International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 2 Issue 6, June - 2013

Adaptable Verbal Ability Testing Tool

Sharanappa S Yarnal
S.J.C.I.T, Chickaballapur.

Ajay N
S.J.C.I.T, Chickaballapur.

Abstract

Verbal ability refers to a person's facility at putting ideas into words, both oral and written. This facility involves possessing not only a strong working vocabulary but also the ability to choose the right words to convey nuances of meaning to a chosen audience. Verbal ability is usually demonstrated as the ability to write and speak well. We describe an approach to automatically generating question for verbal ability of the candidate. Using data from WordNet, we generate a synonym questions. It will have multiple-choice. We present experimental results that suggest that these automatically generated synonym questions give a measure of vocabulary skill that correlates well with subject performance on independently developed human written questions. In addition, strong correlations with standardized vocabulary tests point to the validity of our approach to automatic assessment of word knowledge.

Index Terms–wordnet, Comprehension Index (CI), Familiarity Index (FI), Verbal ability Index (VI), word Knowledge.

1. Introduction

The Comprehension Index is the index to know the candidates how much understand the words in a different way. The Comprehension Index includes four tests. First, Similarities: Abstract verbal reasoning (e.g., "In what way are an apple and a pear alike?"). Second, Vocabulary: The degree to which one has learned, been able to comprehend and verbally express vocabulary (e.g., "What is a guitar?"). Third, Information: Degree of general information acquired from culture (e.g., "Who is the president of Russia?"). Fourth, Comprehension [Supplemental]: Ability to deal with abstract social conventions, rules and expressions (e.g., "What does *Kill 2 birds with 1 stone* metaphorically mean?"). The familiarity index is the index to know the candidates familiar about the words.

In the REAP[10] system automatically provides users with individualized authentic texts to read. These texts, usually retrieved from the Web, are chosen to satisfy several criteria. First, they are selected to match the reading level of the student (Collins- Thompson and Callan, 2004). They must also have vocabulary terms known to the student. To meet this goal, it is necessary to construct an accurate model of the student's vocabulary knowledge (Brown and Eskenazi, 2004). Using this model, the system can locate documents that include a given percentage (e.g., 95%) of words that are known to the student. The remaining percentage (e.g. 5%) consists of new words that the student needs to learn. This percentage is controlled so that there is not so much stretch in the document that the student cannot focus their attention on understanding the new words and the meaning of the text. After reading the text, the student's understanding of new words is assessed. The student's responses are used to update the student model, to support retrieval of furture documents that take into account the changes in student word knowledge.

In this paper, we describe our work an automatic generation of vocabulary assessment of synonym question. We also report results from a study that was designed to comprehension index for a words. In addition to the importance of these assessments in the REAP system, tests of word knowledge are central to research on reading and language and are of practical importance for student placement and in enabling teachers to track improvements in word knowledge throughout the school year. Because tests such as these are traditionally hand-written, development is time-consuming and often relies on methods that are informal and subjective. The research described here addresses these issues through development of automated, explicit methods for generation of vocabulary tests. In addition, these tools are designed to capture the graded and complex nature of word knowledge, allowing for more fine-grained assessment of word learning.

2. Design and Implementation

Design is a phase in which algorithms for each module is specified. Most of the components will require a more detailed discussion. Each subsection of this section refers to or contains a detailed description of a system software component.

In this section, we describe the process used to generate a synonym questions. After introducing the WordNet resource we discuss the synonym question and the forms in which they appear.

2.1 WordNet

WordNet[2][7] is a lexical resource in which English nouns, verbs, adjectives, and adverbs are grouped into synonym sets. A word may appear in a number of these synonym sets, or synsets, each corresponding to a single lexical concept and a single sense of the word (Fellbaum ed., 1998). The word "bat" has ten distinct senses and thus appears in ten synsets in WordNet. Five of these senses correspond to noun senses, and the other five correspond to verb senses. The synset for the verb sense of the word which refers to batting one's eyelashes contains the words "bat" and "flutter", while the synset for the noun sense of the word which refers to the flying mammal contains the words "bat" and "chiropteran". Each sense or synset is accompanied by a definition and, often, example sentences or phrases. A synset can also be linked to other synsets with various relations, including synonym, antonym, hypernym, hyponym, and other syntactic and semantic relations (Fellbaum ed., 1998). For a particular word sense, we programmatically access WordNet to find definitions, example phrases, etc.

2.2 Question Type

In order to retrieve data from WordNet[8], the tool choose the all sense of the word. The system can work with input of varying specificity. The most specific case is when we have all the data: the word itself and a number indicating the sense of the word with respect to WordNet's synsets. When the target words are known beforehand and the word list is short enough, the intended sense can be hand-annotated. More often, however, the input is comprised of just the target word and its part of speech (POS). It is much easier to annotate POS than it is to annotate the sense.

The synonym question has the testee match the target word to a synonym. The system can extract this synonym from WordNet using two methods. One method is to select words that belong to the same synset as the target word and are thus synonyms. In addition, the synonym relation in Word-Net may connect this synset to another synset, and all the words in the latter are acceptable synonyms. The system prefers words in the synset to those in synonym synsets. It also restricts synonyms to single words and to words which are not morphological variants of the target word. When more than one word satisfies all criteria, the most frequently used synonym is chosen, since this should make the question easier. This question could be considered comprehension processing.

2.3 Question Form

The synonym question for the word “verbose” would have the stem “Select the word that is most similar in meaning to the word verbose” with choices “inflammable”, “piping”, matrilineal”, and “long-winded”.

Two issues to consider when creating multiple choice format questions are the wording or appearance of the questions and the criteria for selection of distracters. Author followed the guidelines for good multiple-choice questions[3] described by researchers such as Graesser and Wisher (2001). In accord with these guidelines, they are made questions had 3 choices, although the number of choices is a variable supplied to the question generation software and also considered the most appropriate wording for these questions, leading us to choose stems such as “Select the word that is most similar in meaning to the word plausible” for the synonym question rather than “Choose the word that means the same as the word plausible.” The latter would be problematic when the correct answer is a near-synonym rather than a word with precisely the same meaning.

2.4 Question Generation

Figure 1 shows the relationship between Familiarity Index versus Complexity. This graph shows the word is more familiar then complexity decreases.

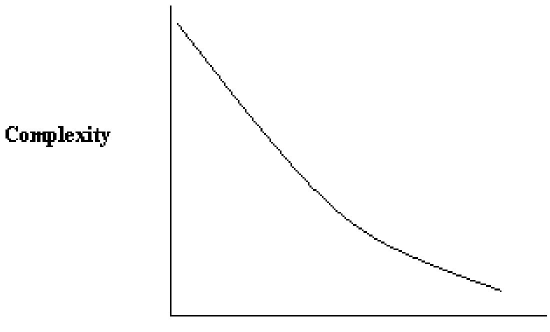

Figure 1: Familiarity vs Complexity

Figure 2 shows the relationship between Verbal ability Index versus Complexity. This graph shows the word is not familiar then complexity increases.

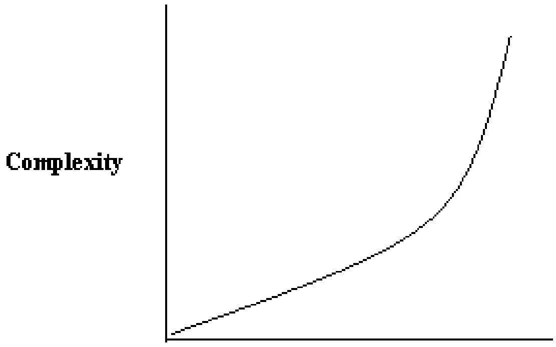

Figure 2 Verbal ability Index vs Complexity

So, Familiarity index and Verbal ability index are inversely proportional to each other.

$$FI=N-(VI-1)$$

Here we keep constant N as 18. The VI must be in the range between 1 and 18. Then we can find the FI by using this formula.

FI is range between 1 and 2: PG level synonym questions.FI is range between 3 and 6: UG level synonym questions.FI is range between 7 and 8: PU level synonym questions.FI is range between 9 and 11: high school level synonym questions.FI is range between 12 and 18: primary school level synonym questions.

When we give the sentence and verbal ability index as inputs. The synonym questions are generated as above said levels.

2.5 Take Test

Conduct a test as any number of candidates. After answer the questions, the candidate know the number of points he/she has got and also number of questions he/she as answered. If the candidates select the correct option he/she gets one point otherwise no point, it means he/she select the wrong option.

2.6 Comprehension Index

When after finishing the test, we can find the comprehension index for words. Here Familiarity Index is inversely proportional to Comprehension Index with respect to complexity.

For each word,

$$FI = \sum_{i=1}^{n} \left[\frac{\text{candidate correctly answered}}{n} \right]$$

Where n=number of candidates taken the test.

Finding the CI is as follows, FI is range between 0.001 and 0.111 then CI is between 1 and 2. FI is range between 0.112 and 0.333 then CI is between 3 and 6. FI is range between 0.334 and 0.444 then CI is between 7 and 8. FI is range between 0.445 and 0.611 then CI is between 9 and 11. FI is range between 0.612 and 1 then CI is between 12 and 18.

3. Analysis

The meaningful sentences are composed of meaningful words; any system that hopes to process natural languages as people do must have information about words and their meanings. This information is traditionally provided through dictionaries, and machine-readable dictionaries are now widely available. But dictionary entries evolved for the convenience of human readers, not for machines. WordNet provides a more effective combination of traditional lexicographic information and modern computing. WordNet is an online lexical database designed for use under program control. English nouns, verbs, adjectives, and adverbs are organized into sets of synonyms, each representing a lexicalized concept. Semantic relations link the synonym sets.

Microsoft Word is the de facto word processor. Whether we work at home, at school or in business, the chances are that we will use MS Word if you need to create your own, or read someone else's document. Word can seem a little frightening at first - especially if you are coming to Word 2007 from previous versions. Word thesaurus can take advantage of to improve our documents. Thesaurus can be used to find synonyms (different words with the same meaning) and antonyms (words with the opposite meaning).Microsoft Word displays the synonym list in the form of most frequently used to least frequently use. For example, we can take a word as ‘develop’, the synonyms are expand, build up, enlarge, extend, increase, widen, and grow. But when compared with the WordNet browser, the synonyms are displayed depends on the senses of word.

4. Conclusion

The paper is described our work in an automatic generation of synonyms questions without using question bank and also designed to capture the graded and complex nature of word knowledge, allowing for fine-grained assessment of word learning. The tool is also finding the comprehension index for words.

Extending our paper to generate different types of question generation, for example, cloze questions, antonym questions etc... And also we can make out tool as a domain specific, provided domain[4] specific database.

Acknowledgement

The authors would like to thank the anonymous reviewers for their comments, which have significantly improved the quality of this paper, and also special thanks to Prof. Nitin V Pujari, presently working as HOD in Dept of CSE, PESIT, for their ideas and guidance for this paper.

References

[1] Gregory Aist. 2001. Towards automatic glossarization: automatically constructing and administering vocabulary assistance factoids and multiple-choice assessment, International Journal of AI in Ed., 2001.

[2] Miller G. A., Beckwidth, R., Fellbaum, C., Gross, D. and Miller, K. J., "Introduction to WordNet: An On-line Lexical Database", International Journal of Lexicography, Vol 3, No.4 (Winter 1990), pp. 235-244.

[3] Jonathan C. Brown, Gwen A. Frishkoff, Maxine Eskenazi, "Automatic Question Generation for Vocabulary Assessment", Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), pages 819–826, Vancouver, October 2005.

[4] S. G. Kolte, S. G. Bhirud, "Word Sense Disambiguation using WordNet Domains", First International Conference on Emerging Trends in Engineering and Technology, IEEE, 2008.

[5] Myunggwon Hwang, Byungsu Youn, Ilyong Chung, Pankoo Kim, "Semantic Measurement of Related degree between Unknown Word and Related Word for Automatic Extension of Lexical Dictionary", Fifth International Conference on Fuzzy Systems and Knowledge Discovery, IEEE, 2008.

[6] Kevyn Collins-Thompson and Jamie Callan. 2004. A language modeling approach to predicting reading difficulty. In Proceedings of the HLT/NAACL 2004 Conference. Boston, 2004.

[7] http://wordnet.princeton.edu/
edu.mit.jwi_2.1.5_manual

[8] Stephanie Chua, Narayanan Kulathuramaiyer, "Semantic Feature Selection Using WordNet", Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI'04), 2004.

[9] Raghuvar Nadig J. Ramanand1 Pushpak Bhattacharyya, "Automatic Evaluation of Wordnet Synonyms and Hypernyms", Proceedings of the 6th International Conference on Natural Language Processing, 2008.

[10] *Michael Heilman, Kevyn Collins-Thompson, Jamie Callan, Maxine Eskenazi,* "Classroom Success of an Intelligent Tutoring System for Lexical Practice and Reading Comprehension" Language Technologies Institute Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America,2004

Mr. Sharanappa S Yarnal is currently pursuing 3rd(6th sem) year Bachelor of Engineering in Computer Science and Engineering at S J C Institute of Technology, Chickballapur, affiliated to Visvesvaraya Technological University, Belgaum, Karnataka, India.

Mr. Ajay.N is an Assistant professor in the Department of Computer Science and Engineering at S J C Institute of Technology ,Chickballapur affiliated to Visvesvaraya Technological University, Belgaum, Karnataka, India.