Journal of Advance Computational Research
Vol. 2 No. 2 (July-December, 2017)

A Comparative Analysis for Exploring Answer Subjectivity

Ani Thomas[1] , Hari Ram Sharma[2], Manoj Kumar Kowar[3] and Sanjay Sharma[4]
[1]Reader, Department of Computer Applications,
[2]Professor, Department of Computer Science and Engineering,
[3]Professor, Department of Electronics & Telecommunications,
[4]Associate Professor, Department of Engineering Mathematics,
Bhilai Institute of Technology, Bhilai House, Durg (C.G.) – 491001 India
E-mails: tpthomas22@yahoo.co.in, hrsharma44@gmail.com, kowar_bit@rediffmail.com
ssharma_bit@yahoo.co.in

Abstract: *The information retrieval systems have been speedily heading towards many of the intelligent evaluation frameworks like subjective-answer assessment systems. The traditional approaches to Information Retrieval systems have been deploying the various models, namely Boolean retrieval models, Vector Space model and Probabilistic models. Each of these models is based on keyword retrieval, which operates at a symbolic, text-matching level, and ignores the semantic and contextual information in the retrieval process, as found so far in much existing work. An alternative view is considered to evaluate candidate answers using machine-generated answer keys that formalize exact and precise concept spaces for queried topics in restricted domain. For the answer keys to be generated automatically, the Bootstrapping Ontologies play a vital role. This paper compares the evaluation process with two already proposed novel techniques that grades the narrative responses on a fuzzy scale, both with respect to precision and recall measures. The evaluation results are found to be very close to the grading level of domain-specific human assessors. The first approach assesses the candidate answers with multi-objective fuzzy decision-making and the second with the utility functions as the evaluation criteria.*

Keywords : *Restricted Domains, Contextual Dependency triples, Question answering Evaluation metrics, Fuzzy Decision Functions, Utility Functions, Precision-Recall Measures.*

1. INTRODUCTION

End-User communities in document processing forums tend to accept Question-Answering (QA) Assessment tools useful if the standards are followed, namely Timeliness, Accuracy, Usability, Completeness and Relevance [1] [8]. The TREC (Text Retrieval Conference) QA track when initiated in 1999 focused on factoid questions for quite some time, while TREC 2003 QA track contained the logic for assessing 'list' and 'definition' questions. TREC 2006 and 2007 concentrated on complex interactive QA that shifted the focus of QA assessment away from "factoid" questions towards more complex information needs that exist within richer user contexts, and to move away from the one-shot interaction model implicit in previous evaluations towards a model based at least in part on interactions with users [2] [6]. In this series, the papers published in TAC (Text Analysis Conference) 2008 introduced many new concepts of text-processing in Knowledge Base Population track-task [14].

In this context, RACAI's QA system appears to be a competent subjective answer generation tool, presented in QA track task of CLEF 2007 campaign, as it aims to extract answers at any lengths for all types of questions. However, answers are generated after undergoing a multitude of Natural Language Processing steps like classifying questions by identifying answer types, decomposing questions into queries, filtering the queries according to P-O-S tags, generating document-indexes for queries (as search terms), scoring document-indexes in order of relevance and then finally by performing best structural match between the linkage of the question an most relevant paragraph sentences, returned in response to automated formulated queries. Well, the work group believes that such a laborious effort could be relieved if questions and sentences could have been matched in a full dependency parsed text-patterns [12].

An inherently underlying assumption that acts as a key to effective retrieval of the answers taken for a typical QA assessment tool is 'reflecting correct meaning of content' [4]. In the present communication, an attempt is made to unfold the Natural Language Semantics with machine-assisted topic acquisition procedures. The precisely chosen text fragments from the domain knowledge, contribute in term-to-term associations between these text fragments which in turn reveals the underlying meaning of inputted sentences. This novel thought automatically extracts and generates the model answers for candidate-answer assessments by making use of bootstrapping ontologies [10]. This inspires further, to provide a decision-making tool that assists the Academicians in assessing the on-line submitted subjective answers at the candidates' end. Even the evaluation process too is governed by the parameters bearing fuzziness between topic-correctness and degree of topic coverage in answers. Here, the assessment phase is compared using two methodologies, namely, fuzzy decision-making and utility functions as the participating evaluation metrics. This could be supported by graphical illustrations of assessment results based upon the mentioned methodologies. Thus, for assessing the subjective answers, the initial requirement has to be the domain knowledge triggered from the text corpus of corresponding domain, the restricted domain being an e-book in this context.

2. THE DOMAIN KNOWLEDGE

The proposed work mainly focuses on context-oriented content retrieval that is the necessary and sufficient condition for generating a precise and ideal answer key. With the aim to design a human-machine interface, it is the suggested domain-specific Ontology that facilitates the meaningful keyword search from the inputted questionnaires of the related subject or theme. However, generalized evaluation prototypes may suffer from unavailability of explicit Ontologies in order to satisfy end-users' interests with varied subject domains. This conceived the idea to take up the self-developing Ontologies by the work group. Further, constructing the above with bootstrapping approach sufficed the complete extraction of topics related to the thematic keywords focused in the questions. The authors would like to recall their past work on constructing of trees of concepts (n-gram pools) from the individually tagged paragraphs following the machine initiated scan upon so formulated search document for extracting the features [10].

It is not at all necessary that a subject of chosen domain encompassing wide range of topics should be taken up in full as a search document for themes about which the question-answering session is to be conducted. To speed up the series of such question-keyword match procedures, the front and the back-indexes of a recommended text can be considered for extracting selective content containing those string patterns. Here, the page accessions are thought to be reliable parameters for finding topics. Hence, the search process of desired content proceeds by defining broad vicinities in the form of page-ranges. The group has currently taken up the design of the page filtering tool that identifies the filtered content as page-ranges [5]. It is this set of pages that serve as the search document for extracting the meaningful answer keys.

3. THE ASSESSMENT ENVIRONMENT

When confining to restricted domain Question-Answering, questions if asked within a context, the respective answers should also be provided surrounding the same context.

3.1. Finding Content Vicinities

A case-study was taken up, where four students studying the Bachelor of Engineering course in Computer Sciences and having sound knowledge of 'Neural Networks', were asked to respond to a question within the mentioned technical subject domain, "Differentiate between Excitatory and Inhibitory connections as inputs to a Processing element / artificial neuron." As discussed earlier, the tool is expected to frontier a human-machine interface panel in order to take up the driving knowledge source, as a text material on "Neural Networks Algorithms, Applications and Programming Techniques" authored by *James Freeman and David M. Strapetus* in machine-readable format. This causes to generate the driving lexicon on the relevant subject from front and back indexes and initiate the search for finding content relevant pages, from where an answer-key can be formulated.

Meanwhile, the system extracts the Noun phrases and Verb phrasal segments from the question and assigns it to set 'Kq' as

Kq = {Processing element, Inhibitory connections, Excitatory connections}

The question is parsed too, using the dependency parser giving dependency triplets of question, 'D_q' that serve as semantic units to infer relevant validation patterns.

Dq = {a_kind_of (inhibitory connections, PE's input)
a_kind_of (excitatory connections, PE's input)}

As the back-indexes of the book-documents are generally expected to give detailed locations of the described topics with page-numbers, it would be a good idea to choose an already available subject-index at the back of the book to search the above phrases, giving out following observations:

Processing elements found on pages 4, 17, 18

Excitatory connections found on page 19

Inhibitory connections found on page 19

As a result, the relevant set of page-numbers after page filtering operation: $S_{pf} = \{17,18,19\}$

3.2. Extraction of Co-related Concepts

In order to construct appropriate succinct model answers to the search keywords of the question, say ‘Kq’, one needs to capture inherent semantic relations among the sentential fragments in text corpus reflecting the meaning of sentences and its related neighbors [4]. This is possible, if each of the participating noun phrases is searched for their existing grammatical role in their existing structural configurations. It has been concluded from an exhausted survey to natural language parsing literature that semantic relations can be smoothly extracted by having access to dependencies between domain-related keywords forming the sentences in varying grammatical roles. The Stanford Natural Processing group gains the maximal popularity for its publicizing freeware (Stanford parser and POS-tagger) to extract semantic components <*subject, predicate, object*>, where the subject and object components are assigned with noun phrases and predicate components depict verb phrases as event-indicators.

Continuing with the pages received in the final set of the experiment taken in section 3.1, the fetched content corresponding to the terms Kq, exhibited in n-gram pools were matched against the validation patterns of the question to generate a precise set of validation patterns approximating to the answer-key.

Dependency Patterns for Paragraphs 1.2.1 / 1.17

- Component of (computational elements, ANS models)
- Synonymy (Computational elements, Artificial Neurons)
- Synonymy (Neurons, PE s)
- Synonymy (Processing elements, PE s)

Dependency Patterns for Paragraph 1.2.1 / 2.17

- Non-similarity (Processing elements, actual biological neurons)
- Representation (Processing elements, group_of_neurons)
- Behaviour_like (systems, actual brain models)
- Ownership (problems, biological structure)
- Ownership (problems, biological structure)

Dependency Patterns for Paragraph 1.2.1 / 3.17

- Similarity (PE, real neuron)
- Criteria (similarity, many inputs)
- Possession (PE, single output)
- Fan output (PE, other PE’s)
- Occupy (PE’s, network)
- Connection (PE, PE)
- Synonym (weight, connection strength)
- Defines (PE’s output, neuron’s firing frequency)
- Defines (weight, synaptic connection strength)

Dependency Patterns for Paragraphs 1.2.1 / 4.19

- Segregation (PE s inputs, various types)
- Category_of (Excitatory connections, Input connections)
- Category_of (Inhibitory connections, Input connections)
- Possession (Excitatory connections, positive weights)
- Possession (Inhibitory connections, negative weights)
- A_kind_of (gain, special-purpose connections)
- A_kind_of (quenching, special-purpose connections)
- A_kind_of (non-specific arousal, special-purpose connections)
- Commonality (Excitatory connections, Input connections)
- Commonality (Inhibitory connections, Input connections)

Dependency Patterns for Paragraphs 1.2.1 / 5.19

- Basis (net-input value, Input connections)
- Calculation_criterion (net input, input values)
- No_calculation_criterion (net input, special_ connections)
- Multiplication (Input values, connection weights)

3.3. Answer key : The Meaningful Content

These validation patterns of page-filtered content are searched for whole or partial pattern-match the validation patterns of the question denoted by D_{ra} set. In this way, only a question-relevant collection of dependency triplets can be generated for the filtered pages, forming set ‘D_a’ containing significant explanation of question keywords [3] [13]. These are illustrated as the selected underlined triples that get matched with Kq or Dq in the mentioned experimental setup to form the model-answer also called answer key ‘Da’.

4. THE CANDIDATE ANSWER DEPENDENCIES

On the other hand,at the Candidate’s end, the candidate-answer passage too gets decomposed into noun phrases say K_{ca} along with their correspondingly extracted dependency relations, D_{ca} which are shown here for the experimental candidates' responses as

Candidate's Response 1: "Processing elements generally have two types of input connections, namely excitatory connections and inhibitory connections. These are categorized on the basis of effect; they produce on the processing neuron. Excitatory connections enhance the effect with positive weights while inhibitory connections deplete the effect with negative weights."

- Seggregation(processing elements, two types)
- Akindof(excitatory connections,input connections)
- Akindof(inhibitory connections,input connections)
- Segregation_criteria(PE s,effect on neurons)
- Possession(exitatory connections,positive weights)
- Possession(inhibitory connections,negative weights)
- akindof(exitatory connections,enhanced effect)
- akindof(inhibitory connections,depleted effect)

Candidate's Response 2: "Processing elements (PEs) are symbolized by a group of neurons which are not similar to actual biological neurons. PE's inputs are segregated into various types. The input connections are either of excitatory or inhibitory type that bear positive or negative weights respectively."

- Synonymy (Processing elements, PE s)
- Symbolization (Processing elements, group_of_neurons)
- Segregation (PE s inputs, various types)
- Akindof(excitatory connections, input connections)
- Akindof(inhibitory connections, input connections)
- Possession(exitatory connections, positive weights)
- Possession(inhibitory connections, negative weights)
- Non-similarity (Processing elements, actual biological neurons)

Candidate's Response 3: "Processing elements can have multiple inputs that fan in the neuron while only one output that fan out to the other PE s in the network. Depending upon the type of the effect they produce, input connections may be excitatory or inhibitory. The former produce positive weights and the latter possess negative weights."

- possession(processing elements, multiple inputs)
- Faninbehavior(multiple inputs, neuron)
- fanoutbehavior(single output, other PEs)
- producingeffect(input connections, types)
- Possession(input connections, positive weights)
- Possession(input connections,negative weights)
- akindof(input connections,exitatory connections)
- akindof(input connections,inhibitory connections)

Candidate's Response 4: "We know that every Processing element is connected to every other PE through a synaptic connection in the network. Multiple inputs from other PE s pass through these input connections. These have either a positive or negative effect on PE. Each connection is associated with a measuring parameter called weight or connection strength. Some connections exhibit pronounced effect called excitatory connections and other exhibit depleted effect called inhibitory connections"

- Synaptic connection(processing elements, PE)
- inclusion(PE, network)
- association(connection, weight)
- synonymy(weight, connection strength)
- Possession(exitatory connections, pronounced effect)
- Possession(inhibitory connections, depleted effect)
- transmission(multiple inputs, other PEs)
- Transmission_media(multiple inputs, input connections)
- Possession(input connections, positive effect)
- Possession(input connections, negative effect)

As the answer-fragments are semantically described by sets of three components, the three parts of the answer dependencies can be adjudged by analyzing three varying degree of similarity measures i.e. one-third, two-third or total dependency-matches between the candidate set and model-answer set.

5. SETTING UP EVALUATION CRITERIA

Extending the Answer evaluation framework for making a rationalized decision making for assessing the keyed-in answers by the candidates, the assessment criteria need to be thought very logically at the end of the evaluator.

Usually, Text miners adhere to unanimously accepted evaluation parameters namely, precision and the other recall. Precision indicates, how much accuracy is reflected by a candidate's answer in its meaning, while parameter recall for a candidate's answer adjudges from the fact that, as to how many points are correctly keyed in by the candidate in context to the asked question, as they exist in the modeled answer-key.

From a small-scale survey, it was noticed that Subject Experts rated a very precise and correctly keyed in answer as 'Accurate':a_1; not very precise but a more or less considerable, as "Relevant':a_2; not at all precise but considerable to some extent as 'Somewhat Relevant':a_3, further more, an answer that not at all precise and irrelevant is generally assessed as 'Vague':a_4. The above opinion is incorporated for the machine-assisted answer evaluation, defined as precision grades. These grades are computed from statistical observations of a candidate's answer (D_{ca}), in form of one-third, two-third or total dependency-match patterns for obtaining the three sets of precision-ratings [9].

The 'Recall' criteria can be described at the evaluator's end as "right recall":a_1', when the candidate recalls all the points for elaborating in his or her answer from the referring books accurately, but if the candidate recalls relatively half of the answer correctly, then it is assessed as "incomplete recall":a_2'. If the candidate is not able to recall and does not write at all or writes other points not at all in context with the asked question i.e., flooded answers, then the recall is "wrong": a_3'. These nominated Recall grades again, find their computations from one-third, two-third or total recall-based dependency-match patterns.

5.1. Evaluation with Fuzzy Multi-Objective Decision Parameters

One of the mathematicians, Ross extends to use the different types of precision-match parameters, to formulate fuzzy evaluation ratings expressed as fuzzy sets in Zadeh's notation, namely O_1, O_2, O_3, explicitly decided at the evaluator's end [7]. These too are illustrated in a graphical plot as shown in figure 1.

$$O_1 = \frac{0.1}{accurate} + \frac{0.3}{relevant} + \frac{0.7}{somewhat\ \ relevant} + \frac{1}{vague}$$

$$O_2 = \frac{0.5}{accurate} + \frac{0.9}{relevant} + \frac{0.4}{somewhat\ \ relevant} + \frac{0.2}{vague}$$

$$O_3 = \frac{1}{accurate} + \frac{0.7}{relevant} + \frac{0.3}{somewhat\ \ relevant} + \frac{0.1}{vague}$$

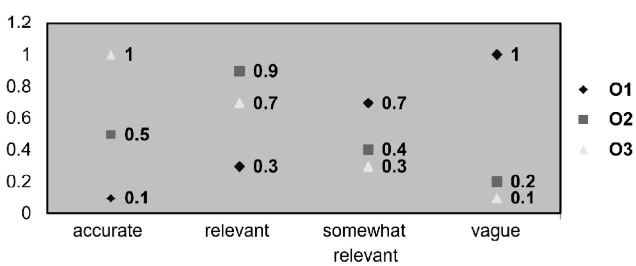

Figure 1: Fuzzy Membership Functions for Precision Evaluation

The membership values in each of these fuzzy sets indicate, up to what extent, the one-third, two-third or totally matched triplets individually contribute to predefined precision-grades. This can be explained by a candidate's answer instance that if total precision-match is found the maximum, then the answer will evaluate close to accurate as compared to that answer with relatively more number of one-third or two-third precision-match patterns [11].

The actual observations of a candidate's answer as to how many triples match the answer-key can be expressed by the

$$\text{Precision} = \frac{correctly\ \ retrieved\ \ relations(D_{ca\_matched})}{relations\ \ retrieved(D_{ca\_total})}$$

so that these values are used in formulating precision evaluation ratings. Let observed precision probabilities = $\{b_1, b_2, b_3\} \rightarrow [0,1]$ These values are computed for one-third, two-third or total match against each candidate answer as shown in table1 and graphically illustrated in figure 2.

Table 1
Observed Precision Ratings for the four Candidate answers

Precision / Candidate	*b1*	$\bar{b}_1$	*b2*	$\bar{b}_2$	*b3*	$\bar{b}_3$
Ca_1	0	1	.625	.375	.375	.625
Ca_2	0	1	.125	.875	.875	.125
Ca_3	.25	.75	.5	.5	.25	.75
Ca_4	.8	.2	0	1	.2	.8

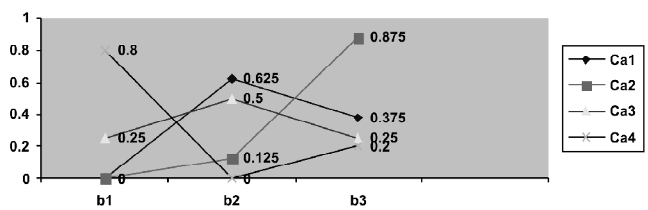

Figure 2: Observed Precision Probabilities

For achieving the proposed outcomes, the Fuzzy decision making technique makes use of an appropriate decision model that computes the joint intersection of 'r' decision measures aiming at both precision-grade and recall-grade based evaluations, expressed as

$$D = \bigcap_{i=1}^{r} (\bar{b}_i \cup O_i)$$

and the optimal solution, D* such that D* = max(D(a_i)) defines the alternative that maximizes D. The resulting decision measures for each of precision-grades are assimilated in table 2, assigning the four candidate answers to respective grades, whose decision-measures were found maximal.

The similar series of fuzzy computations were performed for the set of three Recall evaluation grades as nomenclated in section 5.

Recall grades = {Right, incomplete, wrong} º $\{a_1', a_2', a_3'\}$

Table 2
Decision Functions in Favor of Precision Evaluation Criteria for *Different* Candidate Answers

Decision/ Candidate	$D(a_1)$	$D(a_2)$	$D(a_3)$	$D(a_4)$	D^*
Ca_1	.5	.7	.4	.375	0.7(a2)
Ca_2	.87	.7	.3	.12	0.87(a1)
Ca_3	.5	.75	.5	.5	0.75(a2)
Ca_4	.2	.3	.7	.8	0.8(a4)

For instance, if total similarity of validation patterns are more then the answer is righhe tly recalled as compared to those answers that hold one-thirds or two-thirds similarities. For this, the weighted measures are depicted in the form:

Recall match parameters = {one-thirds , two-thirds , total} that act as the membership values in the fuzzy sets, say $\{O_1', O_2', O_3'\}$.

Hence, below described are fuzzy evaluation ratings, which are fuzzy sets expressed in Zadeh's notation, explicitly defined according to credit point assignment norms thought by human evaluators as illustrated in figure 3.

$$O_1' = \frac{0.1}{right} + \frac{0.7}{incomplete} + \frac{1}{wrong}$$
$$O_2' = \frac{0.6}{right} + \frac{0.9}{incomplete} + \frac{0.4}{wrong}$$
$$O_3' = \frac{1}{right} + \frac{0.3}{incomplete} + \frac{0.1}{wrong}$$

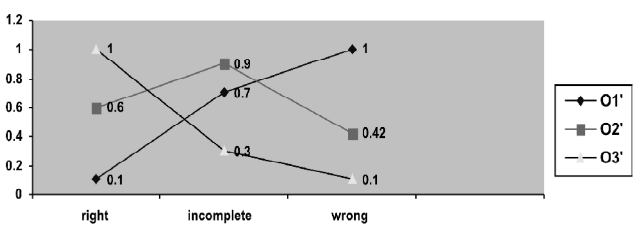

Figure 3: Fuzzy Membership Functions for Recall Evaluation

As the recall probabilities of a candidate's answer can be computed according to the expression:

$$\text{Recall} = \frac{correctly\ \ retrieved\ \ relations(D_{ca\_matched})}{number\ \ of\ \ correct\ \ relations(D_{a\_total})}$$

these probabilities are used in obtaining the Recall evaluation grades.

In the experimental setup, the observed recall ratings = $\{b_1', b_2', b_3'\} \rightarrow [0,1]$ as shown in table 3 and graphically illustrated in figure 4 that formulate fuzzy recall-evaluation ratings, already implemented in the past by the work-group [7].

Table 3 Observed Recall Ratings for the four Candidate answers

Recall / Candidate	b_1'	$\overline{b_1'}$	b_2'	$\overline{b_2'}$	b_3'	$\overline{b_3'}$
Ca_1	0	1	.71	.29	.43	.57
Ca_2	0	1	.14	.86	1	0
Ca_3	.285	.715	.57	.43	.285	.715
Ca_4	.86	.14	0	1	.285	.715

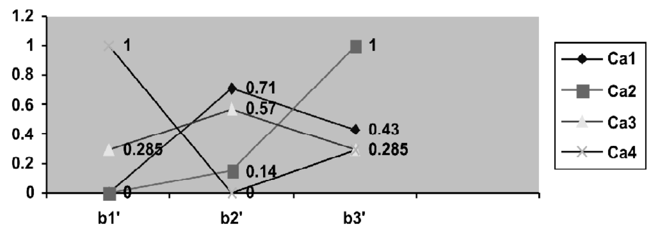

Figure 4: Observed Recall Probabilities

Table 4 shows the computed decision measures with assigned respective recall-grades to the respective four candidate answers with similar min-max fuzzy decision making methodology, discussed in section 5.1.

Table 4
Decision Functions in Favor of Recall Evaluation Criteria for the Four Answers

Decision /Candidate	$D(a_1')$	$D(a_2')$	$D(a_3')$	D'^*
Ca1	.6	.57	.4	0.6(a1)
Ca2	.86	.3	.1	0.86(a1)
Ca3	.6	.715	.43	0.715(a2)
Ca4	.14	.7	.715	0.715(a3)

5.2. Evaluation with Utility Functions

Shifting to another concept of drafting the evaluation framework, the piece of work was carried out in an extended direction. The probability that a candidate's answer triples shall fall into one or more of the states indicating zero, one-third, two-third and total string-pattern match in terms of precision can be conceptualized as parameters, $p(s_0)$, $p(s_1)$, $p(s_2)$, and $p(s_3)$ where $\sum_{i=0}^{3} p(s_i) = 1$. These state probabilities observed for the four candidate answers gives a measure for finding the degree of accuracy as shown in table 5.

This framework reduces the problem to the set of the ordered triplets (S, A, U), where a utility function, $u_{ij} \in U$ defines the value of utility, for taking a decision alternative, a_i where $a_i \in A$, $i = 1, 2, 3, 4$, when the system agrees to the fuzzy state, s_j where $s_j \in S$, $j = 0, 1, 2, 3$. [9]. Consequently, the probabilistic values expressing the relationship between the degree of precise answer ai and the similarity matching of dependency triples sj together constitute a utility matrix U as shown in table 6. These probabilistic values are obtained from the training of candidate samples which were performed by the manual assessment of Subject Experts [9].

Table 5
Precision Probabilities Contributing to Measures of Answer Accuracy and Content Coverage

Candidate	$p(s_0)$	$p(s_1)$	$p(s_2)$	$p(s_3)$
ca_1	0	0	0.625	0.375
ca_2	0	0	0.125	0.875
ca_3	0	0.25	0.5	0.25
ca_4	0.1	0.8	0	0.1

Table 6
The Precision Utility Matrix

	s_0	s_1	s_2	s_3
a_1	-1	1	5	10
a_2	-3	3	9	7
a_3	-7	7	4	3
a_4	-10	10	2	1

The analysis proceeded by computing expected utilities corresponding to above different states help in the decision making process. The expected utility associated with the j[th] alternative $E(u_j)$ is calculated using the following formulae:

$$E(u_j) = \sum_{i=0}^{3} u_{ji} * p(s_i)$$

where u_{ji} is the utility value for a given alternative a_j if the future state of similarity turns out to be state s_i and $p(s_i)$ is the probability of occurrence of possible state s_i.

Taking into consideration the expected utilities, the grading-decision alternative a_j is chosen which has the highest expected utility among the alternatives for any given candidate answer, i.e. the maximum expected utility $E(u^*) = \max_j E(u_j)$.

The expected utilities and the maximum expected utilities for the four candidate answers, are then formulized as shown in table 7, which shows that candidate answer ca_2 is 'accurate', ca_1 and ca_3 are precisely 'relevant' and ca_4 is a 'vague' answer.

Table 7
Expected Utilities of Candidate Answers for Precision

	$E(u_1)$	$E(u_2)$	$E(u_3)$	$E(u_4)$	$E(u^*)$
ca_1	6.875	8.25	3.625	1.625	8.25(a2)
ca_2	9.375	7.25	3.125	1.125	9.375(a1)
ca_3	7.5	13.75	11.25	6.0	13.75(a2)
ca_4	1.7	2.8	5.2	7.1	7.1(a4)

Further, the desired content coverage of the answer with respect to the model answer is calculated using recall metric as discussed in the first approach. The utility matrix for recall depicting the utility functions uij stating that the decision is a_i when the state is s_j is shown in table 8. The expected utilities and maximum expected utilities for recall using the formulae given above as in precision is depicted in table 9.

Table 8
The Recall Utility Matrix

	s_0	s_1	s_2	s_3
a_1	-1	1	6	10
a_2	-5	7	9	3
a_3	-10	10	4	1

Table 9
Expected Utilities of Candidate Answers for Recall

	$E(u_1)$	$E(u_2)$	$E(u_3)$	$E(u^*)$
ca_1	7.5	6.75	2.875	7.5(a1)
ca_2	9.5	3.75	1.375	9.5(a1)
ca_3	5.75	6.3	4.75	6.3(a2)
ca_4	1.7	5.4	7.1	7.1(a3)

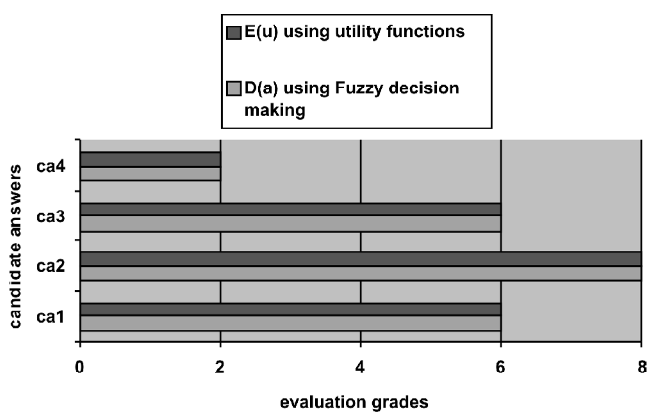

Figure 5: Comparing Candidates' Precision-evaluation Grades

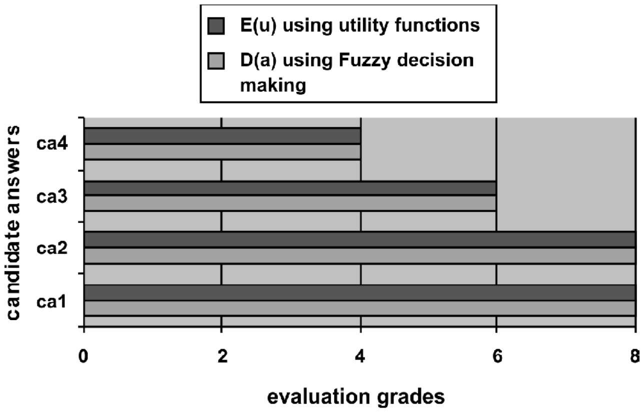

Figure 6: Comparing Candidates' Recall-evaluation Grades

6. CONCLUSION

When the precision and recall values for the four candidate answers were evaluated with the two above mentioned methodologies, the observations as illustrated in figures 5 and 6 were found to give exactly similar assessment (precision and recall) grades to the candidate responses. The training on the sample sets by the human subject experts that one-third, two-third or total dependency match participate in precision and recall measures to provide the results. The more the count of one-third dependency match leads to incomplete answers and the more the count of two-third or total match provides more accuracy of answers. This conveyed the justification of imbibed logic upon the assessable parameters as showing fuzziness for Subjective Question-Answering Assessment systems. These grades prove to be fair and unbiased judgments with respect to the assessments done upon experimented question by human assessors.

7. ACKNOWLEDGEMENT

The authors wish their heart felt gratitude to the Management, Bhilai Institute of Technology, Durg, India for their inspiring encouragement and support towards the completion of the work.

REFERENCES

[1] Burger John, Cardie Claire, Chaudhri Vinay, Gaizauskas Robert, Harabagiu Sanda etal.: "Issues, Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A)", *http://www.nlpir.nist.gov/projects/duc/papers/qa.Roadmap-paper v2.doc*

[2] Dang H. T., Kelly D., Lin J., Overview of the TREC 2007 Question Answering Track, Proceedings of TREC 2007, NIST, (2007).

[3] Diekema Anne R., Yilmazel Ozgur., Liddy Elizabeth D, (2002), Evaluation of Restricted Domain Question Answering Systems. Center for Natural Language Processing, School of Information Studies, Syracuse University, *Proceedings of the 42nd Meeting of the Association for Computational Linguistics*, *Workshop for Question Answering in Restricted Domains*. Barcelona, Spain, 25 July 2004.

[4] Lin Jimmy, Quan Dennis, Sinha Vineet, Bakshi Karun, Huynh David, Katz Boris,. Karger David R. (2003), The Role of Context in Question Answering Systems. *CHI '03: Extended Abstracts on Human Factors in Computing Systems*, pages 1006-1007, April 2003, Fort Lauderdale, Florida. (poster) [DOI: *10.1145/765891.766119*].

[5] Kowar M. K., Sharma Sanjay, Rawal Arpana, Thomas Ani (2006), A Framework for Automatic Answer Evaluation in Restricted Domains. Proceedings of National Conference on Applications of Information technology : Emerging trends, PIMR, Indore, *E-mail: kowar_bit@rediffmail.com*

[6] Matteo Negri, Milen Kouylekov, Bernardo Magnini, Bonaventura Coppola: Reconstructing Diogene: ITC-irst at TREC 2006, Proceedings of TREC 2006 (2006).

[7] Ross Timothy J., a book on "Fuzzy Logic with Engineering Applications", Mc Graw Hill International Editions (Electrical Engineering Series).

[8] Stephan Bloehdorn, Philipp Cimiano, Alistair Duke, Peter Haase, J¨org Heizmann, Ian Thurlow, and Johanna V¨olker "Ontology-based Question Answering for Digital Libraries", *Proceedings of the 11th European Conference on Research and Advanced Technologies for Digital Libraries (ECDL 2007), Budapest, Hungary, September 16-21 2007*, pages: 14–25, Springer, Lecture Notes in Computer Science, September, 2007.

[9] Thomas Ani, Kowar M. K., Sharma Sanjay (2008), "Intelligent Fuzzy Decision Making for Subjective Answer Evaluation", available online at IEEE Xplore release 2.5, DOI: 10.1109/ICETET.2008.216.

[10] Thomas Ani, Kowar M. K., Sharma Sanjay (2008), "Automatic Subject Indexing: A Bootstrapping Approach through Text Mining operations", *International Journal of Engineering Research and Industrial Applications*, **1**(5), 163-174.

[11] Thomas Ani, Sharma H. R., Kowar M. K., Sharma Sanjay (2009), "Recognizing Multi Objective Decision Making Features in Subjective Question Answering", *International Journal of Computer Engineering and Information Technology*, *www.serc.org.*

[12] Tufis D., Stefanesu D., Ion R., Ceausu A., RACAI's Question Answering System at QA@CLEF 2007, Proceedings of CLEF 2007 (2007).

[13] Valenti Salvatore, Neri Francesca, Cucchiarelli Alessandro. An Overview of Current Research on

Automated Essay Grading. DIIGA- Ancona, Italy, *Journal of Information Technology Education*, **2**, 2003, 319-330, *valenti@inform.unian.it, neri@inform.unian.it, alex@inform.unian.*

[14] Ziheng Lin, Huu Hung Hoang, Min-Yen Kan, Long Qiu and Shiren Ye.: NUS at TAC 2008: Using Augmented Timestamped GraphModel and Opinion Snippet Sentence Mining, Proceedings of TAC 2008 (2008).