Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Using Question Classification to Model User Intentions of Different Levels

Yaoyun Zhang, Xuan Wang, Member, IEEE, Xiaolong Wang, Member, IEEE, Shixi Fan, Daoxu Zhang
Intelligence Computing Research Center, Harbin Institute of Technology Shenzhen Graduate School
xiaoni5122@gmail.com, wangxuan@insun.hit.edu.cn, wangxl@insun.hit.edu.cn

***Abstract*—User information need detection is a fundamental issue in automatic question answering systems. Based on real questions collected from on-line question answering communities, this paper proposes a three-level question type taxonomy to model user information need. The three levels are based on interrogative patterns, hidden user intentions and specific answer expectations. One question can have multiple types in level 2&3. Question type assignment of level 2&3 is subjective-orientated, and may vary between different users. Shallow lexical, syntactic and semantic features are used to model the inherent subjectivity of user intentions. Classification experiments are conducted on a corpus of real questions collected from the web. Different machine learning methods are employed. Experimental results are promising. This indicates the capability of modeling user information need and subjectivity statistically, and that strong correlations exist between question types of the same level.**

***Key words*—User Intention modeling, three-level question type taxonomy, pragmatics, question answering, multi-label classification**

I INTRODUCTION

Community question answering (CQA) service has been increasingly popular in recent years. Major CQA portals such as Yahoo! Knowledge[1] and Baidu Knows[2] provide a platform for users to post questions and answer others' questions. This can be seen as an analogy with human-human communication of relative naturalness. The huge archive of questions and answers formed in those portals gives us a chance to examine the complexity and characteristics of user intentions and expectations delivered through questions.

Question type (*QType*), defined to represent the information need (or user intention, these two notions are used interchangeably in this paper) of the question, is often used in question similarity calculation or answer extraction process. Usually, one question is categorized into one certain type. In current question type taxonomy, questions fall into two major categories: 1. Factoid questions asking for simple facts (time, quantity, etc.), and taking named entities as the answer; 2. Complex questions asking for relationships between concepts, procedures, reasons etc., and taking descriptive paragraphs as the answer.

However, when observing abundance questions and accepted answers in CQA service, we are surprised to see that the accepted answers to not only the complex questions but also most of the factoid questions are paragraphs, instead of simple named entities. After analyzing these answers, we summarize three factors need to be considered to interpret and satisfy the user's information needs: 1. *Surface user intention,* which can be told directly by the interrogative patterns of questions. It corresponds to the conventional definition of *QType*; 2. *Coverage of User Intentions*. Many questions contain hidden user intentions untold by the surface text, which should also be detected and met with, to make the answer complete and persuasive; 3. *Specific Expectation from the Answer*. For example, some answers are subject to temporal constraints and vary in different temporal phases; some questions contain detailed personal information, and require customized answers; some answers are subjective-oriented while some are objective-oriented, etc.

This can be illustrated by the question in table 1. Judging from the interrogative pattern, the question is asking for *time*. However, the *reason* why the specific time is picked should be provided as the supporting information to make a convincing answer. Furthermore, the financial crisis should refer to the one happened *recently* instead of other time. *Expertise* is required to give a reliable answer. *Subjective opinions* are also expected to be shared. Clues to these intentions and expectations can be found both in the question and its accepted answer.

This paper models the three essential factors as three-level question type taxonomy (*QTaxo*). As shown in figure 1, for a question, its *QTypes* in level 1(*L1*) represents the surface user intention; and *QTypes* in *L2&3* correspond to *user intention coverage* and *answer expectation*, respectively. A question has only one type in *L1*, but can have multiple types in *L2&3,* as the case in table 1.

TABLE I. EXAMPLE QUESTION FROM CQA SERVICE[3]

Question *Q: How long* will the financial crisis last?
Accepted Answer *Because* the investment rate of America remains to be low…With a low productivity, the income can't keep pace with the increasing population…*In my opinion*, if sensible measures can be carries out, by *the end of the next year*, situation will become better…
L1: time ***L2***: time, reason ***L3***: temporal constraint, subjective, expertise

[1] http://answers.yahoo.com
[2] http://zhidao.baidu.com
[3] http://zhidao.baidu.com/question/77711706.html?si=2

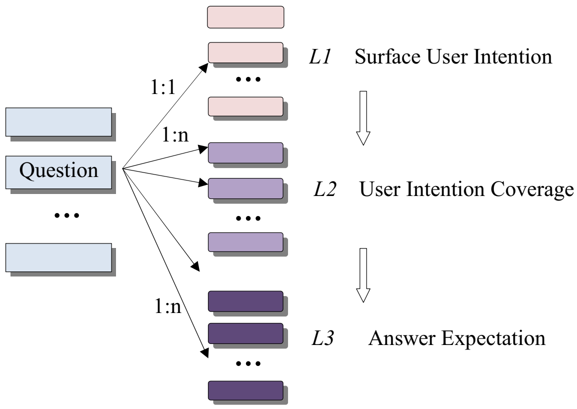

Figure 1. Three-Level Question type taxonomy

The idea of *QTaxo* in *L2 &3* is novel. To the best of our knowledge, no question type taxonomy for user intentions coverage and answer expectations has been proposed before. Nor has a real question been categorized into multiple types before.

As stated before, *L1* of *QTaxo* has already been applied in QA system. *L2 &3* also have a potentially wide range of usages. For example, *L2* can be used to refine the question search space formed by *L1*, or help to make answer summarization strategies. *L3* can be used in quality evaluation of questions and answers, and assigning different weights to information from multiple knowledge sources.

However, the task of accurate recognition of its QTypes in *L2 &3* is non-trivial: a. Real user questions posted in CQA are often complex and ill-formed. b. The mapping between a question and the types in *L2 &3* are subjective-oriented, different users may have different standards and requirements. c. QType recognition in *L2&3* is a multi-label categorization problem, which is inherently much harder than single-label categorizations.

Thousands of real questions are collected from the CQA services and labeled with *QType*s in the *QTaxo*. This paper exploits the capability to build automatic classifiers for each level of the *QTaxo*, with a focus on *L2&3*. Various features, including lexical, shallow syntactic and semantic features are extracted and evaluated. Different types of machine learning methodologies and multiple-label problem transformation methods are implemented and compared. Specifically, this paper aims to answer the following questions:

1. Is there any correlation between question types in the 3 levels?
2. Given a specific question, what features indicate the mapping between it and the question type taxonomy?
3. Is it possible to model the subjectivity in L2 and L3, and to categorize questions into them automatically? If so, to what extend can it be modeled?

The remainder of this paper is organized as follows. Section 2 formalizes the 3-level question type taxonomy and defines each *QType* in it. Section 3 introduces the corpus tagging process and looks into the relations between *QType*s of the 3 levels. Section 4 describes the question categorization experiments and discussions. Related works are reviewed in section 5. Sections 6 are conclusions and future work.

II QUESTION TYPE TAXONOMY

A. Problem Formalization

Define T as the 3-level question type taxonomy,

$T=(C_1, C_2, C_3)$, where $Ci=\{ci_1, ci_2, \ldots, ci_n\}$ is the set of question types in the ith level of T, $i\in\{1,2,3\}$.

For each question, *QTypes* of level 1 represent the surface user intention, which correspond to the conventional definition of *QTypes* in previous work; *QTypes* of level 2 represent all the user intentions, explicit and implicit; *QTypes* of level 3 represent the user's specific expectations from the answer.

Define $Ci(q)$ as the set of types of question q in the ith level of T,

i.e., $Ci(q) \subseteq Ci$, subject to

$|C_1(q)|=1$, $|C_2(q)|\geq 1$, $|C_3(q)|\geq 1$.

The *QType*s in $Ci(q)$ constitute a *type combination*.

B. Question Types Definition

There is no standard question type taxonomy for on-line real questions currently. Extensive observation on raw questions and survey into previous works on question classification and interpretation (see related work in section 5) are conducted. We propose the following question type taxonomy, in an effort to cover the general information needs of users.

a) Question Types in Level 1

14 *QType*s are defined in *L1*: *choice, definition, contrast, description, procedure, reason, absEntity, yesNo, location, person, time, quantity, entity* and *emotional,* as listed in table II. The first 8 types are complex questions, and the latter 5 are factoid questions. The interrogative pattern of *absEntity* is similar to *entity*. However, *absEntity* represents the information need for abstract concepts such as "law", "policy", "theory", etc. *emotional* depicts a special phenomenon of CQA services. Some users treat the CQA service as a BBS, and post emotional statements to elicit public discussion. Since

TABLE II. QUESTION TYPES AND EXAMPLES IN LEVEL 1

Question Type	Example
choice(cho)	Is the version of Office software in Vista OS 2003 or 2007?
definition(def)	What is financial crisis?
contrast (con)	What is the difference between skiing and skating?
description(des)	How is the performance of Vista?
procedure (pro)	A freshman in college, how to manage my time?
reason(rea)	Why do we have to expand domestic demand?
yesNo	Will the price of 3G mobile phone drop?
absEntity (absE)	What is the national policy on higher education?
location (loc)	Where to buy the ¥10 prepaid cards?
person(per)	Who is Kobe Bryant?
time	How long should I rest after exercise to have dinner?
quantity (qua)	How much is the annual fee of the Internet bank?
entity(ent)	Which book is good to read?
emotional(emo)	The medical costs nowadays are so high!.

these postings cannot be counted as questions, they are recognized in *L1* and then filtered out before question categorization in *L2 &3*.

b) Question Types in Level 2

L2 shares 13 types with *L1* expect the "emotional" type.

c) Question Types in Level 3

As illustrated in fig.2, 8 *QType*s are defined in *L3* according to 4 perspectives. 1. Specification: whether the question contains detailed personal information as the context or not; 2. Knowledge source: whether the question requires commonsense or expertise to answer; 3. Temporal constraint: whether the answer is constraint to a specific temporal phase; 4. Orientation: whether the opinions in the answer is subjective-orientated or objective-orientated. Examples of some question types in *L3* are listed in table III.

III QUESTION CLASSIFICATION ALGORITHM

A. Feature Set

Notations:

W_i: the word in question q;

$InteroW_i$: the interrogative word of q;

$Hyper(W_i)$: the direct hypernym definition of W_i in the Chinese electronic dictionary Hownet[4].

Features used for question categorization are as follows:

Word *Uni-*, *Bi-*, *Tri-G*rams (i.e., W_i, W_iW_{i+1}, $W_i\ W_{i+1}W_{i+2}$)

*Long-dis*tance Interrogative *P*atterns: long-distance phrase collocations and question templates collected manually, e.g., "Is … better than … "

*S*tructural *F*eatures

Word SF: $W_{i-1}InteroW_i$, $InteroW_i\ W_{i+1}$, and the first and the last word of q.

*P*art *o*f *S*peech *SF*: $POS_{i-1}InteroW_i$, $InteroW_iPOS_{i+1}$, and the numbers of verbs, nouns and adjectives of q.

*Sem*antic *SF*: $Hyper(W_i)$, $Hyper(W_{i-1})InteroW_i$, $InteroW_iHyper(W_{i+1})$

Feature selection is conducted for each level separately by *IG* [1] provided by *Rainbow*.

TABLE III. EXAMPLES OF QUESTION TYPES IN LEVEL 3

Question Type	Example
highly-specific	My mother was 55. She suffers from chronic gastritis for about 5 years,…, is there any way to alleviate it?
commonsense	What is software design?
expertise	How to make wise investment?
temporal-constraint	How much is the cutting score of the college entrance examination this year?
subjective	What do you think of the recent performance of Yaoming?

[4] a Chinese Counterpart to Wordnet. http://www.keenage.com/

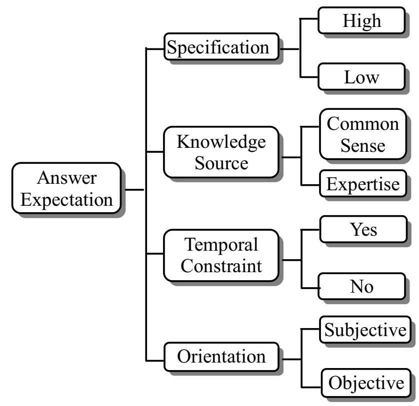

Figure 2. Four perspectives of answer expectations

B. Multi-label Classification Problem Transformation

Define the classifier as H:X→*C*. If |*C*|=1, then H is a single-label classifier; if |*C*|⩾1, then H is a multi-label classifier. Most machine learning classifiers are designed for single-label classification problems. Thus, multi-label classification problems are usually transformed into multi single-label classification problems.

Two common transformation ways are compared in this paper:

1. *TypeCombine*: Treat each *type combination* (defined in section 2.1) of *Li* that appears in the data set as a single type and learn a single-label classifier H:X→TypeCombine (*Ci*);

2. *BinaryC*: Build a binary classifier for each type c_j□C_i, Hc_j: X→{c_j, ¬ c_j} .

TypeCombine and *BinaryC* correspond to the transformation methods *PT3* and *PT4* described in [2], respectively.

C. Employed Machine Learning Algorithms

Three types of classification algorithms are examined: 1. The generative model represented by Naïve Bayes (NB) [3]. 2. The determinative model represented by SVM [4]; 3. The associative rule-based model represented by Classification based on Predictive Association Rules (CPAR) [5].

NB and SVM are implemented via the Rainbow toolkit [6] using default parameters. CPAR is implemented via the IlliMine project[5]. The parameters of CPAR are set as in [5].

IV EXPERIMENT

A. Data Set

Baidu Knows is the largest CQA portal in China. Over time, it has accumulated more than 450 million question and accepted-answer pairs. 25,000 questions of Baidu Knows are chosen randomly from 5 categories: business/financing, computer/internet, education/exam, health care and sports. One posting in the CQA service usually contains two parts: question title and description, which is the detailed extension

[5] http://illimine.cs.uiuc.edu/

of the title. Only the question title is used for analysis and experiments. Some question titles convey no concrete information need, such as “A question about losing weight” and “I am a newcomer in trade”. Some question titles contain multiple questions, such as “What does IT stand for? Is it still hot now? ”. These question titles and redundancies are removed. 17,691 questions are left in the corpus. One question consists of 15 words on average. 14,418 different words appear in the corpus.

B. *Corpus Tagging*

5 people participate in the corpus tagging. Question type definitions and the purpose of building the question type taxonomy are explained to them. The corpus is divided into 5 parts evenly. Each part is tagged by 3 people independently. Each person tags 3 different parts of the corpus.

Level 1 is tagged first. When divergence between the labelers occurs, a fourth people will be asked to make the final decision. 1,321 questions are tagged as the “emotional” type.

For level 2&3, the labelers are free to choose any question type they consider as sensible. The labels from the three persons are combined as the final labels of a question. The reason is that deciding the types of a question in level2&3 is a subjective process. Different people may have different information needs and expectations from the same question. We try to build a corpus general enough to preserve those divergences, i.e., to cover the inherent subjectivity of level2&3. Even if so, question types of the corpus still have strong statistical regulations, as proved later in the experiments.

The training and test collections are selected randomly under the constraint that the number proportion of each question type between the training and test collections is approximately 4:1. Different training and test corpus are formed for each level of the question type taxonomy.

C. *Evaluation*

Different criteria have been used in previous work on multiple-class recognition for performance evaluation. The KDD Cup 2005 [6] evaluation criteria, deemed to be authoritative, are employed in this paper. Precision (P), recall (R) and F1 score are defined as follows:

$$P = \frac{\sum_i \#\,\text{correctly classified as } c_i}{\sum_i \#\text{classified as } c_i} \tag{1}$$

$$R = \frac{\sum_i \#\,\text{correctly classified as } c_i}{\sum_i \#\text{ labeled as } c_i} \tag{2}$$

$$F1 = 2\left(\frac{1}{P} + \frac{1}{R}\right) \tag{3}$$

[6] www.sigkdd.org/kdd2005/kddcup.html

D. *Performance*

a) *Selected Feature Distribution Analysis*

The top 500 features selected by IG for each level are examined. Fig.3 shows the distribution of different types of features.

Feature distribution of *L1&2* are very close. In fact, over 90% features of *L1&2* are the same, indicating the strong correlation between *QType*s of *L1* and *L2.* A sharp contrast between *L3* and the first two levels can be observed. Word n-grams are most useful to *L3*, while for *L1&2* the structural features are more informative. This indicates the pragmatic interpretation of questions, i.e., various user expectations from the answer cannot be told directly from surface interrogative patterns. The low usage of word tri-grams and long-distance interrogative patterns in all three levels can be explained by the irregular formation of questions posted on-line. Features of fewer dependencies are more flexible and effective for such data. Besides, as shown by *Sem-SF*, highly effective *Qtype* indicators center on a small number of semantic categories.

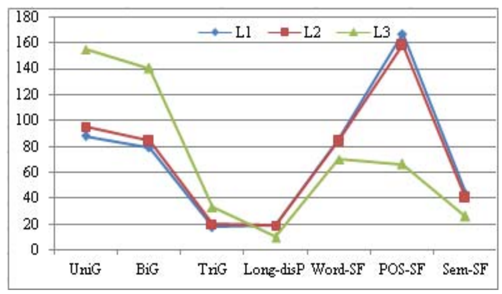

Figure 3. Top 500 features selected by *IG* of 3 levels

TABLE IV. CLASSIFICATION PERFORMANCE OF LEVEL 1

	NB	SVM	CPAR
P	90.76%	91.43%	91.00%

TABLE V. CLASSIFICATION PERFORMANCE OF LEVEL 2

	P	*R*	*F1*
	TypeCombine		
NB	78.50%	65.70%	71.53%
SVM	77.32%	67.27%	**71.94%**
CPAR	75.30%	65.49%	70.05%
	BinaryC		
NB	70.63%	72.25%	**71.43%**
SVM	81.77%	56.09%	66.54%
CPAR	47.09%	77.09%	58.47%

TABLE VI. CLASSIFICATION PERFORMANCE OF LEVEL 3

	P	*R*	*F1*
	TypeCombine		
NB	65.09%	63.43%	**64.24%**
SVM	65.85%	59.17%	62.33%
CPAR	60.73%	54.03%	57.18%
	BinaryC		
NB	74.04%	79.28%	**76.57%**
SVM	73.25%	66.77%	69.86%
CPAR	61.66%	83.43%	70.91%

TABLE VII. CLASSIFICATION PERFORMANCE OF EACH QUESTION TYPE IN LEVEL 1

	cho	def	con	des	pro	rea	yesNo	absE	loc	per	time	qua	ent	emo
P	93.10%	91.67%	67.27%	78.47%	97.00%	97.79%	98.90%	48.33%	97.21%	75.00%	98.78%	98.54%	81.45%	82.04%

TABLE VIII. CLASSIFICATION PERFORMANCE OF EACH QUESTION TYPE IN LEVEL 2 (BINARYC)

	cho	con	def	des	ent	loc	per	pro	qua	rea	absE	time	yesNo
P	65.19%	50.25%	72.62%	70.35%	69.48%	61.67%	70.59%	83.94%	80.20%	50.12%	37.02%	67.01%	92.81%
R	85.12%	73.33%	82.81%	54.95%	75.55%	80.00%	41.38%	76.81%	70.13%	73.54%	46.37%	73.03%	93.53%
F1	73.84%	59.64%	77.38%	61.71%	72.39%	69.65%	52.17%	80.22%	74.83%	59.61%	41.17%	69.89%	93.17%

TABLE IX. CLASSIFICATION PERFORMANCE OF EACH QUESTION TYPE IN LEVEL 3 (BINARYC)

	specification	Knowledge source	temporal constraint	orientation
P	41.09%	72.38%	71.76%	89.26%
R	49.34%	85.13%	82.04%	84.64%
F1	44.84%	78.24%	76.55%	86.89%

a) Question Classification Performance of Level 1

For *L1*, precision equals recall. Only precision values are listed in table IV. The best result is 91.43%. The best result reported in [7] is 92.6%. Nevertheless, the question collection used in [7] is relatively small with less than 2,000 instances. The question taxonomies used are also different.

The precision of each *QType* in *L1* is displayed in table VII. We can see that the majority of *QTypes* obtain satisfying performance. *absEntity* have the lowest precision, namely, 48.33%. The high similarity of interrogative patterns between *absEntity* and *entity* is the main cause to it. More distinguishable features, such as more abstract hypernyms of higher level should be introduced.

b) Question Classification Performance of Level 2

Table V. and table VIII. display the classification performance of *L2*. All three classifiers in *TypeCombine* have comparable results. However, in *BinaryC*, values of the same evaluation criterion vary sharply among the classifiers. When building binary classifiers for each *QType*, SVM has the highest precision, while CPAR has the highest recall. However, NB draws a balance between precision and recall, and has the best F1-value. The overall result of *TypeCombine* is slightly better than that of *BinaryC*. Similar to *L1*, *absEntity* also has the worst results.

Table VI. and table IX. display the classification performance of *L3*. In contrast to the result of *L2*, the performance of *BinaryC* is significantly higher than that of *TypeCombine*. Precisions and recalls of *BinaryC* follow the similar regulation of *L2*. Specification has the lowest classification results. One possible reason is that the judgment criteria of specification are fuzzy. Inconsistency exists among labels in the corpus. More clarifications on the definition of high/low specification should be made.

c) Discussion

As can be seen in table V and table VI, the best F1-values of *L2* and *L3* are 71.94% and 76.57%. These are not as satisfying as that of *L1*. One reason is the essential difficulty of multi-label classification tasks. Another reason is due to the tagging instructions given when building the corpus. All the three workers' labels are combined to be the labels of a question. No consistency check is conducted. Thus, the subjective judgments of different workers are preserved in the corpus. It is possible a question is labeled with two contrary types. Under this condition, the classification performance of *L2* and *L3* is promising. This shows that strong correlations exist between question types of the same level, which can be utilized in our future work for classification improvement.

Each classifier in *BinaryC* displays unique advantages. This can be used in applications of different demands or customizable applications.

V RELATED WORK

A. *Conventional Question Classification*

Question classifications are first proposed to solve the lexical chasm problem between questions, and between questions and answers. The main principle is to remove the influence of different interrogative patterns. FAQFinder [7] is the first system to use machine learning methods for classification of real questions in on-line discussion groups. [8] and [9] propose question taxonomies specifically designed for the encyclopedia service. Other works involving question analysis, such as [10] and [11], define question types of their own and mainly use rule-based methods to do the classification. Their research singly focused on recognition of surface user intention. The question type set of level 1 proposed in this paper inherits part of previous taxonomies.

B. *Coverage of User Intentions*

One finding to support our question type taxonomy concerning comprehensiveness is from the interface design experiment conducted by [12]. It verifies that the majority of users prefer answers to be presented along with context, e.g., other related information as the background, to make the answer reliable and complete.

[8] propose to recognize the "coverage asked" and "coverage would give". The former represents the level of detail of a direct answer the asker expects. The latter represents the level of detail that an information provider would include in a helpful answer. However, exactly what types of information should be covered is not examined..

[10][13][14] analytically extend the intentions and implicatures of complex questions. Their methods rely heavily

on predicate-argument analysis, precise parsers, and large volumes of domain-specific documents to extract topic signatures. Their focus is finding entities and events relevant to the topic of the question. They experiment with regular and well-formed questions, provided by TREC or other organizations. In contrast, we use real user questions of varying qualities from the CQA service. We focus on statistically recognizing the comprehensive types of information needs of a question. More general, domain-independent clues, like shallow linguistic features are exploited.

C. *Answer Expectation*

[15] uses co-training method to detect whether the question searches for subjective or objective information. [16] proposes taxonomies for CQA questions and answers to analyze the reusability of answers. The taxonomy can also detect the question's subjective orientation. However, no method is proposed to recognize the taxonomies. In both paper, other aspects of the answer expectations expect the subjectivity orientation are ignored.

VI Conclusion and Future Work

This paper proposes a three-level question type taxonomy to model user information need. Each level corresponds to (1) the basic information need represented by the surface interrogative form, (2) the types of supportive information needs to make the answer comprehensive and (3) the pragmatic user expectations from the answer, respectively, based on extensive observation of real questions collected from on-line question answering communities. Question type assignment of level 2&3 is inherently subjective and personalized. This paper proposes effective methods of question classification.

The three levels of the type taxonomy can be utilized in different fundamental modules of the question answering system independently or in integration, to facilitate question interpretation or to enhance the quality of summarized answers, etc. Although the corpus used in this paper is in Chinese, the taxonomy is language-independent and can be used universally. Besides, other information access applications can also draw lessons from the design of the taxonomy.

The relations between question types in different levels are to be examined for classification enhancement. More detailed observation on the characteristics of user intentions will also be made.

Acknowledgment

This work is supported by the National High-tech R&D Program of China (863 Program, No. 60703015).

Reference

[1] Tom Mitchell. Machine Learning. McCraw Hill, 1996.

[2] Grigorios Tsoumakas, Ioannis Katakis. Multi-Label Classification: An Overview. International Journal of Data Warehousing and Mining, 3(3):1-13, 2007.

[3] John, G. & Langley, P. . Estimating continuous distributions in bayesian classifiers. Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, Vancoucer, Canada, 1995.

[4] T. Joachims. Text categorization with support vector machines: learning with many relevant features. Proceedings of ECML. 1998.

[5] Xiaoxin Yin and Jiawei Han, CPAR: classification based on predictive association rule. Proceedings of the SDM San Francisco, CA, pp. 369–376, 2003

[6] McCallum, Andrew Kachites. Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering. http://www.cs.cmu.edu/~mccallum/bow. 1996.

[7] Noriko Tomuro. Interrogative Reformulation Patterns and Acquisition of Question Paraphrases. Proceeding of the Second International Workshop on Paraphrasing, pp:33~40,July 2003.

[8] Ingrid Zukerman, Eric Horvitz. Using Machine Learning Techniques to Interpret WH-questions. ACL2001.

[9] Hyo-Jung Oh, Chung-Hee Lee, Hyeon-Jin Kim, Myung-Gil Jang. Descriptive Question Answering in Encyclopedia. Proceedings of the ACL Interactive Poster and Demonstration Sessions, Ann Arbor, pp. 21–24, June 2005.

[10] Sanda Harabagiu, Dan Moldovan, Christine Clark, Mitchell Bowden, Andrew Hickl and Patrick Wang. Employing Two Question Answering Systems in TREC-2005. Proceedings of the Fourteenth Text REtrieval Conference, 2005.

[11] Chuan-Jie Lin, and Chun-Hung Cho, "Question Pre-Processing in a QA System on Internet Discussion Groups", Proceedings of the Workshop on Task-Focused Summarization and Question Answering, Sydney, pp. 16–23, July 2006.

[12] Jimmy Lin, Dennis Quan, Vineet Sinha, Karun Bakshi, David Huynh, Boris Katz, and David R. Karger. What Makes a Good Answer? The Role of Context in Question Answering. Human-Computer Interaction (INTERACT 2003).

[13] Sanda M. Harabagiu and Steven J. Maiorano and Alessandro Moschitti and Cosmin A. Bejan. Intentions, Implicatures and Processing of Complex Questions. HLT-NAACL Workshop on Pragmatics of Question Answering, 2004.

[14] Sanda Harabagiu, Finley Lacatusu and Andrew Hickl. Answering Complex Questions with Random Walk Models. SIGIR, Seattle, Washington, USA.pp220-227, August 6–11, 2006.

[15] B. Li, Y. Liu, and E. Agichtein: CoCQA: Co-Training Over Questions and Answers for Predicting Question Subjectivity Orientation. EMNLP 2008

[16] Yuanjie Liu, Shasha Li, Yunbo Cao, Chin-Yew Lin, Dingyi Han and Yong Yu. Understanding and summarizing answers in community-based question answering services. Proceedings of COLING, Manchester, UK, August 18-22, 2008.