Saarland University

Faculty of Natural Sciences and Technology I

Department of Computer Science

Master thesis

Development of a framework for generating assessment items in a multilingual e-learning environment.

Submitted by

Ramya Rangaswamy

Submitted

15.11.2016

Supervisor

Prof. Dr. Jörg Siekmann

Advisor

Prof. Dr. Sergey Sosnovsky

Reviewers

Prof. Dr. Jörg Siekmann

Prof. Dr. Sergey Sosnovsky

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Statement in Lieu of an Oath

I hereby confirm that I have written this master thesis on my own and that I have not used any other media or materials than the ones referred to in this thesis.

Einverständniserklärung

Ich bin damit einverstanden, dass meine (bestandene) Arbeit in beiden Versionen in die Bibliothek der Informatik aufgenommen und damit veröffentlicht wird.

Declaration of Consent

I agree to make both versions of my thesis (with a passing grade) accessible to the public by having them added to the library of the Computer Science Department.

Saarbrücken, _____ _____

Datum/Date (Unterschrift/Signatare)

Dedicated to

My

Parents & Husband

for their unconditional support and love

Table of Contents

Acknowledgement

It is my pleasant duty to express my sincere thanks to my supervisor, Dr. Sergey Sosnovsky, for having provided me with the opportunity and guidance to pursue this thesis. I would like to thank him for all his advice and constant supervision from the very beginning. I am grateful to Prof. Dr. Jörg Siekmann, for being my supervisor and for his support. I am thankful to Thorsten Hey for giving me an excellent introduction to Interlingua and for all his technical advice and timely help.
Special thanks and appreciation for my husband Balakrishna for his encouragement, constant support, and motivation to accomplish my studies at Saarland University.

My profound thanks and I owe enormous appreciation to the authors whose reference I have cited in my master thesis. Finally, without the support of my family and friends, this journey would not have been possible. They give me their support every day, and when tiredness appears, they always have encouraging words.

Abstract

An effective automatic assessment generation framework is essential for students to evaluate their level of the understanding subject. At the same time, designing and implementing competent E-assessments creation for students is very much required. The added E-assessment generation environment to Interlingua, improve a student's cognizance, skills, and subject understanding through multiple choice questions. It can also provide immediate automatic grading, especially with the objective tests like choice questions and cloze questions (multiple choice). This progresses the complete interactivity of the service and allows students to test own knowledge. E-assessments are utilized in online learning courses as a way for an instructor or manager to identify the level of comprehension of a subject by the students.

This Master thesis designs the framework for E-assessment generation for Interlingua system. The creation of self-assessments supports international students to access learning material in their mother language. In this context student, should navigate along with learning material given in different languages. Here system provides E-assessments to students to improve their understanding in reading comprehension questions and vocabulary in foreign language (English, German and French). More importantly, they can be automatically graded and provides option of frequent testing until the student is familiar with the chapter. This comes handy in cases where the student has a good understanding of a subject in the mother language but faces the problem when learning the same content in a foreign language (Statistics is a learning material being used).

While a student is navigating through the learning material for understanding the subject, at the same time, a student can request tests on the same screen to self-assess or judge the once own understanding of the concepts in that study domain. This improves student's knowledge and confidence level. Each E-assessments test item consists of a short text describing a question or a sentence to be tested and several choices, typically four. In single-response multiple choice question, one of the choices is the correct answer, and the wrong alternatives are called distractors [1]. Likewise, in this context, eight different types of E-assessments will generate to provide good exposure for students. These evaluations are built mainly on vocabulary and reading comprehension type of questions. The attention lies on formative E-assessments and the instructor as well as the student to adapt the learning process based on the feedback.

1. Introduction

1.1 Motivation

The motivation for this work is to assist the students who struggle hard in understanding comprehending subjects in a foreign language. Creation of E-assessment in the Interlingua system benefits international students to have a good understanding of the terminologies in the field that they want to pursue their education at foreign universities. An increasing number of foreign students will be benefitted by implementing a system like Interlingua. The Interlingua system aims at designing an online, ICT (Information and communication technologies) based solution that provides students with relevant learning material in their native language and an added online E-assessment facilitates Interlingua to help students to learn, prepare and progress. The complete interactive interface of the facility lets the students to test individual knowledge by taking E-assessments and supports more student-centred learning (Elementary Statistics is the chosen as a target subject).

Assessment generation is performed in multiple languages which include English, German and French. When students are pursuing education at foreign universities and in foreign languages they often face the daunting tasks like translating technical content to their mother tongue and to adapt to higher technical standards. These issues will be addressed with this system. The Interlingua gives access to students to learn the study material in their mother language. Students whose medium of instruction in the current programme is different from their mother tongue and they are not able to excel in their program as they were able to do earlier. Those students will receive a proper guidance and become capable of resolving their learning difficulties happen due to the low language proficiency and providing self-assessments. These self-assessments would comprise of items that are automatically generated based on learner preferences and the learning material attached to the section that the student is currently browsing. The keywords have been associated with the ontology. Subsequently, the system can produce these concepts across three different languages, and deduce semantic information used in the automatic item generation process.

The practical assessment for the students is an ongoing process that should be carried out in different phases throughout the education, at the same time, designing and implementing effective E-assessments are vital to make it successful. Typically, E-assessment refers to using technology to manage and deliver assessment [2]. Assessment generation takes input from the knowledge source or knowledge base (it is a database of learning material or it is an organized repository of knowledge consisting of concepts, relationships, and specifications). As an output, it produces a reasonable number of questions which assesses students understanding of the learning material given. Each question has a short text describing that question to be tested and a few choices with only one correct answer, and wrong alternatives are called distractors [1]. At the same time, a student can think and choose any of the choices. After answering all the questions and submitting, the response will be saved in the student's database, and the result of that assessment is showed immediately at the bottom of the same screen. The student can retake E-assessments until they get the hang of the subject.

Development of E-assessment types supports the self-assessment of student knowledge, skills and vocabulary proficiency within a selected topic and it ensures great support for foreign international students and encourages them to learn more interactively.

1.2 Problem description

Students now a day are enthusiastic in exploring new opportunities and in widening the horizons of their knowledge. Hence, there is an increase in the influx of international students at various universities. The change in the medium of instruction from their native language at a foreign university makes life difficult for the students who are already experiencing a cultural shock. Instead of investing their time and effort in learning the new concepts, the students are forced to devote more time in translating the course material to the language he/she is comfortable with. Pursuing education in a foreign language is a herculean task and can be a poleaxe to student`s morale and motivation. Hence the crux of the problem is the language of instruction. To overcome these problems, the current thesis implements E-assessments framework (multiple choice questions (MCQ)) for Interlingual system. This provides international students with on demand access to relevant learning material in their mother tongue.

Practice of E-assessments framework solves the problem of the students studying in a foreign language. In this web-based Interlingual E-assessment system focuses on statistics in three target languages. Statistics was chosen based on the fact that it is one of the introductory courses in the department of computer science at many Universities.

Competent assessment generation is a challenging task and a topic of ongoing research. With the help of modern technologies, the automatic generation of test items is created for questions like inline choice questions and cloze questions. The effective item generation has three main challenges they are 1) Selection of principal or critical sentence in the context 2) Identify sense making parts from where to create a gap for fill up the blank 3) The distractor generation [3]. To implement proper assessment item, we require clear requirements and well measurable items. Measuring item generation depends on the learning material used, to analyse and help in predicting the quality of the item.

To implement these technical challenges for effective E-assessments item generation the Interlingua system has used sophisticated algorithms, which are existing Linked Data sets i.e. DBpedia and the Statistics ontology to produce test item templates and ISI glossary is used for translation to provide test items in different languages. The Interlingua system analyses the given slides or the textbook section in various languages and extracts the key concepts and semantically interlinks the relevant learning objects, with the help of all these formative E-assessments are generated and delivered to students in their language of choice.
To accomplish these practices, system uses Natural Language Processing and Information Extraction fields, named-entity recognition, part-of-speech tagging, word sense disambiguation, probabilistic topic modelling.

This thesis delivers furthermost appropriate tests to individual students and generates enough items for important concepts from the textbooks automatically rather than requiring item authors to create items manually. E-Assessment provides a range of possible benefits like user-friendliness for all students including those with difficulties in understanding a foreign language and an immediate feedback to the students for automatically generating items using the semantic input from the client.

1.3 Short description of the approach

The Interlingua system was created as two distinct applications. The repositary and client applications. The repository interface was implemented by DFKI (Deutsches Forschungszentrum für Künstliche Intelligenz) and partner universities who provide to manage existing content and generate models for available documents. These can be accessed by a web-based API. This API provides access to the content and models and supports additions as well as removals. All interactions with students is managed by a client interface which in turn uses the repository's API to support the student with content and intelligent features during their work. The implemented E-assesements framework which is described in this thesis is flexible enough to accommodate various practise scenarios and well-organized to deliver enhanced learning practise to students requiring interlingual support.

The E-assessment generation is a special technology based framework for automatic generation of test items, such as multiple choice, cloze items or match items. Existing Linked Datasets, DBpedia and the statistics domain ontology are used to fetch test item templates and Multilingual vocabularies are provided to test items in different languages. Feedback elements will be based on the answers given by the students and will link additional learning resources. More focus is given to abstract information in elementary statistics, which comprises learning basic terms and their connection to other ideas in the ontology.

1.4 Structure of the thesis

This thesis is organized as follows Chapter 1 introduces the background of Interlingua project, an E-assessment system, in which the proposed approach has been implemented. Chapter 2 continues with a review of related research and background work on Distractors generation, Identification of gaps and Inline and cloze choice items in the text. This chapter also describes a previous tool created for gap detection. Chapter 3 provides the context of the current work including the description of the component implementation. Chapter 4 presents all the details and steps of the proposed approach, along with the necessary technologies such as the implementation details of the draft approach of item generation. Chapter 5 carries out the implementation and screen shots. Chapter 6 presents the conclusion and discusses the future work.

2. Background research and Related work

This section provides an overview and several other models for building Interlingua and the E-assessment framework.

2.1 Literature review of E-assessments generation

Mitkov et. al [4] presents a methodology which can generate multiple choice questions depending on the text corpora in some specific domain. It makes use of various techniques like shallow parsing, term extraction, sentence transformation and computation of semantic distance. This method also uses ontologies such as WordNet [5].

It encompasses three major steps

1. Term extraction concerning frequent concepts inside the text.
2. Stem generation.
3. Distractor selection.

The extraction of the terms occurs by surface level parsing of scanned text corpora. The stem generation filters the clauses and transforms the selected ones to the stem of an item by following simple rules that are aided by WordNet. The final step distractor selection is dictionary based and uses mostly WordNet to get the related distractors. Lin et. al [6] and Brown et. al [7] generated vocabulary questions from the WordNet dataset, which is available as RDF (Resource Description Framework) data.

Holohan et.al,[8] presented the *OntoAWare* system, which provides a set of tools useful for learning content authoring. This employs the semantic web technology with knowledge representation standards and knowledge-processing techniques. One of the key features of the *OntoAWare* system is the generation of questionnaires from ontology elements. Production of learning objects can be achieved by customizing an existing ontology or even by creating a new ontology from scratch. Holohan et.al [9] provided an improvement of the *OntoAWare* system which was aimed to generate E-assessments for problem-solving skills in the domain of relational databases. These E-assessments are produced by utilizing an ontology which describes the domain in question. Students may customize the system to create personalized problems.

2.2 Brief description of the background

For generation of multiple choice questions succeeding three stages are important they are, identifying relevant sentences from which gaps are to be created, creating a gap in the selected sentence and creation of relevant distractors.

The framework of this work involves usage of existing semantic models and Linked Data repositories in the domain of statistics. More importantly, assessment item generation will be implemented for automatic generation of test items. This comprises the analysis of the multilingual learning material, mining of meaningful data from ontology, interlinking of related learning items, generation of formative assessment items and delivering E-assessment platform for students to enrich their knowledge as per their choice. In the following section, we can observe how these technologies can be used together meaningfully.

2.3 Appropriate gap creation in the sentence for cloze item generation

The two important criteria here is to identify the best sentence in the content to create a gap. Shah [10] and Becker et al. [11], have implemented an automatic summarization algorithm to identify meaningful sentences in the content and determine an appropriate gap in those sentences. Heilman [12], Iwata et al. [13], Mostow et al. [14] have produced questions for reading comprehension using NLP(Natural language processing) and machine learning techniques. For inline choice items like multiple choice, Interlingua uses summarization mechanisms and part of speech to identifying gaps.

2.4 Distractor generation

For any question creation either it is choice or cloze, the most important and challenging task is to create relevant distractors as per the question generated. Incorrect answers should look more reachable or sibling to the correct and semantic graph delivers related distractors randomly [15], [16] this helps to make question hardness. The semantic similarity metric is useful to advance distractor generation more strongly [17][18]. Comparison method is based on the semantic and distributional similarity (for similar words). For multiple-term distractors, Mitkov et al. [4] selected noun phrases with the same answers.

Interlingua system uses the distribution of resemblances to improve the generation of distractors because it supports names and proper nouns in different languages according to Smith et al. [19]. Item difficulty and specific matching guidelines are applied to create appropriate distractor. Semantically similar distractors (alternatives) for a correct answer are also valid, but they cannot be correct answers or not correct replacements.

2.5 Quality or features of the item for E-assessments generation

The quality of the item is the prominent issue for E-assessments generation. Gierl et al. [20] introduced item generation to support the identification of psychometric properties for a given template. This distinguishes between similar and non-similar items. Another idea is to evaluate the parameter that affects item difficulty which is subjected to the type of item that we are using. Sonnleitner [21] developed the LLTM model (Linear Logistichen Test-Modells) to recognize partly the elements that contribute to the difficulty of item generation. This also partially helps in reading compression.

2.6 Gap identification and Distractor generations in multiple choice or inline choice item generation

The system also has inbuilt reading comprehension questions that allot metrics (measurability) which support the item difficulty. To identify the important sentence, a summarizer API (Application programming interface) is used. Gaps are carefully chosen from the summary. DISCO (Distributionally related words using co-occurrences) and to retrieve distractors which depends on the distributional similarity algorithm [22] is used It assumes that words with comparable meaning occur in similar contexts.

The distractors with plural and singular items are omitted with the help of stemming algorithm, Soundex algorithm eradicates similar phonetic words (viz., son/sun), and few distractors do not grammatically fit the gap are also eliminated with the help of Stanford Dependency API.[23]

2.7 Text difficulty and Gap dispersity in item generation

Text difficulty applies to item difficulty. Consequently, choice item, gap detection and the relationship between the distractors and correct answers can influence the item difficulty. There are many different metrics to measure text's difficulties in different lan-

guages. One such difficult metrics is Franchois highlights AI readability [24] which analyses the text from lexical, syntactic and semantic level.

Automatically gap generated distractor items are related to correct answers sometimes. For example, correct answer: computer and distractors: hardware, software and workstation have distributional similarities to the right answer. Furthermore, it also measures the distractor relation with each other [23].

3. Context of Interlingua and E-assessment item generation

The context of the work explains the technical detailed implementation of Interlingua System and E-assessments. It outlines the main parameters of Interlingua, assessment item generation and feedback for the E-assessments.

To develop a right system, the Interlingua Interface was designed as two distinct applications. They are a repository and the client interface. Repository application manages the given content and generates models with the help of web-based API, and repository API manages the student interaction. The repository application links learning materials in all three languages and content is processed to extract a structural and semantic model. These are used by client application to process meaningfully. It is domain independent which means it can be extended outside statistics. Interlingua analyses multilingual learning resources and extraction of meaningful information, interlinking of relevant learning objects, generation of formative assessment items, and finally provides supporting material to students in their language and mainly it does not produce automatic language translation instead it produces learning resources such as lecture slides or textbooks. From this, it fetches their main topics and uses this to link relevant learning objects to each other semantically.

Interlingua system is more strengthened by the addition of assessment generation component, thereby making it more a powerful platform. The three most important components of Interlingua system are a front-end facility or client interface to provide students with enriched linked content, E-assessments item generation and providing tests to students. Once the student logs into the client application, the student is motivated to learn. Meanwhile, student can arise a query for semantically equivalent translation in his/her mother language and also can take test in eight different question formats.

In the backend, Interlingua system processes queries, textbook or slides automatically in target languages. A system will be able to recognise all the interior contents of the textbook like chapters, sub-chapters, and index. It represents all these contents semantically and interlinks different textbooks in various languages. Finally, this provides MCQ items, which has proved to be an efficient tool for measuring the progress of student.

3.1 Ontology

Ontology is a model that represents knowledge as a set of concepts within a domain and the relationship between these concepts. It is knowledge management as it captures knowledge within an organization as a model (in our case, it is Statistics ontology) and relevant semantic annotation. It is a dominant learning material that links different textbooks in different languages. It can differentiate between two different terms that have the same meaning. The annotation should be dense and more informative, it should cover all the pieces of learning resource this makes linking components well and more efficiently notice related content across the textbooks.

Due to the recent developments in semantic web technologies a lot of attention is being paid to them in emerging ontology based applications and also in many research areas. One such research area is the field of question generation, a sub-field of artificial intelligence.

3.2 ISI Glossary

International Statistics Institute (ISI) is the organization which has given 3,500 statistical terms (Glossary) in 31 languages and can also translate as per the target language selected. In this case, it is English, French and German. System has fetched all 3.500 terms to define first version of ontology.

SKOS (Simple knowledge organization system) is the semantic web standard used here for statistics ontology.

3.3 Semantic web, Linked Data and DBpedia

The glossary does not convey semantic relations between concepts. To enrich this model additional resources are used. To make it more efficient and enhanced, the DBpedia is used, it has mined a rich knowledge base from Wikipedia and served this knowledge base as Linked Data on the web [25]. During the extraction process, ordered information from the wiki such as infobox fields, groups, and page links to be extracted as RDF triples are added to the knowledge base as properties of the corresponding URI [26]. And more significantly it allows inference for comprehensive concepts and relations be-tween them. This notices linkage among pages and provides information to generating self-assessments. Linked Data has become increasingly popular as a lightweight approach to connecting data. Figure1 shows the Linked Data approach. A major

advantage of using DBpedia as a linking hub because it contains semantic relations bridging different domains. The system uses structured textbook in digital formats for extraction. It recognises the textbook formatting for proper index, contents, sections and subsections and represents as an SKOS model. Finally, maps the index terms to the ontology concepts and then simply follows the links to the correct pages and creates meaningful annotations.

Knowledge bases are playing a major role in enhancing the intelligence of web, enterprise search and in supporting information integration. Today, most knowledge bases cover only specific domains, which are created by relatively small groups of knowledge engineers, and are cost intensive to keep up-to-date as domains change Wikipedia is a central source for big network of linked information. The DBpedia leverages this gigantic source of knowledge by extracting structured information from Wikipedia and by making this information accessible on the web.

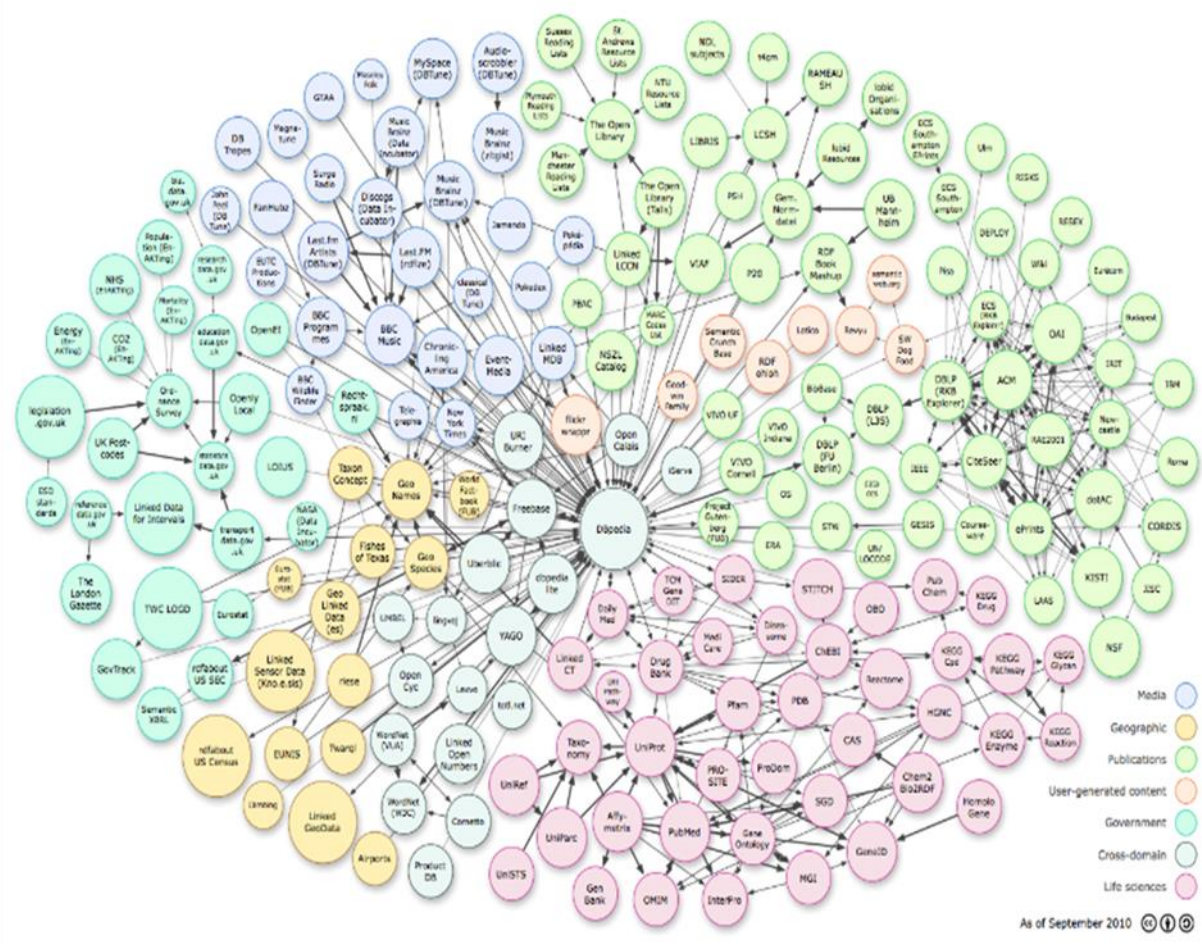

Figure 1: The Linking Open Data cloud diagram. (Image source Wikipedia)

3.4 E-assessment item generation

The system generates reading comprehension and cloze questions. Thus, the system enables the creation, answering and scoring. To provide the most appropriate tests and context, enough items are created for all important concepts from the textbooks and consequently questions are generated.

Questions consist of four main elements: the text stating the question, a set of possible answers called options, the key or the option that is the correct answer, the incorrect options known as distractors. The basic strategy of item generation is to decide on a suitable key concept for the question and then generate distractors. There is a need that distractors should be as close as possible semantically to the key [27].

The system defined the task concept translation and reading comprehension (definition). It translates a concept from a foreign language into the student's mother language. The student should choose between multiple options and identify the correct translation for the items.

To create relevant E-assessments to learners, sufficient items must be generated, and important concepts are identified from the learning material to create items. For reading compression questions system generates a key concept for all definitions as per the chapters. The student should choose between multiple options and identify the correct option for the definition.
Items are produced from the ISI taxonomy representing concepts in multiple languages, and items are created for the following language sets: French – German, German – French, English – German, German – English, English – French, French – English.

The difficulty of item creation depends on the development of a semantic memory representation of the concepts, closeness of the correct translation taken from a concept in the foreign language as given in the stem and closeness to the similar options.
Following are the item quality measurable features: to what extent options are related to the key and how close the stem variable is to the key.

4. Description of the approach

Interlingua is focused on factual and conceptual knowledge in domain statistics, which involves learning basic concepts and their relation to other concepts in ontology.

E-assessments are often required to demonstrate, gauge the level of understanding of an individual once the student learned some new concepts from study materials. In some cases, these assesments are part of a qualification or certification program.

A special technology based implementation is used in this thesis for automatic generation of test items, such as multiple choice, cloze items or match items. Existing Linked Datasets, such as DBpedia ontology are used to instantiate test item templates. Multilingual vocabularies (ISI) will be employed to provide test items in different languages. Feedback elements will be based on the answers given by the students and will link additional learning materials.

One of the interesting research questions was to find how better quality distractors could automatically be generated for multiple choise E-assessments. Some of the very important research works in the field of distractor generation are focused on similarity measures for the automatic generation of better quality distractors using natural language processing techniques [4], as well as lexically structured databases such as WordNet [13]. The quality of multiple choice items is an important topic in education since poor quality items may give away the correct answer or irritate a learner. Distractors should be concise, independent, unambiguous and similar in content, length, and grammar. These guidelines have been considered during the implementation of the automatic distractor generation procedures for each question type.

In the context of assessment generation, we need to get all the concepts introduced in any chapter selected by the student. A concept is an abstract, universal idea, notion or entity that serves to designate a category or class of entities, events or relations. It is a mental picture of a group of things that have common characteristics. Classes delineate concepts in the domain, so they are the focus of most ontology. Semantic relations depict the collaboration of two concepts. Properties describe various features and attributes of the concept. Properties can have different restrictions such as value type, allowed values, the number of values and other features of the values the property can take. In the context of AIG, the crucial part is distractor generation or the incorrect option for a given question.

We use distributional similarity as a mechanism to optimize the generation of distractors because it provides a mechanism to support names and proper nouns in multiple languages. Also, we apply rules related to item difficulty, dispersity, and specific matching rules to improve the distractor quality.

4.1 Multiple Choice Questions

The interaction with students is accomplished by Interlingua client application which in turn uses the repository's API to support the student with content and intelligent features during E-assessments generation.

This is mostly used assessment method can have single or multiple correct answers for both vocabulary and reading compression types. A student may want to use multiple choice questions when there's a clear and correct option, as well as some viable distractors. For example, a teacher or a computer may test a learner's ability to apply knowledge by presenting E-assessments tests. Multiple Choice quiz type of assessment consists of a series of items each containing a question (the stem), a correct answer (the key) and some incorrect answers (distractors).

4.2 True-False

In the case of testing knowledge that has a very clear and definitive answer, true-false questions are a perfect option. Material that works well for true-false-style E-assessments includes facts, language translations, and classification problems. When there's only one clear answer, students can respond quickly.

4.3 Drag and Drop

Drag-and-drop questions let the students sort or group of options. This question format prompts students to interact with objects on a screen because they need to select one option and drag it to another part of the screen.

4.4 Fill up the blanks

Use of fill-up-the-blank questions forces students to recall specific facts directly. When the answer is not listed anywhere on the screen, students must remember it without any prompting. When using fill-in-the-blank questions, misspelled words are marked as incorrect. This asserts that students should know not only the concept but also its exact spelling and synonyms.

5. Implementation

The implementation of Interlingua is made with JAVEE technology. Interlingua is based on client-server architecture. This system provides multiple services to students having different mother tongues. The client-server architecture is a system that provides distinct services to multiple clients by logically dividing the service providers. Several clients may run concurrent instances of any given service simultaneously. A significant advantage of client-server architecture is that it is a distributed architecture providing the possibility of integrating geologically separated components all into one system. Concepts with their translation with the detailed explanations (definitions) in the target language or the language that user intended to learn are to be fetched from a database. The skeleton structure of the whole system is depicted in (Figure 2).

Interlingua offers linked learning resources and also related to the external learning resources. Textbooks and PDF can be added anytime, the contents from the textbook will be linked with three different languages. This model is extracted in client model.

Content manager helps to access the repository via web-based API, complete learning content is generated by client model. The extraction should support formatting and to retrieve the structure properly. This content manager takes care of removing and adding new learning material and stores it in the content base.

The user can log into the client page with his login credentials. Once the user logs in is confirmed, then his source language, as well as study language, are set inside a cookie. After a successful login user is guided to the first page where the explore tab with reading pane presented by default.

Inside the default *explore.html* page left pane allows the user to select the chapter user wants to study, the same time the concepts introduced in that chapter are also fetched and saved inside the array *"conceptArrayAssessment"*. This array is containing all introduced concepts with their translations, and distractors. This array is fetched inside any assessment and then can be extracted to build the desired type of assessment. After the assessment tab allows the user to select different types of E-assessments among eight different types of E-assessments.

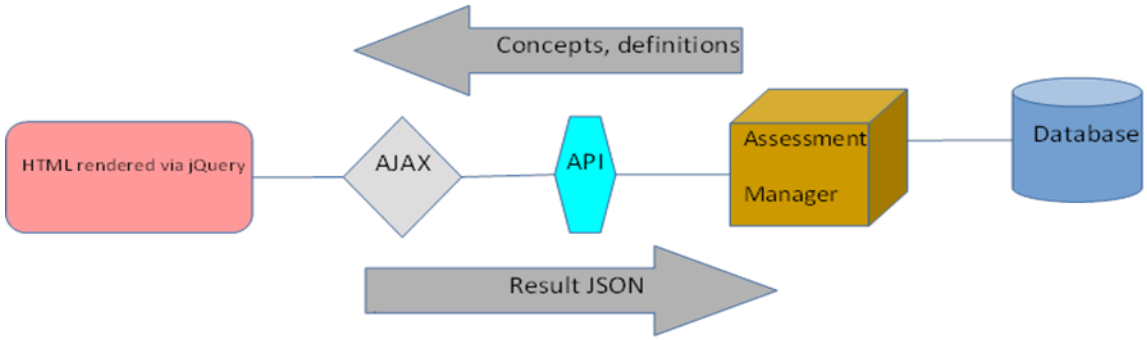

Figure 2: Data flow diagram between repository and the client.

When a student is browsing the document, student can request or issue a query for semantically equivalent concepts in their mother language. With the help tab, the related content is shown in the next separate tab and relevant concepts are shown in mother language. Also, help feature, highlight important concepts in the learning material. It helps students to interact.

Below in (

Figure 3) the Interlingua client system allows students to navigate or browse learning the material and to test their knowledge using the provided Self-Assessment test.

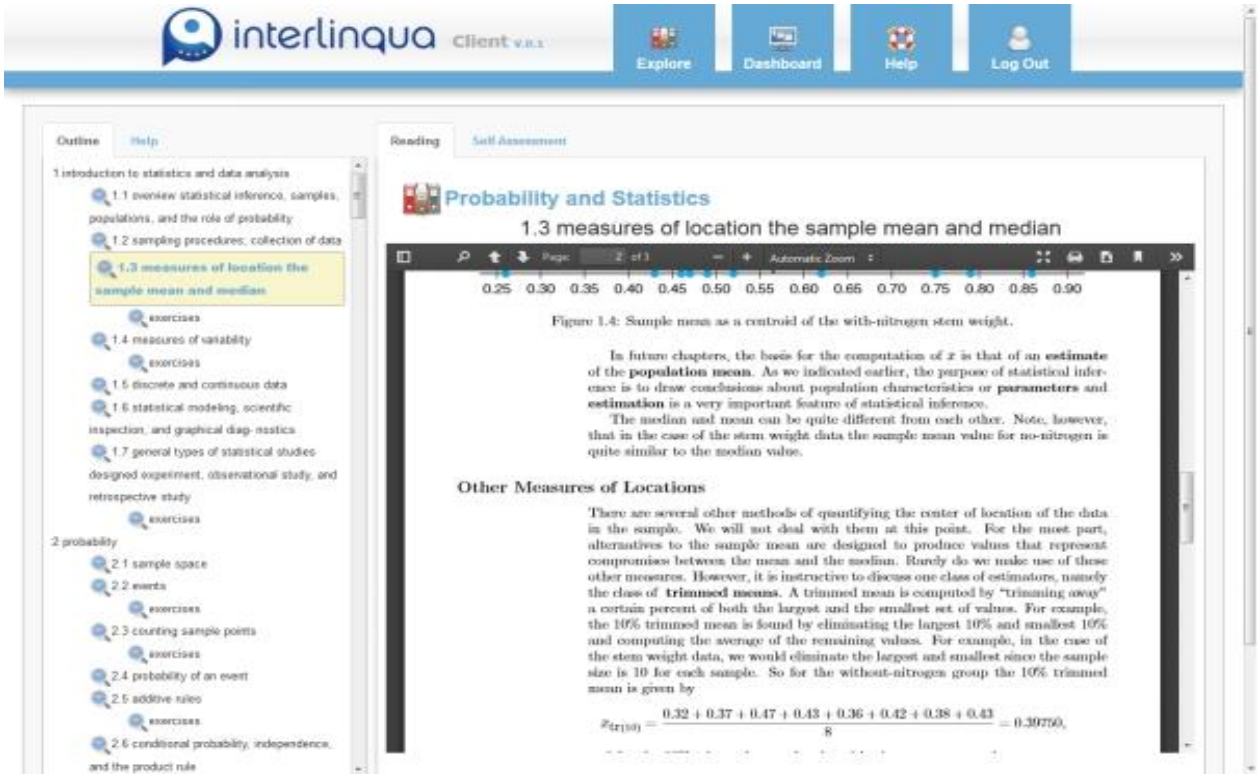

Figure 3: Start page - Client application

The reading pane displays the contents of a specified chapter. The chapter selector is at the left side of the explorer window. The user can select any chapter from the chapter outline viewer. Once the user clicks on the desired chapter, chapter id, book id and the segment ids are fetched then this information is passed via an Ajax query to get all concepts, translation of the concepts and distractors for each concept in a JSON array format. This JSON array is employed in creating different types of E-assessments. The test results for all kinds of E-assessments are bundled into a JSON array and then sent back to the repository which then will be stored in the database. The E-assessments start page lists all types of E-assessments. The user can select desired assessment navigating through a jQuery dropdown as in (Figure 4) below.

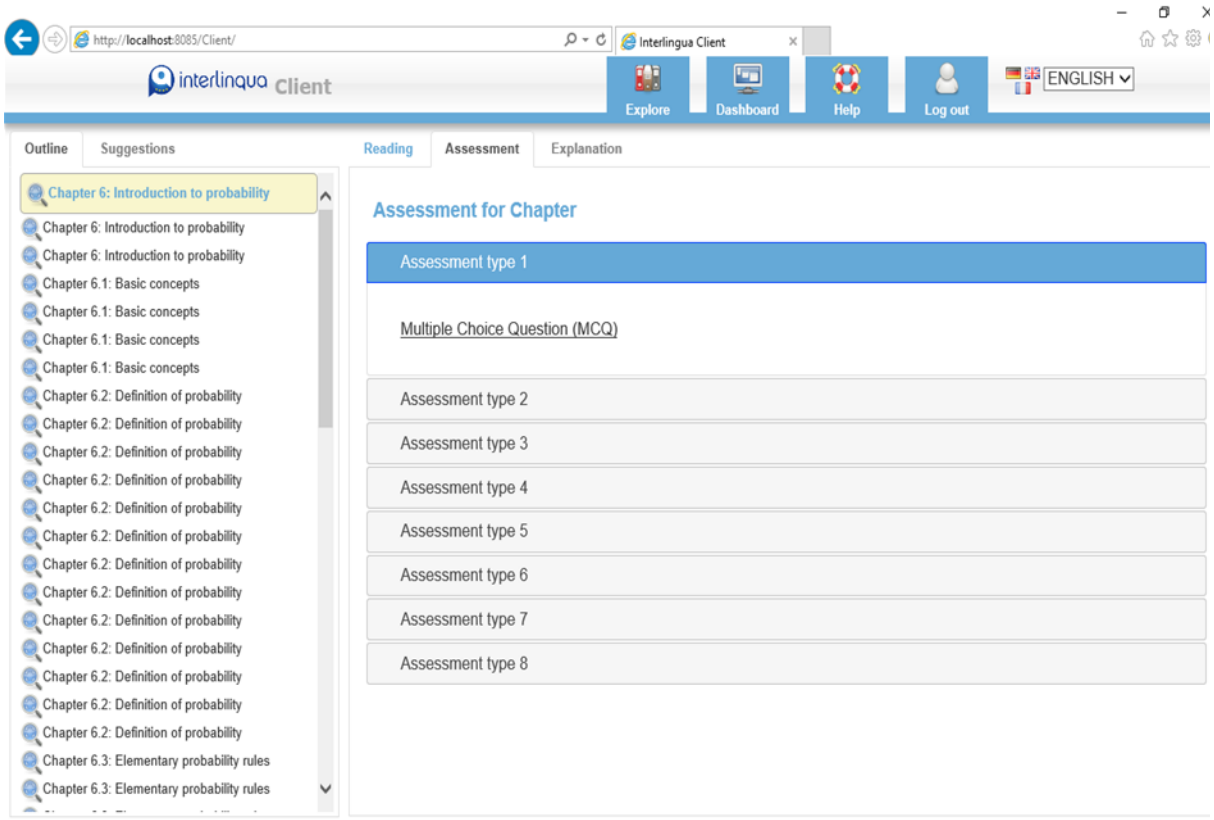

Figure 4: Assessment selector page

5.1 Overview of an Assessment generation

The typical workflow of assessment is as follows. In the first step, an AJAX sends all perfected concepts as data this query initiates a connection to the repository via an API called Repository API which from the database, brings translations and labels for the concepts bundled as a JSON array. This array contains distractors and translations for the concepts that are pre-fetched when the explorer.html loaded. The results of an AJAX request bring a JSON result as shown below:

The contents of this JSON array are unbundled and created the desired type of tests in JavaScript. Each test is assigned to a distinct test ID to keep track of all user activities. After the completion of a test, the user input is bundled into a JSON array containing all the questions, test ID, correct answers, user ID password, source language and the target language. An example of a result JSON is shown the AIG Semantic Library can generate item contents for translation and definition test items, and a REST interface was implemented that can receive requests from the client module and respond with item contents serialized as XML.

{"tes-
tId":"1464987569371","userName":"user","password":"PASSWORD","sourceLanguage":"GER
MAN","targetLanguage":"ENGLISH","questions":[{"questionID":"0","sourceLanguage":"GERMA
N","targe

tLan-
guage":"ENGLISH","userId":"ramya","type":"mcq","correctItems":[{"concept":"conditional","la
bel":" conditional"}],"selectedItems":[{"concept":"conditional","label":"

conditional","isCorrect":"true"}],"items":[{"concept":"conditional","label":" condition-
al","isCorrect":"true"},{"concept":"grid","label":"

grid","isCorrect":"false"},{"concept":"Morgenstern+distributions","label":" Morgenstern distri-
butions","isCorrect":"false"},{"concept":"working+probit","label":" working

probit","isCorrect":"false"},{"concept":"cumulative+process","label":" cumulative

process","isCorrect":"false"}]}]}]}

5.2 Assessment Category 1 - MCQ (Multiple choice questions)

A Multiple choice is a form of an objective assessment in which respondents are asked to select the only correct answer out of the choices from a list. The multiple-choice format is most frequently used in educational testing.

Multiple choice items consist of a stem or a statement of the question or in our case it is the question asking the user for a correct translation of a randomly selected concept of a chapter, the correct answer, keyed alternatives, and distractors. The options are the possible answers that the examiner can choose from, with one or more correct answer called the key and the incorrect answers called distractors. Only one answer can be keyed as correct. In our case, we set only one answer as the correct key which contrasts multiple response items in which more than one answer may be keyed as correct.

There are several advantages to multiple choice tests. It can be a very effective assessment technique If item generated are quality assured, if students are instructed on the way in which the item format works and myths surrounding the tests are corrected, they will perform better on the test. On many E-assessments, reliability has been shown to improve with larger numbers of items on a test, and with good sampling and care over case specificity, overall test reliability can be further increased.

Multiple choice tests often require less time to administer for a given amount of material than would tests requiring written responses. This results in a more comprehensive evaluation of the candidate's extent of knowledge. Even greater efficiency can be created using online examination delivery software. This increase in efficiency can offset the advantages offered by free-response items. That is if free-response items provide twice as much information but take four times as long to complete, multiple-choice items present a better measurement tool.

The result of the assessment is displayed immediately after clicking the submit test button. All the correct answers will be marked green and wrong answers are marked in red, and the correct answer next to the wrong selection will be displayed.

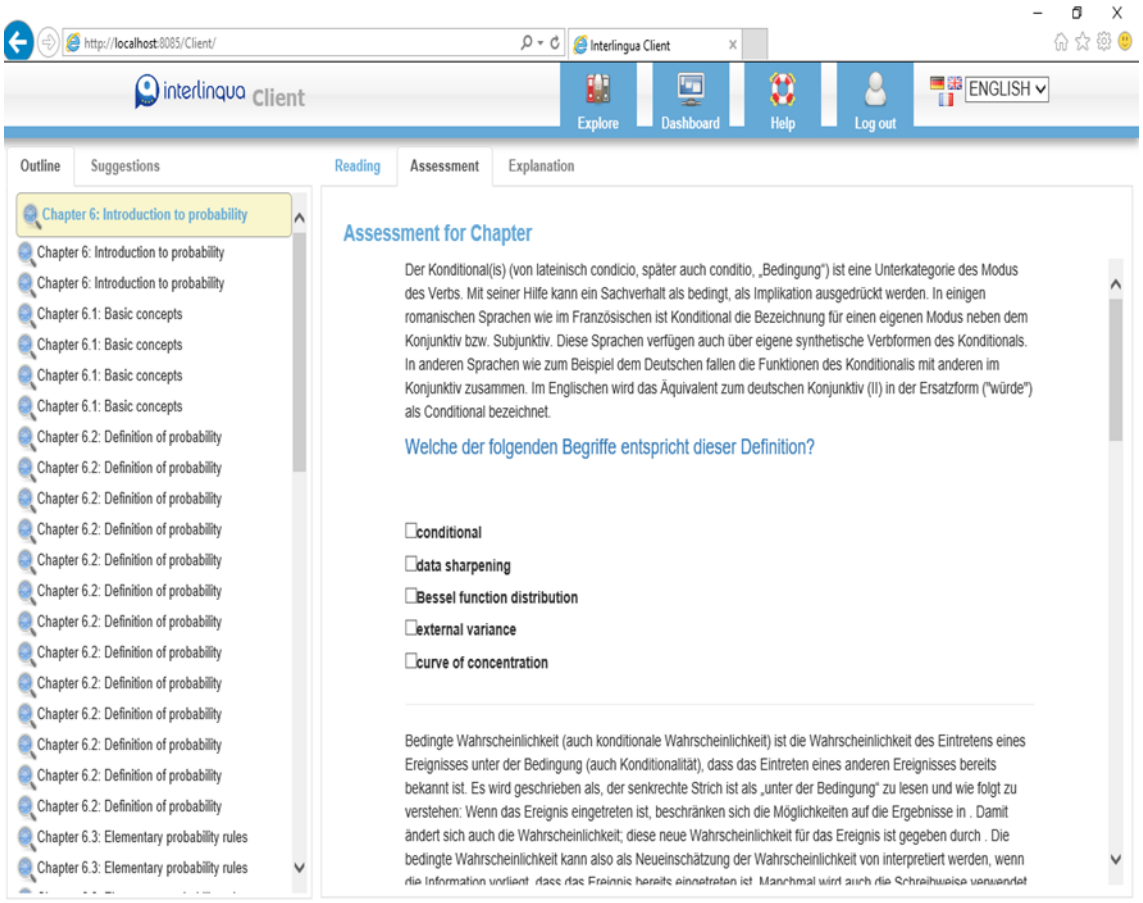

Figure 5: MCQ for description of concepts.

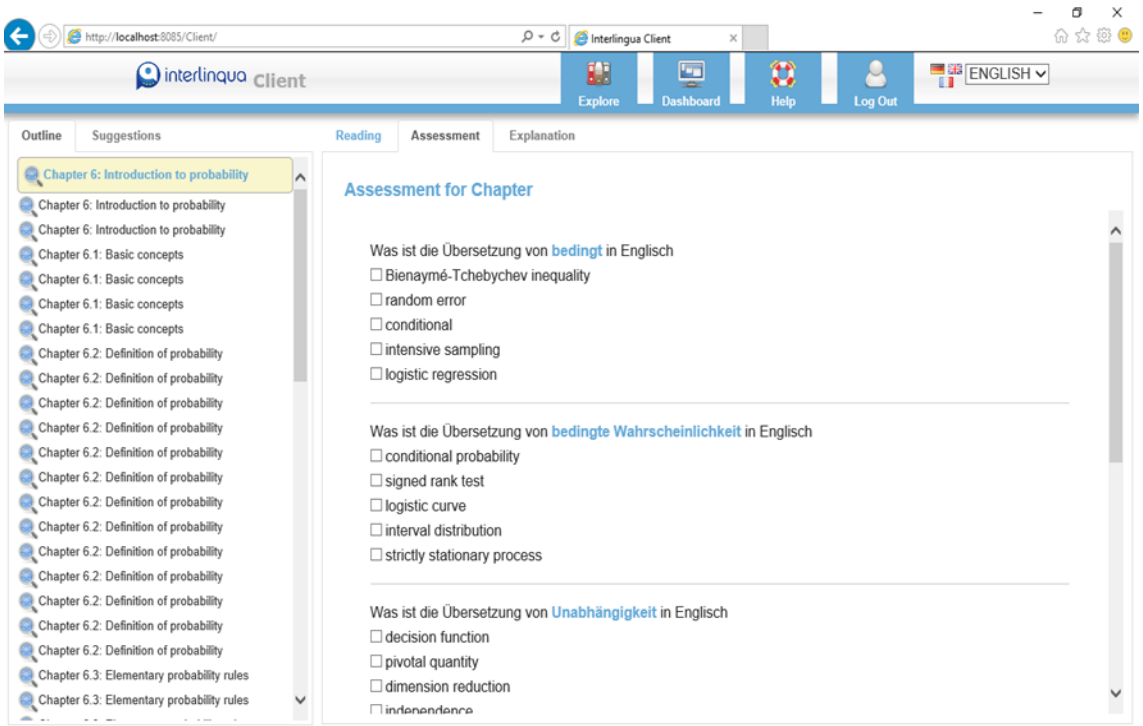

Figure 6: MCQ for translation of concepts.

Multiple choice type of assessment is implemented (Figure 5 and Figure 6) for both finding a correct translation of concepts for a given concept from a collection five presented choices as well as finding an appropriate concept for a given definition of a concept translation.

5.3 Assessment Category 2 - True or False

This type of assessment is also as called binary-choice items and is utilized to assess a student's ability to find out the accuracy of a declarative statement. True-false statements are very useful in measuring a student's ability to differentiate between forced-choices. As such, true-false questions are well suited for measuring knowledge, comprehensive, and application levels of understanding.

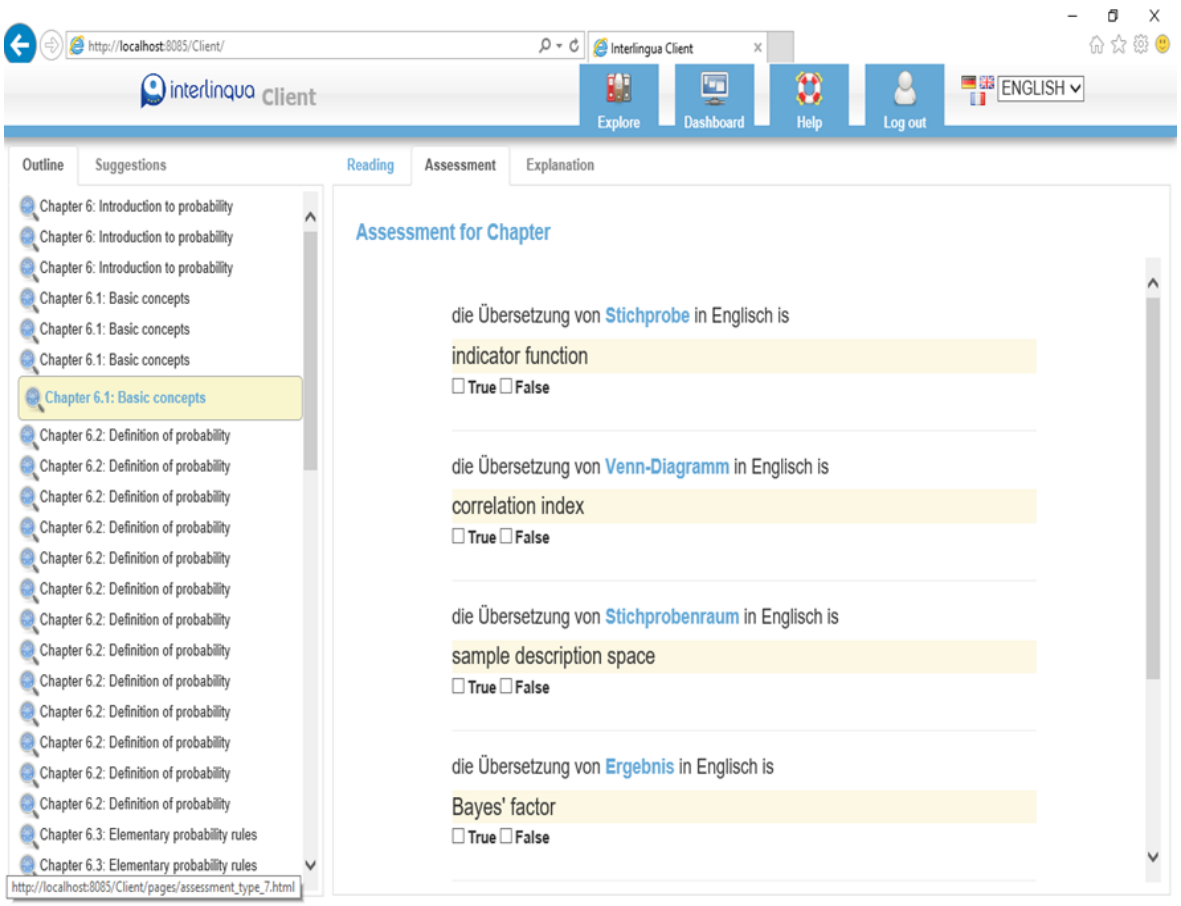

Figure 7: True or False assessment for translation.

A quality true-false item is classified as an objective assessment technique as it will only have one correct answer. In the context of Interlingua, true-false type of assessment is created for finding correct translation of a given concept with a presented translation. The presented translation is chosen randomly from a collection of distractors and correct

translation. The figure shows the implementation of a true-false type of assessment for translation. Once the user submits the test he will be presented with the test results immediately. The wrong inputs will be marked red, and the correct answers will be highlighted in green color.

The true-false assessment type is also developed for selecting right concept for a given definition or detailed explanations (Figure 7 and

Figure 8). The concepts for definitions are selected from a group of exact concept corresponding to the description and as well as some randomly selected distractors. (Figure 8) shows the implementation of a true-false type for finding a correct concept for the given explanation.

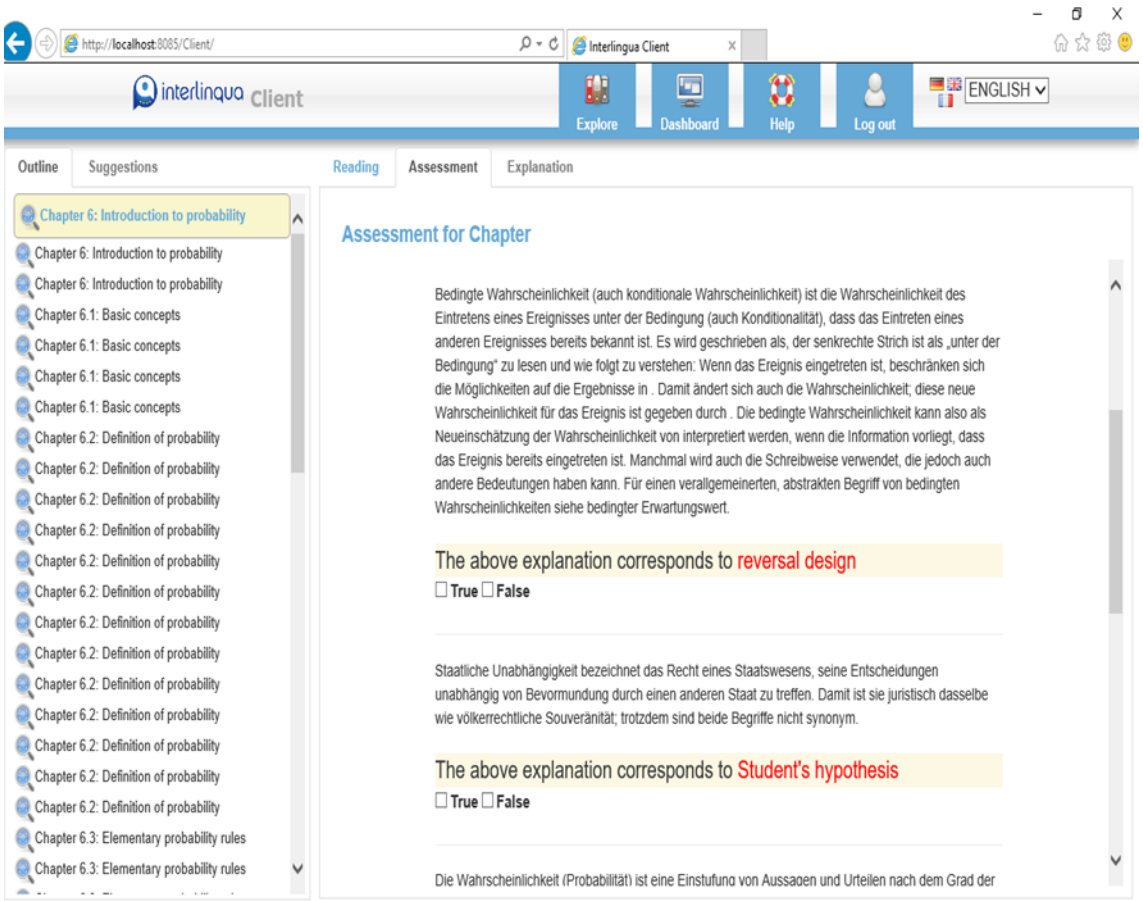

Figure 8: True or False assessment for description of concepts.

5.4 Assessment Category 3 - Fill up the blanks

Fill in the blanks is a common type of assessment and one of the very popular types too. This type of E-assessments is often found in the assessment of grammar and vocabulary. While they do require students to produce language by writing the correct spelling of the answer becomes crucial to this type of assessment, therefore, this is different from multiple-choice questions, they are rather inauthentic regarding language use. Some of the advantages to fill-in-the-blank include high reliability, easier to write, and limits guessing.

Likewise, the previous types of E-assessments here also fill in the blank type is implemented for both translating a concept and choosing an appropriate concept for the given definition. In contrast to the previous types, here there are no choices of the answer to choose. The user is expected to write the correct answer in a text field. This imposes one more constraint that the user must know the concept translation in his target language as well as its correct spelling.

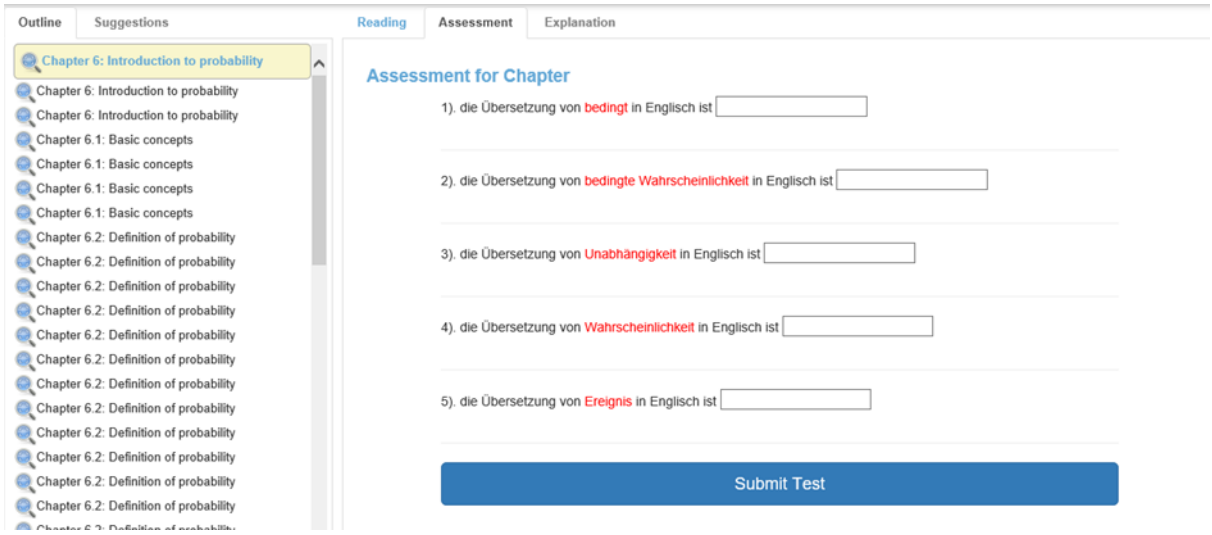

Figure 9: Fill up the blanks type assessment for translation of concepts.

The answer submitted by the user will be evaluated immediately, and the results are displayed once after the submit button is clicked. The background color of the fields where wrong answers are written are turned into red color, and the wrong answers are replaced with correct answers otherwise the fields are turned into green color. Like the previous assessment types, fill-in-the-blank type is also implemented for translating concepts and finding concepts for a given definition (Figure 9 and Figure 10) show the implementation of this kind of assessment.

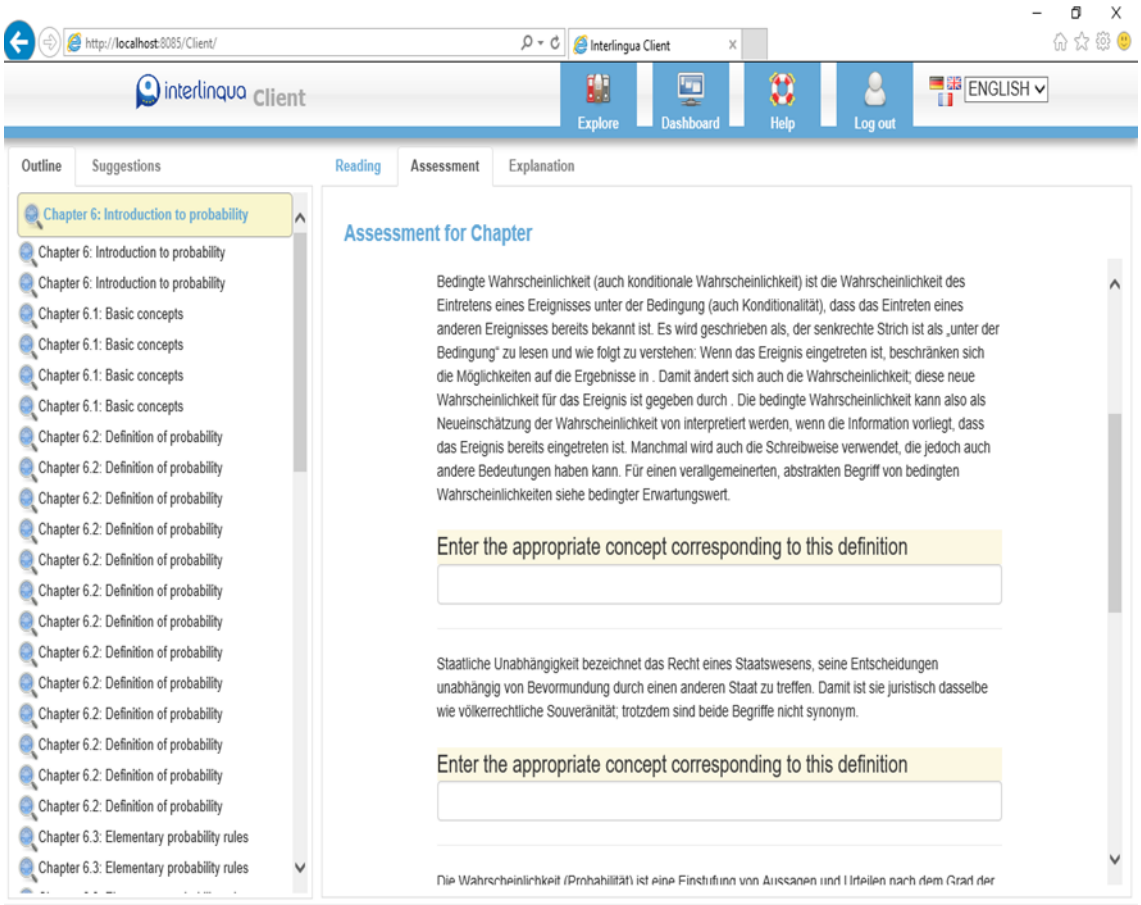

Figure 10: Fill up the blanks type assessment for description of concepts.

5.5 Assessment Category 4 - Content Matching by Drag and Drop

Matching questions are at their best in the case to assess the knowledge gained from a course that features a lot of concepts or connection between items, concepts, etc. In this type of assessment, the user is presented with five cards in two rows each. Cards on the first row display different concepts in the target language. These cards can be dragged and dropped on another five cards in the second row displaying the translation for the draggable cards. However, all these cards are shuffled each time the user loads the test. The screenshot of the implementation is shown in Figure 11. Here two kinds of implementations are done. In the first kind of application, the user cannot drop the card on to a wrong translation in other words a user is constrained to group the cards with the cards showing the exact translation of the draggable cards.

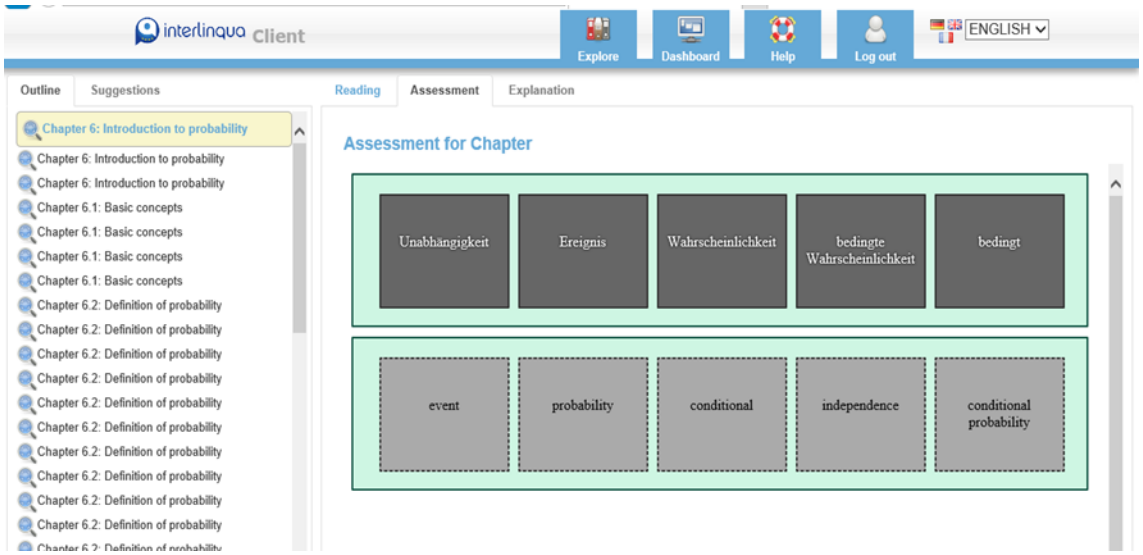

Figure 11: Content matching by drag and drop (type1)

In the second variation of the content matching, the user can drop card onto any other card on the bottom row and submit his answer. Upon submitting the answer, the correct choices turn too green, and those wrong choices are highlighted by changing their color to red.

Figure 12 shows the implementation of the second verity of content matching assessment before the user made any drag and dropped event and

Figure 13 shows the results displayed after the user submits his/her answer.

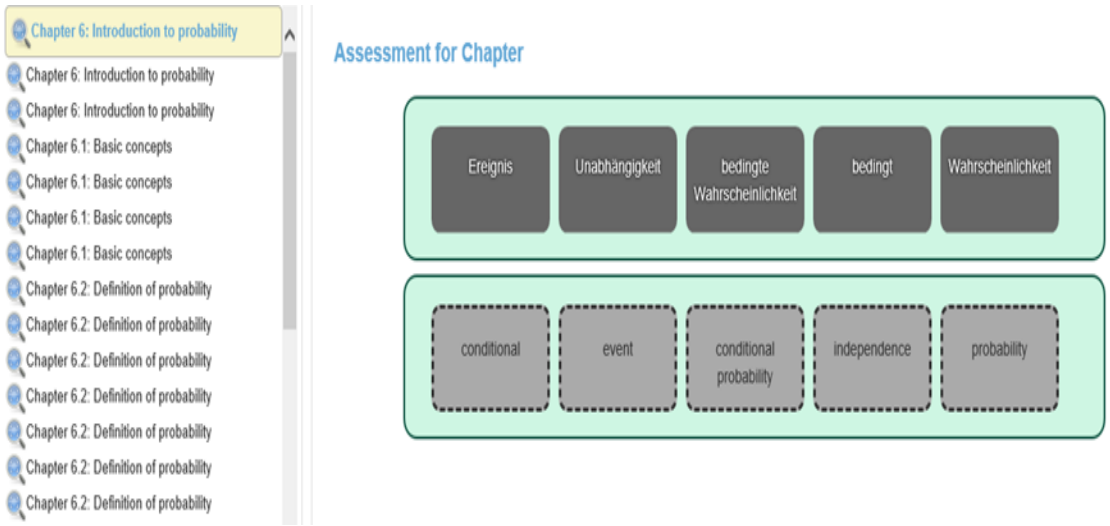

Figure 12: Content matching by drag and drop (type2)

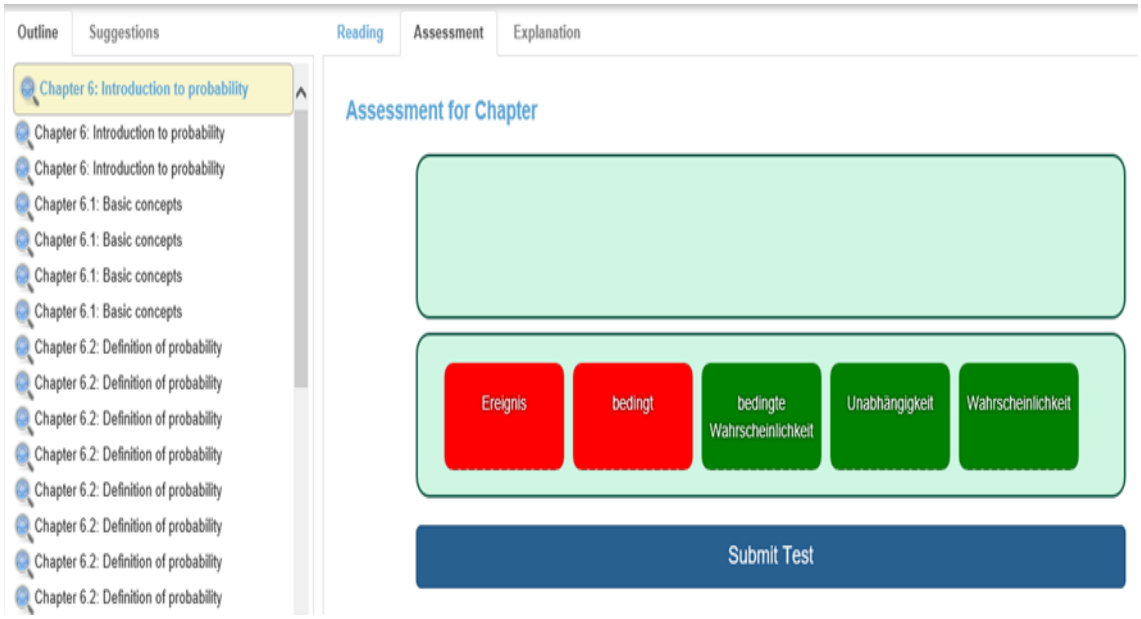

Figure 13: Content matching by drag and drop result displayed.

6. Conclusion and Future work

The Interlingua system is aimed to benefit students who are learning technical subjects in foreign languages. To ensure a steadily increasing learning curve, students must be presented with different types of E-assessments. This thesis describes the Interlingua system in brief and successful implementation of eight different types of E-assessments that are implemented.

E-assessments types like MCQ, fill up the blanks, binary selection, content and matching are implemented to identify appropriate translations and to identify concepts corresponding to descriptions.

A JavaScript framework for eight different types of interactive E-assessments creation is implemented. In the future, this assessment can be improved or more functionalities can be added to make them highly interactive which makes learning fun. Gamification of E-assessments could be a next work for improving the user experience. One of the biggest benefits of Gamification is better learning experience for a learner. The learner can experience “fun” during plying games while maintaining high level of engagement. A good gamification aims at high levels of engagement that leads to an increase in recall and retention. Domínguez et al.[28] have showed that students who completed the gamified experience got better scores in practical assignments and in overall score.

There is room for further improvements. An adaptive hardness of E-assessments can be deployed in future. Results of the evaluation can be analyzed further to increase or decrease the level of difficulty based on the inputs of a user. Students could be presented with a graphical feedback based on the user specific statistics in the form of pie charts bar plots, at the login screen for each user.

Extension of this work finds itself in several prospective directions. One of the immediate possible improvements would be generating equations as cloze items. Another major long term goal would be implementing an intelligent tutoring system into the E-assessment system, that aims to provide immediate and customized feedback to students. An intelligent tutoring system (ITS) is a system based on artificial intelligence. The key roles of an ITS system are teaching, evaluating the student performance with measuring learner’s strengths and weaknesses and also attaching some kind of metric to evaluate descriptive type answers inputs in case when a student presented with a concept and asked to input corresponding definition for it, then It should be possible him or her to

input the explanation in his own words. The system must be capable of detecting the text and compare with the definition in the ontology.

List of figures

Bibliography

[1] A. Papasalouros, K. Kanaris, and K. Kotis, "Automatic Generation Of Multiple Choice Questions From Domain Ontologies.," in *e-Learning*, 2008, pp. 427–434.

[2] T. Alsubait, B. Parsia, and U. Sattler, "Automatic generation of analogy questions for student assessment: An ontology-based approach," in *Research in Learning Technology*, 2012, vol. 20, no. SUPPL, pp. 95–101.

[3] A. Pfeiffer, A. Kuraeva, M. Foulonneau, Y. Djaghloul, E. Tobias, and E. Ras, "Automatically Generated Metrics for Multilingual Inline Choice Questions on Reading Comprehension," in *International Computer Assisted Assessment Conference*, 2015, pp. 80–95.

[4] R. Mitkov, L. A. Ha, A. Varga, and L. Rello, "Semantic similarity of distractors in multiple-choice tests: extrinsic evaluation," in *Proceedings of the Workshop on Geometrical Models of Natural Language Semantics*, 2009, pp. 49–56.

[5] G. A. Miller, "WordNet: a lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[6] Y.-C. Lin, L.-C. Sung, and M. C. Chen, "An automatic multiple-choice question generation scheme for English adjective understanding," in *Workshop on Modeling, Management and Generation of Problems/Questions in eLearning, the 15th International Conference on Computers in Education (ICCE 2007)*, 2007, pp. 137–142.

[7] J. C. Brown, G. A. Frishkoff, and M. Eskenazi, "Automatic question generation for vocabulary assessment," in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 2005, pp. 819–826.

[8] E. Holohan, M. Melia, D. McMullen, and C. Pahl, "Adaptive e-learning content generation based on semantic web technology," 2005.

[9] E. Holohan, M. Melia, D. McMullen, and C. Pahl, "The generation of e-learning exercise problems from subject ontologies," in *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, 2006, pp. 967–969.

[10] R. Shah, "Automatic question generation using discourse cues and distractor selection for cloze questions," International Institute of Information Technology Hyderabad, 2012.

[11] L. Becker, S. Basu, and L. Vanderwende, "Mind the gap: learning to choose gaps for question generation," *... 2012 Conf. North ...*, pp. 742–751, 2012.

[12] M. Heilman, "Automatic Factual Question Generation from Text," *dissertation*, no. June, p. 203, 2011.

[13] T. Goto, T. Kojiri, T. Watanabe, T. Iwata, and T. Yamada, "Automatic generation system of multiple-choice cloze questions and its evaluation," *Knowl. Manag. E-Learning An Int. J.*, vol. 2, no. 3, pp. 210–224, 2010.

[14] J. Mostow *et al.*, "Using Automated Questions to Assess Reading Comprehension, Vocabulary, and Effects of Tutorial Interventions," *Technol. Instr. Cogn. Learn.*, vol. 2,

pp. 103–140, 2004.

[15] F. Linnebank, J. Liem, and B. Bredeweg, “Question generation and answering,” *Science (80-.).*, pp. 1–40, 2010.

[16] M. Foulonneau, “Generating educational assessment items from linked open data: The case of DBpedia,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 7117 LNCS, pp. 16–27.

[17] M. Foulonneau and V. Grouès, “Common vs. Expert knowledge: making the Semantic Web an educational model,” *Linked Learn.*, 2012.

[18] R. Mitkov, L. A. Ha, and N. Karamanis, “A computer-aided environment for generating multiple-choice test items,” *Nat. Lang. Eng.*, vol. 12, no. 2, pp. 177–194, 2006.

[19] S. Smith, P. V. S. Avinesh, and A. Kilgarriff, “Gap-fill Tests for Language Learners : Corpus-Driven Item Generation,” *English*, 2010.

[20] M. J. Gierl and H. Lai, “Using weak and strong theory to create item models for automatic item generation,” *Autom. Item Gener. Theor. Pr.*, vol. 26, 2012.

[21] P. Sonnleitner, “Using the LLTM to evaluate an item-generating system for reading comprehension,” *Psychol. Sci. Q.*, vol. 50, no. 3, pp. 345–362, 2008.

[22] P. Kolb, “Disco: A multilingual database of distributionally similar words,” *Proc. KONVENS-2008, Berlin*, 2008.

[23] T. M. Haladyna, “Automatic item generation: A historical perspective,” *Autom. Item Gener. Routledge, New York*, 2013.

[24] T. François and E. Miltsakaki, “Do NLP and Machine Learning Improve Traditional Readability Formulas?,” in *Proceedings of the First Workshop on Predicting and Improving Text Readability for Target Reader Populations*, 2012, no. Pitr, pp. 49–57.

[25] G. Kobilarov, C. Bizer, S. Auer, and J. Lehmann, “DBpedia-A Linked Data Hub and Data Source for Web and Enterprise Applications,” *Program. Chairs*, p. 4, 2009.

[26] P. N. Mendes, M. Jakob, and C. Bizer, “DBpedia: A Multilingual Cross-domain Knowledge Base.,” in *LREC*, 2012, pp. 1813–1817.

[27] M. Al-Yahya, “Ontology-based multiple choice question generation,” *Sci. World J.*, vol. 2014, 2014.

[28] L. de-Marcos *et al.*, “Social network analysis of a gamified e-learning course: Small-world phenomenon and network metrics as predictors of academic performance,” *Comput. Human Behav.*, vol. 60, pp. 312–321, 2016.