An Intelligent Multi-Lingual Object Oriented Dictionary System

DR. MANSOOR AL-A'ALI & SAMIA HASSAN YOUSIF
Department of Computer Science
College of Information Technology
University of Bahrain
PO Box 32038, Bahrain
Tel. +973- 39607724

Abstract:- This paper presents a new approach for the development of an intelligent Object Oriented dictionary system called TOOT, which has been demonstrated to have a number of useful and practical applications for teaching school children a variety of language related features. Each word in the dictionary belongs to a class and may have one or more subclasses. Subclasses inherit all the attributes of their super class and this concept was utilized in processing the dictionary functions. Some words have associated with them an image, the associated actions it can do, for example a human can write, and the actions that can be carried out on it, for example a pen can be used for writing. The applications of this intelligent dictionary include translation from English to Arabic, checking phrase semantics, helping in writing a child story with pictures, etc. One of the major benefits of this dictionary is that words are placed in their corresponding class or subclass and the same word may exist in more than one class and subclass; this makes translation for example more accurate due to semantic considerations. The paper presents the design and applications of the TOOT dictionary.

Keywords: Object Orientation, Intelligent, Teaching, Semantics, Syntax, Translation, Dictionary, Language, English, Arabic.

1 Introduction

Research into the use of computing for teaching the English language is long established and there is plenty of reported literature in this direction [2], [3], [4] & [6]. Teaching a native or a foreign language to students is clearly a challenging task and attempts to automate this process have clearly not yielded the appropriate results. We have analyzed this problem and considered it from totally different direction, which is building the language semantics into a knowledge based dictionary system containing a number of semantic features. Until now dictionaries are considered as tools containing words, their descriptions and probably the translated meaning in another language. We take the concept of a dictionary a little further by considering it as a hub of words, their meanings, their associations, the possible actions relating these words, the flow of these actions and most importantly, their abstractions and inheritance issues. Words as we know them are in reality sets of classes and their subclasses and these classes can inherit knowledge and actions from each other and hence the concept of building semantics.

There are many types of dictionaries which are used for different purposes, such as Language Translation, Encyclopedias, Language Dictionaries, Universal Dictionaries, Multiple Access Dictionaries, Mathematical Dictionaries, Technical Dictionaries, Medical Dictionaries, Biological Dictionaries, Business Dictionaries, etc. There is a variety of human knowledge accumulated in a variety of fields which is kept in the form of dictionaries. Improvements and developments of dictionaries will continue to progress to suit new target applications. Most improvements will be focused on the dictionary features such as: Semantics, Multiple Languages, Imaging (Fingerprint dictionary, Signature Dictionary, Blood testing Dictionary), voices and sounds (Used in telephones or international conferences).

There are reported attempts to develop special purpose dictionaries to hold information about specific industrial, commercial or scientific items or definitions which have been reported in the literature [8]. Some researchers focused on language issues in bi-lingual dictionaries. Some researchers proposed dictionaries for capturing phrases for those who wish to learn English as a foreign language [9]. Others focused on using dictionaries for reading comprehension [9]. Some studied the way in which learners of English as a foreign language read and interpret dictionary entries [5] and [14]. Some researchers studied the dictionary use by language learners and translators [14]. Laufer & Kimmel [7], studied the issues of how learners really use dictionaries. Some researchers focused on the ambiguity of translating prepositions [10]. Some researchers focused on language dictionaries to keep information about language words, phrases and terms as a

reference[3] and [10]. Some researchers presented formats of dictionaries to cross link the words of two languages which in most cases is English with another language[1] & [11].

Most researchers have always resorted to Natural Language Processing and the techniques of Artificial Intelligence to deal with language issues including understanding and translation. Natural language processing research dealt with issues of understanding and querying text in the same language as well as for the purpose of translating from one language to another.

Little or no attention has been given to build the intelligence in the dictionary rather than processing the syntax and semantics of the language itself. After thoroughly searching the literature, the author could not find any similar attempt to develop an 'intelligent' dictionary to be used especially as a multi-purpose educational tool for young school children which is based on a scientific approach in computer science. The concept of building an Objected Oriented dictionary which incorporates the rules of inheritance and semantics of the words of any language has not been reported anywhere.

In the case of language learning and teaching by young school children, the dictionary should not be looked at as a simple database of words of the same language or as a cross reference of words to be used between two languages, but as a source of knowledge which can be engineered to house a range of intelligent features which supersede the static dictionary cross reference features. Young children especially, need to feel more of the knowledge relating to the word rather than reading a static meaning or a description which does not link the word to its real life interpretation.

2 The TOOT Dictionary concepts and design

TOOT is an intelligent multi-purpose semantic dictionary which holds 'knowledge' about each word. TOOT keeps the following about each word:

Object-Oriented Class

Objects are individual instances of a class. For example, you may create an object called Spot from the class Dog. The Dog class defines what it is to be a Dog object, and all the dog-related messages a Dog object can act upon. Other objects can be created for this class and could be called Fido, Rover, etc (Fig. 2). The Dog class defines messages that the Dog object understands, such as bark, bite and chase.

1. Actions associated with the word: these are actions which are generated by the word and actions which can be applied on the word. For example, the word boy can have the actions read, write, eat, etc.
2. Inherited knowledge from its parent (super class) nodes: these are actions generated by the word's upper class and actions applied on the word's upper class. For example, the word living provide a downward inheritance of the actions breath, eat, etc.
3. Subjects and Objects associated with the word: these are objects for each action generated by the word and subjects of the action applied on the word.
4. The Description of each word (meaning).
5. The word translation in Arabic although it can be in any language.
6. A representative image of the word and its associated actions.
7. The voice pronunciation of the word.

Object Orientation (OO) is a technique used in computer science to build inheritance between classes; words at the higher level pass all or some of their features to their subclasses. Using the OO techniques makes TOOT Dictionary Objects more semantically related and hence better intelligence can be implemented. The following describes the basic concepts of Object Orientation in relation to TOOT.

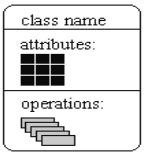

Figure 1 Classes

Objects

An Object is a black box that receives and sends messages, e.g., book, dog, girl, bird.

Classes

A Class is a description or definition of an object or group of similar objects. Classes are templates or blueprints used for building objects and they determine everything about an object. A class defines the data contained in an object and the processing (operations) done by an object, see Fig. 1.

Figure 2 Dog Class and its object

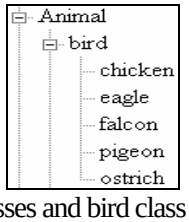

Figure 3 Animal classes and bird class and the children

Classes versus Instances

Each object created from a class contains the same attributes and has the same methods of that class. However, values assigned to an object's attributes may be different for each instance of the class. All the Birds look similar. They all have two wings and two legs. The Bird is the class. Ostrich, Chicken and other birds are objects or instances of bird. They can have different colors, different sizes, some of them big and others are small, and some can fly and other can not.

Object Components

Since a class is a description or definition of an object, a class defines the components of an object. Some description of object components and how they can be implemented in TOOT dictionary will follow.

Attributes

Attributes are data values or variables stored in an object, for example Height, width, etc.

Methods

Each object has methods. A method is the action(s) that a message carries out. Arguments are often supplied as part of a message, for example, the eat message might contain an argument that says what to eat, (meat, food, wood, etc.).

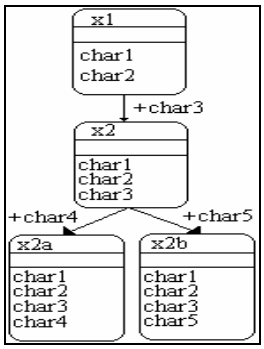

Figure 4 Inheritance concepts

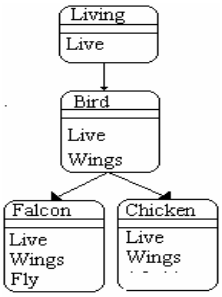

Figure 5 Example of Inheritance

Inheritance

Inheritance is the ability to define new classes from existing classes, see Fig. 4 & Fig. 5. No need to start from scratch when adding a new class or object. A descendant (sub-class) class inherits all the attributes and methods defined for its ancestor class or classes. Inheritance is a way of incorporating, without copying. You can define new attributes and methods in the new class. The descendant class can also be extended or added to the behavior of methods. From figure 4, x2 inherits characteristic of x1 in addition to its own characteristic.

For example, after creating the class living, a subclass Bird can be made, which defines Birds and will get the live action from living with its own attribute, the wings. Bird class can have Falcon and Chicken as a subclass, Falcons can fly where Chickens cannot. Both of those classes Falcon and Chicken will inherit live and wings from the upper class bird. (Figure 5)

Encapsulation

Encapsulation begins when grouping data together into an object. It is also a form of information hiding, where an object hides its internal implementation details from view. Encapsulating an object in TOOT Dictionary is implemented by making its attributes private. The object's public interface can be implemented by defining public events or functions in the class. This can be done in TOOT Dictionary by two levels of scope for variables: Public and private. Public are the inherited ones, which are private for the parent or the ancestor or parent of ancestor and so on.

Polymorphism

Polymorphism is derived from a Greek word meaning many forms. In object oriented it means different objects are able to respond differently to the same message. For example,

Form 1: The Ostrich is a Bird
Form 2: The Ostrich is an Animal
Form 3: The Ostrich is a Living thing.
Form 4: The Ostrich is a thing.

Objected Oriented Representation of the TOOT dictionary entities and relationships

It is helpful to visually diagram the relationships between the TOOT entities. An Entity relationship Diagram (ERD) shows the way entities records in one table relate to entities in another.

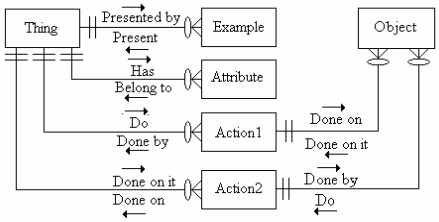

Figure 6 Objected Oriented Representation of the TOOT entities and relationships

Analysis of OO representation, see Fig. 6:

1. *Thing and action1 have a one-to-many relationship. (Each Thing Does many Actions of type Action1, but Action1 must be done by only one Thing)*

2. *Thing and action2 have a one-to-many relationship. (Each Thing is Done on it many Actions of type Action2, but Action2 must be done on only one Thing)*
3. *Thing and Attribute have a one-to-many relationship. (Each Thing may have many attributes, but Attribute may be contained in only one Thing)*
4. *Thing and Example have a one-to-many relationship. (Each Thing has many Example, but Example belong to only one Thing)*
5. *Action1 and Object have a one-to-many relationship. (Each Action1 can be done on many Objects, but Object is done on it only one Action1)*
6. *Action2 and Object have a one-to-many relationship. (Each Action2 done by many Objects, but Object do only one Action1)*

3 TOOT tools and Interface Design

TOOT provides the basic required features for adding, deleting and updating the knowledge in the dictionary. The main interface form is the starting point for the child and the teacher to navigate through TOOT. Any of the tools implemented can be invoked under the supervision of the teacher or by the child after he/she is made familiar with TOOT tools. A brief descript of each of TOOT applications or tools is given.

The Story

The story in TOOT Dictionary acts as an educational tool that helps the child to understand English words by displaying the corresponding images of the words and hence the child can relate to the English word in terms of its text and picture, see Fig. 7. TOOT holds a picture for each word where appropriate. After writing a sentence in the given edit box, TOOT Dictionary will generate the picture like story according to the written sentences. Words that have no pictures are displayed in written format but in much bigger and colored font, see Fig. 7.

Spelling Editor

The child or the teacher writes any set of simple sentences in the given empty box and after pressing on the correction button, the correction of the written sentences will appear. The correct spelling of the words is taken directly from Ms-Word via OLE Link.

Fig. 7 the story form

Figure 8 Quiz Form

The Quiz tool

This is an educational tool that examines the child's knowledge of English vocabulary, it functions by giving a question with a little hint on how to solve it, and marks are given to each answer in order to help the teacher and the child with the status and achievement of the child, see Fig. 8.

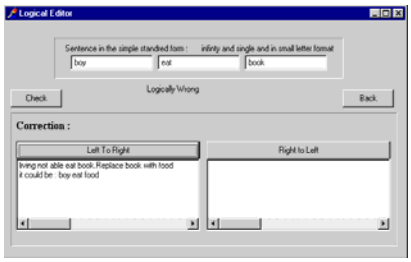

Figure 9 Semantic Checker Form

The Language Semantic Checker

The semantic Checker checks the semantics of simple sentences entered by the child. This is made possible because of the intelligent features of the OO approach and the actions associated with each word in TOOT. The young child may not appreciate that the sequence of words around a given verb has an effect on the meaning of the sentence or may not appreciate the meaning of the verb itself. For example, the child may write the sentence 'boy eat book' or the teacher may give the child this sentence in order to test the child. This sentence will be rejected by TOOT as a semantically incorrect sentence, see Fig. 9. To semantically check a simple sentence, all rules of OO classes and inheritance will be applied. In the example sentence 'Living eat food'; this sentence will contain the name of the thing, its private action (not the inherited) and the private object (not the inherited). In the example sentence 'boy eat food' shown in figure 10; this sentence

will contain the name of the thing, its private actions (not the inherited) and the general object (the inherited). In the example sentence 'living eat meat' shown in figure 11; this sentence will contain the name of the thing, its general action (the inherited) and the general object (the inherited).

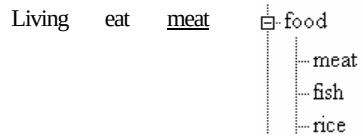

Figure 11 Inherited Actions

4 The TOOT Tree Structure

The Tree is the most essential and attractive component in TOOT because it represents the classification of the word in a good fashion which gives a clear understanding of the meanings to the child. The tree provides a full path of parents and children classes in one line structure rather than separate fields. The tree makes it easier to do forward and backward navigation to different levels. The tree structure tool is used for searching, storing and displaying words.

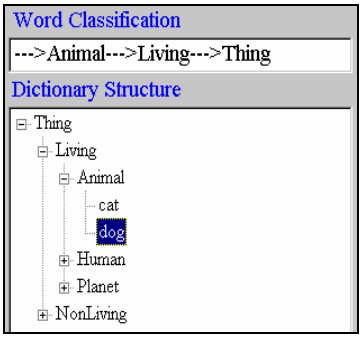

Figure 12 Tree Structure

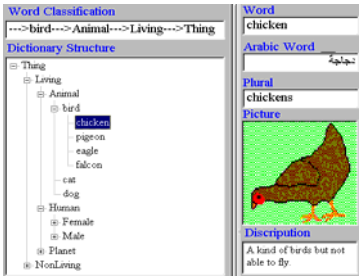

Fig. 13 Part of the main form (birds under the bird class)

The three directions of the levels shown in Figure 12, same level direction, previous level (super class), which is the parent and next level (sub-class) which is the children. These directions increase the benefit of the Tree Structure. Showing the words that are at the same level, gives a better understanding of the word because it presents the property of contradiction. It also presents the differences by giving the properties like the actions, attributes and pictures for each word in the same level gives a better understanding of the meaning, see Fig. 13. For example, to know the

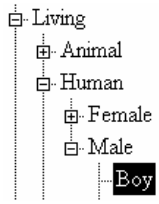

Figure 10 Inherited Actions

chicken amongst the birds, you must know some of its details which is different from other birds, or the other birds' details which the chicken does not have. The child only needs to remember the name of a thing by just knowing its group.

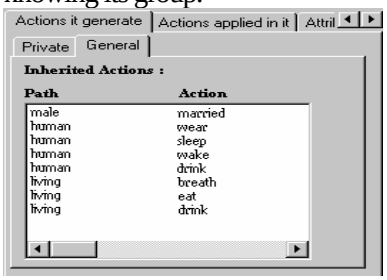

Figure 14 Inherited Actions for boy

Presenting the words in the previous level is needed. One reason is the provided inheritance concepts of the word's actions, for example the word (human) will inherit the actions of its parent the word (living) as a general set of actions like breath, drink and eat, and therefore no duplicate information will be stored. Going back to the previous levels will provide the inherited actions in order to get a better idea about those actions and the related words. The word boy will inherit some actions from male and from what male has inherited from human and what human has inherited from living and so on, Figure 14.

5 Conclusion

This paper presented a new insight into the way we should think of on-line dictionaries and especially in terms of their use for teaching school children. An online dictionary was developed and used as a source of knowledge, which was engineered to house a range of intelligent features, which supersede the standard static dictionary cross-reference features. This paper presented a new approach for the development of an intelligent Object Oriented dictionary called TOOT, which has been demonstrated to have a number of useful and practical applications for teaching. The applications of this intelligent dictionary include checking simple phrase semantics, helping in displaying a child story with pictures to represent words where appropriate, etc. One of the major benefits of this dictionary is that words are placed in their corresponding class or subclass and the same word may exist in more than one class and subclass; this makes translation for example more accurate due to semantic consideration. Research is on going to incorporate more knowledge and

features in TOOT to help in teaching English language and especially in understanding the semantics and applications of words and sentences.

References:-

1. Agirre, E., Arregi, X., Artola, X., De Ilarraza, A.D., Sarasola, K., Soroa, A., 1999. MLDS: a translator-oriented multilingual dictionary system, *Natural Language Engineering*, Vol. 5, Iss. pt.4, pp. 325-53.
2. Cla´ udia B.M.J. Andrea V. Steilb, Jose´ L. Todescoc. Martinsa, Factors influencing the adoption of the Internet as a teaching tool at foreign language schools *Computers & Education*, Volume 42, Issue 4, May 2004, pp. 353–374.
3. Columbus G., 2000, Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making, *System*, Volume 28, Issue 4, pp. 623-626
4. Francis Mangubhai, Perce Marland, Ann Dashwood and Jeong-Bae Son, Teaching a foreign language: one teacher's practical theory, *Teaching and Teacher Education*, Volume 20, Issue 3, April 2004, pp. 291-311.
5. García J. S., The Use and Abuse of EFL Dictionaries. How Learners of English as a Foreign Language Read and Interpret Dictionary Entries, *System*, Volume 29, Issue 2, June 2001, Pages 306-308
6. Kenneth Ruthven Sara Hennessy & Sue Brindley. T., Teacher representation of successful use of computer-based tools and resources in secondary-school English Mathematics and Science, *Teaching and teacher education*, Volume 20, Issue 3, April 2004, pp. 259-275.
7. Laufer B. and Kimmel M., Bilingualised dictionaries: How learners really use them, *System*, Volume 25, Issue 3, 1997, pp. 361-369
8. Lovis C., Baud R., Rassinoux A. M., Michel P. A. and Scherrer J. R., Medical dictionaries for patient encoding systems: a methodology, *Artificial Intelligence in Medicine*, Volume 14, Issues 1-2, 10, 1998, pp. 201-214
9. McAlpine J. and Myles J., Capturing phraseology in an online dictionary for advanced users of English as a second language: a response to user needs, *System*, Volume 31, Issue 1, 2003, pp. 71-84
10. Ochi, Y., Yano, Y., Wakita, R., Polite expression dictionary system deriving examples from retrieval resources, *Transactions of the Institute of Electronics*, Information and Communication Engineers D-I. Vol. J83D-I, Iss. 6, 2000, pp. 710-18.
11. Ohtsuka, Y., Yamaguchi, H., Sagawa, Y., Horiba, I., Sugie, N.,. An intelligent electronic English-Japanese dictionary system can disambiguate the senses of prepositions appearing in a sentence just as language experts do, *Transactions of the Institute of Electrical Engineers of Japan*, Part C. Vol. 120-C, Iss. 12, 2000, pp. 1809-13.
12. Ohtsuka, Y., Yamaguchi, H., Yamamoto, S., Sugie, N., An intelligent electronic English-Japanese dictionary system can disambiguate the senses of prepositions appearing in a sentence just as language experts do, IEEE International Conference on Systems, Man, and Cybernetics, 1999, (Cat. No.99CH37028), IEEE, Piscataway, NJ, USA. pp. 820-3 vol.5.
13. Ronald J., Research on dictionary use in the context of foreign language learning: focus on reading comprehension: Yukio Tono. Max Niemeyer Verlag, Tübingen, 2001, xii+257, System, Volume 30, Issue 2, pp. 251-254.
14. Schneider K. P., Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators, *System*, Volume 29, Issue 1, 2001, Pages 153-159.