International Symposium on Communications
and Information Technologies 2004 (ISCIT 2004)
Sapporo, Japan, October 26- 29, 2004

Evaluation of the New Feature Types for Question Classification with Support Vector Machines

Marcin Skowron* and Kenji Araki*
* Graduate School of Information Science and Technology
Hokkaido University, Sapporo, 060-8628, Japan
Tel: +81-11-7067-389, Fax: +81-11-7067-389.
E-mail: {ms,araki} @media.eng.hokudai.ac.jp

***Abstract*— Question classification is of crucial importance for question answering. In question classification, the accuracy of Machine Learning algorithms was found to significantly outperform other approaches. The two key issues in classification with a ML-based approach are classifier design and features selection. Support Vector Machines is known to work well for sparse, high dimensional problems. However, the frequently used bag-of-words approach does not take full advantage of information contained in a question. To exploit this information we introduce three new feature types: Subordinate Word Category, Question Focus and Syntactic-Semantic Structure. As the results demonstrate, the inclusion of the new features provides higher accuracy of question classification compared to the standard bag-of-words approach and other ML based methods such as SVM with the Tree Kernel, SVM with Error Correcting Codes and SNoW. A classification accuracy of 84.6% obtained using the three introduced feature types is, as of yet the highest reported in the literature.**

I. Introduction

With the rapid growth of text available on the Internet, it has become more difficult for users to find specific information. The standard approach of querying an Internet search engine often returns thousands of results, containing a ranked list of documents along with their partial content (snippets). For an average Internet user, it is often time-consuming and laborious to find requested information. Often to find the searched information, a user has to connect to several servers and scan through dozens of documents to locate it. We think that for a human being the most natural and straightforward approach to such a task is to ask a question in a natural language form. The output ought to be a correct answer resembling as much as possible those given by human beings. The realization of this task is an active research field in the current Question Answering (QA) systems.

In order to provide a correct answer to a question from a large collection of documents, like that of the Internet, one needs to impose some constraints on the scope of the possible answers. A constraint frequently used in QA systems is a question category. Question classification assigns a category to a given question based on the type of answer entity the question represents [12]. The outcome of the question classification serves to decrease the number of answer candidates. Consequently, a computer system does not need to verify all candidates found in the retrieved documents to decide if it is a correct answer to a given question. Because a verification based exclusively on the expected-answer type is often sufficient to find a correct answer, question classification is of prime importance for QA systems.

This paper describes the automatic method of question classification using Support Vector Machines (SVM)[6] [21] in a taxonomy that includes 6 coarse-grained and 50 fine-grained categories. We introduce and evaluate 3 new feature types (Subordinate Word Category, Question Focus and Syntactic-Semantic Structure) that help to exploit additional information that is useful for question classification, which is overlooked by the standard, bag-of-words approach. As the results demonstrate, the inclusion of these feature types helps to achieve a higher accuracy in a question classification task, compared to that obtained using the bag-of-words approach. Furthermore, the accuracy achieved using the set of the introduced feature types is the highest result reported in the literature so far for this taxonomy and dataset.

II. Question Classification

Question classification is of crucial importance for QA Systems. Question classification is defined as the task that, given a question, maps it to one of *k* classes, which provide a semantic constraint on the sought-after answer [13]. This information, typically with other constraints on the answer, is used in a downstream process that leads to selection of a correct answer from among several candidates. As described in the literature, a QA system that is able to classify a question with more detailed taxonomy and use this information to extract and verify answer candidates, achieves higher overall accuracy [5] [16]. Additionally, in some systems question category information is also used in a question category dependent query formation process [18]. As the results show, such a query retrieves a less distorted set of documents, where a correct answer appears more frequently, compared to a set retrieved with a query formed in the standard keyword-based approach.

TABLE I
THE COARSE AND FINE GRAINED QUESTION CATEGORIES

Coarse	Fine
ABBR	abbreviation, expansion
DESC	definition, description, manner, reason
ENTY	animal, body, color, creation, currency, disease, event, food, instrument, language, letter, other, plant, product, religion, sport, substance, symbol, technique, term, vehicle, word
HUM	description, group, individual, title
LOC	city, country, mountain, other, state
NUM	code, count, date, distance, money, order, other, percent, period, speed, temperature, size, weight

III. TAXONOMY AND DATASET

In recent years, numerous question taxonomies have been defined, but there is no one standard used by all the systems. For example, this is the case of the systems participating in the TREC QA-Track. Most of them implement their own question taxonomy. Moreover, the used taxonomy is frequently redefined on a year-to-year basis. Usually the systems use a taxonomy consisting of less than 20 question categories. However, as demonstrated by several QA systems, employing a more detailed one consisting of a fine-grained category definition is beneficial in the process of positioning and verifying answer candidates.

In our work, we used hierarchical, two-layered taxonomy proposed by Li and Roth in [13] consisting of 6 coarse-grained and 50 fine-grained categories, [Table I]. Recently, this taxonomy was employed also in other QA systems, and different approaches to automatic question classification were evaluated based on it [4] [7] [12] [13] [25]. We decided to use this taxonomy because of its effective overall coverage of question types that are usable by the answer candidates verification module of our QA system, and a freely available training dataset. Using it we could also compare the question classification results of our SVM based classifier to other methods that used the same dataset.

For the training and evaluation of our question classifier, we use the publicly available dataset provided by USC [8], UIUC [13] and TREC[22] [23] [24], which consists of 5,500 classified questions for the training set, and 500 more for testing. The test data are a set from the Question Answering Track of TREC 10. The training set is assembled from previous TREC questions as well as from archives of online question answering systems [Li, Roth 2002]. All the questions from these datasets have been manually labeled using the taxonomy presented in Table I, by UIUC [13].

IV. APPROACHES TO THE QUESTION CLASSIFICATION

The approaches to question classification can be discriminated into the following, three main groups: rule-based, language modeling and machine learning based[1].

[1]We do not include an explanation of the language modeling approach, due to its low performance using a detailed taxonomy. For more information refer to [4] [12].

A. Rule Based Classification

In the rule based approach, hand-written grammar rules and a set of regular expression are employed to parse a question and to determine the answer type [Van Durme, 2003]. However, with this approach the researches have faced several limitations:

- Hand-writing the rules and preparing the efficient regular expressions is a difficult and time-consuming process.
- Hand-written rules have limited coverage and is fairly complicated to broaden the scope of answer categories to include more detailed ones.
- In order to adopt a new taxonomy, many previously prepared rules have to be modified or completely rewritten.

Considering these limitations, most of the systems that use hand-written rules are bound to use a limited number of question type categories. Consequently, question category information is limited to its use, which as previously described, influences the performance of the whole QA system [5] [16].

B. Machine Learning-Based Classification

In the machine learning approach, expert knowledge is replaced by a sufficiently large set of labeled questions. Using this collection, a classifier is trained in a supervised manner. Possible choices of classifiers include but are not limited to: Neural Network, Naive Bayes, Decision Tree and Support Vector Machines. The machine learning approach addresses many limitations of the rule-based method, which were presented above. The advantages include:

- Short creation time.
- No need for expert knowledge (automatic creation of a classifier).
- Broader coverage; can be obtained by providing new training examples.
- If required, the classifier can be flexibly reconstructed (retrained) to fit to a new taxonomy.

At present, the results achieved using the machine learning approach represents a state of the art in question classification. The different machine learning methods presented below utilized the same taxonomy and dataset as described in **III**.

V. STATE OF THE ART IN QUESTION CLASSIFICATION

Currently, the primary machine learning algorithm used for question classification is Support Vector Machines (SVM) [7] [19] [25]. Researchers apply SVM to question classification because it constantly outperforms other machine learning techniques in several applications including text classification, which is similar to question classification [9] [17] [20]. However, as the results presented in the literature demonstrate, the highest accuracy was obtained using the SNoW learning architecture-based classifier.

The research of Zhang and Lee [25] presented work on question classification using the Support Vector Machines, and compared its results to these obtained by other machine learning approaches like Nearest-Neighbors (a simplified version of well-known kNN algorithm), Naive-Bayes, Decision Tree and

TABLE II

THE QUESTION CLASSIFICATION ACCURACY FOR THE FINE-GRAINED CATEGORIES OBTAINED BY THE STATE OF THE ART SYSTEMS

	SVM (BOW) [25]	SVM (BSH) [7]	SNoW [13]
P1	80.2%	82.0%	84.2%

Sparse Network of Winnows (SNoW). All the classifiers were trained using the same dataset. The SVM classifier achieved the highest results compared to other machine learning based classifiers, both in the bag-of-words and the bag-of-bigrams approaches. The advantage of the SVM was especially significant under the fine-grained category definition [2]. The research proposed also a specific kernel function called the tree kernel, to enable the SVM to take advantage of the syntactic structures of question. Unfortunately, its application to the classifier under the fine-grained category definition did not bring improvements. The highest accuracy reported in this work for the first classification, under the fine-grained category definition was achieved using the bag-of-words (BOW) features. This and other results of the state of the art systems, obtained using the same dataset, for the first classification (P1) under the fine-grained category definition [13] are presented in Table II.

Similar results was reported in later work that used the SVM classifier with the bag-of-words features [7]. The authors performed the experiments after dimensionality reduction by computing the term space transformation using singular value decomposition (SVD) and applying BCH codes to convert a multi-class classification problem into a number of two-class problems. The accuracy improvement to 82.0%, was reported in a bag-of-bigrams approach, after the inclusion of the name entity based features for the seven selected Named Entity categories [2].

The work of Li and Roth [13] described the system that obtained the highest question classification accuracy achieved up to date for the presented taxonomy and dataset, using the classifier based on the SNoW (Sparse Network of Winnows) learning architecture. The classifier was trained using a rich selection of features including: part-of-speech (POS) tags, non-overlapping phrases (chunks), named entities (NEs), head chunks, semantically related words, conjunctive (n-grams) and relational features. The total number of features used is approximately 200,000; for each question, up to a couple hundred are active.

As presented in Table II, though SVM was found to outperform other machine learning approaches in several applications, the highest result obtained so far for the question classification task was achieved using the SNoW learning architecture. We think that the high performance of SNoW classifier is the result of the sensible selection and effective application of a rich set of features, especially those based on the semantic analysis. Up to date, no SVM based classifier was able to successfully employ a similar number of features to provide such detailed representation of questions, helpful in the classification task.

VI. SUPPORT VECTOR MACHINES FOR QUESTION CLASSIFICATION

Support Vector Machines [6] [21] is based on the Structural Risk Minimization principle from Computational Learning Theory [21]. The SVM in the basic form learns the linear hyperplane that separates a set of positive examples from a set of negative examples with maximum margin (the margin is defined by the distance of the hyperplane to the nearest of the positive and negative examples) [14]. By using appropriate kernel functions, SVM can be extended to learn polonymical classifiers, radial basic function (RBF) network, and three-layer sigmoid neural nets.

The selection of this classifier was based on the following observations, concerning the properties of the SVM and the question classification task [3]:

- **High dimensional input space**.
  In the experiments to be discussed later, the number of used features is close to 9900. However, since the SVM uses overfitting protection, it can work well with such large feature space.
- **Dense concepts and sparse instances**.
  As the previous results demonstrated, the effective SVM based classifier should combine many features (learn a "dense" concept). The feature types introduced in this work, provide such an additional density to a used questions representation.
- **Multi-class classification problems**.
  Question classification can be linearly separable and handled by the SVM binary classifier (see below for more detailed description).

A. Binary Classifier for the Multi-class Problems

The objective of our experiment is to classify a given question to one of 50 possible categories. Although the SVM is inherently binary classifier, it is possible to extend its use to a multi-class problems like that of question classification. This is performed by reducing the multi-class problem to multiple binary classifications [1]. There are two popular alternatives: one-against-all and all-pairs. We used the former approach, constructing 50 separate classifiers trained on data where the questions from one question category formed one class and all the remaining questions from other categories created the second one. The SVM Light [10] implementation of SVM is used in the following experiments.

VII. FEATURE SELECTION

The feature selection is required to find a balance between the need to provide sufficient information to the classifier and the danger of providing them in excess. In the former, because of a lack of sufficient information the classifier is not able to effectively discriminate the test questions based on the learned model. On the other hand, providing too

[2] For the details of the evaluation of several machine learning approaches in the question classification task see [25].

[3] Similar reasons were presented in [11] for the justification of the SVM application to the text categorization task.

many features leads to overfitting during a training process with sparse data, introduces noise in the feature space, and inflicts higher computational complexity. A frequently used solution is dimensionality reduction. Here, care has to be taken to minimize the loss of features that are useful for the classification.

As demonstrated in previous work, the feature selection is of crucial importance for a wide spectrum of classification task, that use machine learning [13] [19] [20]. Question classification to some extent is similar to text categorization. The goal in the latter is to assign a given text to a previously defined class. In question classification, a given text is usually question sentence, a few words long. As shown in [13], question classification requires more complicated features than text categorization. However, in spite of SVM robustness to handle large feature sets, as of yet there are no similarly effective applications of such a rich set of features for the SVM based classifier. Motivated by this, we decided to introduce new feature types for the SVM based classifier and to evaluate their impact on the accuracy of question classification.

VIII. New Feature Types for Question Classification

The bag-of-words approach is frequently used in a number of classification tasks including question classification. However, in our opinion, with this approach the classifier is not able to take full advantage of information contained in a question. In the bag-of-words approach, a word can be used only directly by checking if it exists in a feature space or not. Similarly, in the training process, the model is created without using the semantic information contained in question words. A word position in a sentence is another overlooked information, similar to information on syntactic-semantic structures. To address these limitations we introduce three new feature types for a question classification task. These are: Subordinate Word Category, Question Focus and Syntactic-Semantic Structure.

A. Subordinate Word Category

In the bag-of-words and similar approaches (eg. bag-of-ngrams), information contained in a word can be used only directly. In the training process of a classifier as well as during the classification of test questions, other types of information existing on different layers (eg. semantic) are not utilized. Consequently, without providing a representation of a given word in a higher, more general level, the words that less frequently occur in a dataset are used only to a very limited extent, if used at all. We think that these words possess valuable semantic information, which is useful for question classification. In several cases, the remaining words exist at the same time, in several question categories, and as such do not provide sufficient information to the classifier to correctly assign a question category. For example in the test question "What is the proper name for a female walrus ?" the words "What", "is", "proper", "for" or "female" can be found in several categories, while the word "walrus" did not appear in training data. In this situation the word "walrus", the only one that could potentially provide really useful information to a classifier can not be used in the bag-of-words approach, thus it is difficult to correctly discriminate such questions.

To capture semantic information contained in a word on a higher level of representation we propose a new feature type, the Subordinate Word Category. This feature type is realized by assigning a WordNet [15] hyponym to a common nouns found in a given question. The list of selected hyponyms includes 25 categories like: animal, plant, vehicle, quantitative relation, length, body part, land, water, people, etc. If found, these hyponyms are assigned for all common nouns found in a given question and add as a new entry to a feature space. Additionally, a common category "YEAR" is assigned for cardinal numbers consisting of four digits, and used to substitute the original word. Similarly, the category "NUMBER" is used for all the remaining cardinal numbers.

B. Question Focus

For the purpose of the question classification task, the question focus can be defined as a phrase in the given question that disambiguates it and emphasizes the expected answer type. In the bag-of-words approach all words are treated equally without considering their position in a question. Question focus word, which is often a valuable indication of question category is another type of information that cannot be used in this approach. To exploit this additional, useful for classification information we introduce the Question Focus feature type.

In the experiments that follow a question focus word is recognized using a set of the regular expressions applied to a POS tagged question. For example, one of the regular expression searches for the first common noun appearing after the word "What". For instance, in the question: "What county is Chicago in?" the word "county" is recognized to be the question focus word. After applying this feature a few questions from the "LOC:other" category, both in training and test data, gain the additional common feature. Similarly, if discovered the question focus words are assigned for the remaining questions from this category, as well as for the questions contained in the other categories from the dataset. As the results demonstrate the inclusion of this feature type leads to the improvement in the accuracy of question classification.

C. Syntactic-Semantic Structure

Our analysis of the dataset revealed that some syntactic-semantic structures that frequently exist in questions from one category do not appear in the others. In our opinion, the ability to exploit these structures provides a valuable information for a classifier that is overlooked in the standard bag-of-words approach. To construct highly distinguishable patterns, the syntactic-semantic structures need to be general enough to allow variation of different questions that belongs to one category, and at the same time, strict enough to capture the differences among questions from one category and the others. In this work the structures were automatically generated based on the training dataset, with the following processing:

TABLE III
THE QUESTION CLASSIFICATION ACCURACY USING BASE-LINE APPROACH (BOW-LINEAR KERNEL) WITH DIFFERENT NUMBER OF FEATURES (F1-F3), LOWERCASED LETTERS (LC) AND POS TAGS ASSIGNED (POS)

	BOW F1	BOW F2	BOW F3	LC F1	POS F1
P1	**80.2%**	79.8%	79.4%	79.4%	79.6%

- Using the set value of TFIDF, select and later preserve in the original form the collection of "categories important nouns".
- Substitute remaining nouns with the tokens that respect the surface feature of a given word.
- Substitute the cardinal numbers with one, common token.

If such a structure is found to exist at least twice in one and only one question category it is stored and assigned as an additional, common feature to questions that share it.

IX. EVALUATION

As explained in [13] the authors were aware that using their taxonomy, the classification of some questions may be ambiguous between few question categories. In their works, the classifier is permitted to assign a multiple labels to one question, if the classifier confidence level is low. Although this approach can be beneficial in practical application to a QA system, for the sake of achieving a strict measure of classification accuracy we decided to count the precision of correctly classified questions using only the first answer category assigned by the classifier.

Our experiments, as well as the results presented in [25] demonstrated, that under the fine-grained category definition the SVM based classifier achieves the highest accuracy with the linear kernel, using the bag-of-words, compared to ones obtained with other kernels, and using bag-of-bigrams approach. Hence, in the experiments that follow, the results obtained using the linear kernel with the bag-of-words features are considered as a base-line for the results comparison. Additional experiments, which results are presented in Table III show that the usage of the different number of features (set obtained after excluding the words that appeared more than: 1,000 times (F1), 700 times (F2) and 1,200 times (F3)), normalized words by converting all letters to lower case (LC), and the POS tagged words (POS), did not bring improvement.

The classifier was trained on 5 different size training datasets and tested on the TREC10 questions. Table IV shows the accuracy of question classification for the fine-grained categories, achieved using the bag-of-words approach (BOW), as well as the results obtained after extending the BOW with the new feature types (SWC - Subordinate Word Category, QF - Question Focus, SSS - Syntactic-Semantic Structure). The classification accuracy is measured as the proportion of the correctly classified questions among all test questions. As the results demonstrate, the inclusion of each of the proposed feature type contributed to a higher accuracy compared to the bag-of-words approach. The biggest improvement of 3.0% was achieved after the inclusion of the Subordinate Word Category feature type. As the results show, the SVM handles large set of features without overfitting; the accuracy grows evenly along with the larger training set provided.

TABLE IV
THE QUESTION CLASSIFICATION ACCURACY FOR THE FIRST CLASSIFICATION UNDER THE FINE-GRAINED CATEGORIES USING DIFFERENT FEATURE TYPES

		New Feature Types		
	BOW	SWC	QF	SSS
1000	66.8%	69.6%	68.8%	69.4%
2000	71.4%	75.2%	73.8%	73.4%
3000	75.0%	76.8%	76.2%	76.2%
4000	77.8%	78.0%	79.0%	79.2%
5500	**80.2%**	**83.2%**	**82.6%**	**81.4%**

TABLE V
THE QUESTION CLASSIFICATION ACCURACY FOR THE FIRST CLASSIFICATION UNDER THE FINE-GRAINED CATEGORIES USING DIFFERENT SETS OF FEATURE TYPES

		Set of Feature Types	
	BOW	SWC QF	SWC SSS
1000	66.8%	70.6%	70.6%
2000	71.4%	76.4%	76.8%
3000	75.0%	79.0%	78.2%
4000	77.8%	79.6%	78.8%
5500	**80.2%**	**84.4%**	**84.2%**
	BOW	QF SSS	SWC QF SSS
1000	66.8%	70.4%	71.2%
2000	71.4%	74.6%	77.4%
3000	75.0%	77.8%	80.2%
4000	77.8%	80.0%	80.6%
5500	**80.2%**	**82.6%**	**84.6%**

The results obtained after adding various sets of the feature types are presented in Table V. The highest accuracy of 84.6% was achieved in the run using all the proposed feature types (SWC QF SSS), bringing approximately 22% error reduction compared to the base-line approach. This result, obtained by the SVM based classifier, is higher than those reported in the previous researches [4] [7] [12] [13] [25], for the same training and test data collection.

The closer analyze of the misclassified questions revealed that some of them are the result of inconsequent labeling of questions in the dataset. For example, the questions " Where is Amsterdam ?" from the training data and similar question " Where is Milan?" from the test data have different labels, "LOC:other" and "LOC:city ", respectively. Similarly, the test questions "What county is Modesto, California in ?" and "What county is Phoenix, AZ in ?" are labeled "LOC:city" while the training question "What county is Chicago in ?" is labeled "LOC:other". In the test set, five questions with the inconsistent labels were discovered. In the run that used the corrected labels for this questions, the accuracy improved of 1% was found for all the sets of feature types, improving the highest noted accuracy to 85.6%.

X. DISCUSSION

The research confirmed that the high-performance question classification requires to employ much richer set of features than this available on the word level. The introduction of the new feature types supplied additional information to the SVM based classifier that could not be exploited in the standard bag-of-words approach. Additionally, using these features, the classifier could "learn faster", from a smaller set of training data; a similar accuracy to this obtained in the base-line approach using 5,500 training questions, was achieved using a set of 3,000 questions. Using the whole set of the presented feature types the classifier, achieved the result of 84.6%, for the first classification under the fine-grained categories definition. This result demonstrates that semantic and structural information contained in a question can provide highly discriminative features that help to classify a given question to a correct category. All the presented feature types are based on the freely available tools, like POS tagger [3] and WordNet [15], constructed automatically, which is not always a case in the other methods (eg. the good performance of the SNoW based classifier, depends heavily on the feature called "RelWords" (related words), which are constructed semi-automatically).

XI. Conclusions and Future Work

This paper presented a machine learning approach to question classification task using the Support Vector Machines. We proposed three new feature types that address the limitations of the bag-of-words and similar approaches (eg. bag-of-ngrams) frequently used in several classification tasks. The experimental results demonstrate that the inclusion of the new features types Subordinate Word Category, Question Focus and Syntactic-Semantic Structure was useful for improving the performance of the classifier over the bag-of-words approach. Using the set of three feature types, a result of 84.6% was achieved, bringing the error reduction of 22% compared to the base-line approach. A comparison with the state of the art systems has shown that using these features, the classifier was able to achieve better accuracy than any other machine-based classifier. The additional advantage of this approach is the fact that the new feature types were created automatically, using only the freely available tools like POS Tagger and WordNet. Our future work includes further tests and refining of the introduced feature types, especially the Syntactic-Semantic Structure, which in our opinion, posses the potential to provide higher coverage of various question categories.

References

[1] E. Allwein, E. Schapire and Y. Singer, Reducing multiclass to binary: A unifying approach for margin classifiers. Journal of Machine Learning Research, 1:113-141, 2000.

[2] D.M. Bikel, R.L. Schwartz and R.M. Weischedel, An algorithm that learns what‘s in a name. Machine Learning, vol. 34, no. 1-3, 211-231, 1999.

[3] E. Brill, Transformation-Based Error Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging. Computational Linguistics 21(4), 543-566, 1995.

[4] J Brown, Entity-Tagged Language Models for Question Classification in a QA System, Report on: http://www-2.cs.cmu.edu/˜jonbrown/IRLab/Brown-IRLab.pdf, 2003.

[5] C. Cardie, V. Ng, D. Pierce and C. Buckley, Examining the Role of Statistical Knowledge Sources in a General-Knowledge Question-Answering Systems. In Proceedings of the 6th Applied Natural Language Processing Conference, 180-187, 2000.

[6] C. Cortes and V. Vapnik, Support-Vector Network, Machine Learning 20, 1-25, 1995.

[7] K. Hacioglu, W. Ward, Question Classification with Support Vector Machines and Error Correcting Coded, In the Proceedings of HLT-NACCL 2003, 28-30, 2003.

[8] E. Hovy, L. Gerber, U. Hermajakob, C. Lin and D. Ravichandran, Towards Semantics-based Answer Pinpointing. In Proceedings of the DARPA Human Language Technology conference (HLT), 1999.

[9] T. Joachims, Text Categorization with Support Vector Machines: Learning with Many Relevant Features. In Proceedings of European Conference on Machine Learning (ECML98), 137-142, 1998.

[10] T. Joachims, Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning, MIT-Press, 1999.

[11] T. Joachims, Advances in Kernel Methods: Support Vector Learning, chapter Making large-Scale SVM Learning Practical. Chap. 11. MIT-Press, 1999.

[12] W. Li, Question Classification Using Language Modeling. CIIR Technical Report, 2002.

[13] X. Li and D. Roth, Learning Question Classifiers. In Proceedings of the 19th International Conference on Computational Linguistics (COLING'02), 556-562, 2002.

[14] L. Marquez, Machine Learning and Natural Language Processing, " Seminar: Industrias de la lengua / La ingeniera Linguistica en la sociedad de la informaci3on.", 2000.

[15] G. Miller G., WordNet: a lexical database for English. Communications of the ACM 38(11), 39-41, 1995.

[16] M.A. Pasca and S.M. Harabagiu, High Performance Question/Answering. In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 366-374, 2001.

[17] J.D.M. Rennie and R. Rifkin, Improving multiclass text classification with the support vector machines, MIT Artificial Intelligence Laboratory Publications, AIM-2001-026, 2001.

[18] M. Skowron and K. Araki, What Can Be Learned from Previously Answered Questions? A Corpus-Based Approach to Question Answering. Intelligent Information Systems 2004. New Trends in Intelligent Information. Proceedings of the International IIS: IIPWM04 Conference, 379-387, 2004.

[19] J. Suzuki, H. Taira, Y. Sasaki and E. Maeda, Question Classification using HDAG Kernel, Workshop on Multilingual Summarization and Question Answering 2003, post-conference workshop in conjunction with ACL-2003, 61-68, 2002.

[20] H. Taira and M. Haruno, Feature Selection in SVM Text Categorization. In Proceedings of the 16th Conference of the American Association for Artificial Intelligence (AAAI99), 480-486, 1999.

[21] V. Vapnik, The Nature of Statistical Learning Theory. Springer, 1995.

[22] E. Voorhees, The TREC-8 Question Answering Track Report. In Proceedings of the 8th Text Retrieval Conference (TREC8), 77-82, 1999.

[23] E. Voorhees E., Overview of the TREC-9 Question Answering Track Report. In Proceedings of the 9th Text Retrieval Conference (TREC9), 71-80, 2000.

[24] E. Voorhees, Overview of the TREC 2001 Question Answering Track Report. In Proceedings of the 10th Text Retrieval Conference (TREC10), 157-165, 2001.

[25] D. Zhang and W.S Lee, Question Classification using Support Vector Machines. Proceedings of the 26th ACM SIGIR, 26-32, 2003.