Question classification in Persian language based on conditional random fields

Ali Mollaei
Islamic Azad university of Mashhad, Iran
Ali.mollaei@iran.ir

Rahati quchani Saeed
Islamic Azad university of Mashhad, Iran
Rahati@mshdiau.ac.ir

Estaji Azam
Ferdowsi University of Mashhad
Estaji@ferdowsi.um.ac.ir

Abstract— The question classification system is one of the important subsystems in the Question Answering Systems (QAS). In such systems through retrieval methods and information extraction the texts are retrieved in order to get to a correct answer. The current study is designed to present the architecture of question classification (QC) in Persian based on the Conditional Random Fields (CRF) machine learning model and evaluate effects of various features on its accuracy. In this study, sentences were classified into two levels of coarse and fine classes based on the type of the answer to each question. After extracting features and setting sliding window on the CRF model, CRF question classifier (QC) is train. Then, the QC predicts labels for every token in question. Next, a majority voting on the question classification output, is used to extract a unique label for each question. Further, the effects of different features on the ultimate accuracy of the system were evaluated. Finally results of this question classifier, illustrate a satisfactory accuracy. (Abstract)

Keywords-component; question answering system, question classification, conditional random fields, majority voting (key words)

I. INTRODUCTION *(HEADING 1)*

All Today, it's possible to access to a high level of contextual information on the internet, and it is essential for users to search information accordance with their needs. However, the searching engines, using the traditional methods, in answer to the user's short question often come up with thousands of pages, which might have been arranged according to commercial goals. Such system provides a large set of possible answers arranged based on the keywords of the user's question and it's the user who should browse through this massive set and find the true answer, if there is any. Frequently, retrieved information differs vastly with the users intended meaning. On the other hand, it's hard on most users to find the appropriate answers to their questions from among the massive information and it's necessary that they have the required skill and experience for changing a question into a few key-words. In contrast to this technology, there might be a QAS which is able to get the user's question as a question in a natural language and extract the appropriate answer with a minimum redundancy and a maximum accuracy. From one perspective, QAS can be divided into two categories of open and specific fields. The open-field QAS changes are yearly reflected in the text retrieval conference (TREC). By definition, it should be able to answer the general questions with referring to a predetermined large set of texts. In contrast, the specific field systems are used for specialized fields such as medical dataset. QC is one of the important processes in QASs, that is semantically categorizes each question according to the types of the answers. For example, the question "who first went to the moon?" ("چه کسی برای اولین بار به کره ماه رفت؟") is categorized as the "human" type of answer because it is perceived based on its answer, and then it is labeled as classified. The same is correct for, other questions about places, colors, animals, etc. After classifying a question, the searching system browses for those kinds of paragraphs, which belong to the related type of answer so that the appropriate answer is extracted.

This study is to present a QCS for Persian language. First, to train machine learning model a dataset is needed. Therefore, after gathering Persian questions, in supervised learning method, questions must be classified by an expert into fine and coarse categories. On the feature extraction step, different features part of speech (POS), question informer (QI), question word, tokens of question and their positions in each sentence, are extracted from each question. The labels used in the current study are mentioned in Table I. To train the QCS the CRF machine learning model has been used. Each word in a question is placed on a separate line and in front of each token, there are their corresponding features on different columns. On the last column, the appropriate answer type which has been chosen by an expert is mentioned repeatedly for each token of the question. the next question is inserted at the same rules after rendering a blank line in dataset. Throughout the dataset the answer-type labels are repeated for every token in a question. However, in this situation some ambiguity occurs, when the classifier trays to predict unique label for each question. In other words, due to the existing uncertainty conditions, there might be different answer-types to some questions. As a solution, we use a majority voting on the predicted labels. The one which is repeated the most frequently is chosen as the final label. Thus it is possible to have a unique answer-type label for each question. Finally, after the data was provided, the CRF machine learning model is trained, and its accuracy is evaluated by intended testing dataset.

TABLE I : COARSE AND FINE GRAINED QUESTION CATEGORIES.

Abbreviation(14)	Abbreviation(3) , explanation(11)
Description(2781)	Manner(464), Reason(728), definition (437),description(1152)
Human(412)	Title(14), person(238) , Job(1), Speech(31), group(111),other(17)
Location(404)	City(60), country(53), Sea(33),side(1),state(9), mountain(27),other(221)
Numeric(437)	Num(114),date(155), period(30), length(27), percent(37), weight(4), money(28), Temperature(6), Size(2), Rank(3), height(1), distance(4), count(12), code(5),other(9)
Entity(952)	Word(12), vehicle(4), tools(39),term(67), object(3), multimedia(137), material(102), linguistic(18),language(23), knowledge(1), Literature(1), Symbol(9), Symbol(8), Religion(64), Product(33),plant(18),food(20), event(37), dolor(8), body(48), animal(34), medicine(33), method(32),other(201)

Wei in [1] proposed a classifier based on support vector machine (SVM), and he mentions following features as the features which are used in the classifier: "interrogative word; primary sememe, which is in HowNet, of first-degree and second-degree dependent word of interrogative word and named entity and singular/plural features" as well in feature extraction phase. Dongwei in [2] presented a QC method based on improved rules, by selecting seven common types of questions. Zhang in [3] used a SVM machine learning model in QC with sentence word, POS, named entity and semantics features for training classifier. Xia in [4] adapted three strategies to extract classification features: (1) using focused words in a question, (2) Using a domain attribute, (3) Using the binding of two domain attributes. They have reported a two-step classifier, encompassing rule based and SVM classifier. Nguyen in [5] used bag-of- words as features for their all experiments. They proposed semi-supervised learning for improving the accuracy of QC. Hejazi in [6] used an ontological rule-based classifier for determining answer-type labels to questions. The labels are considered as a question target and the identification of the question types is emphasized in relation to Persian ontology. Mohammadi-janghara in [7] used a combination of "uni-gram" and "bi-gram" model in a QAS in the field of "biography". The words have higher differentiation and they are related to special categories, such as birth dates, are extracted from the question after being weighed precisely. Then, they turn into some special dictionaries form files. Next they are saved as key-words for each question type. Based on different answer-types namely, short, descriptive and listing-Persian questions they are classified. In [8] VSNOW machine learning model is used, based on a hierarchical classification of English questions with six coarse and 50 fine categories. Wang [9] used semantic grams and SVM learning model for classifying Chinese questions. They reported 20 percent accuracy increase with using semantic Uni-grams and Bi-grams, as compared to the usual form using N-grams (Uni-grams and Bi-grams). Instead of using a binary vector, Huang in [10] used a SVM, based on the word-weighing method. Word-weighing is applied with pre-processing step on the data, according to the idea of entropy in information retrieval. Lee in [11] managed to classify questions by using the SVM learning model and features like question informer (QI), Bi-grams, the first word of the question, the first two words of the question, and wh-question words. In [12], some actions such as word segmentation, key-word and head phrase extraction as well as some semantic features such as the HowNet and some syntactic features, are suggested as important steps in feature extraction. Metzler in [13] suggested n-gram functions, parts of speech (POS), semantic Word Net and Name Entity Recognition as important features for system training. In [14], a QCS is trained, using some two-layer forward neural network with back propagation. The Significant features used in that study are the Query-Text Relevance, the average word and phrase frequency, question length and word and phrase variance diffusion. Li and et al. has proposed by [21] the classification of the what-type questions. They just consider the nouns as semantic words. Each word in question tagged as label using conditional random fields model, and the head noun's label is chosen as the question category. The features such as words, part-of-speech, chunker, parser information, question length, name entity, hypernym, synset and transition features are used for training English what-type questions classifier.

The rest of the paper is organized as below. Section II describes background of paper. Section III presents feature extraction step. In Section IV, we present our classifier structure and Section V describes discussion and results. The Conclusion is presented in section VI.

II. BACKGROUND

A. *Question Classification*

Question Classification in the current study means the process of writing $g{:}X \rightarrow \{c_1, c_2, \ldots, c_n\}$, where each $x \in X$ is a question and it mapped into one of the n classes [8]. Each of the classes is distinguished on the basis of semantic limitation. Questions are classified in two levels according to the type of the answer to the question as illustrated in Table I. The first level (coarse level) consists of six coarse categories. For example, the question “In which city is Azady Square located?”, is about the “location” on the first level and it is about ‘the name of the city’ on the second level (fine level). Thus, regarding the expected answer to this question, i.e., “Azady square is located in Tehran (“میدان آزادی در تهران قرار دارد.”) it would be classified as: (Location: City).

One of the probable problems in question classification is lack of determined limitations on the range of the possible answers to a question, which might cause ambiguity. For instance, the question “What is Diabetes?” or the question “What is the PH level of water?” can belong to both the “Definition” and “Medicine” categories. So, these two questions cannot be seen as belonging to a unique category, but they can be a member of both “Definition” and “Number” categories. Therefore, these questions do not belong to a single category. Throughout the current study, it has been assumed that there is only one answer type to each question. To solve this problem, however, it is possible to set the classifier so as to make it consider several types of categories for each question. Nevertheless, according to classifier type, a classifier can select classes with higher weighing label in the classification in semantic sense.

B. *Conditional Random Fields Machine Learning Model*

Structures such as the combination of sliding windows with artificial neural networks or Hidden Markov Models are among the most straight-forward machine learning models in the field of sequential data, while there is internal interaction within the data and there is a high dependency in the observed sequence elements. However, these kinds of machine learning models show less efficiency than their next generations. To meet these limitations, Markov’s Maximum Entropy Model is introduced. There is also some sort of limitation to this model, so called “Label Bias”. As a result, the CRF Machine Learning Model is introduced to cover the limitations which existed in the previous models. Lafferty [15] used the CRF structures as a framework for segmentation and labeling sequential data.

A CRF designs the probability p(x|y) using following: a Hidden Markov Model nodes corresponding with elements of the “y” object, and potential functions. Potential functions are considered as conditions on “x” features. In this case, the training process takes place with (x, y) pairs being used for setting different parameters for the likelihood maximization. The CRF is mostly used in sequential training issues, such as NP chunking, POS, Tagging and Name Entity Recognition. Recent studies [15,16] and [17] support the idea that the CRF is superior to other structures such as Markov’s Hidden Model and Maximum Entropy [18] when facing redundancy features.

In the case of Machine Learning Models e.g., the SVM binary vectors should be used for system training. In other words, vectors cannot be used with sentence tokens. Therefore, all of the sentence tokens should be converted into binary rates. On the other hand, there is internal coherences for all of the tokens in a sentence, their sequence in a sentence, and the corresponding features of the tokens are of high importance. Unfortunately, when a sentence converted to a binary vector, some of the existing coherence among the tokens and corresponding features would be lost. So is necessary to use the models that would be able to learn the most coherence between existing features. Regarding sequential data, models such as CRF has a superior over other machine learning models.

C. *dataset*

Like other supervised learning models, we need a dataset for training classifier, in this case, data collected in two steps. The first step is to gather the initial dataset and the second step, is to provide the final dataset. The initial dataset includes questions on each row, and there are three labels namely, the coarse category, the Fine Category and the QI for each question on the same row they are extracted manually. The initial dataset, however, is not appropriate for the training machine learning model because it cannot have essential features. Thus, another dataset would be needed; The final dataset is one to which we can add the needed features by using feature extracting functions. There are a row and several columns for each token. In other words, each token is placed on a row and in front of that, on the next column, there comes its corresponding features. On the last column, the label corresponding with the whole sentence is repeated for all the tokens of the question. In this dataset a matrix is formed using the features and labels of each question. In addition, every question is separated from the other ones with a blank line.

In this study, six coarse grained classes of question as well as fine grained classes are used as they have been defined in TREC. Of course in this research we extend numbers of fine categories with respect to gathered Persian questions. Nearly 70% of the Persian questions are gathered from the primary and junior high school materials, and the rest of them, from FAQ in several websites. Table (1) illustrates the Coarse and Fine level labels, and also the frequency of the sentences in each category.

III. FEATURE EXTRACTION

Different features are used to train question classifier, such as term weighing, co-occurrence, and key-term extraction, which are all considered as statistical features. Generally, regarding the question classification issue, the mere use of above mentioned statistical features are less preferred than of semantic and syntactic features [13]. The following features are used in the present study for training CRF classifier.

A. Question Informer

While answering a question, it is classified based on a few words. For example, when hearing the question “Who wrote ‘Shahname’?” (“شاهنامه توسط چه کسی نوشته شده؟”) The listener prompts that he should look for ‘a person’ who has written ‘Shahname’. Similarly, he knows that the answer to the question “In which city is Azady Square located?” is a ‘city’ where Azady Square is located. Therefore, there is always a word or some words in a reasonable question that would determine the answer-type or its categories. These certain words suggest the target of the question and generally show what the question is asked about, which is called QI. For extracting QI in Persian, we have used the tool that implemented in [19]. There are rules about manual QI extraction, which are mentioned below:

- Whenever the question word appears as a interrogative adjective, the related noun or noun phrase with it question word, would be chosen as the QI or its substantive. Example: “In which city is Azady Square located?” (“میدان آزادی در کدام شهر قرار داد؟”) “City” is considered as QI.
- In case of using such interrogative pronouns such as ‘when’, ‘who’, ‘where’, ‘why’, ‘who’, they are used as QI themselves. Example: “When did the European Economic Depression begin?” (“بحران اقتصادی اروپا کی شروع شد؟”) “When” is considered as QI.
- In Persian language, in case of other question words or whenever the verb comes immediately after the question word, QI would be chosen through semantic relation analysis. Usually the head of noun phrase in the question is concerned. Example: “what does the definition of ‘Morphology’ say?” Of course, in Persian the word order is different from that in English. The same question would be as follows: “The definition of ‘Morphology’ what says?” (“تعریف مورفولوژی چه می گوید؟”). In this case, “morphology” is considered as QI.

B. Words in a Question

The words in a question and their arrangement form the main structure of the question. A change in the arrangement of the words may lead to change the whole meaning of the question. So, in addition to the significance of the words meaning, their arrangement is very important in feature matrixes. Other features would be extracted through syntactic and semantic analysis from the question.

C. Question Words

Question words are considered to be another important semantically features. All of the words in a question would be verified against a list of existing question word and if a word is confirmed to the existing list, it word would be chosen as a question word. In case there are two question words, in Persian the one on the nearest right would be superior.

D. N-gram

Normally, every word in a sentence is separated from its previous and the next words by a space. Frequently, however, there are two or more words coming one after another, separated by spaces, and altogether they have just one single meaning. Therefore, there is a need to identify such words in a question, which are composed from many morphemes, but in fact they are considered a word with a unique meaning that is defined as N-gram. The N-gram feature would be used to extract such cases. In the current study, the Bi-gram and Tri-gram features have been used.

E. Part of speech, before and after a word

Part of speech (POS) is one of the most frequently used features in extraction, information retrieval, and data-analysis systems in the field of natural languages. This feature specifies syntactically structure of word in sentence. In the current study, the syntactic labels of each word, as well as those of its previous and its next words have been used in the training system. We have used the tool mentioned in the [20] for this purpose.

F. Other features

Besides the mentioned features, some others are used as feature matrix as well including the first-level classifier output (the Coarse Category) which is used as a feature for training the second level classifier (the Fine Category), the position of each token and interrogative word in a question (refer to table II). Like other features, these ones are considered as complementary to the Matrix of Features.

IV. CLASSIFIER STRUCTURE

A. CRF Training and Its Combination with Voting

The classifier receives a Persian question as the input, and according to received learning from the training dataset, classifies the question to identify the answer type. Generally, it does the map g: x → $\{c_1, c_2, \ldots, c_n\}$, Where X stands for the set of questions and $\{c_{۱}, c_{۲}, \ldots, c_{n}\}$ are a number of pre-determined answer-types. In this study, the CRF model has been used for every classifier.

There are two levels for answer types. The first level (The Coarse Category) which consists of six elements is semantically more general than the second level (The Fine category) which consists of 58 elements. Therefore, regarding the hierarchical semantic relationship, some sort of special classifier is needed in which the output on each level would be used as the input on the next level. The first classifier provides the coarse class labels and the second classifier, provides the Fine class labels. In the data collection phase, each question was classified by an expert into two levels of Coarse and Fine. Using mere question words for training the CRF, learning model would not be precise enough. Thus, other features also needed to be extracted from the question. Each token of question in the dataset is placed on a separate row and on the first column. The extracted features of the token are placed in different columns on the same row. The features which are specific for a question will be repeated in all columns corresponding to question tokens. The answer-type label of question is also repeated in the last column for all tokens. At this stage, there are many matrixes, one for each question in the dataset, each consisting of a question with its features. The feature matrixes are separated from each other with a blank line. Next, the original dataset is divided into two sets of "Training data" and "test data" and they consist 4500 and 500 questions, respectively.

In this system, The Fine level classifier is nearly the same as the coarse level in terms of features, except for the fact that the coarse level labels are considered as a new feature column for the Fine level classifier training. That's because the Fine level labels are semantically considered to be a sub category of the coarse level. Therefore, the use of the coarse level labels in the Fine level classifier feature columns would lead to a higher level of accuracy. With paying attention to "predicted Fine label" column of Table II, the classifier predicts a label for any token of question. These labels may not be properly equal to each other for all tokens in a question. Thus, for solving this problem we use majority voting method by considering probability nature of CRF model. This means that the label is selected which has a high frequency in interrogative sentence. On implementing stage, we use majority voting method in two classifiers for extracting final label.

TABLE II : MATRIX OF FEATURES FOR ONE QUESTION

Question tokens	**Token position**	**POS**		**Interrogative word position**	**QI**	**Interrogative word**	**Coarse label**	**Predicted Fine label**	**A Label selected by majority voting**	
میدان	0	N	SING	3	شهر	کدام	Loc	City	→	City
آزادی	1	ADJ	CMPR	3	شهر	کدام	Loc	Other	→	City
در	2	P	-	3	شهر	کدام	Loc	City	→	City
کدام	3	N	SING	3	شهر	کدام	Loc	City	→	City
شهر	4	N	SING	3	شهر	کدام	Loc	City	→	City
قرار	5	ADJ	SIM	3	شهر	کدام	Loc	Other	→	City
دارد	6	V	AUX	3	شهر	کدام	Loc	Other	→	City
؟	7	DELM	-	3	شهر	کدام	Loc	City	→	City

V. DISCUSSION AND RESULTS

Table III illustrates the effects of different features on the classification accuracy. The results suggest that regarding the predetermined categories, lack of using the semantically rich features, would lead to a significantly decreased accuracy. For example lack of using QI feature which is more semantically valuable than the other ones, would lead to more decrease of accuracy, while in the case of the syntactic feature, the amount of the accuracy decrease would not be the high. However, it is not suitable to remove any of them in the features Matrix since some of them may be effective although so slightly.

TABLE III : EFFECTS OF DIFFERENT FEATURES IN CLASSIFICATION ACCURACY

Not used features in training and test	**Accuracy %**
QI	75.26
Token in the question	76.92
Interrogative word	79.17
3-gram	79.71
Next and before POS of word	79.72
POS	79.73
2-gram	79.85
All features	79.84

The first column on Table IV illustrates the results of implementation in both Coarse and Fine categories, which equal to 75.26% and 70.20% respectively. The Percentage of a predicted correct label to total predicted labels is used as a criterion for accuracy assessment. As it is shown, second level classification accuracy is less than that of first level. It is because of the fact that second level classification needs more semantic features extraction, such as the name entity recognition and the list of the semantically words dictionary related to the Fine labels.

TABLE IV : RESULT OF CLASSIFIERS ACCURACY

Classifier	Without QI feature	With QI feature	Using majority voting
Coarse classifier	75.26	80.35	87.20
Fine classifier	70.20	78.71	85.31

The second column of Table IV provides the Persian question classification outputs regarding the QI, which suggest the desired effects of this feature on the classification accuracy.

Its effects are more significant on increasing the Fine category accuracy, which confirms the sensitivity of this

Classifier to name entities because the QI focuses more on the noun phrases of the question. The third column on Table IV illustrates the output as integrated the CRF Model with a majority voting. There should be a unique label corresponding to each question. But, the CRF model specifies the answer-type label of any question for each token in that question. Thus, according to the probability-based nature of this model, different labels can be predicted for the words in a question. This may lead to some ambiguity in choosing the final label. As a solution, a majority voting is applied on the CRF Classification output, which leads to ambiguity remove, and the final system accuracy increase.

VI. CONCLUSION

In comparison with similar research such as [21], our research focused on Persian question informer, open-field classification and integrating majority voting with CRF to improve the accuracy of the coarse and fine Persian classifiers.

The results suggest that QI would have very satisfactory effect on the final accuracy of Persian question classifier. In addition, improving the sliding window of features template in the CRF model improves the final accuracy, too. The semantic Persian WordNet, name entity recognition, manual list of related words to each category, and synonyms are very important features in this field.

Accuracy in POS tagging, identifying QI, spelling, manual labeling and a few noises would influence in the final accuracy of the classifier directly. Another improving factor for the system is the enrichment of the data. Some of the categories in the dataset have very few training examples and it decreases the data comprehensiveness. Thus, it is needed to add some new question to the dataset for some categories, especially on the Fine level.

ACKNOWLEDGMENT

The authors would like to thank Mrs. Mahbobeh Behrooz, Mr. Majid Ghafuri in Islamic Azad university Mashhad branch for providing some applied programs.

REFERENCES

[1] WEI, Z., JUNJIE, C., AND YANQING, N. 2011. Research on Chinese Question Classification Based on Hownet and Dependency Parsing. intelligent systems and applications (ISA), 3rd international workshop. Wuhan.

[2] DONGWEI, G., CHENHUI, J., MIAO, L., AND KANGPING, W. 2011. Chinese Question Answering System Based on Question Classification and Similarity. International Conference on Information Science and Technology, Nanjining, Jiangsu, China.

[3] ZHANG, K., AND ZHAO, J. 2010. A Chinese Question-Answering System with Question Classification and Answer Clustering. International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). Yantai, Shandong.

[4] XIA, L., TENG, Z., AND REN, F. 2008. An Integrated Approach for Question Classification in Chinese Cuisine Question Answering System. Second International Symposium on Universal Communication IEEE DOI. Osaka, Jpan.

[5] NGUYEN, T. T., NGUYEN, L. M., AND SHIMAZU, A. 2007. Improving the Accuracy of Question Classification with Machine Learning. Research, Innovation and Vision for the Future, IEEE International Conference on 04 June 2007. Hanoi.

[6] HEJAZI, M., MIRIAN-HOSSIENABADI, M., NESHATIAN, K., AND AFGHAHI, E. 2004. Question answering system based on ontology in telecommunication field with automatic classification and extraction documents. journal of research and scientific institute in computer engineering.

[7] MOHAMMADI-JANGHARA, M., AND ANALUEI, M. 2009. proposing a question answering system with question categorization and candidate sentence by using of n-gram. 14th international computer society of Iran computer conference (CSICC). Tehran.

[8] LI, X. AND ROTH, D. 2004. Learning Question Classifiers. Proceedings of the 19th International Conference on Computational Linguistics.

[9] WANG, L., ZHANG, H., WANG, D., AND HUANG, J. 2009. Chinese Question Classification Based on Semantic Gram and SVM. 1 State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China ,International Forum on Computer Science-Technology and Applications.

[10] HUANG, P., BU, J. J., CHEN, C., AND QIU, G. 2007. An Effective Feature-Weighting Model for Question Classification. College of Computer Science, Zhejiang University, Huangzhou, China IEEE DOI.

[11] LEE, C. W., DAY, M. Y., SUNG, C. L., LEE, Y. H., JIANG, T. J., WU, C. W., SHIH, C. W., CHEN, Y. R., AND HSU, W. L. 2007. Chinese-Chinese and English-Chinese Question Answering with ASQA at NTCIR-6 CLQA. Proceedings of NTCIR-6 Workshop Meeting, May 15-18. Tokyo, Japan.

[12] Liu, B., HAO, Z., YANG, X., AND LIN, X. 2006. Chinese question classification with support vector machine. international journal of computer science and network security (IJCSNS).

[13] METZLER, D. AND CROFT, W., B. 2005. Analysis of Statistical Question Classification for Fact-Based Questions. Springer Science + Business Media, Inc. Manufactured in The Netherlands. University of Massachusetts, Amherst- Information Retrieva. 8, 481–504.

[14] FEI, T., AND HENG, W. J. 2003. Question Classification for E-learning by Artificial Neural Network IEEE. Institute for Infocomm Research National University of Singapore.

[15] LAFFERTY, J., MCCULLUM, A., AND C.N.PEREIRA, F. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. Proceedings of the Eighteenth International Conference on Machine Learning, pp. 282-289.

[16] SARAWAGI, S., AND COHE, W. W. 2004. Semi-markov conditional random fields for information extraction. Advances in Neural Information Processing Systems. 17. pages 1185–1192.MIT Press, Cambridge.

[17] FEI, S. AND PEREIRA, F. 2003. Shallow parsing with conditional random fields. Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (HLT-NAACL 2003). Edmonton, Canada.

[18] MCCALLUM, A., FREITAG, D., AND PEREIRA, F. 2000. Maximum Entropy Markov Models for Information Extraction and Segmentation. Proceedings of the 17th International Conference on Machine Learning (ICML). pp. 591-598.

[19] MOLLAEI, A., RAHATI-QUCHANI, S., AND ESTAJI, A. 2010. persian question clasification acuracy improvement by using question informer based on conditional random fields model. the 4th iran data mining conference (IDMC). sharif university of technology. Tehran, Iran.

[20] RAHATI-QUCHANI, S., AZIMZADEH, A. AND ARABMOHAMMAD, M. 2007. part of speech tagging based on hidden markov model. 13th international computer society of Iran computer conference (CSICC). Island Kish, Iran.

[21] LI, F., ZHANG, X.., YUAN, J. AND ZHU, X. 2008. "Classifying What-type Questions by Head Noun Tagging", Proceedings of the 22nd International Conference on Computational Linguistics. pp. 481-488.