ISSN:1306-3111
e-Journal of New World Sciences Academy
2010, Volume: 5, Number: 4, Article Number: 1A0118

ENGINEERING SCIENCES
Received: August 2010
Accepted: October 2010
Series : 1A
ISSN : 1308-7231

Ünal Çakıroğlu
Karadeniz Technical University
cakiroglu@ktu.edu.tr
Trabzon-Turkey

DATA MODELLING BY USING SEMANTIC NETWORKS AND FRAMES

ABSTRACT

Computer understanding of texts in Natural Language consists of morphological, syntactic and semantic analysis phases. In this study understanding and solving arithmetic problems is taken a base study in Natural Language Processing. Data is modelled as semantic networks and problem texts are shown as frames in semantic analysis. This model is tested by students' problem solving strategies with a correspondence tool and it is seen that the performance of the system is achieved over accuracy 90% in Turkish Language.

Keywords: Natural Language Processing, Semantic Networks, Frames, Data Modeling, Arithmetic Problems

ANLAMSAL AĞLAR VE ÇERÇEVELER İLE VERİ MODELLEME

ÖZET

Doğal Dilin bilgisayar tarafından anlaşılması, morfolojik, sözdizimsel ve anlamsal analiz aşamalarını kapsar. Bu çalışmada aritmetik problemleri anlama ve çözme üzerinde durulmaktadır. Çalışmada very anlamsal ağ biçiminde modellenerek problem ifadeleri anlamsal ağ yapısıyla çerçeveler şeklinde gösterilmiştir. Uygulanan modelin performansı ilköğretim öğrencileriyle aynı sorulara verilen cevaplarla karşılaştırıldığında Türkçe metinler için sistemin %90 'ın üzerinde başarı sağladığı görülmüştür.

Anahtar Kelimeler: Doğal Dil İşleme, Anlamsal Ağlar, Çerçeveler, Veri Modelleme, Aritmetik Problemler

1. INTRODUCTION (GİRİŞ)

Natural language processing (NLP) is the engineering of systems that investigates written or spoken natural language. A natural language processor understands human language. "Understanding" in this context means the computer can accept human language input from the keyboard and then perform the required computing tasks. Human languages are complex and incorporate many features that produce ambiguity: different meanings for the same words and different meanings produced by sentence structures, idioms and metaphors. Many problems must be solved in transforming them for use by a computer. So it is not easy writing a program of computer understanding [1,2]. For example; In Turkish in the sentence "Patron işçileri sürdü." (The boss banished the employee.) The verb "sürmek" means (to banish) or (to spread something on somewhere for sticking) in Turkish.

One of the criteria about meaning extraction is the test that is answering the questions and giving reasons for answers according to a knowledgebase [1,2]. This kind of explaining capability is a feature that separates real expert systems and the others. If the linguistic analysis is achieved exactly by the computers, many useful tools can be designed. Some of them are listed below:

Automatic translation of written texts, designing of question-answer machines, automatic talking and command understanding systems, speech synthesis, speech generation, text summarizing, problem solving, presenting ideas and alternative solutions to problems.

There are some systems for problem solving in English and in other linguistics. In Weizenbaum's ELIZA, the syntactic and semantic are used together in human-computer dialogue. In Winograd's SHRDLU, data is stored as semantic networks, frames or scripts. Also in Cullingford's SAM Mechanism is an example for computer language understanding systems. In 1965 Daniel Bobrow designed the first rule-based system STUDENT [3]. This system can solve primary school algebraic problems. "Bill's father's uncle is twice aged as Bill's father. After two years Bill's father's age will be three times of Bill's age. Now, all of their ages are totally 92. Find Bill's age?" is an example that STUDENT can solve. These kinds of systems could find the correct answers by context free approach. In this approach the concepts in sentences are not known by the system. System only solves the problem by the given rules. There is no system that models whole concepts in sentences. In order to design more functioning understanding systems, all the concepts in the sentences must be identified to the systems [5]. It is clear that, it will take more time for modeling whole language.

In this study, a semantic analyzer program for solving arithmetical problems in Turkish is generated. In semantic analysis of the system data is stored as semantic networks and the semantic effect of morphological structures are taken in to consideration. After examining primary school 1st, 2nd and 3rd grade school mathematics books, it is seen that most of the problems can be solved by using the basic four arithmetic operations; addition, subtraction, multiplication and division. This study includes generating and applying algorithms for solving these kind of problems. The rest of the paper is about determining the performance of the algorithms.

2. RESEARCH SIGNIFICANCE (ÇALIŞMANIN ÖNEMİ)

According to achieve the understanding and solving process, system has to deal with all steps of NLP studies; morphology, syntax and semantic [6]. In detail these steps consist of morphological analysis of words, syntactic analysis of the sentences, constructing a knowledgebase using data gained from previous phases, formulization of meaningful data, solving equations related to the formulas and presenting the solutions.

2.1. Morphological Analysis (Morfolojik Analiz)

Morphological analysis recognizes punctuations, possessives, proper names, short forms, words, roots and suffixes. Turkish has a rich morphological structure, so our program has to carry out a morphological analysis of each word in the input before proceeding with the syntactic analysis [5]. This is the first step of NLP tasks consist of analyzing roots, afixes and suffixes of all words. In NLP studies the main aims of morphological analysis are:

- Determining the types of words: In Turkish, word types may be in the form of noun, adjective, verb, pronoun, particle or others. The meaning sets of word sets are generally defined by language grammar [4,10]. So, determining the word types can help to obtain some hints about the meaning sets of the words in sentences.
- Searching affixes and suffixes of the words: In this process, word is divided into affix morphemes that are designated and accepted by the linguists. So that, determining affixes and suffixes and roots can help putting words in the correct meaning sets.
- Determining affix types: In Turkish, sometimes different word types can be formed by the same affixes. Ex: In the sentences "Armudu yedim." and "Ali'nin armudu …" the suffix (-u) has different duties. In first sentence its duty is (suffix of object), in the second one the duty is its duty is (determinative suffix). According to separate these kind of differences, it is necessary to search for suffix types.

The morphological phase of the system for given problem texts is shown below:

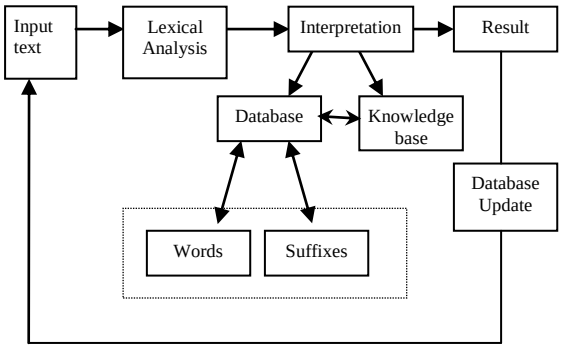

Figure 1. Morphological analysis level
(Şekil 1. Morfolojik analiz düzeyi)

Here, the most commonly used Turkish words and suffixes are included in the system's database. There are three steps in morphological analysis. First the root of input word is determined, then morphological tests are realized and the morphemes of the input word are determined [2,6]. In lexical analysis step, the spelling errors are found and corrected. Then by analyzing the whole sentences, after corresponding roots and suffixes; the missing roots and suffixes are presented to the user. After users define, the root and suffix database is updated. After this module the updated sentence is back propagated for new lexical analysis. By the module "Result" the morphological analysis is finished and the result of morphological phase is passed to the syntactic analysis phase in the form of one for each sentence of one for each words. This data is valuable for understanding of problem texts by semantic analysis. Ex: The problem; "Bir bakkal (da) 500 kg. şeker vardır. Bakkal (a) 100 kg. daha şeker geldi. Hasan bakkal (dan) 5 kg. şeker aldı. Bakkal (da) kaç kg. şeker kaldı. " consists of the word "bakkal (shop)" and the suffixes joined to this root.

Here the suffix (-da) in (bakkalda) has a meaning of "to be present" , in the word (bakkala) , the suffix (-a) adds the meaning of (to the shop), the suffix (-dan) in the word (bakkaldan) means (from the shop) and adds the meaning of "leaving somewhere". After this phase an example morphological analysis of a problem "Ahmet'te kaç elma oldu?" (How many apples did Ahmet have?) text is given below:

"Ahmet'te" → Ahmet: noun(root) ; -te : locative
"kaç "" → adjective (root)
"elma" → noun (root)
"oldu" → ol :verb (root); -du : tense (past)

In this system, the data about the types of words and the duties of suffixes related the meanings are listed in table1.

Table 1. Morphological analysis of example text
(Tablo 1. Örnek bir metnin morfolojik analizi)

Word	Word Type	Suffix Type
Bir	adjective	
Bakkal -da	noun	Locative (LS)
500	adjective	
Kg	noun	
Şeker	noun	
Var -dır	verb	Inform suffix (IS)
Toptan –cı	noun	Suffix makes noun from noun (NNS)
Bakkal –a	noun	Locative (LS)
Getir -di	verb	Past Tense (PTS)
Hasan	noun	
Al –ır -sa	verb	Present Tense (PS), Condition Tense (CTS)
Bakkal -da	noun	Locative (LS)
Kal -ır	verb	Present Tense (PS)

2.2. Syntactic Analysis (Sözdizimsel Analiz)

Syntactic analysis is comparing the syntactic or morphological units of sentences with the hierarchical syntax rules. By this way, it is tested if the sentence is suitable for the language. This process is useful for eliminating the meaningless correspondence in semantic analysis phase. The key observation in the theory of syntax is that the words in a sentence can be more or less naturally grouped into what are called what are called "phrases", and those phrases can often be treated as a unit. Phrase structure trees are often used to represent the configuration of sentences. These can show how the structural elements are related, and the relations among nodes in the tree can be used to describe constraints that have to hold. One approach to characterizing syntactic structure involves giving rules to describe how phrases can be generated. For example here are some such rules:

S → NP VP
NP → Det {Adj} Noun
VP → Verb {NP} {PP}
PP → Prep NP

As sentences are finite units, languages have finite sets of sentence variants. So the researchers generally use Finite State Machines (FSM) for identifying sentences. Determining the sentence units and their tasks presents valuable hints when the addition of morphological analysis is evaluated. In Turkish semantic, the considered unit of the sentence is located close to the verb of the sentence [5]. Therefore the subject of the sentences is located at the end of the sentence for getting better performance in semantic analysis.

2.3. Semantic Analysis (Anlamsal Analiz)

In the theory of "conceptual dependency" the claim is that the relations among complex events by composing them out of more simple events. These observations lead to the theory of case frames. A case frame is a representation of an action or event, along with its participants. The reason they are called case frames, nouns are assigned case depending on the role that the referent of the noun phrase plays in the sentence. The idea of case frames is that each verb is associated with a specific case frame, and a set of "role mappings" which indicate how the syntactic arguments of the sentence are assigned to the participant slots in the case frame. Here are some typical slots in case frames: agent, object, location, source, goal, beneficiary.

For example the verb "buy" might be associated with a "purchase" case frame with a buyer and seller and an thing bought. So we will assume that it uses the "source" slot for the seller, the "goal" slot for the buyer, and the "object" slot for the thing bought [11].

3. METHOD (YÖNTEM)

In this system's semantic analysis phase the knowledgebase is transformed in to an interpretable form. The positions of meaningful words and suffixes are determined and the relations between these units are identified in this module. These relations should be used in meaning extraction or solving problems. It is impossible to store all words and all other units of sentences in the system database. So in this study, the necessary data is stored as semantic networks. A program is developed for constructing this network. This program forms a knowledgebase that can be assigned the main part of the system. This knowledgebase is a union of objects (words) and the types of relationships between the objects. This knowledgebase can be updated by the system, it self. So, limited vocabularies are used in system database. Some problem frames are constructed and used for obtaining this knowledgebase.

Here, first we searched the primary school 1st, 2nd, 3rd class mathematics books in Turkish schools. Then the problem texts are grouped as addition, subtraction, division, multiplication, etc. These groups are:

- **1. Type:** (X+Y=F) and (X-Y=F). Here the amount of one object is increasing or decreasing.

"Bir çiftlikte 25 inek vardır. Çiftliğe 4 inek daha geldi. Kaç inek oldu? " (There are 25 cows in a farm. 4 cows are brought to the farm. How many cows are there in the farm now?)

"Bir iş yerinin mevcudu kırk kişidir. 3 kişi işi bırakırsa iş yerinde kaç kişi kalır?"(The population of a company is forty. If 3 employees leave the company, how many people will be in the company?)

- **2. Type:** (X1+Y1=F1), (X2+Y1=F2). In this type of problems amount of two objescts are effected.

"Ahmet'in 15 balonu vardır. Mehmet'in 5 balonu vardır. Ahmet 4 balonunu Mehmet'e verirse, Mehmet'in kaç balonu olur? "(Ahmet has 15 balloons and Mehmet has got 5. How many balloons does Ahmet have if Mehmet give 4 of his balloons to Ahmet?)

- **3. Type:** (X+Y1+Y2+Y3+…Yn=F), (X-Y1-Y2-Y3-…Yn=F). In this type an object is effected more than one object.

— "Metin'in çantasında 8 kalem vardır. 4 kalemi Esra'ya, 2 kalemi Ali'ye verirse Metin'in kaç kalemi kalır? " (There are 8 pencils in Metin's case. If he gives 4 pencils to Esra and 2 pencils to Ali, how many pencils remains in Metin's case?)

- **4.Type:** (X*Y=F). These are problems that need multiplication operations.

— "Bir bakkalda bir günde 50 ekmek satılıyorsa, 8 günde kaç ekmek satılır?" (If 50 breads are sold in a day, how many breads can be sold in 8 days in a shop?)

- **5. Type:** (X/Y=F). This type consists of problems that can be solved by division operation.

— "Bir baba 50$ parasını 5 çocuğuna bölüştürdü. Bir çocuğun ne kadar parası oldu?" (Father divided his money into his 5 sons equally, how much money does a child have?)

In all types of problems, many kinds of problems can be obtained by changing the "given" and "wanted" parts of the problems. It is seen in mathematics books that; there are many alternatives of problem texts that can be solved by multiplication and division. One of the disadvantages is; this kind of problems consist different structures and different kinds of word sets. By the way many of the word sets must be identified to the system for solving multiplicative and division problems. In "Related Units Analysis" phase, the words, suffixes and roots are used to construct a base for semantic network. Semantic networks are the main part of the system. The relations are stored as networks here. A semantic network is union o objects and the relations between objects by interaction with knowledgebase. In this system problem texts are divided into small meaningful parts and semantic networks are constructed.

According to construct network, all the examined objects and relations are noticed, th "Frame Based Relation Table (FBRT)" realized. A part of FBRT is shown in Table2.

Table 2. A Part Of Fbrt
(Tablo2. Fbrt ye ilişkin bir bölüm)

Word Group	Related Units	Relation	Frame
Var Sahip olmak Elde olmak Mevcut olmak Kalmak …	Subject-Object	has	A subject has an object at the beginning
Almak Eklemek Artmak İlave Toplamak Katmak …	Subject-Object	increase	Objects of subject are increase
Azalmak Eksilmek Satmak Yemek Vermek Silmek….	Subject-Object	decrease	Objects of subject are increase
Olmak Bulunmak	Subject-Object	to be	A subject has an object at the end
…	…	…	

In order to determine the addition effect of morphology, another knowledgebase called Morphological Analysis Knowledgebase (MAKB) is generated by using morphemes and their relations. A part of MAKB is shown in Table3.

Table 3.A Part Of Makb
(Tablo 3. Makb ye ilişkin bir bölüm)

Suffix/word	Duty	Relation
-de	Locative	has
-e	Locative	increase
-den	Locative	decrease
-se	conjunction	Active action
-nin	determinitive	has
daha	stiffer	increase
…	…	…

The MAKB is used to support relations of FBRT for constructing semantic network of a problem. According to this operation the sentence "Bir bahçede 50 ağaç var. Bahçeye 5 ağaç dikildi. Bahçede kaç ağaç oldu? (There are 50 trees in a garden. If 5 are planted, how many trees will be in the garden?) " has a network shown below:

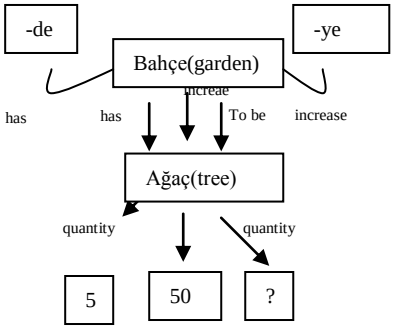

Figure 2. A semantic network for selected problem
(Şekil 2. Seçilen bir problem için anlamsal ağ)

In the problem solving phase, all the wanted and given data classified in the problem types (equations) about the selected problem. Then the answer of the problem is communicated to the user easily comprehensible fashion. Also if there are logical mistakes in the input problem texts, system presents these and addresses the location of mistakes. Example:
Input Text:.Davut'un 15 oyuncağı var. Hasan'ın 5 meyvesi vardır. Toplam kaç armut vardır? (Davut has 15 toys. Hasan has 5 fruits. How many pears are there?)
Program Message: "'toys' and 'fruits are not same kind of objects, can not do operation!'"

4. FINDINGS AND DISCUSSIONS (BULGULAR VE TARTIŞMALAR)

According to test the program, another part of a study is realized in a Turkish primary school 1st, 2nd and 3rd grade students who are 6,7,8 years old. For this study every kinds of problem texts are tried to be selected for comparing all kinds of texts. Here are some examples about the problems.

- "Ali 5 elma daha alırsa 9 elması olur. Ali'de kaç elma vardır?" (If Ali takes 5 apples, the number of his apples become 9. How many apples does Ali have?)
- "Bir öğrenci günde 5 sayfa kitap okuyor. 10 günde kaç sayfa okur?" (A student reads 5 pages of a book in a day. How many pages does he read in 10days?)

- "Bir çiftlikte 200 tavuk mevcuttur. Otuz beş tavuk ölürse kaç kalır?" (There are 200 chicken in a farm. If thirtyfive die, what will be the chicken population?)
- "Bir kırtasiyeci yüz yirmi iki kalem alırsa beş yüz kalemi oluyor. Kaç kalemi vardır?" (If a retailer buys one hundred and twenty two pencils, he will totally have five hundred pencils. How many pencils did the retailer have initially?)

According to selected problems, the primary school students and the system performance obtained by calculating the average correct answers, through 20 problems.

Table 3. Students and system comparison due to selected problems
(Tablo3. Seçilen problemler için öğrenciler ve sistemin karşılaştırılması)

1.grade	2.grade	3.grade	system
14	17	20	18

The performance of the system is higher than 1st, and 2nd grade students, and it is close to the 3rd grades for selected 20 problems. The system performance, in addition and subtraction problems if 90%. These kinds of problems are classified in X+Y=F, X-Y= F and other variances of addition and subtraction groups that the reasons are explained above. The system performance is 61% in multiplication and division problems through selected 400 problems.

5. CONCLUSIONS AND RECOMMENDADITONS (SONUÇLAR VE ÖNERİLER)

In this paper an implementation of a problem solver system is presented. This method describes how to use semantic networks to store data. And how to obtain relations by semantic and frame structures of the language. The main goal is understanding and solving the artimetic problems in mathematics books. The words and relationships in knowledgebase is limited. In future work implementation of knowledge system would be employed to maximize the words and relationships. The various phases of the program are developed quite independent each other. So they can be adapted to the similar NLP studies. These kinds of systems are prototypes for modeling of understanding. The statistics that are made for the sentence structure and word orderings in Turkish will provide good conclusions for a general understanding model. This system can help solving arithmetical problems and also learning Turkish grammatical rules. Natural language is ambigious by nature. A simple sentence can be interpreted in many different ways. In order for the computer to process the sentence, it needs to know the exact meaning of the sentence. Thus, the logical form is needed as an intermediate unambigious representation of the meaning of the sentence. The results obtained in this work can be used in research which aims to presenting reasons, to offer alternatives by semantic networks.

REFERENCES (KAYNAKLAR)

1. Woods, W.A., (1970). Transition network grammars for natural language analysis, Communications of the ACM, vol. 13, 591-606,
2. Lee, M.C. and Leong, R.V., (1992). NLUS – A Prolog Based Natural Language Understanding System, In Proceeding of the 4th International Conference on Computing and Information, pp. 204-207
3. James, A., (1995). Natural Language Understanding (second edition), The Benjamin / Cummings Publishing Company, Inc.
4. Banguoğlu, T., (1990). Türkçe'nin Grameri, Türk Tarih Kurumu Basımevi, Ankara.
5. Pakkan, M. and Say, C., (1994) "Qualitative collision detection in the roller world" , Proc. Ninth Turkish Symposium Computer and Information Sciences, Antalya, pp. 550-557

6. Eryigit, G. and Adali, E., (2004). An Affix Stripping Morphogical Analyser For Turkish, Proceedings of the IASTED International Conference on Artificial Intelligence and Applications, Innsbruck, Austria.
7. Budy, A., Byrd, L., Luger, G., Mellish, C., Milne, R. and Palmer, M., (1979), "MECHO: A program to solve mechanics problems" Working Paper 50, Department of Artificial Intelligence, Edinburgh University.
8. Bozşahin,C., Zeyrek, D., and Pembeci, İ., (1995), Computer-aided Learning of Turkish Morphology", Proceedings of Joint Conf. Of ACH and ALLC, Debrecen, Hungary, 1998
9. Güngördü, Z., and Oflazer, K., "Parsing Turkish using the Lexical-Functional Grammar Formalizm, Machine Translation", Vol.10, Number 4.
10. Vagifoğlu, V., (2004), "Doğal Dil İşleme", Yapay Zeka Problemler-Yöntemler-Algoritmalar, Seçkin Yayıncılık, Ankara.
11. Schank, R.C. and Abelson, R.P., (1977), Scripts, Plans, Goals, and Understanding, Lawrence Erlbaum, Hillsdale, NJ