6 AUTOMATIC EVALUATION & GRADING OF NON-FACTOID TYPE QUESTIONS BASED ON BE METHOD IN HKDA INTELLIGENT PAPER EVALUATION SYSTEM

There has recently been a significant increase in the number of community-based question and answer services on the Web, where people answer other peoples’ questions. These services rapidly build up large archives of questions and answers, and these archives are a valuable linguistic resource.

Question Answering is a technique to find information from a huge text base using a given question. Several Web-based question answering systems have been also designed during last two decades. However, the most widely used question types of the evaluation are multiple objective types, limited to yes/no pattern, multiple choice/single or multiple answer, and fill-in-blanks questions with a fixed string and numeric answer.

One of the major tasks in a essay type question is that while giving answer by students/learner is to automatic evaluate in the archive that a semantically similar to a answer service. The goal of the learning and testing such automated examination system is to improve a student’s writing skill is to write (answering essay questions), receiving feedback (marks too) from the teachers/tutor (after correction).

Therefore, researcher focuses on answer extraction method for non-factoid questions. Researcher classified non-factoid type questions into three types: why type, definition type and how type. Researcher analyzed each type of questions and developed answer extraction patterns for these types of questions. Basic Element (BE) method is originally proposed by Hovy et.
and researcher applied this BE method for question answering evaluation. Evaluation is done by comparison between BEs of system answer and BEs of correct answers.

In this thesis, researcher proposed a web based Intelligent Tutoring System (Human Knowledge Discovery System Intelligent Paper Evaluation System, HKDA-IQPES) for automatic evaluation & grading of Essay Questions (descriptive type answers). HKDA-IQPES can be considered as a self sufficient module for evaluation and grading of Essay Questions in an environment where the students learn the things through regular class room based instructions & write exams online.

6.1 Introduction
The increasing demand for student/trainer to acquire multi-disciplinary knowledge and skills has led to increased pressure on academics to ensure transfer of adequate knowledge and skills to all student/trainer. Several Web-based learning and evaluation systems have been also designed during last two decades, such as WebCT [159], QUIZIT [157], ASSYST [69], PILOT [150] and e-manipal as proposed name for the interactive Learning System (iLS) [126].

Such system provides learning materials in a better-understood format as well as evaluation of content understanding. However, the most widely used question types of the evaluation are multiple objective types, limited to yes/no pattern, multiple choice/single or multiple answer, and fill-in-blanks questions with a fixed string and numeric answer [130, 154, 108, 170]. One of the major tasks in a essay type question is that while giving answer by students/learner is to automatic evaluate in the archive that a semantically similar to a answer service.
The goal of the learning and testing such automated examination system is to improve a student’s writing skill is to write (answering essay questions), receiving feedback (marks too) from the teachers/tutor (after correction). In addition to these, such systems were designed without keeping in mind the individual capability and capacity of understanding. Due to this reason it is almost impossible to satisfy individuals’ needs.

6.2 Question Types of the HKDA Examination System

The main purpose of the question bank is to eliminate or reduce to a considerable extent, the defects on setting a question paper. At present, examiners and paper setters usually adopt an ad hoc procedure. The selection of questions is not done systematically and it is not uncommon for paper setters to just repeat questions from the last few years question papers.

There is no attempt to work out a table of specifications and to select/write questions to represent all the cells in such a table. A systematically developed pool of questions of evaluation, namely a question bank, offers a chance to set up valid and reliable question papers to compare the performance of a group of students with that of the wider population and to match the examinations to the curriculum that is taught, instead of the other way round.

In this proposed HKDA System, researcher has classified the question bank into types of question patterns categories of base response types according to the response styles of answers.

This allows rapid learning of test questions to assess students’ problem solving knowledge. Researcher represent a very promising and expected evolution for the knowledge based agents for question paper generation for
facilitating its reusability to the teacher/tutor, student/trainer in this proposed system.

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Description</th>
<th>Response Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>True/false</td>
<td>Selecting a response from the choices ‘True’ and ‘False’</td>
<td>LID</td>
</tr>
<tr>
<td>Single response</td>
<td>Selecting a single response from the choices</td>
<td>LID</td>
</tr>
<tr>
<td>Multiple response</td>
<td>Selecting multiple responses from the choices</td>
<td>LID</td>
</tr>
<tr>
<td>Order</td>
<td>reordering the choices that are displayed initially</td>
<td>LID</td>
</tr>
<tr>
<td>Associate</td>
<td>pairing up the choices that are displayed initially</td>
<td>GRP</td>
</tr>
<tr>
<td>Match</td>
<td>pairing up choices from a source set into a target set</td>
<td>GRP</td>
</tr>
<tr>
<td>Gap match</td>
<td>filling gaps from an associated set of choices</td>
<td>GRP</td>
</tr>
<tr>
<td>inline choice</td>
<td>filling gaps from a shared stock of choices</td>
<td>LID</td>
</tr>
<tr>
<td>Text entry</td>
<td>filling gaps by constructing a simple piece of text</td>
<td>STR/NUM</td>
</tr>
<tr>
<td>Extended text</td>
<td>entering an extended amount of text</td>
<td>STR/NUM</td>
</tr>
<tr>
<td>Hot text</td>
<td>Selecting choices embedded within a surrounding context</td>
<td>LID</td>
</tr>
<tr>
<td>Hot spot</td>
<td>Selecting areas (hotspots) in the graphic image</td>
<td>LID</td>
</tr>
<tr>
<td>select point</td>
<td>Selecting points in the graphic image</td>
<td>XY</td>
</tr>
<tr>
<td>graphic order</td>
<td>reordering the choices that are presented as hotspots on a graphic image</td>
<td>LID</td>
</tr>
<tr>
<td>Graphic associate</td>
<td>pairing up the choices that are presented as hotspots on a graphic image</td>
<td>GRP</td>
</tr>
<tr>
<td>graphic gap</td>
<td>a graphical interaction of filling gaps from an</td>
<td>GRP</td>
</tr>
</tbody>
</table>
match set of choices
position object positioning a given object on the image XY
Slider Selecting a numerical value between a lower and upper bound NUM
Drawing using a common set of drawing tools to modify a given graphical image FILE
Upload uploading a pre-prepared file representing the response FILE

Table 6-1: Question Types Supported in the HKDA System

These question types can be classified into six categories of base response types according to the response styles of answers. They are:

- **LID (Logical Identifier)**
  - A category of response styles that presents various choices and provides a mechanism for the test taker to select one or more choices.

- **XY (X-Y Co-ordinate)**
  - A category of response styles that presents an image, or various images, for the test taker to select a position on the image or images to indicate their choice.

- **STR (String)**
  - A category of response styles that allows the test taker to enter text.

- **NUM (Numeric)**
  - A category of response styles that allows the test taker enters a number to indicate their choice. The entered number can be integer or float data type.

- **GRP (Logical Group)**
A category of response styles that allows a test taker to group objects together to indicate their choice. It can be classified into 2 sub-categories: pair and directed-pair.

• FILE

A category of response styles that allows a test taker to upload a file.

Richer types of answers in this system need to be graded manually by the Teacher/Tutor and later on automatically.

This thesis is concerned with the design and construct a heterogeneous environment, development in an open architecture and implementation of cost-effective ITS to teach procedural knowledge as well as facilitating the acquisition of conceptual knowledge, in multiple subjects, self sufficient module for evaluation and grading of Essay Questions in an environment where the students learn the things through regular class room based instructions & write exams online.

6.3 Overview of HKDA System Intelligent Paper Evaluation System (HKDA-IQPES)

The HKDA-IQPES System is based on web based service oriented architecture, Domain Independent Information (DII) and Domain Specific Information (DSI) which can be easily updated. DII such as personal information, Background & Experience, Goals, Preferences etc. and DSI such as Overall competence level, Competence level for each structure unit, Competence level for each subject node, Misconceptions for each subject node are updated using initialization and adaptation algorithm.

This system is helpful to classify new user, set initial values for overlay model, distinguishes several typical group of Users such as Learner, Level-1 ... n and Expert-1 ... n. The System is allowing agents including learner
to add questions from their specific domain in the Question Bank (QB). Those agents who are trying to add questions in other domain Question Agent will take care of it. Similarly agents have their own privileges to update the system or the QB or review on demand. HKDA-IQPES System is shown in Figure 6.1, where various agents playing their role in different areas of functions.

**Examination Designing:**

This functionality is managed by Teacher/Tutor, Exam Agent (EA), QA. Auto generated question are also prepared if agents are busy with other work.

**Examination Maintenance:**

Major role is played by Domain Advisory Agent (DAA) and passed it to Knowledge Agent (KA): The agents maintaining the Examination and they may Teacher/Tutor, EA, QA.

**Examination Supervision:**

Dedicated and Knowledgebase managed by our YAW Agent of system named as You Are Watching System. From the beginning to end of the exam, monitoring the student/trainer’ conditions on the teacher's screen.

**Question Paper Answering:**

This system is not specific to Student/Trainer but also any level of users who can go for test to improve in different domain.

**Paper Scoring and Review:**

Paper result is declared by the Scoring Agent (SA) and passes it to AA, KA, EA, DAA, Teacher/Tutor, and Student/Trainer. Review Agent (RA) presents all reviews result and report to all such agents.
All the questions, materials and answers are transmitted in bits stream format after encoding, not in file format. It can ensure not only security in the transmission process, but also protect from invasion.

When one student starts his/her test, the system will distribute him/her a password produced according to his/her client’s hardware and operation system environment. It can prevent cribbers to login from other clients imitating this student.

The newest method to prevent the distant student from leaving the client or is making use of face detection and recognition technology. Before the start of the exam, the student is asked to sit in front of the computer camera, which can monitor the whole scene. Then the system will automatically recognize the student’s face and his identification. As soon as the validation is performed, the student must stay in front of the camera until the exam is finished.

If there is a change of power failure or Internet disconnect then it can also track by the YAW Agent.
Standardization of Intelligent Tutoring System for Adaptive Information Retrieval and Knowledge Discovery by Monitoring Human Interaction
6.4 HKDA System Intelligent Question Answering Paper Evaluation System (HKDA-IQPES)

Question Answering is a technology to find information from a huge text base using a given question [12, 58]. With regard to the appropriateness of writing style, researcher manually constructed a set of lexico-syntactic patterns that match with key sentences and assign score of appropriateness to matched sentences. With regard to the relevance to the question, researcher proposed a method to retrieve passages related to topic using lexical chain. The retrieval method also assigns relevance score to each passage.

A set of answer candidates consists of not only the extracted passages but also the key sentences. Each answer candidate is assigned score of answer, which is a combination of the score of appropriateness and the score of relevance as well as the score of document retrieval, in order to select “good” answer candidates. The process is summarized as follows.

6.4.1 Process Flow of Question Answering

The process is summarized as follows.

Step 1: Classify the input question into a type of question using a set of manually constructed lexico-syntactic patterns for question classification. Each of patterns detects an interrogative and a question focus in a question and classifies the question into one of types shown in Table 6-2 when the pattern is matched. In the Table 6-2, other represents the type of non-factoid questions other than definition, why and how. It should be noted that the questions of the type factoid are treated as the type of other in the system with a different architecture.

Step 2: Analyze the question to extract keywords that are content-bearing words or compound nouns, which are series of noun phrases and
undefined words. When the question type is definition, also extract a question target, which is a topic noun phrase, and the head of the target, which is the last Bonesets segment of the topic noun phrase. With regard to why, how, and other, extract a clue predicate, which is a last verb or a last adjective.

Step 3: Retrieve documents with keywords and additional keywords. The additional keywords are obtained from the result of Web search.

Step 4: According to the type of the given question, apply a set of manually constructed lexico-syntactic patterns of writing styles to each sentence in retrieved documents. Each sentence is given a score of appropriateness that is associated with a matched pattern.

Step 5: Filter out irrelevant documents according to the document score and the sum of appropriateness score of sentences in the document.

Step 6: From remaining documents, retrieve passages related to the topic of question using the information about lexical chains related to the topic, and assign a relevance score to each retrieved passage.

Step 7: Calculate a unified score for each answer candidate by combining the score of appropriateness, the score of relevance, and the score of document retrieval. Then, select a set of final answer candidates by using clustering-based summarization technique.

<table>
<thead>
<tr>
<th>Name of type</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why</td>
<td>(What is the reason for)</td>
</tr>
<tr>
<td>Definition</td>
<td>(What is the meaning of DBMS?)</td>
</tr>
<tr>
<td>How</td>
<td>(How do I do)</td>
</tr>
<tr>
<td>Factoid</td>
<td>(Who is ) (Where is)</td>
</tr>
</tbody>
</table>
Factoid is treated as other in the system with a different architecture.

### 6.4.2 Why-type Question

As for why-type question, researcher will use the following extraction patterns and non-extraction patterns. If one sentence matches extraction patterns, this sentence will be extracted as answer candidate. But this candidate will be removed from candidate list if it matches non-extraction patterns.

**Extraction patterns**
- Verb + “(tame)”
- Noun + “(tame)”
- (tame) + Postposition “(ni)”
- (tame) + “—”
- (tame) + Auxiliary Verb “(da)”

**Non-extraction patterns**
- Pronoun + Postposition “(no)” + “(tame)”
- Verb + “(tame)” + Postposition “(no)”
- Noun + “(tame)” + Postposition “(no)”

The semantic clue words are the words, which mean reason, cause and background.

### 6.4.3 Definition-type Question

Definition-type questions require word meaning, term definition, description of term and so on. For e.g., in the question “What is World Heritage Convention?”, it requires definition of “World Heritage Convention” which is the most important element in this question. Researcher calls the important element Main Keyword.
In order to choose Main Keyword, researcher firstly checks blanketed word or named entity, then modifier of topic word, and finally, topic word. In the question “What agreement is World Heritage Convention?”, the word “agreement” is also important as well as Main Keyword. Researcher calls this kind of word Attributive Word ‘ ’. Attributive Word is the word, which composes noun phrase with an interrogative. Extraction patterns are shown as follows:

Main Keyword + “(ha)” — “(ga)” — “(mo)”
⋯ “(ga)” + Main Keyword + “(wo)”
⋯ “(suru)” + Main Keyword
⋯ “(no)” + Main Keyword
Main Keyword + “(toha)”
⋯ “(no)” + Attributive Word
⋯ “(suru)” + Attributive Word

If a matched sentence includes Main Keyword, the whole sentence will be an answer candidate. If a matched sentence includes Attributive Word, its modifying element will be an answer candidate.

6.4.4 How-type Question

How-type question is inquiry of some procedure, method or conditions of action. Verbal expressions in a question sentence will be clue to recognize answer for this type of questions. For example, in the question “How is World Heritage decided?”, the verb decide will be important clue for answer extraction.

Extraction pattern for How-type question, researcher will uses the main verb (Main Verb) of a question sentence and Main Keyword, which is clue for definition type question. Extraction patterns are shown as follows:
6.5 BE Method Applied in HKDA-IQPES

BE method proposed by Hovy et. al. was used for automatic evaluation of text summarization applied in HKDA System Intelligent Question Answering Paper Evaluation System (HKDA-IQPES). BE is defined as a minimal semantic unit which consists of two elements and relation (head-modifier-relation) between these elements [99, 100]. This relation names are mainly from parse tree.

In order to evaluate HKDA-IQPES System summary using BE method, each sentence of HKDA-IQPES System summary and reference summary will be parsed and parse tree of each summary will be broken into BEs. Evaluation is done by comparison between BEs of reference summary and BEs of HKDA-IQPES System summary. If BEs of each summary are similar, HKDA-IQPES System summary will be a good summary.

Researcher will show an example of BE breaking using the following sentence.

| Two Libyans were indicted for the Lockerbie bombing in 1991. |

In this sample sentence, word “two” modifies “Libyans” and they are connected by relation “nn” (a sequence of nouns). Words “Libyans” and “indict” have relation verb-object. The results of BE breaking will be shown in Table 6-3.
There are several levels of BE matching proposed by Hovy.
1. exact matching at lexical level
2. matching at the level of word original form
3. matching at the level of synonym
4. matching with paraphrase at phrase level
5. matching at semantic level

Moreover, there will be partial matching of BE elements and reference resolution of BE elements. However, current implementation of BE breaking and matching is at the level of lexical and word original form level.

6.6 BE-based Evaluation of HKDA-IQPES

In BE-based evaluation, HKDA-IQPES System answers are scored by comparison between BEs of HKDA-IQPES System answer and BEs of correct answers. Score between one system answer and one correct answer is calculated in F-measure as follows:

\[
\text{Precision}(P) = \frac{\text{matched BEs}}{\text{number of BEs of system}}
\]
Score of one system answer will be the max score in all the scores calculated by the above F-measure for all correct answers because correct answer which has the max score will be recognized as the most similar one to the system answer. In this evaluation, if a small part of system answer is almost same as one correct answer, score of this system answer will be low. When size and contents of answers are almost the same, score will be high.

6.7 Conclusion

In this thesis, researcher describes answer extraction method for non-factoid questions. Researcher classified non-factoid type questions into three types: why type, definition type and how type. Researcher analyzed each type of questions and developed answer extraction patterns for these types of questions.

For automatic evaluation, researcher has proposed BE based evaluation HKDA-IQPES System for answers of questions. BE method is originally proposed by Hovy et. and researcher trying to apply BE method for question answering evaluation. Evaluation is done by comparison between BEs of HKDA-IQPES System answer and BEs of correct answers.