Performance Evaluation of SVM-Based Information Extraction using τ Margin Values

By Kurnia Muludi

International Journal on Electrical Engineering and Informatics - Volume 2, Number 4, 2010

Performance Evaluation of SVM-Based Information Extraction using τ Margin Values

Kuspriyanto[1,2], Oerip S Santoso[2], Dwi H Widyantoro[2], Husni S Sastramihardja[2], Kurnia Muludi[2,3], and Siti Maimunah[2,4]

[1]Computer Engineering Research Group,
[2]School of Electrical Engineering and Informatics,
Bandung Institute of Technology, Jl. Ganeca 10 Bandung, Indonesia
[3]Soil Science Department, Agriculture Faculty - University of Lampung, Indonesia
Jl. Sumantri Brojonegoro No. 1 Bandar Lampung 35145
[4]Information System Dept., Information Tech. Faculty,
Surabaya Adhitama Institute of Technology
Jl. A.R. Hakim No.100 Surabaya, Indonesia

Abstract: The rapid growth of Internet causes the abundance of textual information. It is necessary to have smart tools and methods than can access text content as needed. One of the success methods is Support Vector Machine (SVM). This paper will discuss how the performance of the SVM-GATE algorithm on extracting information from Indonesian language corpus in response to τ margin variation. Experimental results show that there is optimum τ margin for both Indonesian corpus of Vegetable Market and Seminar Announcement Corpus. The best Performance of SVM-GATE obtained at the τ Margin of 0.5 and the Window Size of 4x4.

Keywords: Information Extraction, Support Vector Machine, Bahasa Indonesia Corpus, NLP, GATE, optimum margin.

1. Introduction

Along with the rapid Internet development, the volume of textual information is also incredibly growing. Currently Information Retrieval technology alone is not able to provide specific information needs because this technology only provides information on the level of document collection. Development tool and intelligent methods that can access the content of the document are therefore a crucial issue.

Information extraction is the process of getting information about the pre-specified events, entities or relationships in the text such as news articles (Newswire) and web pages. Many research of information extraction are focused on named entity recognition. In general information extraction task can be regarded as an entity recognition task in the text. Extraction of information is very useful in many applications such as business intelligence, automatic annotations on web pages, and knowledge management.

Extraction of information can be approached through a classification problem where the text is split into tokens and grouped into the appropriate class. *Hidden Markov Models* are a popular method for the task, but this method cannot handle multiple tokens with attribut[1]].

One of successful machine learning methods in the extraction of information is the *Support Vector Machine* (SVM), which is part of the supervised machine learning algorithms. This algorithm has achieved the performance state-of-the-art in various classification tasks, including named entity recognition [1, 2].

SVM classifier can predict where a type of tag (token classes) begins and ends in the text. Classifier is trained from a text that has been annotated. SVM classifier is used to distinguish items of one class against another class based on attributes of training examples. These attributes are called features. The simplest classification problem is to distinguish between positive and negative examples of concepts. Problems in extraction of information is how to

Received: March 8, 2010. Accepted: October 12, 2010

determine whether the text position is the beginning of a tag (token class) or not and the end of a tag or not.

In this paper we will discuss how the performance of the SVM algorithm on extracting information from Indonesian language and English corpus in response to τ margin variation and will show experimental results in detail.

The rest of this paper is organized as follows. Section 2 presents an overview of related work. Section 3 explains Support Vector Machine in general. Section 4 discusses SVM Based Information Extraction. The evaluation of the experimental result is covered in Section 5. Finally we conclude in Section 6.

A. Related Work

Text categorization is one of research areas where SVM has been implemented successfully [4, 5, 6]. In the area Information Extraction Systems, several researches have been conducted very well. Isozaki et al. [3], for examples, used SVM based- information extraction systems and trained four classifiers using sigmoid function to transfer the output of SVM into probabilities and applying Viterbi algorithm to determine the optimal sequence of labels for a sentence. The system is evaluated on the Japanese-language corpus using window size of 2. The results show that this system has better performance than systems based on Maximum Entropy and Rule Learning. This system also describes an efficient implementation for the quadratic kernel SVM. Other researcher, Mayfield et al.[7], applied SVM with lattice-based approach to the cubic kernel for calculating the lattice transition probabilities. By using window size of 3, satisfactory results are obtained like Li et al., [8].

The normal SVM treats positive and negative training examples equally, which may result in poor performance when applying the SVM to some very unbalanced classification problems due to very sparse positive example in dataset [9]. A few approaches have been proposed to adapt the SVM to classification problem with uneven dataset. One was presented in Morik et al. [10], where the cost factor for positive examples was distinguished from the cost factor for negative example to adjust the cost of false positive vs. false negative.

Li et al. [9] introduced the uneven margin parameter (τ margin) in the SVM algorithm for document categorization for small categories. Uneven margin parameter is the ratio of negative margin to a positive margin. By using this parameter, SVM can handle imbalanced data better than normal SVM model. Other researchers such as Shawe-Taylor et al., [11] have successfully implemented uneven margin SVM for document filtering while Gao and Huang [12] have a good result in implementing it for extracting acronym from plain text.

Even though SVM with uneven margin has been successfully implemented for Information Extraction, none of the above researches discuss what performance trend we will get in response to τ margin variation. Li et al. [9] use *cross-validation* method to get best margin which takes long time so it is not popular [13]. In this paper we will explore relationship between τ margin variation and SVM-GATE Information Extraction System performance so that we will have deeper understanding in setting up SVM parameter for Information Extraction.

2. SVM

Support vector machines (SVM) are based on the Structural Risk Minimization principle [14] from statistical learning theory. The idea of structural risk minimization is to find a hypothesis *h* with the lowest true error. Vapnik connects the bounds on the true error with the margin of separating hyperplanes. In their basic form, support vector machines find the hyperplane that separates the training data with maximum margin. SVM is a useful technique for data classification. A classification task usually involves with training and testing data which consist of some data instances. Each instance in the training set contains one “target value" (class labels) and several “attributes" (features).

The goal of SVM is to produce a model which predicts target value of data instances in the testing set which are given only the attributes. Given a training set of instance-label pairs

$(x_i;y_i)$; i = 1, …. , *l* where $X_i \in R^n$ and $y \in \{+1,-1\}$, the support vector machines [15, 16] require the solution of the following optimization problem:

$$\begin{aligned} &\min_{w,b,\xi} \frac{1}{2} w^T w + C\sum_{i=1}^{m} \xi_i \\ &\text{s.t } yi(w^T \phi(x_i) + b) \geq 1 - \xi_i, \\ &\xi_i \geq 1 \end{aligned} \tag{1}$$

where w is a linear separator with offset b, C is a constant, ϕ is a vector function, and ξ is slack variable.

The SVM with uneven margins are obtained by solving the quadratic optimization problem [9] :

$$\begin{aligned} &\min_{w,b,\xi} \langle w,w \rangle + C\sum_{i=1}^{m} \xi_i \\ &\text{s.t } \langle w,x_i \rangle + \xi_i + b \geq 1 \quad \text{if } y_i = +1 \\ &\quad \langle w,x_i \rangle - \xi_i + b \leq -\tau \quad \text{if } y_i = -1 \\ &\quad \xi_i \geq 0 \quad \text{for } i = 1,\ldots,m \end{aligned} \tag{2}$$

It can be seen from the above equation that there is an addition of uneven margin parameter τ (tau margin). τ is the ratio of negative-class margin to positive class margin, and will be equal to 1 on the standard SVM. In case of imbalanced datasets, a larger margin for the positive class than for the negative class is used, as can be seen in Figure 1. Therefore, in the SVM with uneven margins the value of τ is $0 < \tau < 1$.

GATE-SVM system is a variant of the SVM with uneven margins. In the usual SVM, positive and negative examples are treated the same way that margin hyperplane to the negative examples with a margin equal to the negative examples. However, the imbalanced training data where the positive examples are much less, then the SVM is not always appropriate representing the actual positive examples distribution. Therefore SVM with the positive margin greater than the negative margin is a better SVM model.

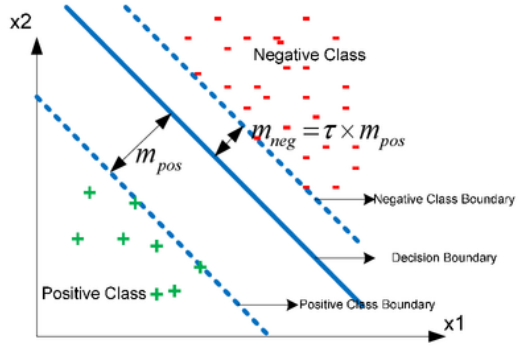

Figure 1. Illustration SVM with uneven margins (x1 & x2 are features).

3. SVM Based Information Extraction

Information Extraction is a token classification task rather than a Text Classification task. With Information Extraction we are working with texts but the basic units that we are looking for to classify are tokens in the text rather than the entire text. With Text Classification we are

seeking to identify whether an entire text is a member of particular category, while in Information Extraction the categories are start and end, and the objects we seek to assign to these categories are the individual tokens. Additional information about the token are encoded to enable SVM learning algorithm to generalize. Several features of the token as well as relational information about the surrounding tokens can be encoded.

SVM Based Information Extraction in this paper is employing GATE (General Architecture for Texts Engineering) toolkit [2]. Machine learning process in GATE is based on SVM Light Wrapper [17].

Two steps of this Information Extraction are training and extraction. Training process is to build a classifier for certain tag. This process is started with annotating sample corpus with targeted slots. In order to build the SVM model, we require start and end annotations for each class of annotation. SVM parameters are selected to build the model. Once the model becomes available, SVM classifier classifies the unseen documents which creates the annotation for start and tag over the text

On training process for SVM, first corpus is stored in a GATE format and annotated in accordance with the token type (e.g. *commodity*) and this document is used as an input to build SVM model (which is needed <*commodity*> as start tag and </*commodity*> as end tag on the object text/token in question). SVM models are generated and stored in an external file for later use.

In this experiment to build the features vector of tokens, several NLP features are used. First, *Case/Orthography,* which is the use of uppercase and lowercase letters by the token. Second, *Types of tokens* i.e. words, numbers, symbols, or punctuation. Third, *Entity*, the output module of *named entity recognition* standards owned by GATE. The window size is the number of tokens before and after the target token. It is also used as an input for SVM.

In SVM-GATE system, two SVM classifiers for each type of entity, one classifier for the start and another one for the end word, are trained. One-word entities are regarded as both start and end. In contrast, [3] trained four SVM classifiers for each named entity type – besides the two SVMs for start and end, also one for middle words, and one for single word entities. They also trained an extra SVM classifier to recognize words that do not belong to any named entity. [7] trained an SVM classifier for every possible transition of tags so that may lead to a large number of SVM classifiers. As our SVM classifiers only identify the start or end word for every target class, some post processing is needed to combine these into a single tag. To extract information from a new document, the system requires the SVM model produced in the learning process. SVM classifiers then will annotate text with the initial tag and the end tag that match existing models. In the next stage, the start and the end tags are combined in an appropriate token.

For example, given this annotated text:

There will a talk by <*speaker*>Laura Petitto</*speaker*> tomorrow morning.
On learning phase, each word will be used as an example of following classes:
not start tag = {“there”, “will”, “a”, “talk”, “by”, “Petitto”, “tomorrow”, “morning”}
not end tag = {“there”, “will”, “a”, “talk”, “by”, “Laura”, “tomorrow”, “morning”}
start tag = {“Laura”}
end tag = {“Petitto”}
Given this new text,
To remind you that Bill Gates will arrive at noon.
on extraction phase, SVM Classifier will classify the new text as follow:
not start tag = {“To”, “remind”, “you”, “that”, “Gates”, “will”, “arrive””, “at”, “noon”}
not end tag = {“To”, “remind”, “you”, “that”, “Bill”, “will”, “arrive””, “at”, “noon”}
start tag = {“Bill”}
end tag = {“Gates”}
After post processing, start tag and end tag will be combined into a single tag <speaker>Bill Gates</speaker>.

To measure the performance of the algorithm in extracting the information, Precision, Recall and F-Measure are used.

$$precision = \frac{\#correct\_extracted\_filler}{\#total\_extracted\_filler} \quad (3)$$

$$recall = \frac{\#correct\_extracted\_filler}{\#total\_correct\_filler\_should\_be\_extracted\_} \quad (4)$$

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} \quad (5)$$

Precision represents proportion of correct filler returned, while recall represents the proportion of returned filler that are actually targets. F-measure is a composite measure of precision and recall.

4. Evaluation

A. *Dataset*

In order to evaluate SVM-based Information extraction in different languages, Vegetable Market corpus and Seminar Announcement Corpus [18] are selected. Vegetable Market Corpus are obtained from crawling the internet and consists of 210 documents (web pages) in Indonesian language that contains news about the change in vegetable prices in cities in Indonesia, while the latest corpus consist of 485 documents.

For the purposes of learning and testing, each token corresponding to Vegetable Market documents manually annotated with eight classes using labels (slot) as follows:

- *Date*: The date when the news was written on a web page
- *Location*: Place of occurrence
- *Commodity*: Commodity Type of vegetables
- *Price_before*: vegetable commodity prices before the price changes
- *Price_latest*: vegetable commodity prices after the price change (current)
- *Unit*: Units that are used in trading
- *Price_change*: commodity price fluctuations
- *Event*: Event-related causes commodity price changes.

Vegetable corpus statistics are presented in Table 1, while the sample web page that has been Annotated shown in Figure 2.

Table1. Vegetable corpus statistics that are used as a dataset

Slot	frequency	Example
Date	282	12 November 2008, 10/12/2009
Location	341	Pasar Induk Keramat Jati, Kabupaten Bandung
Event	182	Banjir, kemarau
Price_before	561	5000
Price_latest	1427	6000
Commodity	1409	Kacang panjang, mentimun, kol
Price_change	106	Naik seribu rupiah
Unit	1088	Kg, Ikat, Liter

The total number of tokens in the first dataset is 390,226. Only 5396 (1.4%) of the token is positive example. The ratio of positive and negative data in the dataset is shown in Figure 3.

MAGELANG, SELASA - Harga sayur mayur di **<location>**Kabupaten Magelang**</location>**, kini turun secara signifikan. Pada berbagai jenis sayuran, penurunan harga terjadi bervariasi, mulai dari Rp 500 per kilogram (kg), hingga Rp 1.500 per kg.

Sumartini, salah seorang pedagang sayur di Pasar Muntilan, mengatakan, **<commodity>**kacang panjang **</commodity>** misalnya mengalami penurunan harga dari Rp **<price_before>**3.500**</price_before>** per **<unit>**kg**</unit>**. menjadi Rp **<price_latest>**2.500**</price_latest>** per **<unit>**kg**</unit>**. Begitupun, harga **<commodity>**seledri **</commodity>** yang semula Rp **<price_before>**2.500**</price_before>** per **<unit>**kg**</unit>** sekarang menjadi Rp **<price_latest>**1.500 **</price_latest>** per **<unit>**kg**</unit>**.. Untuk **<commodity>**tomat**</commodity>** dan **<commodity>**wortel**</commodity>**, masing-masing turun harga Rp **<price_before>**500**</price_before>** per **<unit>**kg**</unit>** menjadi Rp **<price_latest>**1.500 **</price_latest>** per **<unit>**kg**</unit>** dan Rp **<price_latest>**1.000**</price_latest>** per **<unit>**kg**</unit>**.

Menurutnya, kondisi ini dimungkinkan terjadi karena melimpahnya persediaan sayur di **<event>**musim panen **</event>**. Namun, karena pasar sedang sepi, saya pun tetap membatasi pembelian dari petani dan pedagang grosir, ujarnya, Selasa (**<date>**12/8**</date>**). Penurunan harga ini sudah berlangsung selama seminggu terakhir.

Figure 2. The Annotated webpage example.

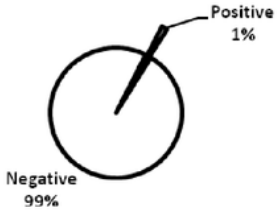

Figure 3. The ratio of positive and negative data in the dataset of Vegetable Market Corpus.

The summary of Seminar Announcement Corpus is presented in Table 2. The detail of this corpus statistics can be evaluated from [18].

Table 2. Number of examples for each entity/slot type, together with the number of non-tagged words of Seminar Announcement Corpus.

Slot				Non-entity	% of Positive Example
Stime	Etime	Speaker	Location		
980	433	754	643	157647	1.8

To get better results, the experiment is run for ten trials (10-fold cross validation) for each variation of input of the tested models. The model used by the SVM kernel is SVM linear with *reciprocal weighting*. While the chosen classification technique is *one-against-all* [17].

To seek the best performance, influence of window size is evaluated first, and then the best windows size is chosen for evaluating SVM performance on τ margin variation. Finally, the combination of the best window size and τ margin are compared to other non-optimal combinations. Since τ margin is the ratio of negative-class margin to positive class margin and its values are between 0 and 1, then the value of 0.1, 0.2, …1.0 is chosen to be combined with the best window size.

B. Experimental Result

Based on our preliminary experiment, it shows that the performance of SVM to extract information improved with the increase of Window Size. However, this increase not significant

at Window Size greater than 10 and the optimum one is reached at 4x4. This windows size is used for the rest of the experiment.

Figure 4 shows the relationship between τ margin variation and their precision and recall on Vegetable Corpus and Seminar Announcement Corpus. Precision is increasing as τ margin getting bigger, while recall is on the contrary become less and less on both corpus.

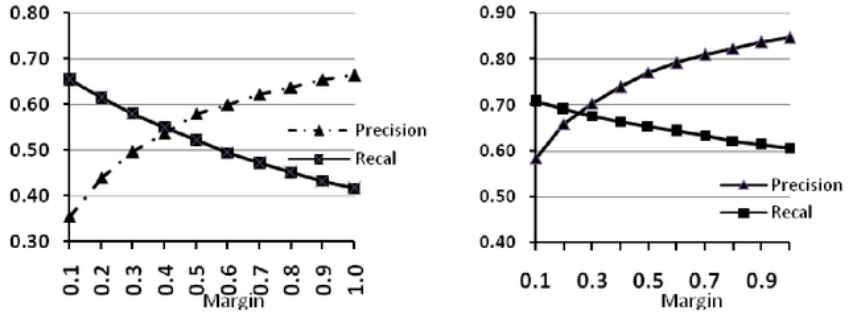

Figure 4. The relationship between τ margin variation and precision and recall on Vegetable Corpus (left) and Seminar Announcement Corpus (right).

Figures 5 shows that generally uneven τ margin ($0 < \tau < 1.0$) is better than even τ margin ($\tau = 1.0$) on F-measure on both corpus. Optimum margin is reached on $\tau = 0.5$, although the optimum margin is not so obvious in the Seminar Announcement corpus. From both figures we can say that for certain corpus there is a specific best τ margin. If we compare between optimum τ margin and uneven magin, there are improvements of F-measure as much as 7.47 % and 0.04 % for Vegetable Corpus and Seminar Announcement Corpus repectively.

In order to give clear insight, the best windows size (4x4) and τ margin of 0.5 are selected and compared to other combinations. Figure 6 shows that with the increasing number of documents samples for learning; F-Measure is increasing as well. These results are also similar to that result obtained by Li et al. [9] on the Job Corpus for small samples. Increasing the number of sample documents cause classifier quality is getting better so that their performance in the extraction of information also become better.

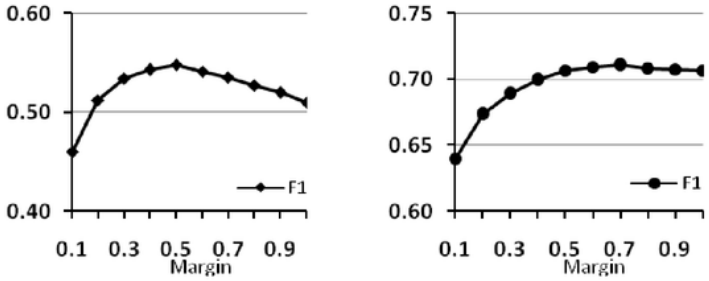

Figure 5. The relationship between τ margin variation and F-measure on Vegetable Corpus (left) and Seminar Announcement Corpus (right).

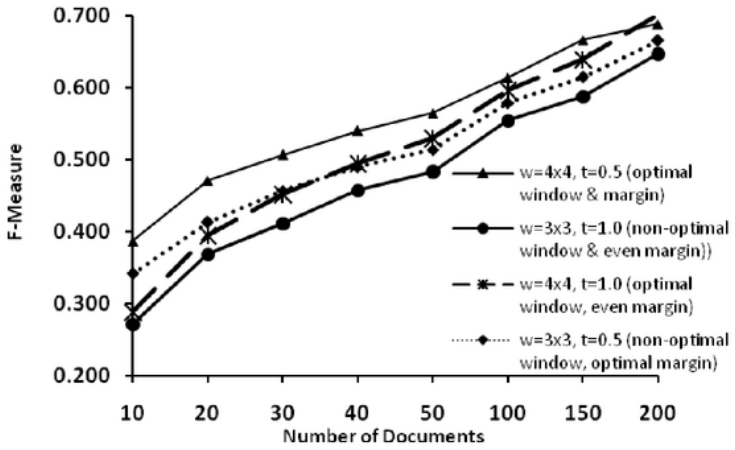

Figure 6. The relationship between the number of sample documents for learning and F-Measure on several composition of window size (w) and t (τ margin) using SVM GATE.

When we are varying several windows size and τ margin, the best F-measure and precision are obtained at window size of 4 and the τ margin of 0.5 followed by a window size of 3 and the τ margin of 0.5 and by even margin setting result. This result is different from the results of Li et al. [9] and Paramita [19] who recommend Window Size of 3 and τ margin of 0.6. The difference is expected due to the differences in the distribution of positive and negative examples that differ between the corpus in Paramita [19] and that we use.

5. Conclusion And Future Research

A. Conclusion

Through a classification problem, Information extraction can be approached effectively. In this SVM-based Information Extraction system, two SVM classifiers for each type of entity, one classifier for the start and another one for the end word, are trained. SVM classifiers then will annotate new document with the initial tag and the end tag that match existing models. Some importance parameters of the SVM are windows size and τ margin.

The performance of SVM-GATE tends to increase as *Window Size* increased, but the increased performance at Window Size greater than 10 is not significant. There is an optimum τ margin for both Indonesian corpus of Vegetable Market and Seminar Announcement Corpus. The best Performance of SVM-GATE obtained at the τ Margin of 0.5 and the Window Size of 4x4. Using optimal τ margin and windows size and compare it to even margin, we find that there are improvements of F-measure as much as 7.47 % and 0.04 % for Vegetable Corpus and Seminar Announcement Corpus respectively.

B. Future Research

To further improve the performance of SVM-GATE, NLP enrichment is needed such as the use of Part of Speech Tagger for Indonesian. It should be further investigated how to optimize τ Margin in relation to the corpus statistics.

References

[1] Bouckaer, R.R., "Low level information extraction. A Bayesian network based approach", *In Proc. TextML*, 2002.

[2] Cunningham, Hamish, D. Maynard, K. Bontcheva, V. Tablan, "GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications", *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics* (ACL'02), 2002. URL: http://gate.ac.uk/sale/acl02/acl-main.pdf.

[3] Isozaki, H. and H. Kazawa, "Efficient Support Vector Classifiers for Named Entity Recognition", *In Proceedings of the 19th International Conference on Computational Linguistics* (COLING'02), pages 390–396, Taipei, Taiwan, 2002.

[4] Joachims, T., "Text categorization with support vector machines: Learning with many relevant features", *In Proceedings of the European Conference on Machine Learning*, pages 137-142, Berlin. Springer, 1998.

[5] Tong, S. and D. Koller, "Support Vector Machine Active Learning with Applications to Text Classification", *Journal of Machine Learning Research,* p 45-66, 2001.

[6] Yang, Y., "A study on thresholding strategies for text categorization", *In Proc. of ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR'01), p137-145, New York. The Association for Computing Machinery, 2001.

[7] Mayfield, J., P. McNamee, and C. Piatko, "Named Entity Recognition Using Hundreds of Thousands of Features", *In Proceedings of CoNLL-2003*, pages 184–187. Edmonton, Canada, 2003.

[8] Li, Y., K. Bontcheva, and H. Cunningham, "SVM Based Learning System For Information Extraction", *Sheffield Machine Learning Workshop*, Lecture Notes in Computer Science, Springer Verlag, 2005.

[9] Li, Y. and Shawe-Taylor, J., "The SVM with uneven margins and Chinese document categorization", *In Proceedings of The 17th Pacific Asia Conference on Language, Information and Computation* (PACLIC17), pages 216-227, Singapore, Oct, 2003.

[10] Morik, K., P. Brockhausen, and T. Joachims, "Combining statistical learning with a knowledge-based approach - a case study in intensive care monitoring", *In Proc. 16th Int'l Conf. on Machine Learning* (ICML-99), pages 268-277, San Francisco, CA. Morgan Kaufmann, 1999.

[11] Shawe-Taylor, J., et al., "Kernel Methods for Document Filtering", *In: The Eleventh Text Retrieval Conference* (TREC 2002), 19 - 22 November 2002.

[12] Gao, Y.M. and Y.L. Huang, "Using SVM with uneven margins to extract acronym-expansions", *International Conference on Machine Learning and Cybernetics*, 2009 (3) 1286 – 1292.

[13] Ageev, M.S. and Dobrov, B.V., "Support Vector Machine Parameter Optimization for Text Categorization Problems", *Proc. of Information Systems Technology and its Applications*, International Conference ISTA, 2003.

[14] Vapnik, V., "Statistical Learning Theory", Wiley, 1998.

[15] Boser, B. E., I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers", *In Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp. 144-152. ACM Press, 1992.

[16] Cortes, C. and Vapnik, V., "Support Vector Networks", Machine Learning 20: 1-25, 1995.

[17] Cunningham, H, D. Maynard, K. Bontcheva, V. Tablan, "Developing Language Processing Components with GATE Version 4 (a User Guide)", The University of Sheffield, 2001-2007. URL: http://gate.ac.uk/sale/tao/

[18] SAIC, *Proceedings of the Seventh Message Understanding Conference* (MUC-7), 1998. http://www.itl.nist.gov/iaui/894.02/-related projects/muc/index.html.

[19] Paramita, "Penerapan Support Vector Machine untuk Ekstraksi Informasi dari Dokumen Teks", Tugas Akhir Program Studi Teknik Informatika STEI Institut Teknologi Bandung, 2008.

Kuspriyanto, He graduated from his Bachelor degree program in Electrical Engineering ITB; his Master and Doctoral program from USTL, Montpellier, France. He is now a lecturer of School of Electrical Engineering and Informatics, Bandung Institute of Technology. His research interest is in real time system.

Oerip S. Santoso, He graduated from his Medical degree in The University of Indonesia; He got M.Sc. degree in Computer Science from University of Wisconsin, Madison, USA and Doctoral degree in Mathematical Programming from Unitersity Pierre Marie-Curie (Paris VI) Paris, French. He is now a lecturer of School of Electrical Engineering and Informatics, Bandung Institute of Technology. His research interest is in numerical computation and medical informatics.

Dwi H. Widyantoro, He graduated from his Bachelor degree program in Computer Science ITB; his Master and Doctoral program in Computer Science from Texas A&M University, College Station, TX, USA. He is now a lecturer of School of Electrical Engineering and Informatics, Bandung Institute of Technology.

Husni S. Sastramihardja, His Doctoral program in Informatic Engineering from ITB. He is now a lecturer of School of Electrical Engineering and Informatics, Bandung Institute of Technology. His research interest is in information system.

Kurnia Muludi, He received his B.Sc. degree in Soil Science from Agriculture Faculty, The University of Lampung, Indonesia in 1987. He got M.Sc. degree in Soil Science from The University of Ghent, Belgium in 1994. Currently, he is studying doctorate degree at Bandung Institute of Technology, Indonesia. He is a lecturer at The University of Lampung, Indonesia.

Siti Maimunah, She received her Bachelor degree in Computer Engineering and Master degree in Informatic Engineering from 10 Nopember Institute of Technology Surabaya, Indonesia. Currently, she is studying doctorate degree at Bandung Institute of Technology, Indonesia. She is a lecturer at Surabaya Adhi Tama Institute of Technology, Indonesia.

Performance Evaluation of SVM-Based Information Extraction using τ Margin Values

ORIGINALITY REPORT

30%

SIMILARITY INDEX

MATCHED SOURCE

1	gate.ac.uk Internet	218 words — 6%
	★gate.ac.uk Internet	6%

EXCLUDE QUOTES	OFF	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	ON		