

Conceptual analysis studies in Turkish tests and suggestions to solve the computer network terms lexicon

Yesim Aktas *, Department of Computer Engineering, Suleyman Demirel University, 32260, Isparta, Turkey
Abdulkadir Cakir, Department of Electric-Electronic Engineering, Suleyman Demirel University, 32260, Isparta, Turkey

Suggested Citation

Aktas, Y. & Cakir, A. (2017). Conceptual analysis studies in Turkish tests and suggestions to solve the computer network terms lexicon. *Global Journal of Information Technology: Emerging Technologies*. 7(1), 175-180

Received November 19,2016; revised January 23, 2017; accepted April 20, 2017.

Selection and peer review under responsibility of Prof. Dr. Dogan Ibrahim, Near East University, Cyprus.

©2016 SciencePark Research, Organization & Counseling. All rights reserved.

Abstract

Conceptual inference on texts with natural language processing is one of the most important topics of artificial intelligence. The meaning extraction process with artificial intelligence algorithms is frequently done in foreign languages, and in recent years studies for Turkish have been increasing. In this study, concept-meaning extraction studies made for Turkish have been researched and extraction algorithms which can be done on the computer network concept dictionary have been investigated.

Keywords: artificial intelligence, wordnet, algorithms, dictionary, network.

* ADDRESS FOR CORRESPONDENCE: **Yesim Aktas**, Department of Computer Engineering, Suleyman Demirel University, Isparta, Turkey. E-mail address: yes07m@gmail.com

1. Introduction

Natural language processing (NLP) is one of the lower branches of artificial intelligence, and language works with areas such as shape and meaning. Together with developing information projects, work in the field of NDP is increasing rapidly. The increase of the studies made possible the meaning relations in the language features by calculating them with concept relations.

When the language structure is resolved and the semantic properties are removed, the largest conceptual dictionary created is Wordnet (Miller, Beckwith, Fellbaum, Gross & Miller, 1993). Wordnet was tried to be associated many times after manual creation, but ultimately the first experiments of automatic association were done by Hearts (Hearts, 1998). Wordnet was originally developed for the English language. As usage increases, it has begun to be developed for various languages (Vossen, 1998). The Balkanet project has been developed as a Wordnet subproject with Balkan languages and Turkish (Bilgin, Cetinoglu & Oflazer, 2004).

This study focuses on two main reasons. Firstly; Studies conducted for Turkish Wordnet and related texts have been examined. Studies on Turkish sound structure and automation algorithms have been investigated. The second study aims to present concepts on the works that can be used on ontological (conceptual) dictionary created in another study by extracting concepts from Turkish Wordnet, Turkish Informatics Society, Turkish Language Association dictionary and computer network books. Turkish has the ability to create new vocabulary along with every addition it adds (Eryigit & Oflazer, 2006). In studies conducted for the Turkish language (Oflazer, Say, Hakkani-Tur & Tur, 2003; Oflazer, 2003; Hakkani-Tur, Tur & Oflazer, 2002), the words are grouped into shooting groups. The study aims to separate the words on the quoted text and to work on their roots.

2. Studies

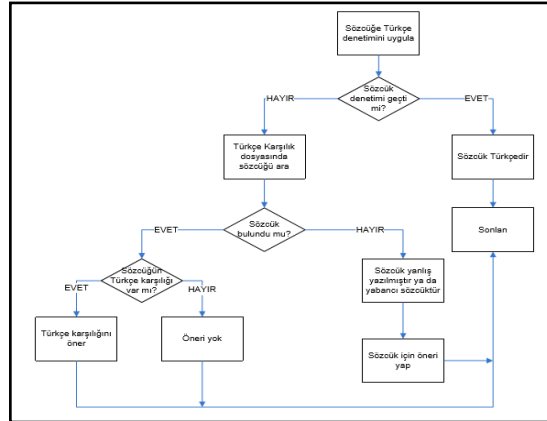
2.1. Wordnet Anaysis Studies

One of the works done for Turkish Wordnet is the work done by using automatic translation and template methods. WordNet for translation has been translated into Turkish. As a template method; One template for each relationship between the concepts was created. Whether the words around the templates belonged to that class were examined; The samples obtained are divided into two classes (Amasyali, 2011). In his work Gungor worked with dictionary based algorithm and achieved successful results. A single feature of the Wordnet dictionary has not been exploited, many features have been looked at. Most of the algorithms developed are; Is an algorithm developed by removing the concept from irrelevant words in the definition citation (Aydin, Erkan, Gungor & Takci, 2014).

In a study prepared considering the conceptual relations of the Turkic people, the words are divided into additions and roots. The endings of the roots, endless numbers and ordering, have been worked on in a certain part, rather than working in a large dictionary, making it difficult to use. After separating the words, their suffixes and their roots, the suffixes formed additional pairs among themselves. Since there is more than one sequential sound rule, a basic statistical based study is done. For error correction, the additions and roots of the kelmen were scanned in their resource pools and replaced by replacing the roots with the appropriate resource set. In this context; Fuzzy logic simulations, misconceptions made in the past, and encounter rates encountered during resource training (Dilsiz, 2005). In another surveillance exercise, firstly it was checked whether or not heel was spelled. Non-spoofable words are not accepted in Turkish and added to the dictionary to be created by the user. Ignorable words are resolved according to the generated hyphenation algorithm. This analysis is based on the Turkish sound rules, and at least one word that does not fit is accepted as a Turkic word entered from foreign languages. In the dictionary, Turkish equivalents are suggested for

words with corresponding Turkish equivalents. Words that cannot be analyzed are not in Turkish or they are accepted as erroneous and the suggestions are listed. The flow diagram of the word suggestion algorithm of the work done in Figure 1 is shown (Delibas, 2008).

Figure 1. Word proposition algorithm flow diagram (Delibas, 2008).



2.2. Computer Network Terms And Resolution Recommendations

In a previous study, the merging of the knowledge dictionary was done in the prepared TXT file; Turkish Wordnet, Turkish Language Association TDK Synonyms dictionary Turk Dil Kurumu Ana Sayfası(2015) Turkish Informatics Society information dictionary (Bilisim Sozlugu, 2014). Es Anlamyakın Arama Sozlugu(mythes-tr,2014) and the created dictionary were used. Because the generated dictionary is ontological base, the use of Turkish concept mining and analysis algorithms in their projects can increase the success (Aktas, Yilmaz, Cakir & Kutlu, 2016). In Figure 2, used dictionaries and word counts are shown.

Figure 2. Dictionaries and word counts used

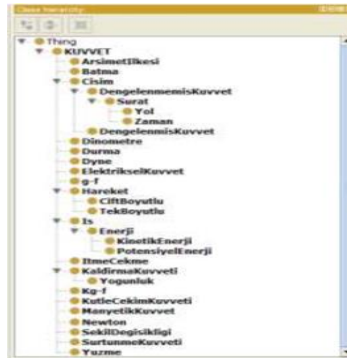
<u>Sözlük</u>	<u>Kelime Sayısı</u>
Wordnet	14.796
TDK Eş Anlamlar Sözlüğü	125.009
Eş Anlam-Yakın Anlam Sözlüğü	82.938
TBD Bilişim Sözlüğü	Yaklaşık 12.000
Kendi Oluşturduğumuz Sözlük	210.001

The dictionary structure, unlike the other dictionaries, consists not only of synonyms or semantics, but also of multiple situations. With the word itself; Synonym, side meaning, opposite meaning, definition tags. The implication of the side of Keliman can be explained as follows; For example, "network" is synonymous with "network", and "unix" is connected to "network" with a meaning. When you get out of the created dictionary; The dictionary can be analyzed using ontological analysis algorithms. One of the analysis works for the Turkish Language Institution is to look at the 30-word concept sequence for deriving meaning of the word. A set of contexts has been removed from the words on the right and 15 left of Kelimenin. For example, in the case of the word "RAM" in the text, it is deduced that the text is related to "computer" (Aydin, Erkan, Gungor & Takci, 2014). This can be the

main theme of one of the concepts, sub-concepts, including a top branch of the tree, by going out of the way, checking the concept's synonyms and close meaning labels. With the concealed semantic indexing method (GAD), more meaningful results were obtained than the GAD method, which was formed from the terms of 2-gram and 3-gram letters with n gram supported GAD study. Generated terms are words accepted and meaningful associations are established (Güven, 2007).

In Hung's work, he used additional semantic category information in the SOM (self-organizing map) model to improve field clustering performance. Three new vector presentation approaches have been proposed, namely the extended meaningful vector model, the hypernymic significant vector model and the hybrid vector space model. It has been seen that clustering performance has been improved

Figure 3. Class hierarchy of ontology-based concept map (Gultepe & Memis, 2014)



by removing symbolic information from Wordnet ontology with the hypernyms significant vector model emerging from these three models. Information from an ontology, such as WordNet ontology, has been shown to improve SOM clustering performance (Hung & Wermtner, 2004). The concept can be deduced by studying the relationship between the concept and the concept map with the semantic tag.

In Gultepe study, course planning, implementation and planning concept maps were created ontology basis. For this, Protégé used the ontology development editorial. The graphical interface of this ontology, 6-7-8. He visualized the concept maps of class mechanical topics(Gultepe & Memis, 2014). The display on the ontology-based concept map of the ordering of these classes is shown in Figure 3. In the study which was tested on ODTU-Sabancı Wooden Structure, statistical results were obtained based on decision support machines as a classifier and double dependency probabilities in loyalty analysis for Turkish (Eryigit, Adali & Oflazer, 2006).

and the hybrid vector space model. It has been seen that clustering performance has been improved by removing symbolic information from Wordnet ontology with the hypernyms significant vector model emerging from these three models. Information from an ontology, such as WordNet ontology, has been shown to improve SOM clustering performance (Hung & Wermter, 2004). The concept can be deduced by studying the relationship between the concept and the concept map with the semantic tag.

In Gultepe study, course planning, implementation and planning concept maps were created ontology basis. For this, Protégé used the ontology development editor. The graphical interface of this ontology, 6-7-8. He visualized the concept maps of class mechanical topics (Gultepe & Memis, 2014). The display on the ontology-based concept map of the ordering of these classes is shown in Figure 3. In the study which was tested on ODTU-Sabancı Wooden Structure, statistical results were obtained based on decision support machines as a classifier and double dependency probabilities in loyalty analysis for Turkish (Eryigit, Adali & Oflazer, 2006).

Results and Recommendations

In this study, the meaning and concept extraction studies on Turkish Wordnet and conceptually related dictionaries were examined.

Firstly, considering the studies, it is seen that the algorithms constructed in the field of NLD are basically the same as the conceptual inference of synonyms and semantic labels and that the definition label is effectively used when used as a side label.

Secondly, two studies which can be used for concept inferences have been investigated using the ontological dictionary, which has the largest dictionary feature made in the previously created Turkish language.

Future work will be done with artificial intelligence algorithms on concept inference for computer networks, and the success of existing algorithms will be examined.

Acknowledgments

The authors are grateful to BAP, who provided financial support with the project numbered 4534-YL1-15 at the time of the work.

References

- Aktas, Y., Yilmaz, I. E., Cakir, A., & Kutlu, A. (2016). Wordnet ve Bilgisayar Ağ Terimleri Sözlüğünün Oluşturulması, Akademik Bilisim 16, Adnan Menderes Üniversitesi, Aydın.
- Amasyali, M., F. (2011). Türkçe Wordnet'in Otomatik Oluşturulması, Bilgisayar Mühendisliği, Yıldız Teknik Üniversitesi, İstanbul.
- Aydin, C.R., Erkan, A., Gungor, T., & Takci, H. (2014). Sözlük Kullanarak Türkçe için Kavram Madenciliği Metotları Geliştirme, Akademik Bilisim, Mersin.
- Aydin, C.R., Erkan, A., Gungor, T., & Takci, H. (2014). Sözlük Kullanarak Türkçe için Kavram Madenciliği Metotları Geliştirme, Akademik Bilisim.
- Bilgin, O., Cetinoglu, O., & Oflazer, K. (2004). Building a WordNet for Turkish. *Romanian Journal of Information Science and Technology*, 7(1-2), 163-172
- Bilisim Sözlüğü. (2014). Retrieved from; <http://www.tbd.org.tr/index.php?sayfa=sozluk> Türkiye Bilisim Derneği.
- Delibas, A. (2008). Doğal Dil İşleme ile Türkçe Yazım Hatalarının Denetlenmesi (Yüksek Lisans Tezi), İstanbul Teknik Üniversitesi

- Aktas, Y. & Cakir, A. (2017). Conceptual analysis studies in Turkish tests and suggestions to solve the computer network terms lexicon. *Global Journal of Information Technology: Emerging Technologies*, 7(1), 175-180
- Dilsiz, S. (2005). *Bulanık Mantık ve Yapay Sinir Aglari Ile Turkce Yazim Denetleyicisi*(Yuksek Lisans Tezi), İstanbul Teknik Üniversitesi.
- Eryigit, G., & Oflazer, K. (2006). Statistical dependency parsing of Turkish. In: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, Italya.
- Eryigit, G., Adali, E., & Oflazer, K. (2006). Turkce Cumlelerin Kural Tabanlı Baglilik Analizi. Retrieved from; <http://research.sabanciuniv.edu/1210/1/301180000624.pdf>
- Gultepe, Y., & Memis, E. K. (2014). Kavram Haritalarının Ontoloji Tabanlı Olusturulması: Kuvvet Konusu Uygulama Ornegi, *Journal of Instructional Technologies & Teacher Education*, 24-33, 2014.
- Güven, A. (2007). *Türkçe Belgelerin Anlam Tabanlı Yöntemlerle Madenciligi*(Doktora Tezi), Yıldız Teknik Üniversitesi, İstanbul.
- Hakkani-Tur, D.Z., Tur, G., & Oflazer, K. (2002). Statistical morphological disambiguation for agglutinative languages. *Computers and the Humanities*, 36(4), 381—410
- Hearts, M. A. (1998). *Automated discovery of Wordnet Relations*, In C. Fellbaum, editor, *Wordnet: an Electronic Lexical Database*, MIT Press.
- Hung, C., & Wermter, S. (2004). Neural Network-based Document Clustering Using WordNet Ontologies, *International Journal of Hybrid Intelligent Systems*.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1993). Introduction to WordNet: An On-line Lexical Database.
- mythes-tr (2014). Turkce Es Anlamlar Sozlugu. Retrieved from; <https://github.com/maidis/mythes-tr> .
- Oflazer, K. (2003). Dependency parsing with an extended finite-state approach. *Computational Linguistics* 29(4).
- Oflazer, K., Say, B., Hakkani-Tur, D.Z., & Tur, G. (2003). Building a Turkish treebank. In Abeille, A., ed.: *Building and Exploiting Syntactically-annotated Corpora*. Kluwer Academic Publishers.
- Türk Dil Kurumu Ana Sayfası(2015). Retrieved from; <http://www.tdk.gov.tr> , Türk Dil Kurumu
- Vossen, P. (1998). EuroWordNet: A Multilingual Database with Lexical Semantic Networks. *Computational Linguistics*, 25, 427–430