annotation scheme as provided by Table 1. It adapted classification features as proposed by Teufel [13]. It employed several classifications. The results were good in terms of accuracy (all above 76%) for Multi-Homogenous Classifier for each algorithm. However, the F-measure was really low; no single classification strategy could obtain more than F1-measure 0.25. Differ from previous work, current work only address classification to capture sentences which yield AIM and OWN\_MTHD categories. Moreover, previous research worked on full-paper classification, whereas current work only addresses abstract section sentences classification.

### 3. Proposed Method

An article title usually represents the aim and specific method of research (to clearly indicate research subject) [2-5]. Title will be generated from abstract sentences that have rhetorical category of aim and own method. Therefore, we need to categorize each sentence into three rhetorical categories as shown in Table 2.

Table 2. Annotation Scheme Adaptation

Category	Description	Sentence Example
AIM	Statement of specific research goal, or hypothesis of current paper	We present an efficient multi-level chart parser that was designed for syntactic analysis of closed captions (subtitles) in a real-time Machine Translation (MT) system.
OWN_MTHD	New knowledge claim, own work: methods	In order to achieve high parsing speed, we divided an existing English grammar into multiple levels.
NR	Other information that is not belong to AIM or OWN_MTHD category	The parsing time favorable compares with a Tomita parser and a chart parser parsing time when run on the same grammar and lexicon.

AIM and OWN\_MTHD categories represent the specific research aim and method, while the information on other rhetorical categories in Table 1 is argued not important to be presented in the title. As an example, Table 3 shows annotation for an abstract of scientific paper "Analysis of Japanese Compound Nouns by Direct Text Scanning".

Supervised learning was employed to build classification model. We utilized several classification algorithms. At first, we annotated papers' abstract section sentences with corresponding rhetorical category. Then, we moved to decide the features. The features were needed in supervised learning to represent input object. The features should contain enough information to accurately predict the output. As for classification features for sentence classification, we examined and judged that several features used in previous researches then adapted them specifically for abstract sentences classification purpose [21, 24, 25]. The features explanation could be seen in Table 4.

Table 3. Annotation Example of An Abstract

Position: annotation	Sentence text
1: AIM	This paper aims to analyze word dependency structure in compound nouns appearing in Japanese newspaper articles.
2: 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.
3: NR	The non-segmentation property and unregistered words cause initial segmentation errors which result in erroneous analysis.
4: 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.
5: 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.
6: NR	This makes it possible to correct initial over-segmentation errors, and leads to higher accuracy.
7: OWN_MTHD	The accuracy of the method is evaluated using the compound nouns of length 5, 6, 7, and 8.
8: OWN_MTHD	A baseline is also introduced and compared.

Table 4. Classification Features

Group Name	Feature Name	Description	Value
Position Feature	Position	Location of sentence in abstract per abstract length. Adaptation of explicit structure and location feature [24], and Contractor et al. [21].	Numeric {real value}
	AIM lexicon	Consist of formulaic lexicon that are relevant to research aim. Adaptation of semantic feature [24].	
	OWN_MTHD lexicon	Consist of formulaic lexicon that are relevant to research method. Adaptation of semantic feature [24].	Yes or No
Content- Semantic Feature	NR lexicon	Consist of formulaic lexicon that are relevant to research background, previous works, conclusion, and research comparison. Adaptation of semantic feature [24].	
	Weight Presentation	Sum of TF weight for each word per length of sentence. Adaptation of content feature and TF-IDF values [24] and Contractor et al. [21]	Numeric {real value}
Sequential Feature	Previous Rhetorical Category	Previous rhetorical category. Value "NO" indicates there is no sentence before current sentence	Nominal {NO, AIM, OWN_MTHD, NR}

On deciding these features, we analyzed the two characteristics of datasets beforehand, namely writing pattern in abstract (rhetorical pattern) and formulaic lexicon statistics (to determine best formulaic lexicon to use). We also argue that sentence classification should respect to previous rhetorical category. Abstract has certain information zones and ordering, similar as what illustrated in Figure 1. We adopted expression that is relevant to AIM, OWN\_MTHD and NR from action and concept lexicon in [24]. We also added our own analysis of relevant expression to support the lexicon. Details of employed lexicons could be seen in Table 5.

Table 5. Formulaic Lexicons, Adaptation from [25]

Lexicon	Formulaic Lexicons
AIM lexicons	aim, goal, intention, theme, belief, we present, this paper, this work, in this paper, in this work, address, we investigate, we investigated, investigate/investigates/ investigated, study, studied, approach, framework, introduce
OWN_MTHD lexicons	apply, employ, make use, utilize, modification, refine/ refinement, incorporate/incorporates/incorporating, implement/implements/implemented
NR lexicons	enhance, defeat, improve/ improves, perform better, outperform, outweigh, surpass, compare, compete, accuracy, baseline, comparison, competition, evaluation, inferiority, performance, precision, optimum, recall, superiority, accomplishment, achievement, benefit, breakthrough, improvement, proof, remedy, success, triumph, verification, observed, achieve/ achieves/ achieved, state-of-the-art, experimental, result, evaluative, superior, inferior, better, best, worst, greater, larger, faster, weaker, over, effective, quality, significant, significantly, stronger, found, find/ finds, show/ shows, previous work, previous research, to our knowledge, conclude, conclusion

#### 4. . Experiments, Results, and Analysis

Our dataset came from two domains: Computer Science (CS) and Chemistry (GaN). We used conference paper abstracts from 71 papers in Widyantoro et al [20], 217 papers in Computational Linguistic Conference (COLING) 2014, 186 papers in Human Language Technology Conference (HLT) 2015, and 176 long papers in International Joint Conference in Natural Language Processing (IJCNLP) 2015. For Chemistry dataset, we used 488 GaN paper abstracts in Putra & Fujita [22]. In total, we have 3648 sentences in CS dataset and 2755 sentences in GaN dataset. We used two different kinds of domains to test whether our method could work across domains.

We employed WEKA [26] tool for experiment, our classification algorithms consisted of WEKA implementations: Naïve Bayes, C45 (named as J48), SMO, and Simple Logistic. Naïve Bayes provides a simple probabilistic knowledge for classification. It predicts the most probable label by measuring the probability of labels for given classification features [27]. C45 produces decision tree by observing data and inducing the information gain [27]. SVM is a learning algorithm that constructs a hyperplane with maximal margin between classes [27]. Logistic Regression (Simple Logistic) assumes that probability of each label can be computed as a linear combination of features and some problem-specific parameters [27].

In case of imbalanced dataset, filters are often used. We employed two filters: SMOTE (Synthetic Minority Oversampling Technique) and Class Balancer. SMOTE is performed by creating extra data for minority data [28]. It will cause classifier to create less specific model. On the other hand, Class Balancer reweights the instances in the data so that each class has the same weight [26].

##### 4.1. Rhetorical Writing Pattern

The top 5 abstract's rhetorical writing pattern for each dataset could be seen in Table 6. For example, (AIM, NR) pattern means abstract text consisted of one or several AIM sentence(s) then followed by one or several NR sentence(s). Abstract which presented in Table 3 has "AIM, NR, OWN\_MTHD, NR, OWN\_MTHD" writing pattern. From Table 6, we actually can see "AIM, OWN\_MTHD, NR" sequence occurs in more than one pattern with some additions.

Table 6. Abstract Rhetorical Pattern

Computer Science (CS) Dataset			Chemistry Dataset (GaN) Dataset		
Pattern	Count	Percentage	Pattern	Count	Percentage
NR, AIM, OWN_MTHD, NR	159	24.54	AIM, NR	202	41.39
AIM, OWN_MTHD, NR	129	19.90	AIM, OWN-MTHD, NR	124	25.41
NR, AIM, NR	105	16.20	AIM, NR, OWN_MTHD, NR	51	10.45
AIM, NR, OWN_MTHD, NR	56	8.64	AIM, OWN_MTHD, NR,	30	6.15
AIM, NR	33	5.09	OWN_MTHD, NR		
			NR, AIM, NR	18	3.69

There are several patterns overlapping among these 2 domains, however differ in rank. It implies the writing style among CS and GaN datasets are different. Annotation result for each dataset (bigram table) could be seen in Table 7 and Table 8.

Table 7. Annotation Result, CS Dataset

PREV \ NEXT	AIM	OWN_MTHD	NR
NO	283	4	361
AIM	18	372	271
OWN_MTHD	15	417	502
NR	363	173	869
Total	679	966	2003

Table 8. Annotation Result, GaN Dataset

PREV \ NEXT	AIM	OWN_MTHD	NR
NO	434	8	46
AIM	14	171	300
OWN_MTHD	7	112	280
NR	40	115	1128
Total	495	406	1854

Certain rhetoric sentence is usually followed by certain rhetoric as well. It could be seen that the first sentence of abstracts is usually AIM or NR for both datasets. Our annotation result confirms Séaghdha and Teufel abstract's writing style [23].

## 4.2. Experiments

There are three dataset combinations in our experiment: CS+GaN, only CS, only GaN. We would like to see whether the models could be domain-independent. We built the models for each domain of dataset. Filters (SMOTE, Class Balancer) were employed for each domain of dataset to handle imbalanced issue. SMOTE filtered CS dataset has 2037 AIM instances, 1932 OWN\_MTHD instances, and 2003 NR instances. SMOTE filtered GaN dataset has 1980 AIM instances, 1624 OWN\_MTHD instances, and 1854 NR instances.

Table 9. Weighted Avg. F Measure, Models Based on CS training data (SMOTE)

Model	10-Fold-Cross Validation (SMOTE)	GaN Test Data (SMOTE)	10-Fold-Cross Validation (NO FILTER)	GaN Test Data (NO FILTER)
Naïve Bayes	0.721	0.684	0.776	0.693
J48	0.786	0.568	0.771	0.683
SMO	0.732	0.693	0.762	0.697
Simple Logistic	0.742	0.668	0.765	0.697

For testing, we used 10-fold-cross validation and test data (CS dataset-trained model tested using GaN dataset and vice versa). We found out that model with SMOTE filter performed best. Table 9 and Table 10 shows the performance of each models for CS training dataset and GaN training dataset. The best weighted average F-measure for CS and GaN dataset respectively are **0.786** and **0.797** using 10-fold-cross validation. When using no filter, F1-Measure was not quite different; the models classify more OWN\_MTHD as NR. However, the models classify more NR as OWN\_MTHD when using SMOTE filter. Since different algorithms performed much or less similar, algorithm choice does not affect performance considerably.

Table 10. Weighted Avg. F Measure, Models Based on GaN Training data (SMOTE)

Model	10-Fold-Cross Validation (SMOTE)	CS Test Data (SMOTE)	10-Fold-Cross Validation (NO FILTER)	GaN Test Data (NO FILTER)
Naïve Bayes	0.767	0.595	0.803	0.541
J48	0.797	0.620	0.779	0.457
SMO	0.757	0.576	0.749	0.440
Simple Logistic	0.762	0.655	0.777	0.447

We also tested the previous models performance on-the-run to evaluate the error propagation. We found out that the model performed very poor when tested on-the-run with another domain dataset, but much or less have similar for same domain dataset. We argue that poor on-the-run performance when tested using different domain of dataset is affected by different general writing style on both datasets (see Table 6). This result is very logical, since writing patterns can vary across datasets. In CS dataset, the most important feature is AIM lexicon, followed by previous rhetoric category, and position. On the other hand, for GaN dataset, the most important feature is position, followed by previous rhetoric category.

We also tested models performance when using both domains of dataset as training set, and combined with SMOTE filter as well. Models performed better when the dataset filtered using SMOTE. Result could be seen in Table 11. Results show that the classifier will perform better when the models were constructed separately for each domain dataset.

Table 11. Weighted Avg. F Measure, Models Based on CS+GaN Training Set

Model	10-Fold-Cross Validation	CS Dataset as Test Set	GaN Dataset as Test Set
Naïve Bayes	0.739	0.668	0.753
J48	0.760	0.715	0.761
SMO	0.708	0.576	0.735
Simple Logistic	0.735	0.683	0.722

On the other hand, we also tested whether number of instances in dataset would have significant influence in the performance of the models on-the-run. In this case, we experimented using J48 model (dataset filtered using SMOTE) with CS dataset and tested on the run on tree iterations of incrementing number of instances. Result could be seen in Table 12. Our experiment shows the increment in number of dataset would not affect the performance significantly.

Table 12. Performance on Test Data On-the-Run

Traning Data	Testing Data (Target)	Misclassified Sentences
71 Dwi	217 COLING	0.284
71 Dwi + 217 COLING	186 HLT	0.365
71 Dwi + 217 COLING + 186 HLT	174 IJCNLP	0.316

We took some misclassified sentences example from annotating CS dataset on-the-run by J48 which is constructed using CS dataset (filtered by SMOTE) in Table 13. Misclassified sentences usually occur when an abstract has **rare pattern**.

Table 13. Misclassified Sentences Example

No	Sentence	Correct Label	Classified As
1	We present here a method of calculating person name match probability using a language model derived from a directory of legal professionals.	AIM	OWN_ MTHD
2	In this research, we developed a support system for revising titles.	AIM	NR
3	To achieve these goals, we combine two supervised machine learning paradigms, online and multitask learning, adapting and unifying them in a single framework.	OWN_ MTHD	AIM
4	In order to achieve high parsing speed, we divided an existing english grammar into multiple levels.	OWN_ MTHD	NR
5	Our contribution is a large-scale user study with 121 participants using the netspeak search engine to shed light on this issue for the first time.	NR	AIM
6	The final corpus and the annotation guidelines are freely available to encourage future research in argument recognition.	NR	OWN_ MTHD

For example, first sentence is misclassified since OWN\_MTHD is not usually followed by AIM. In second example, NR sentence is usually followed by NR in the beginning if there is no AIM lexicon detected. Third sentence is misclassified since NR sentence is more likely followed by AIM sentence at the beginning of abstract. These three misclassifications results are very logical according to the statistics presented in Table 6. In fourth example, it is misclassified due to precence of NR lexicon in the beginning of abstract. In fifth example, it is classified as AIM due to emergence of AIM lexicon. As for the last example, it is classified as OWN\_MTHD because the effect weight presentation, and does not contain any formulaic lexicons. Based on these facts, we argue that our models are overfitting to the most general writing patterns of the dataset. Other variations of writing patterns aside of the incommons contributed significantly to the 0.30~0.36 error rate when on-the-run.

As opposed to recent most similar research [20], despite cannot be compared straightforwardly, our models performed well and stable. Widyantoro, et al [20] obtained high accuracy, but suffered in recall. It makes the previous model impractical to be used. On the other hand, our models were stable in both terms. Learning from Widyantoro et al, the 15 categories annotation scheme as proposed by Teufel et al [13] may need to be improved to produce satisfiable classifier. One way is to adjust the scheme, specific to the task (like what we did). We also find that the annotation scheme may not be applicable for complex sentence. For example, "we present here a method of calculating person name match probability using a language model derived from a directory of legal professionals" contains research purpose and research method information. This suggest providing multi-label annotation scheme, possibly improve classifier performance.

## 5. Conclusion

This work introduces novel rhetorical corpus and classification models that specifically could be incorporated in automatic title generation task. A title usually contains information that is specific to the research, especially aim and method of the research. In our method, we develop rhetorical classification as sequence labeling with three categories: AIM, OWN\_MTHD, and NR. Our models have much or less reasonable performance compared to 10-fold-cross validation when tested using same datasets as training set on the run (0.30-0.36 error rate at best). We argue the models tend to fit the most common pattern(s) of datasets. We also performed experiments to analyze whether increasing the number of datasets will result in better performance. Our models performed better when they were constructed independently for each dataset, than domain-mixed dataset. It means rhetorical classification may not be applicable as domain-independent classification using our proposed features.

Further research needs to be done to find richer domain independent features. As for near future, we will investigate whether applying sentence rhetorical classification for automatic title generation task can produce good result, and how this strategy should be applied in the task.

## References

- [1] H Xu, E Martin, A Mahidadia. *Extractive Summarization Based on Keyword Profile and Language Model*. The 2015 Annual Conference of the North American Chapter of ACL. Denver, USA. 2015; 123-132.
- [2] HR Jamali, M Nikzad. Article Title Type and Its Relation with the Number of Downloads and Citations. *Scientometrics*. 2011; 88(2): 653-661.
- [3] Carlos EP, Joao Paulo da SNL, Bianca Sakamoto RP. Articles with Short Titles Describing the Results Are Cited More Often. *CLINICS*. 2012; 67(5): 509-513.
- [4] A Letchford, H Susannah M, T Preis. The Advantage of Short Paper Titles. *R. Soc. Open Sci*. 2015; 2.
- [5] T Akman. Selection of Authors, Titles and Writing a Manuscript Abstract. *Turkish Journal of Urology*. 2013; 39 (Supplement 1): 5-7.
- [6] C Lee Giles, Yves Petinot, Pradeep BT, Hui Han, Steve Lawrence. *eBizSearch; A Niche Search Engine for e-Business*. 26<sup>th</sup> ACM Annual SIGIR Conference on Research and Development in Information Retrieval. 2003; 413-414.
- [7] C Lopez, V Prince, M Roche. *NOMIT: Automatic Titling by Nominalizing*. Human Language Technology (The 2012 Annual Conference of the North American Chapter of ACL). Montreal, Canada. 2012; 274-283.
- [8] CA Colmenares, M Litvak, A Matrach, F Silvestri. *HEADS: Headline Generation as Sequence Prediction Using an Abstract Feature-Rich Space*. The 2015 Annual Conference of the North American Chapter of ACL. Denver. 2015; 133-142.
- [9] K Sheng-yi, W Chien-chi, K Ko-chien, L Lin-shan. *Automatic Title Generation for Chinese Spoken Documents with A Delicately Scored Viterbi Algorithm*. Spoken Language Technology Workshop (SLT). Goa, India. 2008; 165-168.
- [10] R Jin, A Hauptmann. *Automatic Title Generation for Spoken Broadcast News*. Human Language Technology. San Diego, USA. 2001; 1-3.
- [11] SC Chen, LS Lee. *Automatic Title Generation for Chinese Spoken Documents Using an Adaptive K-Nearest-Neighbor Approach*. European Conference of Speech Communication and Technology. Geneva, Switzerland. 2003; 2813-2816.
- [12] S Teufel, M Moens. Summarizing Scientific Articles—Experiments with Relevance and Rhetorical Status. *Journal of Computational Linguistics*. 2002; 28(4): 409-445.
- [13] S Teufel, A Siddhartan, C Batchelor. *Towards Discipline Independent Argumentative Zoning: Evidence from Chemistry and Computational Linguistics*. Conference on Empirical Methods in Natural Language Processing—09. Singapore. 2009; 1493-1502.
- [14] A Wahib, AZ Arifin, D Purwitasari. Improving Multi-Document Summary Method Based on Sentence Distribution. *Journal of Telecommunication Computing Electronics and Control (TELKOMNIKA)*. 2016; 14(1): 286-289.
- [15] PV Amoli, O Sojoodi Sh. Scientific Document Clustering Based on Text Summarization. *International Journal of Electrical and Computer Engineering (IJECE)*. 2015; 5(4): 782-787.
- [16] S Merity, T Murphy, JCurran. *Accurate Argumentative Zoning with Maximum Entropy Models*. ACL Workshop on Text and Citation Analysis for Scholarly Digital Library. Athens, Greece. 2009; 19-26.
- [17] K Hirohata, N Okazaki, S Ananiadou, M Izuka. *Identifying Sections in Scientific Abstracts using Conditional Random Fields*. International Joint Conference on Natural Language Processing (IJCNLP). Hyderabad, India. 2008; 381-388.

- 
- [18] S Teufel, M Moens. *Sentence Extraction and rhetorical Classification for Flexible Abstracts*. AAAI Spring Symposium on Intelligent Text Summarization. Stanford, USA. 1998; 16-25.
- [19] H Ren. A Comparison Study on the Rhetorical Moves of Abstracts in Published Research Articles and Master's Foreign-language Theses. *Journal of English Language Teaching*. 2011; 4(1): 162-166.
- [20] DH Widyantoro, ML Khodra, B Riyanto, EA Aziz. A Multiclass-based Classification Strategy for Rhetorical Sentence Categorization from Scientific Papers. *Journal of ICT Research and Applications*. 2013; 7(3): 235-249.
- [21] D Contractor, YF Guo, A Korhonen. *Using Argumentative Zones for Extractive Summarization for Scientific Articles*. International Conference on Computational Linguistics (COLING). Mumbai, India. 2012; 663-678.
- [22] JWG Putra, K Fujita. *Scientific Paper Title Validity Checker Utilizing Vector Space Model and Topics Model*. Konferensi Nasional Informatika. Bandung, Indonesia. 2005; 69-74.
- [23] DÓ Séaghdha, S Teufel. *Unsupervised Learning of Rhetorical Structure with Un-Topic Models*. International Conference on Computational Linguistics. Dublin, Ireland. 2014; 2-13.
- [24] S Teufel, M Moens. *Discourse Level Argumentation in Scientific Articles: Human and Automatic Annotation*. Towards Standards and Tools for Discourse Tagging ACL 1999 Workshop. 1999.
- [25] S Teufel. *Argumentative Zoning: Information Extraction from Scientific Text*. PhD Thesis. Edinburgh: Postgraduate University of Edinburgh; 1999.
- [26] M Hall, E Frank, G Holmes, B Pfahringer, P Reutemann, IH Witten. *The WEKA Data Mining Software: An Update*. Special Interest Group in Knowledge Discovery and Data Mining (SIGKDD) Explorations. 2009; 11(1).
- [27] CM Bishop. *Pattern Recognition and Machine Learning*. New York: Springer. 2006: 21-24, 179-220, 325-356, 663-666.
- [28] NV Chawla, KW Bowyer, LO Hall, WP Kegelmeyer. SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*. 2002; 16: 321-357.