Enhancing Object-Oriented UML for Developing an Intelligent Semantic Knowledge Dictionary

Mansoor Al Aali, Samia Yousif
College of Information Technology
Ahlia University
Manama, Kingdom of Bahrain
Emails: {malaali, syousif}@ahlia.edu.bh

Abstract—This research presents a new SEMANTIC dictionary system developed by utilizing the Object-Oriented (OO) approach. SEMANTIC dictionary system contains English words, their Arabic meanings, associated actions, semantic relationships, inherited actions and attributes, exceptional relationships and other semantics and characteristics. SEMANTIC dictionary system utilizes Object-Oriented major features, such as objects, classes, aggregation, inheritances, encapsulation and polymorphism. SEMANTIC dictionary is a knowledge base, which can be considered as an intelligent language model and can be used for many language teaching purposes. This research shows how simple phrases can be generated or validated to be semantically correct. In the process of using OO UML to represent semantic knowledge, we have made enhancements and additions to UML itself.

Keywords—Semantic; Dictionary; English; Object Orientation; Teaching

I. INTRODUCTION

Research into natural language processing has received great attention and will continue to do so because of the importance of such topic to humanity [1][18]. Natural language processing presents two main challenges: representing semantics and querying these semantics. Until now, no one claims to have the solution to store semantics of complex sentences and link all its words together with the meaning of words, phrases of other sentences [2][12]. This research presents a novel approach by storing the semantics of English simple phrases using the Object Oriented techniques for teaching and learning English. Object oriented technique is a method of viewing knowledge at different levels of abstraction [19].

English is the first or second language in most countries educational systems. Students in many countries study the English language starting from the 1st grade of elementary schools [9][13][14]. Mastering English in both writing and reading is essential and crucial to the success of students in high schools and colleges for them to be able to find a good job in the future [11]. Furthermore, the students need to feel more of the knowledge relating to the words to improve their writing skills in English. Students studying English as Second Language (ESL) need to use dictionaries in order to know what words mean and what they can do with them in order to understand and communicate in English [6].

Dictionaries are one of many important tools [10] that can increase the students’ understanding of the language and develop their vocabulary. Furthermore, Learners’ Dictionaries are used to practice English correctly, effectively and appropriately [17]. Consequently, dictionaries must have maximum semantic information for words related to semantic classification that will assist the students to understand English language, as well as improve and enhance their writing in English.

Several developers have attempted to develop inclusive and unique dictionaries with special features that may help the dictionary to become more useful and popular. We have conducted a survey on online dictionaries and elicited that almost all dictionaries come under three categories: learner, translator and search. Each dictionary has a number of features, for example, some dictionaries provide words definition, examples, pronunciation, words translation, text translation, pictures, rhymes, synonyms and words functions like noun, adjective, etc. Many companies have cashed in on the concept and have developed dictionaries for a variety of purposes such as Cambridge Learner’s Dictionary, Word Central, Ultralingua Language Software, RhymeZone, Wordsmyth Children's Dictionary, McGraw-Hill Children's Dictionary, OneLook Dictionary Search, Freesearch, Sakhr, Tarjin, WordNet and ALWafi.

Writing fluently in English language requires knowledge of the conventional contexts, strong vocabulary, and collocations surrounding the word. While this information may be presented implicitly in dictionaries to help ESL learners, ESL students often have difficulties finding the right meaning for unfamiliar words or phrases in their dictionaries. In response to these difficulties, some researchers have proposed new ways to access the English lexicon [8] and presenting typical phraseology rather than words in isolation [6]. In addition, some researchers have suggested a representation, with a tree-based model of runtime dictionaries, allowing a very fast morphological analysis with semantic recognition, for example, the path <SPORTS/Water Sports/Swimming> describes the document field <Swimming> as a sub-field of <Water Sports>, which is a sub-field of super-field <SPORTS> [3].

This research presents some aspects of semantics for simple sentences and phrases, for example, the simple phrase (boy eats book) is grammatically correct while it is wrong in real world meaning. In turn, SEMANTIC dictionary system
rejects this sentence and displays the reasons for the rejection and provides a dynamic correct example such as (boy eats apple). This is made possible because of the intelligent features of the Object-Oriented approach and the actions associated with each word (noun). Some researchers proposed a practical retrieval system for semantic representation by building relationships between verbs and nouns in derivation frames [3]. In SEMANTIC dictionary, administrators create relationships between actions (verbs) and classes (nouns) to provide important information about the meanings of nouns. For instance, the action ‘eat’ could be the relationship between the super classes ‘human’ and ‘food’ which means ‘human eat food’.

Teaching English needs to be changed or improved by adding various tools and new strategies that will assist students to comprehend English in the best way [5]. Many researchers continue to study a variety of new approaches in teaching English [1][2][7][11][12][13][14][15]. Communicative language teaching approach advocates the development of communicative competence as a primary goal through the extensive use of the second language as a means of communication during classroom lessons. Using computers raises students’ motivation and their performance as well as placing students in a learning state with full participation, concentration and enjoyment [4][5].

SEMANTIC dictionary system is a web-based dictionary system. Each word in this dictionary belongs to a class and may have one or more subclasses. Subclasses inherit all the public attributes and operations of their super class and this concept is utilized in all types of processing on the SEMANTIC dictionary system. For example, the super class 'human' inherits the public operation 'eat' to its subclass 'boy' and therefore we can generate or check the simple sentence 'boy eats food' or 'boys eat food' with semantic meaning.

The work in this research signifies that SEMANTIC dictionary system assists teachers to make the English subject more interesting and raise student's motivation and contribution, as well as it enables students to understand the meaning of words and helps them to organize sentences and build them correctly with semantic meaning.

The rest of the paper consists of a number of sections. Section II presents the enhancements to UML for SEMANTIC Dictionary System and the enhancement of UML diagram for SEMANTIC dictionary system. Section III presents a summary discussion for the subject-only concept and supporting scenarios. Section IV presents the conclusions and future work.

II. APPLYING THE OBJECT-ORIENTED TECHNIQUES TO INCORPORATESEMANTICS IN THE SEMANTIC DICTIONARY

Object-Oriented (OO) is a paradigm for creating software systems using objects. Objects are tangible and conceptual things we find in the real world. Using OO makes SEMANTIC dictionary system objects more semantically related and hence better intelligence can be implemented. OO has major notions such as objects, classes, inheritances, encapsulation and polymorphism. The OO Unified Modeling Language (UML) is the industry-standard language for specifying, visualizing, constructing, and documenting the artifacts of software systems. It simplifies the complex process of software design, making a "blueprint" for construction. These OO concepts were used as a model for representing SEMANTIC knowledge base rather than SEMANTIC development. The main reason for using UML is to model the simple semantics of the language phrases which can then be modelled internally in our software system depicting the SEMANTIC DICTIONARY system. Modelling using UML means the ability to directly represent the knowledge using an Object oriented language. The reasons for expanding or extending the UML is that the original UML does not have certain semantic concepts suitable for our purpose as will be illustrated in the next section.
A. Enhancing UML for SEMANTIC Dictionary System

The public attributes and operations are expressed with a plus sign (+). The private attributes and operations are expressed with a minus sign (-). The exceptional operations are expressed with a less than sign (<). The “Subject-Only” operations are expressed with (&). The letter “b” is added to these signs (+, -, < or &) when the type of the operation is “Action By” while the letter “o” is added to them when the type of the operation is “Action On”. Notice that both “Exceptional” and “Subject-Only” operations are not a part of UML class diagram, but we added them, which means that the signs are according to our view as well as the addition of the letters “b” and “o” to the signs. The signs and the letters are utilized to distinguish between the actions.

B. Enhancement of UML Diagram for SEMANTIC Dictionary System

Figure 1 demonstrates two super classes with their subclasses, instances and relationships. The dotted lines mean that the diagram can be extended depending on the system. The first super class is called Super class A and it has Subclass Aa. This subclass contains three instances: Instance Aa1, Instance Aa2 and Instance Aa3. The second super class is called Super class B and it has Subclass Ba. This subclass consists of three instances: Instance Ba1, Instance Ba2 and Instance Ba3. The super classes inherit the public operations (Actions) to its subclasses. Accordingly, these subclasses inherit its public Actions to its subclasses or instances. Super classes are illustrated with rectangles divided into two compartments. Place the name of the class in the first partition (centered, bolded, and capitalized each word), list the subclasses or instances in the second partition. The following are the relationships between classes and the drawings of these relationships, as shown in Figure 2.

Normal Association (Public Operations): It is a connection between classes, a semantic connection between objects (instances) of the classes involved in the association. It is normally bidirectional, which means that if an object is associated with another object, both objects are aware of each other. First direction is Action By (+b Action), and second direction is Action On (+o Action). The multiplicity (0 .. *) and (1 .. *) are used to present how many objects are linked. This association is inherited to the subclasses. It is sketched by a wide straight line connected between two super classes. The Action By (+b Action) is put above this line while the Action On (+o Action) is placed below it. Moreover, the multiplicity and directions’ arrows are presented on this line.

Recursive Association (Public Operations): It is a connection between a class and itself. The association still presents a semantic connection between objects, but the connected objects are of the same class. This is also bidirectional and it has multiplicity like normal association. This association is inherited to the subclasses. It is expressed by a wide elbow connector between the super class and itself. The Action By (+b Action) is placed above this connector whereas the Action On (+o Action) is put below it. Furthermore, the multiplicity and directions’ arrows are presented on this connector.

Private Association (Private Operations): It is a connection between objects (instances). It is bidirectional; First direction is Action By (-b Action), and second direction is Action On (-o Action). This association can’t be inherited. It is sketched by a thin elbow connector between two instances. The Action By (-b Action) is presented above this connector while the Action On (-o Action) is placed below it. In addition, directions’ arrows are demonstrated on this connector.

Exceptional Association (Exceptional Operations): It is a connection between some objects and classes. Exceptional Action By (-b Action) means that the inherited action that the object can’t do and Exceptional Action On (-o Action) means that the inherited action can’t be carried out on the object. Consequently these actions are excluded from this object. It is demonstrated by a thin straight line between the instance of a super class and another super class, the short vertical line is put before the end up of this straight line. The Action By (-b Action) is placed above the straight line whereas the Action On (-o Action) is put below the straight line.

Subject-Only Operations: It means that the action is done by the subject (class) without an object. It is always Action By (&b Action) as well as it is inherited to the subclasses. It is sketched by a wide straight line connected to the super class and end up by oval arrow. The Action By (&b Action) is placed above this line.

In SEMANTIC dictionary system, the relationships or associations (public operation Action By (+b Action)) are created between two super classes manually; subsequently the operations Action On (+o Action) are generated between them automatically. The normal line is sketched between two super classes (Super Class A and Super Class B) indicates the public relationship (Action By) which is created manually between them while the dotted lines are sketched between them and its classes signify the public relationship (Action On) which is generated automatically as well as it shows the inherence public operations Action By (+b Action) and Action On (+o Action) as illustrated in Figure 2.

For example, Super class A inherits the public operations (Action By and Action On) to its Subclass Aa and its instances (Instance Aa1, Instance Aa2 and Instance Aa3). Notice that the sketched lines of relationships between classes in Figure 2 are hidden in SEMANTIC dictionary system. The other relationships that were applied in SEMANTIC dictionary system can be exhibited in the same manner.

Mainly, English simple sentence consists of three major parts are The Subject, The Verb and The Object. Now, will describe the Public Associations (Public Operations) and will explain how these relationships are achieved in the SEMANTIC Dictionary System as well as will present the intelligent aspects of it. Furthermore, it will provide more details about enhanced UML for drawing this type of relationships. The administrator or the teacher will create the public relationships (associations or operations) between the super classes utilizing simple present verbs manually. SEMANTIC Dictionary System has some intelligent aspects, for example, super classes will inherit these public
relationships within present verbs to its subclasses and it will produce other cases of public relationships via different verbs, such as: past verbs, future verbs and present continuous verbs automatically. The super classes will inherit these produced cases of public relationships to its subclasses as well.

Figure 1. Enhancement UML Diagram (Meta Diagram)

Figure 2. Classes Relationships (Meta Diagram)

Figure 3. English Simple Sentence Meta Diagram (Active Simple Sentence within Present Verb)
Consequently, SEMANTIC Dictionary System will generate several English simple sentences automatically, for example: active simple sentences and passive simple sentences within present verbs, active simple sentences and passive simple sentences within past verbs, active simple sentences and passive simple sentences within future verbs, active simple sentences and passive simple sentences within present perfect verbs and active simple sentences within present continues verbs. In addition, SEMANTIC dictionary system will generate these sentences using singular subject, plural subject, singular object and plural object. The following demonstrates the public relationships utilizing present verbs to generate active simple present sentence.

Figure 3 illustrates the Meta diagram of active simple sentence within a present verb. Depending on Figure 3, we have drawn the enhanced UML Meta diagram; see Figure 4. One of "Class A" or more can carry out the Action By (+b Action) on one of "Class B" or more. On other hand, one of "Class A" or more can perform the Action By (+b Action). Here, the Action By is in the present. Similarly, "Class A" and "Class B" have many attributes. Figure 5 is an example in the SEMANTIC Dictionary System; a Boy is a human and his gender is masculine, an Apple is a vegetarian food and its gender is neuter. The public relationship between the Boy and the Apple is the Action By, which is the present verb (eat), that they have already inherited from their super classes (Human and Vegetarian Food respectively). This is the intelligent aspect of the SEMANTIC dictionary system. Hence, SEMANTIC will generate numerous sentences automatically, as presented in the following: A boy eats, Boys eat, A boy eats an apple, A boy eats apples, Boys eat an apple, Boys eat apples.

Passive simple present sentence means that "The Subject" does not perform the action of the present verb in the sentence - the action is performed on it. Figure 6 illustrates the Meta diagram of passive simple sentence within a present verb. Regularly, "The Subject" and "The Object" are nouns and they have adjectives, whereas when we write a simple sentence, we can write "The Subject" and "The Object" without any adjective. Moreover, we can write them in singular and plural. Depending on Figure 6, we have drawn the enhanced UML Meta diagram, see Figure 7. Action On (+o Action) is the action that can be carried out on one of "Class B" or more by one of "Class A" or more. Similarly, "Class A" and "Class B" have many attributes. Figure 8 is an example in the SEMANTIC Dictionary System; a Boy is a human and his gender is masculine, an Apple is a vegetarian food and its gender is neuter. The super classes (Human and Vegetarian Food) inherit the public relationship
(present verb: eat) to their subclasses (the Boy and the Apple respectively). Since SEMANTIC dictionary system has intelligent aspects, it will produce the Action On between the subclasses (the Boy and the Apple), which is the action (eaten by). Furthermore, SEMANTIC will generate various sentences automatically, as presented in the following: An apple is eaten, Apples are eaten, An apple is eaten by a boy, An apple is eaten by boys, Apples are eaten by a boy, Apples are eaten by boys.

III. SUBJECT-ONLY EXPLANATION AND SCENARIOS

The “Subject-Only” operations are expressed with an ampersand (&). The letter “b” is added to these signs (+, -, < or &) when the type of the operation is “Action By” while the letter “o” is added to them when the type of the operation is “Action On”. Notice that both “Exceptional” and “Subject-Only” operations are not a part of the original UML class diagram, but we added them, which means that their signs are according to our view, as well as the addition of the letters “b” and “o” to the signs. The signs and the letters are utilized to distinguish between the actions. In addition, Figure 9 presents the following:

- **Classes**: SEMANTIC dictionary system has several super classes and many subclasses whereas Figure 9 presents three Super classes: Human, Vegetarian Food and Information Media. In addition, it shows six subclasses, such as (Reading) of super class Information Media and (Family and People) of super class Human.

- **Instances**: In SEMANTIC dictionary system, each subclass has many instances, for example (Man, Woman, Boy and Girl) of subclass People. Another example: the subclass Fruit has many instances that are applied in SEMANTIC dictionary system while Figure 9 shows (Apple and Banana) as an example.

- **Aggregation**: A lot of aggregations are applied in SEMANTIC dictionary system, for example the super class Information Media consists of class Reading and the class Reading consists of objects or instances (Book, Magazine and Newspaper).

- **Public Operations**: SEMANTIC dictionary system has many public operations (Action By): for example; buys, sells and browses, for example: a girl buys an apple. As well as the public operations (Action On) which are generated automatically: such as; bought by, sold by and browsed by, for example: an apple is bought by a girl.

- **Private Operations**: SEMANTIC dictionary system has a private operation (Action By): “engages” for object Man. For example: a man engages a woman and a private operation (Action On): “engaged to” for object Woman. For example: a woman engaged to a man.

- **Subject Only Operations**: SEMANTIC dictionary system contains several Subject Only operations (Action By): such as Sits, Grows, Sleeps, Swims and Weeps. For example, a boy sits.

- **Exceptional operations**: SEMANTIC dictionary system has an exceptional operation (Action By): Cooks. For example, a boy cannot cook a cloth. Also, an exceptional operation (Action On): Cooked by. For example, a football is not cooked by a human.

- **Inheritance**: In SEMANTIC dictionary system the super classes inherit the public operations (Actions) to its subclasses. For example, the super class Human inherits its public operations such as (eat, cook, read and write) to its subclasses (People and Family) and consequently these classes inherits these operations to its subclasses or instances like (Man, Woman, and Boy).

- **Multiplicity**: Different multiplicities are applied in SEMANTIC dictionary system. For example: “a man eats an apple”, “three men eat an apple”, “a man eats two apples” and “four men eat five apples”.

- **Normal Association**: SEMANTIC dictionary system has many normal associations, for example, first direction: a boy reads many (zero or many) books. Second direction: a book can be read by many (one or many) boys.

- **Recursive Association**: SEMANTIC dictionary system has a recursive association, first direction: (one or many) woman feeds (one or many) sons. Second direction: (one or many) son fed by (one or many) women.

IV. CONCLUSION AND FUTURE WORK

This paper presented a new approach for the development and evaluation of an intelligent English-Arabic Object Oriented dictionary system called SEMANTIC. This research is the first to effectively use the knowledge concept and enhance UML to deal with English language semantics for simple phrases. SEMANTIC Dictionary System was tested on the 6th class of an elementary school for girls in kingdom of Bahrain to test its effectiveness in teaching some aspects of the English language and to test its level of contribution to students learning and the tests are very promising [23].
We are planning to compare our enhanced UML with OWL for the representation of natural language semantics especially for the application of translation from English to Arabic.

ACKNOWLEDGEMENT

We would like to thank Ahlia University for providing the support and our students for working on developing and testing some of our research ideas.

Figure 9. Example of Enhanced UML Diagram in the SEMANTIC Dictionary System.
REFERENCES


