AUTOMATIC CONFIRMATION OF CONNOTATION WORDS

Atilla Suncak
Kastamonu University, Araç Rafet Vergili Vocatioal Highschool,
Department of Computer Technologies, Kastamonu

Özlem Aktaş
Dokuz Eylül University, Department of Computer Engineering, İzmir

Çağdaş Can Birant
Dokuz Eylül University, Department of Computer Engineering, İzmir

Abstract

Smart-phones and tablet computers are undoubtedly the most useful invention of our time and used for numerous kinds of internet transactions, leisure activities and entertaining online / offline games, especially by younger. Developing technologies and some games which use these technologies may affect students' vocabulary (Turkish vocabulary in the context of this study) and pronunciation learning. The purpose of the study is to develop a software architecture which can enable the students to learn the Turkish words, their meanings, the semantic relations between them such as synonymy and antonym, through entertaining games running on mobile operating systems; and to enable researchers to gain insight about how people make inferences about a word when they first see it via free connotation information which is tried to be performed automatically by generating an algorithm of 'Automatic Confirmation of Connotation Words'.

Çağrışım (Connotation) Project is a mobile project that uses the Synonym and Antonym dictionaries prepared for the Turkish Language Association by Dokuz Eylül University Natural Language Processing Work Group. It is aimed at teaching Turkish vocabulary, with synonyms, antonyms and related words, via an entertaining activity. In this software, a randomly selected word appears on the screen and the user enters the first word that comes into the mind and the system grades the response according to the semantic relations defined in the software. At the end of one turn, all semantically related words, their meanings and even clipart of some words are presented to the user. The responses by the users which were stored enable the software to conduct some statistical analyses on them and therefore reveal the intellectual statuses and development of the users. At the end of the session the data will be sent to the main database.

What is more, an automatic confirmation of connotation words is performed. The aim of this performance is to develop a better algorithm for connotation words gathered in Çağrışım project and represent.

Keywords: vocabulary, antonym, synonym, semantic relations, mobile applications, educational software, connotation.

1.Introduction

The rapid popularization of smart-phones and tablet computers especially among secondary school students has reached to a degree that it would affect their learning, thus their future. Therefore, the quality of the applications installed on those devices has become an important issue. It is a well-known fact that these applications, especially games affect the development of teenagers.

Educational games have become one of the essential methods used in the education of children. The aims for the development are to increase children's mental capacities, to facilitate their learning and thus making education more productive and efficient. According to Gee, a member of the National Academy of Education in the USA, the tasks in the games, especially strategy games, arise interest in children, cause them to strive and support their mental development [1][2]. In this respect, Gee advises schools to adapt some of the strategies used in the games into the education systems in order to educate the children better and to determine their abilities.

It has been suggested that the students liked game-based learning environments since these environments reduced anxiety, helped individual learning and supported learning visually [3]. Both learning is facilitated, and the remembering rates are increased by a game, designed considering the characteristics of the students; therefore, instructors in the schools preferred game-based learning [4].

Rapid development and spread of the technology nowadays and the interest of individuals in games have caused software developers to publish various games that can be played on websites and mobile devices. These games, encountered almost everywhere, have engaged attention and win the favour of the students who are in their developmental stages; even they have become an addiction.

In this context, it is significant that the games are designed and developed in a way that they would affect the development and creativity of the individuals positively, not causing them to idle their time away. Vocabulary games are the games that facilitate the learning of new words in a language, and reinforce memory. There are many vocabulary games developed for almost all languages in the world. It is possible to find various vocabulary games for Turkish to play on the internet or on mobile devices such as finding the word from a given meaning, correcting jumbled words, constructing words from random letters, hangman, anagram, word hunt, etc. Although, most of those games are designed with simple interfaces to enable the input, some of the games have entertaining animations and graphics.

Some games, especially vocabulary games, presented to the teenagers may influence the learning of Turkish words and by extending their ability to express themselves appropriately.

2. Overview of Çağrışım Project

This study aims at introducing "Çağrışım" (Connotation) by presenting the architecture used for implementing the software. The main purpose of "Çağrışım" game project - a subproject of FATIH(stands for 'Studies of Increasing Opportunities and Reforming Technology') Information Access and Management Systems project- is to implement a game to run on the smart-phones and tablet computers which have the following properties: (i) to ensure learning of Turkish words with their synonyms, antonyms etc. in an entertaining environment supported with visual materials, (ii) to enable collection of information about which connotations the students notice when they first see a word, thus enabling school administration and psychological counsellors to evaluate the students' psychological statuses and ability to relate concepts by informing them about the connotations the students entered in the game as reply to a word they encountered. These aims are in parallel with the aim of FATIH project, which is "to develop innovative education technologies for enabling students to learn more efficiently and effectively."

In addition to these, this game will contribute to the construction of creating mind maps, thus enabling the permanence of knowledge. The present games which can be found on the internet are mostly internet based games with a limited inventory of words but Çağrışım uses three dictionaries, a synonyms dictionary, an antonyms dictionary and a dictionary of related words which was prepared by The Turkish Language Institution by Dokuz Eylül University Natural Language Processing Group; therefore this is a verified and comprehensive vocabulary list.

"Çağrışım" software is played on smart devices without the need of an internet connection; however the scores and the results, some recommendation based on the answers given by the students, some statistical analyses and the development statuses are sent to the parents, teachers, counsellors or the administration through some interactive panels. The game gives alerts to the internet interfaces used by the parents, teachers and counsellors about the achievement of the student or if any occasion arises that would be an indication of a psychological problem, such as entering word connotations related to hatred, violence and any kinds of abuse; thus the system will enable quick interactions between the school and the parents and quick responses from these.

Sample interface of the game is given in Figure 1.

Figure 1 : Sample Interface

When the game ends, the software sends the answers and other game data to the central computer. Using these data, a "Turkish Connotations List (or Dictionary)" is aimed at build as a support to the "Turkish Semantic Network" still being developed by Dokuz Eylül University Computer Engineering Department.

The vocabulary list used in the software is gathered from the Synonyms Dictionary and Antonyms dictionary published by the Turkish Language Association (TDK), other printed dictionaries of this kind, and new words will be added to the list, if necessary, by linguists participating in the project. The short texts and clipart describing the words in the vocabulary list, which do not infringe copyrights, have been either collected from the internet or prepared by the project members.

It is thought that, at the end of the project, a co-occurrence analysis of the words and their connotations obtained from the users during the use of the software, may yield new meaning relations between words [6]. Thus the system can learn new relation with the results obtained from different users. The response times of the students are evaluated using the decision cubes, and the concepts on which the students have difficulty can be determined with regard to regions, schools and classes. The concepts causing difficulties can be reinforced by more frequent appearance in the software.

3. Automatic Confirmation of Connotation Words

Çağrışım is a game that is totally open to be developed and worked on. As the developer group, we managed to enable the software to determine over two words whether they are connotation-words or not.

As it is mentioned before, some words in the list do not only have their synonym / antonym equivalents, but also have their info-texts and clipart. So, we decided to work on those words and their info-texts. At first, those texts have to be analyzed, and then compared whether they have anything in common or not.

For this reason, text mining technology is the best solution for this purpose. But, most text mining algorithms are for English language. As Turkish language is not similar to English, even not close, we have to create our own algorithm or try to perform a hybrid some of them.

The main plan is that; the first word's info text is mined. Just like in English, the text has to be cleared from the unrelated words, prepositions or pronouns. After that performance, remaining words has to be simplified till their roots. In text mining, root is the basic element and the most important element as 'it' has a meaning. But we have more than one or two roots, so we have to compare those roots among each other and only related ones will be used. After the whole operation of the first word, the second word's info text is also performed those operations. And finally we are going to have two root groups. Number of common words or related-meaning words between two groups helps the software to determine the relation between the first and second word.

3.1 Analyzing Processes of Connotation Words

As it is mentioned before, we have Turkish antonym, synonym and close-meaning vocabulary list. By using this list, we tried to capture a meaning-relation between two words with help of their info-texts. To explain that, let us give an example :

1- Un = Buğdaydan elde edilen, ekmek yapmaya yarayan beyaz renkli toz madde.

   (Flour : A dust-form material made from wheat which is white and used for making bread.)

2- Şeker = Şeker pancarından elde edilen, beyaz renkli tatlı bir toz madde.

   (Sugar : A dusty material made from beet which is white and sweet.)

Now we have two words with their info-texts. At first, we should mine those texts and get meaningful roots.

First group = Buğday(wheat), ekmek(bread), beyaz (white), toz (dust)

Second group = Şeker pancarı (beet), beyaz (white), tatlı(sweet), toz (dust)

To analyze and determine relation between those group;

1. First phase: Compare roots with other roots. If those two group's words have common word(s), then we can say that they have a connotative relation because of that common word(s).

2. Second phase: If first phase declines, then we start comparing each root's antonym/synonym/close-meanings of 1.Gruop with 2. Groups' words. If no result, then we perform the same operation to 2.Gruops'words and compare them with 1.Groups' words.

3. Third phase: If those phases both decline, then we start analyzing deeper. We find all antonym/synonym/close-meaning words of both groups' roots. Then we now have another two group which consist of antonym, synonym and close meaning words of 1. and 2.Group roots. Then, we try to find common words.

According to the example, first phase has the strongest meaning-relation of two words. Second phase and third phase has relation too but not as strong as first phase. So, when we compare "UN (FLOUR)" and "ŞEKER(SUGAR)", we create 1.Group and 2.Group. When system check those groups, it finds BEYAZ(WHITE) and TOZ(DUST). So, system interprets the situation as '*Flour and sugar has a connotative relation as they have two common words (white, dust) found in first phase*'.

Another example;

Elma : Ağaçta yetişen, kırmızı, yeşil ve sarı renkleri olan, tatlı veya ekşi yuvarlak bir meyve.

(Apple : Round-shaped fruit that is grown on tree, may be red,green or yellow and may taste sweet or sour)

Domates: Kırmızı renkli, yuvarlak, yemeklerde kullanılan bir sebze.

(Tomato: A kind of vegetable that is used in meals , red and round.)

1. Group: round, fruit, tree, red, green, yellow, sweet, sour

2. Group: vegetable, meal, red, round

As it is seen above, the system can find two common words in the first phase. For this reason, it is interpreted that 'Apple *and tomato has a connotative relation as they have two common words (red, round) found in first phase*'.

As we are dealing with meanings of words, there may become some problems in analyzing. For example, if info-texts are not prepared properly, then system may not find any relation or leads to be analyzed wrongly. What is more, if stemming operation is not performed properly, then system may not find any relation either. Therefore, for the best result from system, whole operation (every step) has to be performed properly so that the next step works properly.

4. Conclusion

ÇAĞRIŞIM game software is directly related with the aims of the FATİH Information Access and Management Systems Project, especially with the aim stated as "developing innovative education technologies for enabling students to learn more efficiently and effectively" and makes a great contribution to this aim. ÇAĞRIŞIM is a mobile learning application which offers the game-based learning environment to the students, enables them to learn new words, their meanings, the words related to these words (antonym, synonymy, etc.) with the support of visual materials.

Also, the connotative data entered by the students during the game will enable teachers, administrators and counsellors to evaluate the knowledge, the conceptual association abilities and the psychological statuses of the students. The development and achievement of the students will be presented to the teachers and the parents after some filtering and statistical analyses, thus enabling the student-teacher and parent interaction.

The students will improve their Turkish language skills as they play the game; they will learn the correct uses of language and develop their language understanding and production capabilities; they will enrich their conceptual maps, thus contributing to their own academic, psychological and social developments. The game will facilitate the students in encoding correctly and meaningfully their newly acquired knowledge in any area. Also they will become more proficient and more aware in their language and will become more efficient readers. In this respect, it is thought that the project will contribute in bringing up individuals who think critically, questioning the information they encounter, learning to learn, using their knowledge to overcome problems.

Generating an algorithm of automatic confirmation of connotation words will undoubtedly lead others to develop better ones, because this is about meaning of words; a computer is being forced to understand a meaning of a word. Therefore, it is pretty obvious that other researchers will benefit this algorithm. What is more, this is a big step to manage to generate a Dictionary of Turkish Connotation Words.

5. Acknowledgements

References

[1] Gee. J. P., Learning by Design: good video games as learning machines, E-Learning 2(1), pp. 5-16, (2005).

[2] Gee, J., P., Morgridge, T., Being a Lion and Being a Soldier - Learning and Games, University of Wisconsin-Madison, http://www.jamespaulgee.com/publications (2007).

[3] Yağız, E., Oyun-tabanlı Öğrenme Ortamlarının İlköğretim Öğrencilerinin Bilgisayar Dersindeki Başarıları ve Öz-yeterlik Algıları Üzerine Etkileri, (Yüksek Lisans Tezi), Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar ve Öğretim Teknojileri Eğitimi Anabilim Dalı Ankara, (2007)

[4] Facer, K., Joiner, R., Stanton, D., Reid, R., Hull, R., Kirk, D. Savannah: Mobile Gaming and Learning, Journal of Computer Assisted Learning, 20(6), pp. 399-409, (2004).

[5] Sommerville, I. Software Engineering, Addison-Wesley Press Pearson Education Inc., ISBN: 978-0137035151, Massachusetts, (2011).

[6] Kroeger, P. Analyzing Grammar: An Introduction. Cambridge: Cambridge University Press, ISBN 978-0521016537, (2005).