Estimation of FAQ Knowledge by Classifying Questions and Answers

JUN HARADA[1], MASAO FUKETA[2], KAZUHIRO MORITA[2], TORU SUMITOMO[2],
ELSAYED ATLAM[2] AND JUN-ICHI AOE[2]

[1]Dept. of Intelligent Mechanical Systems Engineering, Kochi University of Technology,
185 Miyanoguchi, Tosayamada-Cho, Kami-Shi, 782-8502, JAPAN
[2]Dept. of Information Science and Intelligent Systems, University of Tokushima
2-1 Minami-Josanjima-Cho, Tokushima-Shi, 770-8506, JAPAN

Abstracts:- Question and answering (QA) systems in the CRM scheme require both the user's satisfaction related to the quality and the amount of questions to be managed, it depends on the cost. This paper presents an estimation method of the FAQ service by introducing the following measurements: 1) user's disrepute for products which defined by four types of classifying questions (IMPOSSIBLE, SIDE EFFECT, INSUFFICIENT and UNCLEAR) and the degree for each type is defined; 2) kindness for solutions replied which is defined by four types of classifying answers (ACTION, CONFIRMATION, EXPLANATION, and NO PROBLEM) and the degree for each type is defined; 3) sufficiency for the whole FAQ service that introduced by the 1) and 2).

Key-Words:- Question and answering, Frequently Asked Questions, Semantic Expressions.

1. Introduction

Question and answering (QA) systems in the CRM scheme require both the quality relating user's satisfaction and the mount of questions to be managed, that is to say, it depends on the cost. There are many QA researches for large text databases, but they are not relation to CRM schemes [1] [2] [3]. Useful for the CRM researches includes answering opinion questions by [7], good and bad expression understanding by [5], sentence subjectivity by [6], and estimating sentence types by [4].

One solution is to control the QA systems by human operators like a call center. In fact, operator can achieve response with good quality for users, but it is impossible to manage many questions because of high human expenses. A FAQ（Frequently Asked Questions）scheme is a well known approach that users can find appropriate answers. It is difficult to determine whether the FAQ service is sufficient, or not. Hammond et al. [8] presented knowledge navigation of FAQ systems, but there was no discussion about evaluation of the FAQ service. The behavior of the FAQ scheme is similar to the QA system that takes questions as the input and replies answers because FAQ knowledge bases are static in general. Therefore, it is very important study to analyze and to classify questions and answers of FAQ knowledge.

This paper presents a measurement of quality of the FAQ service by introducing the following measurements. 1) Measurement of user's disrepute for products is defined by classifying questions. Questions are classified by four types IMPOSSIBLE, SIDE EFFECT, INSUFFICIENT and UNCLEAR, and the degree for each type is defined.
2) Measurement of kindness for solutions replied is defined by classifying answers. Answers are classified by four types ACTION, CONFIRMATION, EXPLANATION, and NO PROBLEM, and the degree for each type is defined. 3) Measurements of sufficiency for the whole FAQ service is introduced by the 1) and 2). This degree is defined by an integer and it becomes very easy to estimate the FAQ service.

2. Semantic Expression of FAQ Knowledge Bases

In the FAQ dialogue, a questioner (a user or a customer) expects that a respondent (a company person) provides useful answers resolving his/her claim. This section discusses a formal definition for FAQ dialogue systems by defining a Q-Class attribute. The Q-Class attribute means the degree of questioner's disrepute and it is defined by four kinds of classes. For examples, "Can not print out" means Q-CLASS is [IMPOSSIBLE], "printer is noisy" means Q-CLASS is [SIDE EFFECT], "Printing character is unclear" [INSUFFICIENT] and "The red

lamp of the printer has been lighted up" means Q-CLASS is [UNCLEAR].

We can define the degree of user's disrepute by using Q-CLASS. By introducing additional attributes, the formal description of semantic expressions for questions and answers are as follows:

[Definition 1]

The question semantic expression SEMANTIC (p) for question *p* is defined as [C(SITUATION); x(OBJECT); i(CLAIM); a(Q-CLASS)]. Consider question *p* ="The printed character is unclear".

The question semantic expression is [[[printing],[SITUATION]];[[character],[OBJECT]]; [[unclear],[CLAIM]];[[INSUFFICIENT],[Q-CLASS]]].

For the answer semantic expression, a A-CLASS attribute is introduced. It is four values of answers. For example, "Change a cartridge" is [ACTION], "Check a printer cartridge" is [CONFIRMATION], "Printed characters becomes unclear if there is little ink residual quantity in the cartridge" is [EXPLANATION] and "It is not a failure" is [NO PROBLEM].

[Definition 2]

The answer semantic expression for question semantic expression SEMANTIC(*q*) is defined as TRANSFORM(SEMANTIC(*p*)).

Consider question semantic expression SEMANTIC(*p*) = [[[printing],[SITUATION]];[[character],[OBJECT]]; [[unclear],[CLAIM]];[[INSUFFICIENT],[Q-CLASS]]]. It means that [the user hopes that the printed character becomes clear] and one of the expected answers should be [recommend change the cartridge]. The answer semantic expression TRANSFORM(SEMANTIC(*p*)) becomes [[[printing],[SITUATION]];[[cartridge],[OBJECT]]; [[change],[SOLUTION]];[[ACTION],[A-CLASS)]]]. We can define the degree of answers by using A-CLASS.

[Definition 3]

For the semantic expression *r*, SURFACE(*r*) defines a set of surface sentences.

Consider the question semantic expression *r* = [[[printing],[SITUATION]];[[cartridge],[OBJECT]]; [[change],[SOLUTION]];[[ACTION],[A-CLASS)]]], SURFACE(*r*) includes sentence "The printed character is unclear". By the same manner, for the answer semantic expression *r* = [[[printing],[SITUATION]];[[cartridge],[OBJECT]]; [[change],[SOLUTION]];[[ACTION],[A-CLASS]]], SURFACE(*r*) includes "Please change a cartridge".
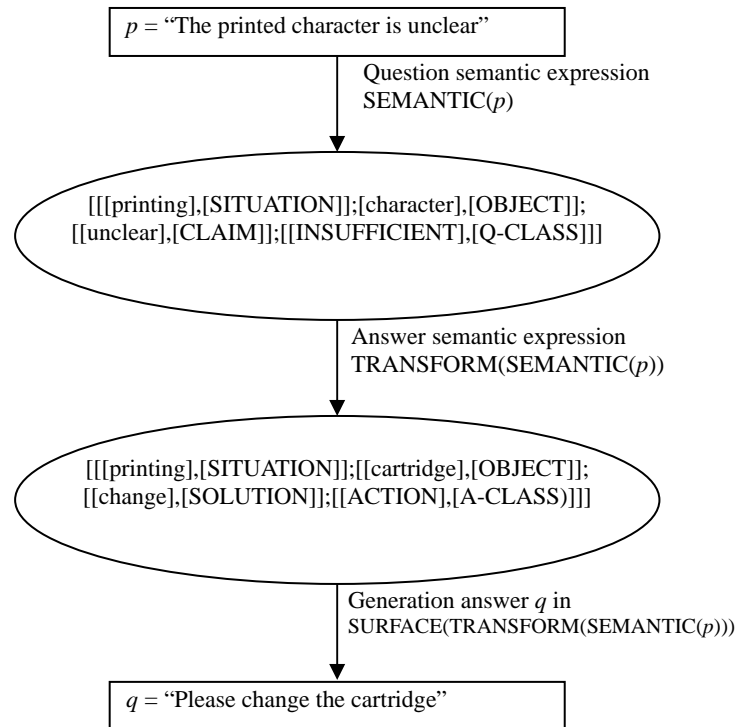

Figure 1 Understanding process for question

Figure 1 shows an illustration of understanding process for the question "The printed characters are unclear" and the answer semantic expression.

In this case, value [printing] of attribute SITUATION is kept as the same value in the answer semantic expressions, but there are many semantic expressions to be changed. Consider question *p'*= "Paper is got blocked in printing" SEMANTIC(*p'*) = [[[printing],[SITUATION]];[[paper],[OBJECT]]; [[block],[CLAIM]];[[SIDE EFFECT],[Q-CLASS]]] and TRANSFORM(SEMANTIC(p')) includes [[[paper setting],[SITUATION]]; [[sheet holder],[OBJECT]];[[fixed],[SOLUTION]]; [[ACTION],[A-CLASS]]] as one of answer semantic expressions. This case means the situation of question p' is focusing on the detailed situation in the answer.

3. Estimation Measurements of FAQ Knowledge Bases

3.1 Degree of disrepute for questions

In the question understanding process, affective information (user's tone, sentence style and so on.) are considered and Q-CLASS defines in the question

semantic expression. This section defines the degree of the user's disrepute from questions as follows:

1) DISREPUTE([[IMPOSSIBLE],[Q-CLASS]])=4
Value **[IMPOSSIBLE]** means that the function which should be committed essentially does not work, so the degree of user's disrepute is the highest level. This point is defined by 4,

2) DISREPUTE([[SIDE EFFECT],[Q-CLASS]])=3
Value **[SIDE EFFECT]** means there is a bad phenomenon unrelated to the original function, so the degree of user's disrepute is in the second level. This point is defined by 3,

3) DISREPUTE([[INSUFFICIENT],[Q-CLASS]])=2
Value **[INSUFFICIENT]** means a function is lower than the expected performance, so the degree of user's disrepute is in the third level. This point is defined by 2,

4) DISREPUTE([[UNCLEAR],[Q-CLASS]])=1
Value **[UNCLEAR]** means the operating method and the results are unclear, so the degree of user's disrepute is the lowest level. This point is defined by 1.

Table 1 shows examples about the degree of **DISREPUTE** with all values (IMPOSSIBLE, SIDE EFFECT, INSUFFICIENT, UNCLEAR)

Table 1 Examples of the degree of DISREPUTE

Q-CLASS	Disrepute	Examples
(IM)	4	The picture does not appear. The modem does not auto answer.
(SE)	3	The screen is shaking. I hear strange noises from refrigerator.
(IN)	2	The screen is too bright or too dark. The image is too light or dark.
(UN)	1	What does this DPOP error message mean? The colon in the time display keeps blinking, is this normal?

(IM)=IMPOSSIBLE;   (SE)=SIDE-EFFECT;   (IN)=INSUFFICIENT; (UN)=UNCLEAR;

3.2 Degree of kindness for answers

The quality of answers depends on FAQ knowledge bases, so it is impossible to resolve all troubles in products, this is why some answers are insufficient for users. This section defines the degree of kindness for answers, by using the attribute A-CLASS in the answer semantic expression as follows:

1) KINDNESS([[ACTION],[A-CLASS]])=4
Value **[ACTION]** means the concrete countermeasures for a question are shown, so the

degree of KINDNESS for users is the highest level. This point is defined by 4,

2) KINDNESS([[CONFIRMATION],[A-CLASS]])=3
Value **[CONFIRMATION]** of attribute [A-CLASS] means a trouble is solved by checking, so the degree of KINDNESS for users is in the second level. This point is defined by 3,

3) KINDNESS([[EXPLANATION],[A-CLASS]])=2
Value **[EXPLANATION]** means a trouble situation is explained, but no solution is provided, so the degree of KINDNESS for users is in the third level. This point is defined by 2,

4) KINDNESS([[NO PROBLEM],[A-CLASS]])=1
Value **[NO PROBLEM]** of attribute [A-CLASS] means no troubles are confirmed, so the degree of KINDNESS for users is in the lowest level. This point is defined by 1.

Table 2 shows examples about the degree of KINDNESS for answers with all values (ACTION, CONFIRMATION, EXPLANATION, NO PROBLEM).

Table 2 Examples of the degree of KINDNESS for answers.

A-CLASS	Kind ness	Examples
(AC)	4	Adjust the H-Size, H-Phase controls. Set the S0 register to 0. ATS0=0.
(CO)	3	Confirm the jumper settings on the drive. Check all cards.
(EX)	2	The auto answer mode is determined by the S0 register.
(NO)	1	Wash basket not straight may be not be a problem.

(AC)=ACTION;   (CO)=CONFIRMATION;   (EX)=EXPLANATION; (NO)=NO PROBLEM;

3.3 Degree of sufficiency for FAQ

From degrees DISREPUTE, KINDNESS, we can define the degree of FAQ service sufficiency, denoted by SUFFICIENCY(q) for question p as follows:
Algorithm: Determination of degree of SUFFICIENCY(q)
Input: Question p
Output: SUFFICIENCY(q)
[Method]
(Step 1) For [[i], [Q-CLASS]] in SEMANTIC(p)
for a given question p selected by a user,
determine H = (5-DISREPUTE([[i],[Q-CLASS]])).
(Step 2) For all [[j],[A-CLASS]] in
 TRANSFORM(SEMANTIC(p)),

determine K= KINDNESS([[j],[A-CLASS]]).
Let n be the number of [[j],[A-CLASS]] in TRANSFORM(SEMANTIC(p)).
Suppose that there are values K1, K2, … and Kn.
(Step 3) Calculate total values KTOTAL
by summing up K1, K2, …, and Kn.
(Step 4) Set KTOTAL/n to KAVERAGE.
(Step 5) Determine
SUFFICINECY(p) by H* KAVERAGE.

(End of Algorithm)

Consider question p =“The printed character is unclear”.
The question semantic expression is
[[printing],[SITUATION]];[[character],[OBJECT]]; [[unclear],[CLAIM]];[[INSUFFICIENT],[Q-CLASS]]].
   In Step 1,
H=(5-DISREPUTE([[INSUFFICIENT],[Q-CLASS]])) =3.
For [[j],[A-CLASS]] in
TRANSFORM(SEMANTIC(p)),
suppose that there four answers.
1) “Change cartridge”
with attribute [[ACTION],[A-Class]].
2) “Check printer cartridge”
with attribute [[CONFIRM],[A-Class]].
3) “Printed character becomes unclear if there is little ink residual quantity in the cartridge”
with attribute [[EXPLANATION],[A-CLASS]]
4) “It is not failure” with attribute [[NO PROBLEM],[A-CLASS]].
   Step 2 determines the following values:
K1= KINDNESS([[ACTION],[A-CLASS]])=4
K2= KINDNESS([[CONFIRM],[A-CLASS]])=3
K3=KINDNESS([[EXPLANATION],[A-CLASS]]) =2
K4=KINDNESS([[NO PROBLEM],[A-CLASS]])=1
KTOTAL is 10 in Step 3 and KAVERAGE is 2.5 in Step 4.
SUFFICIENCY(p)=H*KAVERAGE=3*2.5=7.5
in Step5. In this example, If there is only the first answer,
SUFFICINECY(p)=H*KAVERAGE =3*4=12.
For only the last answer,
SUFFICIENCY (p)=H*KAVERAGE=3*1=3.
By using SUFFICIENCY(p), FAQ systems can be automatically estimated.

4. Experimental Observations
4.1 Experimental data and their properties
For FAQ data, 4,538 questions and 5,356 answers

have been prepared for six kinds of products (computers, telephones/facsimiles, digital cameras, AV equipments, home electronics and cars). The 1,513 questions and answers have been collected from FAQ web pages and FAQ documents of products
[http://kadenfan.hitachi.co.jp/q_a/index.html][http://www.mitsubishielectric.co.jp/cs/index.html][http://www.toshiba.co.jp/digital/support/index.htm][http://livingdoors.jp/qa/], where Japanese has been translated into English. The remaining 3,025 questions have been produced by ten expert persons (10 Ph.D. Students). Let N(Q) be the number of questions obtained from FAQ data, let N(Q+) be the number of questions produced by ten expert persons, let N(A) be the number of answers. **Table 3** shows information about the FAQ data.

Table 3 Information about FAQ data

Product	\<Computers>	\<Telephones/Facsimiles>	\<Digital cameras>	\<AV equipments>	\<Home electronics>	\<Cars>
N(Q)	155	213	136	567	249	193
N(Q+)	310	425	274	1133	497	386
N(A)	1,032	739	425	1,702	839	619

4.2 Question semantic expressions
Classified results by question semantic expressions in the FAQ data are discussed.
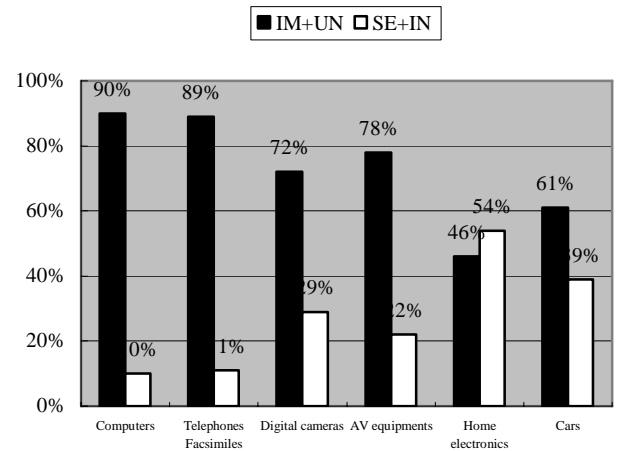

F**igure 2 Classified results by question semantic expressions.**
   (IM)=IMPOSSIBLE; (SE)=SIDE-EFFECT; (IN)=INSUFFICIENT; (UN)=UNCLEAR

   From **Figure 2**, the following observations are obtained:
1)   The total rate 90% of IMPOSSIBLE and UNCLEAR for field \<Computers> are larger than the total rate 46% for field \<Home Electronics>.

2)   The total rate 54% of SIDE EFFECT and INSUFFICIENT for field <Home electronics> are larger than the total rate 10% for field <Computers>. Operations associating with field <Computers> are generally difficult for users, so the IMPOSSIBLE and UNCLEAR questions increase. In other words, users can not determine whether their claims for field <Computers> are concrete classes SIDE EFFECT and INSUFFICIENT, or not. Consequently, many questions of field <Computers> belong to IMPOSSIBLE and UNCLEAR. However, products for fields <Home electronics> must be safe and complete because it is easy to discover the dissatisfaction of the home electronics used in everyday life. Therefore users are affective for SIDE EFFECT and INSUFFICIENT. This observation can be reflected on other fields, for example, fields <Home Electronics> and <Cars> have the similar tendency.

4.3 Answer semantic expressions

Figure 3 shows the classified results by answer semantic expressions.
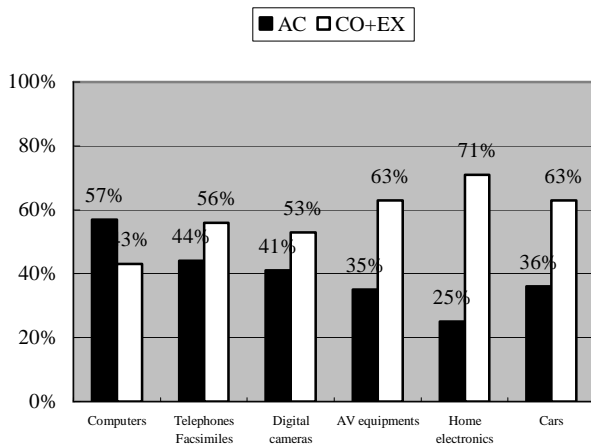

Figure 3 Classified results by answer semantic expressions.
(AC)=ACTION; (CO)=CONFIRMATION; (EX)=EXPLANATION; (NO)=NO PROBLEM

From **Figure 3**, the following observations are obtained. The rate 57% of ACTION for field <Computers> is the highest in the fields. It means that there are concrete answers for many IMPOSSIBLE and UNCLEAR questions in that field as discussed in section 4.2.2. The fields <Telephones /Facsimiles> and <Digital cameras> have the similar property while the rate of ACTION is just less than that of field <Computers>. For the total rate 71% of

CONFIRMATION and EXPLANATION of <Home Electronics> is the highest in the fields. The reason is that product must be safe and complete as discussed in section 4.2.2. In other words, we can say that there are many indirect questions requesting CONFIRMATION and EXPLANATION answers for <Home Electronics>. The fields <AV equipments> and <Cars> have the similar property.

4.4 KINDNESS and SUFFICIENCY
Table 4 shows KINDNESS and SUFFICIENCY by question and answer semantic expressions.

Table 4 KINDNESS and SUFFICIENCY

	<Computers>	<Telephones /Facsimiles>	<Digital cameras>	<AV equipments>	<Home electronics>	<Cars>
K	3.3	2.9	2.6	2.7	2.6	2.8
S	8.2	7.3	6.2	6.6	6.4	6.9

K=KINDNESS; S=SUFFICIENCY

In **Table 4** the degree 3.3 of KINDNESS for <Computers> is the highest because there are many pairs of questions for the UNCLEAR class and answers for ACTION class. On the other hand, the rate 2.6 for <Digital Cameras> and <Home electronics> is low because there are many pairs of questions for the SIDE EFFECT classes and answers for EXPLANATION and NO PROBLEM classes. These desirable features reflect on the degree 8.2 of SUFFICIENCY for <Computers>.

Figure 4 shows bar graphs (5-degree of DIREPUTE), KINDNESS, and SUFFICIENCY for each feature, where the scale of SUFFICIENCY is reduced to 1/4. From **Figure 4**, we can obtain the following observations:
1) For features "battery/power" and "pictures", (5-degree of DIREPUTE) of "battery/power" is lower than that of "pictures", but SUFFICIENCY of "battery/power" is higher than that of "pictures" because FAQ of "battery/power" has good KINDNESS rather than that of "pictures".
2) For features "LCD monitors" and "other features", (5-degree of DIREPUTE) of "LCD monitors" is similar to that of "other features", but SUFFICIENCY of "LCD monitors" is very lower than that of "other features" because FAQ of "LCD monitors" has no good answers. For example, "There is no trouble". In order to improve SUFFICIENCY of "LCD monitors", answers of FAQ must include explanations of phenomenon and the

reasons even if it is difficult to solve the problem in the current products.

3)   (5-degree of DIREPUTE) of features except "pictures" is low, but the degree of KINDNESS is the same level. Thus, there are no differences for the resulting SUFFICIENCY between "pictures" and "other features".
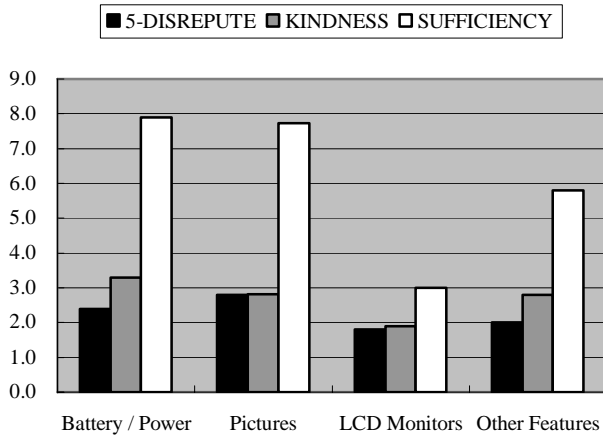
■5-DISREPUTE ▨KINDNESS ☐SUFFICIENCY

Figure 4 Degrees for 5 questions and for each features

In conclusion, the degree of DISREPUTE can be utilized to take directly the tendency of user's disrepute for products and the degrees of KINDNESS is useful to improve answers in FAQ knowledge bases. Moreover, the degrees of SUFFICIENCY can be applied to surveillance of the whole FAQ service.

5. Conclusions

This paper has been presented an estimation method of the FAQ service by introducing the following measurements: 1) user's disrepute for products which defined by four types of classifying questions (IMPOSSIBLE, SIDE EFFECT, INSUFFICIENT and UNCLEAR) and the degree for each type is defined; 2) kindness for solutions replied which defined by four types of classifying answers (ACTION, CONFIRMATION, EXPLANATION, and NO PROBLEM), and the degree for each type is defined; 3) sufficiency for the whole FAQ service that introduced by the 1) and 2).

The presented approaches have been evaluated by the FAQ data with 4,538 questions and 5,356 answers. Moreover, the real time simulation to estimate user's sufficiency has been computed. From this evaluation, it turned out that the presented approach is useful and effectiveness.

References
[1] Kadoya, Y., Morita, K., Fuketa, M., Oono, M., El-Sayed Atlam, Sumitomo, T. and Aoe, J. "A Sentence Classification Technique by Using Intention Association Expressions.", *Computer Mathematics*, Vol.82, No.7, 2005, pp.777-792.
[2] Ferret, L., Grau, B., Plantet, M. H., Illouz, G., Jacquemin, C., Masson, N., and Lecuyer, P. "QALC: the Question-Answering system of LIMSI-CNRS.", *In The Ninth Text Retrieval Conference (TREC-9)*, 2001, pp.235-244.
[3] Fukumoto, J., Kato, T., and Masui, F. "Question Answering Challenge (QAC1) Question answering evaluation at NTCIR Workshop 3.", *in Working Notes of the Third NTCIR Workshop Meeting: QAC1*, 2002, pp.1-10.
[4] Tokunaga, H., El-Sayed Atlam, Fuketa, M., Morita, K., Tsuda, K. and Aoe, J. "Estimating sentence types in computer related new product bulletins using a decision tree.", *Information Sciences*, Vol.168, No.1-4, 2004, pp.185-200.
[5] Fuketa, M., Kadoya, Y., El-Sayed Atlam, Kunikata, T, Morita, K., Kashiji, S. and Aoe, J. "A Method of Extracting and Evaluating Good and Bad Reputations for Natural Language Expressions.", *Information Technology & Decision Making*, Vol.4, No.2, 2005, pp.177-196.
[6] Hatzivassiloglou, V., and Wiebe. J. M. "Effects of adjective orientation and gradability on sentence subjectivity." *in Proceedings of the 18th International Conference on Computational Linguistics (COLING)*, 2000, pp.299-305.
[7] Yu, H., and Hatzivassiloglou, V. "Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Sentences.", *in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2003, pp.129-136.
[8] Hammond, K., Burke, R., Martin, C., & Lytinen, S. "FAQ Finder: A Case-Based Approach to Knowledge Navigation." *in Proceedings of the 11th Conference on Artificial Intelligence for Applications*, 1995, pp.80-86.