



Düzce Üniversitesi Bilim ve Teknoloji Dergisi

Derleme Makalesi

XML Anahtar Kelimeleri Yardımıyla Türkçe Aritmetik Problemlerin Anlaşılması ve Çözülmesi

Süleyman EKEN^{a,*}, Ekin EKİNCİ^a, Ahmet SAYAR^a

^a Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Kocaeli Üniversitesi, Kocaeli, TÜRKİYE

* Sorumlu yazarın e-posta adresi: suleyman.eken@kocaeli.edu.tr

ÖZET

Doğal Dil İşleme; matematik, bilgisayar bilimi, yapay zekâ ve dilbilimi alanlarındaki çalışmaları kapsamaktadır. Yapay zekâ alanındaki çalışmaların hedeflerinden biri doğal dilde söyleneni anlamaktır. Doğal dili anlamaya yönelik birçok çalışma mevcuttur. Bu çalışmalardan bazıları matematik, fizik ve mekanik alandaki problemlerin anlaşılması ve çözülmesi ile ilgilidir. Bu çalışmada, ilkökul seviyesindeki matematik problemleri anlayabilen ve çözebilen bir aritmetik problem çözücü sistem önerilmektedir. Uygulama, kelimelerin morfolojik analizi, problem tipinin tanımlanması (toplama, çıkarma, çarpma, bölme) ve sonucun hesaplanması adımlarını içermektedir. Problem tipleri bir XML dokümanı yardımıyla belirlenmekte, bu doküman ise problem tiplerine karşılık gelen anahtar kelimeleri içermektedir. Bu uygulamanın temel katkısı, öğretmenlerin öğrencilerini daha iyi değerlendirmesi açısından kullanacakları eğitim materyallerini hazırlamalarına yardımcı olmaktadır.

Anahtar Kelimeler: Doğal dil işleme, Problem çözme, Aritmetik problemler, Yapay zeka

Understanding and Solving Turkish Arithmetic Problems via XML Keywords

ABSTRACT

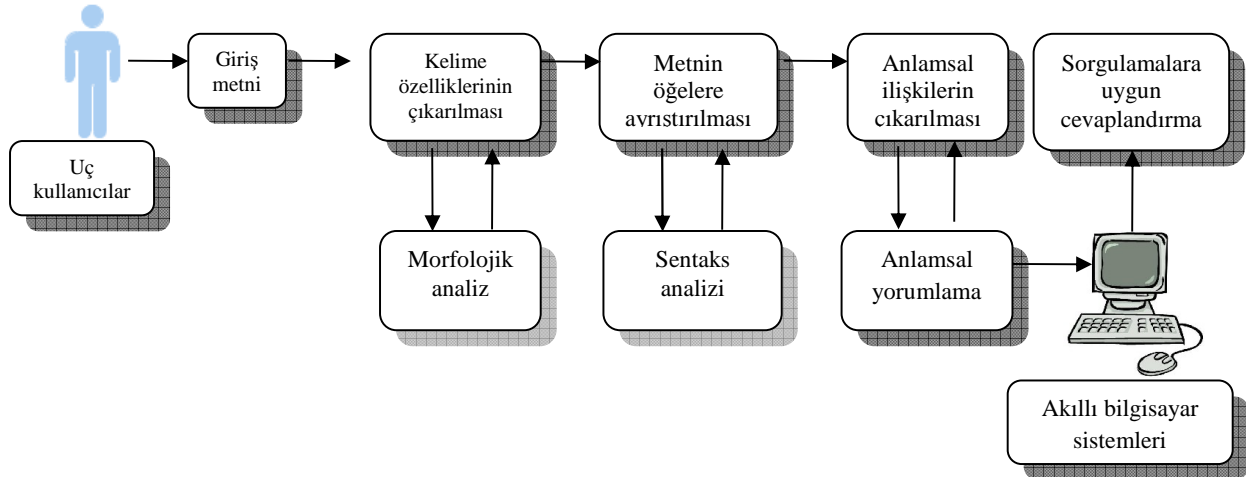
Natural Language Processing covers the studies in the areas of mathematics, computer science, artificial intelligence, and linguistics. One of the goals of the studies in artificial intelligence is to understand what is said in a natural language. There are various applications to understand natural languages. Some of these applications are related to understanding and solving the problems in mathematics, physics, and mechanics. In this paper, an arithmetic problem solver system that can interpret and solve primary school-level arithmetic problems is presented. The steps of application contain morphological analysis of words, definition of problem types (addition, subtraction, division, multiplication), and calculating the results. Problem types are determined by forming and using an XML document, this document has keywords corresponding problem types. The main contribution of this application is to help teachers in preparing their teaching materials for better evaluating their students.

Keywords: Natural Language Processing, Problem Solving, Arithmetic Problems, Artificial Intelligence

I. GİRİŞ

DOĞAL Dil İşlemenin (DDİ) temel fonksiyonlarından biri, herhangi bir dili çözümlene, anlama, yorumlama ve çözüm üretme işlemlerini gerçekleştirebilen bilgisayar modeli sistemlerinin oluşturulmasıdır. Doğal dildeki metinlerin yapısal ve anlamsal çözümlenmesini gerçekleştiren DDİ [1] 50'li yılların başında yapay zekânın küçük bir alt alanı olarak ortaya çıkmıştır. Gelişen teknolojinin de etkisiyle gerçekleştirilen çalışmalarda bu alana yönelik ihtiyaç her geçen gün artmış ve elde edilen büyük başarılar sonucunda DDİ bilgisayar bilimlerinde başlı başına bir disiplin haline gelmiştir. Nabiyev'in [2] de belirttiği gibi DDİ alanındaki temel araştırmalar; doğal dillerin fonksiyon ve yapısının daha iyi anlaşılmasını, bilgisayarla insanların arasındaki iletişim için köprünün kurulmasının sağlanmasını ve bilgisayarla dil çeviri sistemlerinin gerçekleştirilmesini kapsamaktadır.

Doğal dilde verilmiş bir metnin bilgisayarda modellenip analiz edilmesi metnin sistemin yorumlayabileceği şekle dönüştürülmesi demektir. Bu yorumlama işlemini gerçekleştiren dil çözümleyicisinin (Fig. 1'e bakınız.) bir takım fonksiyonları vardır: (i) doğal dilde ifade edilen metnin doğruluğunun incelenmesi, (ii) muhtemel imla hatalarının tespiti ve düzeltilmesi, (iii) kelime özelliklerinin çıkarılması-morfolojik analiz , (iv) metnin/cümlenin öğelere ayrıştırılması, (v) anlamsal ilişkilerinin çıkarılması, (vi) problemi ifade eden sentaks ve semantik analizlerin birleştirilerek sorgulamalara uygun cevaplandırmanın yapılması.



Şekil 1. Problem metninin modellenmesi

Doğal dili çözümlene aşamalarından ilki biçimbilimsel (morfolojik) analizdir. Bu aşamada dilde biçimi oluşturan bileşenlerin türleri (isim, sıfat, fiil, zamir, edat vb.) tanımlanır ve sınıflandırması yapılır. Ayrıca kelimedeki köklerinin ve o köklere gelen eklerin görevlerinin ayrıştırılması işlemi gerçekleştirilir. Morfolojik analiz sonucunda elde edilecek bilgiler problem metninin anlaşılması açısından oldukça önem arz etmektedir. Çalışmada bu adım, yapılacak işlemlerin, anahtar kelimelerin ve bu anahtar kelimelerin hangi şekilde saklanacaklarının (kök ya da gövde) belirlenmesi aşamasında kullanılmaktadır. Kısaca anlam taşıyan kelimelerin belirlenmesi ve sınıflandırılması için gerçekleştirilmiştir. İkinci aşamada, gramer adı verilen tanımlayıcı şekilde dilin yapısı gösterilmekte ve bu yapıya göre gramer temelli genel yapısı (ögelere ayrıştırılması) oluşturulmaktadır. Bu adımda cümlenin öğelerine ayrılmasından ziyade söz dizimsel sıra ve cümlelerin yapısı belirlenmiştir. Üçüncü aşama ise anlam taşıyan kelimelerin sınıflandırılması ve sentaks analizi işleminden sonra

gelen anlamlandırma sürecidir. Son aşamada sentaks ve semantik analizler birleştirilerek sorgulamalara uygun cevaplandırmalar yapılmaktadır.

Literatürde problem çözme ve insan- bilgisayar etkileşimine yönelik sistemler üzerine yapılmış birçok çalışma mevcuttur. Daniel Bobrow [3] tarafından geliştirilen STUDENT isimli program cebir problemlerini analiz eden ve çözen ilk programdır. İlköğretim aritmetik problemlerini İngilizce doğal dilde anlayıp denklemler kuran bir program olan STUDENT, kelime eşleştirme (pattern matching) tekniğine göre çalışmaktadır. Program her cümlenin bir denklem olduğunu varsaymaktadır. Programa önceden bazı denklemleri oluşturmasına yardımcı olacak belirli kelimeler hakkında bilgiler yüklenmiştir. STUDENT, verilen aritmetik problem içindeki “eşittir”, “fazladır”, “eksiktir” gibi kelimeleri operatör olarak tayin edip bu kelimeler etrafındaki diğer kelimeleri değişkenler olarak atamaktadır. Program bu istenilen anlamlar ile sözdizimini (sentaks) çözümlenmektedir. Eugene Charniak [4] tarafından LISP ve CONVERT dilleri ile geliştirilen CARPS (CALculus Rate Problem Solver) sistemi ise sadece oran-orantı problemlerini çözebilmektedir. Bundy ve arkadaşları [5] mekanikle ilgili özellikle makara ve hareket problemlerini başarıyla çözebilen MECHO isimli programı önermişlerdir. Verilen problem içindeki mekanik ile ilgili kelimeler (örneğin “makara”) bulunmakta ve bu bulunan kelimeler ipucu kabul edilerek bu kelimelerle ilgili diğer kelimeler de sözlükten bulunmaktadır. Cümleler denklemler halinde ifade edilmektedir. Literatürde internet tabanlı aritmetik problemlerin çözümüne yönelik çalışmalar da vardır. Bunlardan bazıları şunlardır: Huang ve arkadaşları [6] düşük başarılı ilköğretim seviyesindeki öğrencilere yönelik ünlü matematikçi Polya’nın [7] problem çözme sürecini işleten internet tabanlı bilgisayar destekli matematiksel problemleri çözen bir sistem geliştirmişlerdir. İlgili öğrencilerin temel toplama-çıkarma problemlerini çözme becerilerini geliştirmeyi ve öğrenme isteklerinin artmasını sağlamayı hedeflemişlerdir. Chang ve arkadaşları [8] MathCAL isminde; problemin anlaşılması, bir plan yapılması, planın işletilmesi ve çözümün gözden geçirilmesi süreçlerine dayanan bilgisayar destekli bir sistem önermişlerdir. 130 beşinci sınıf öğrencisi üzerinde sistemin etkinliği denenmiş ve öğrencilerin problem çözümüne yaklaşımlarını iyileştirdiği görülmüştür. Ilany ve Margolin [9] ise çözümü özet matematiksel yapılar ile dilbilimsel durumlar arasındaki geçişe dayalı matematiksel kelime problemler için bir eğitim-öğretim modeli sunmuşlardır. Türkçe aritmetik problemlerin anlaşılması ve çözümüne yönelik ise iki çalışma vardır: ALİ (Aritmetikçi Lisan İşleyici) ise Say [10] tarafından Turbo Prolog dili kullanılarak geliştirilen Türkçe aritmetik problemleri çözen ilk programdır. 23 tane yazı içeren Türkçe aritmetik problemi ilköğretim 3.sınıf Türkçe matematik kitaplarından seçerek sistemin kullanılabilirliğini test etmiştir. Çakıroğlu [11] ise çalışmasında Türkçe için bir bilgisayar anlaması modeli oluşturulmaya çalışılmış ve ilköğretim 1, 2, 3.sınıf düzeyindeki matematik problemlerin bilgisayar programı tarafından anlaşılıp çözülmesi hedeflemiştir.

İlköğretimde okutulan matematik kitapları üzerinde yapılan incelemelere göre sorular beş gruba ayrılmıştır. Bu gruplar toplama, çıkartma, bölme ve çarpma olarak ele alınmıştır. Bu gruplara ilişkin örnek problemler aşağıda sıralanmıştır:

1. Tür $(X+Y=F)$, $(X-Y=F)$: Bir nesneye yapılan ekleme veya nesne miktarından azalma söz konusu olduğu problemlerdir. "Ali'nin 10 kalem var. 4 kalem daha aldı. Kaç kalem olur?"
2. Tür $(X_1-Y_1=F_1)$, $(X_2+Y_1=F_2)$: İki nesnenin etkilendiği soru tipleridir. "Ayşe'nin 7 tokası vardır. Fatma'nın 5 tokası vardır. "Ayşe 4 tokasını Fatma'ya verirse, Fatma'nın kaç tokası olur?"
3. Tür $(X+Y_1+Y_2+...+Y_n = F)$, $(X-Y_1-Y_2-...-Y_n = F)$: Bir nesnenin birden çok nesne tarafından azalma veya artma yönüyle etkilendiği problemlerdir. "Aziz'in 5 bilyesi var. Aziz, Mehmet'ten 6 bilye, Ahmet'ten 7 bilye aldı. Aziz'in bilye sayısı kaç oldu?"

4.Tür ($X*N=F$): Çarpma işlemi gerektiren soru tipleridir. "Bir marangoz günde 3 masa yapıyor. 5 günde kaç masa yapar?"

5.Tür ($X/N=F$): Bölme işlemi gerektiren problemlerdir. "Bir adamın 25 TL mirası kalmıştır. Mirası 5 çocuğuna eşit olarak dağıtmıştır. Çocuklarda ne kadar para oldu?"

Gerçekleştirilen bu uygulamada yukarıdaki problem türlerinden sadece 1., 2., 4. ve 5. tür çözülmektedir. Makalenin geri kalanı şu şekilde organize edilmiştir: II. bölümde uygulamanın mimari hakkında bilgi verilip geliştirilen sistemin çalışabilirlik testleri yapılmıştır. Son bölümde ise sonuçlar değerlendirilip gelecekte ne gibi çalışmalar yapılacak onlardan bahsedilmiştir.

II. ARİTMETİK PROBLEM ÇÖZME MİMARİ YAPISI VE DURUM ÇALIŞMALARI

Türkçe aritmetik problemlerin bilgisayarla çözümlenmesi için geliştirilen uygulamada izlenen temel adımlar şöyledir:

- Kullanıcının girdiği soru içindeki tüm kelimelerin ayrıştırılıp liste içinde tutulması.
- Soru içindeki sayıların ayrıştırılıp dizi içinde tutulması.
- Problem türünün belirlenmesi.
- Problemin çözümü.
- Çözilemeyen problemin kullanıcı tarafından bilgisayara öğretilmesi ve ilgili kelimelerin XML olarak veri dosyasında tutulması.

A. PROBLEMDEN SAYILARIN ELDE EDİLMESİ VE PROBLEMİN ÇÖZÜLMESİ

Tasarlanan uygulama şu şekilde çalışmaktadır. İlk önce kullanıcı tarafından çözülmesi istenen problem cümlesi, kelimelere ayrıştırılarak bir kelime listesi şeklinde tutulur. Uygulama sadece yukarıda da belirtildiği üzere belirli soru kalıplarını çözümleyip sonuca varabilmektedir. Bu kalıplar en fazla üç cümleden oluşmalı, problem en fazla iki sayı içermelidir. Eğer soru kalıbı tek cümle şeklinde ise sayıların ikisini de içermelidir, iki cümle ise bir sayı ilk cümlede diğer sayı ikinci cümlede bulunmalıdır, üç cümleden oluşuyorsa da sayılar iki cümlelik soru kalıplarında olduğu gibi birinci ve ikinci cümlede bulunmalıdır. Bu sayıları elde ederken bir önceki adımda elde edilen kelime dizilerinde her bir kelimenin 0'dan 9'a kadar herhangi bir rakam içerip içermediğine bakılmaktadır. Eğer içeriyorsa o kelime bir sayıdır. Sayılar bazı ek ve kesme işareti gibi noktalama işaretleri içerebilmektedir. Bunlar sayıdan arındırılarak sade haliyle kaydedilir. Sayılar elde edildikten sonra problemin hangi gruba girdiği araştırılmalıdır. Burada soru cümlesi kelimelere ve sayılara ayrılıp toplama-çıkarma mı yoksa çarpma-bölme mi olduğuna karar verilmektedir. Bu araştırma işlemleri XML'de saklı anahtar kelimeler yardımıyla yapılmaktadır. XML'de saklı olan anahtar kelimeler kök veya gövde durumunda olduğu için eldeki soru cümlesindeki kelimelerin de kökleri ve gövdeleri bulunmalıdır. Bu işlem için ise Türk dilleri için geliştirilmiş ve açık kaynak kodlu doğal dil işleme kütüphanesi olan Zemberek'ten yararlanılmıştır.

2004 yılında geliştirilmeye başlanan Zemberek, kullanıcısı için biçimsel analizden kelime tabanlı yazım denetimine, kelime eki bulmadan hecelemeye, hatalı yazımlar için öneriden istatistiksel analize kadar pek çok temel doğal dil işleme görevlerini yerine getirmektedir [12]. Genel yapısı incelendiğinde ise Zemberek; çekirdek, dil gerçekleştirimleri ve veri olmak üzere üçe ayrılmaktadır. Çekirdekte temel doğal dil işleme algoritmaları, dil gerçekleştirimlerinde çalışma kapsamındaki Türk dilleri, veri kısmında ise harf, kök ve ek bilgileri mevcuttur [13].

Eldeki soru cümlesinde bulunan ve çarpma işlemini tanımlayan anahtar kelimeler "saniye, dakika, saat, gün, hafta, ay, yıl" gibi zaman kavramlarıdır. Eğer problemin çarpma işlemi olduğu sonucuna varılırsa çarpma işlemi gerçekleşmektedir. Aksi takdirde problem içinde farklı problem grubuna giren anahtar kelimelerin bulunmasına göre farklı problem türleri çözülmektedir. (Örnek çıktılar için Fig.2 ve Fig. 3'e bakınız) Bu anahtar kelimeler, dört işlem türü için aşağıda verilen XML dokümanında olduğu gibi gruplanmıştır.

```
<?xml version="1.0" encoding="utf-8"?>
<DATA>
<ITEM>
<AD> al</AD>
<GRUP>toplam</GRUP>
</ITEM>
<ITEM>
<AD> ekle</AD>
<GRUP>toplam</GRUP>
</ITEM>
<ITEM>
<AD> ilave</AD>
<GRUP>toplam</GRUP>
</ITEM>
<ITEM>
<AD> art</AD>
<GRUP>toplam</GRUP>
</ITEM>
<ITEM>
<AD> azal</AD>
<GRUP>fark</GRUP>
</ITEM>
<ITEM>
<AD> eksil</AD>
<GRUP>fark</GRUP>
</ITEM>
<ITEM>
<AD> sat</AD>
<GRUP>fark</GRUP>
</ITEM>
<ITEM>
<AD> dağıt</AD>
<GRUP>böl</GRUP>
</ITEM>
<ITEM>
<AD> pay</AD>
<GRUP>böl</GRUP>
</ITEM>
<ITEM>
<AD> gün</AD>
<GRUP>çarp</GRUP>
</ITEM>
<ITEM>
<AD> hafta</AD>
<GRUP>çarp</GRUP>
</ITEM>
```

Ancak çarpma işlemi gerektiren problem türlerinde iki cümlede de geçen anahtar kelimeler aynı veya ilk cümledeki zamanın ikinci cümledekinden küçük olması gerekmektedir. "Bir bakkal günde 50

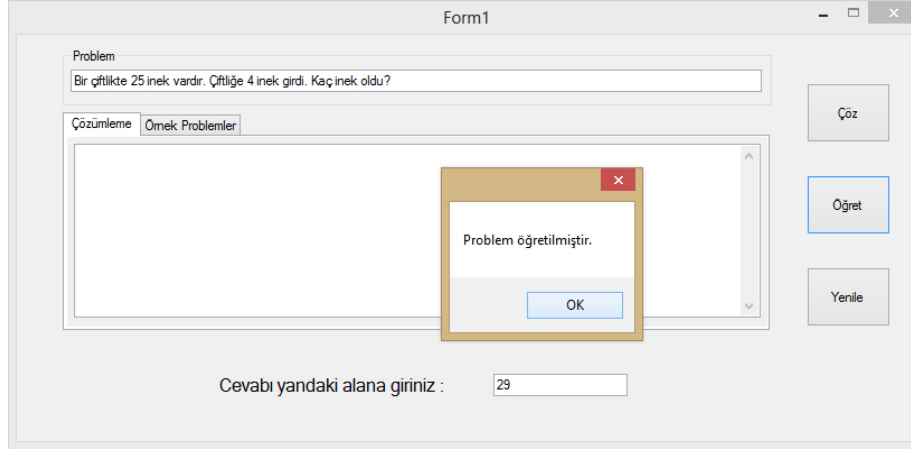
ekmek satıyor. 8 günde kaç ekme  satar?" veya "Bir bakkal günde 50 ekme  satıyor. 1 ayda kaç ekme  satar?" sorularında olduĐu gibi. İlk cümledeki zamanın ikinci cümledekenden büyük olması durumunda ise bu problem 5. türe girmektedir. "Bir bakkal ayda 60 ekme  satıyor. 1 günde kaç ekme  satar?" sorusunda olduĐu gibi.

Şekil 2. Örnek bir çıkarma problemi

Şekil 3. Örnek bir çarpma problemi

B. ÇÖZÜLEMİYEN PROBLEMLERİN SİSTEME ÖĞRETİLMESİ

Problemler XML’de tutulan anahtar kelimelere göre çözüldüĐü için bu kelimeler dışında farklı kelime içeren sorular çözülememektedir. Dört işlem gerektiren anahtar kelimelerin sınırı yoktur. Kullanıcının problemi sistem tarafından çözülemediĐinde kullanıcıya bir uyarı verdirilerek yeni anahtar kelimenin öğretilmesi istenir. Kullanıcı bu durumda cevabı yazar ve kullanıcının girmiş olduĐu cevaba göre sistem, anahtar kelimenin türünü belirleyip dört işlem türüne uygunluĐu da kontrol ederek XML veri dosyasına kaydeder. Anahtar kelimenin belirlenmesi işleminde yine Zemberek doğal dil işleme kütüphanesinin kelime türünü belirleyen fonksiyonundan yararlanılmalıdır çünkü öğretilen problem çarpma ise soru cümlesinden seçilecek olan anahtar kelime bir zaman kavramı olmalıdır, toplama ya da çıkarma ise de bu sefer anahtar kelime bir fiil olmalıdır. Eğer Böylece kullanıcının çözülemeyen bir toplama işlemi cevabını çıkarma işlemi olarak yani yanlış girerse sistem bunu engelleyip tekrar kullanıcıdan girdi isteyecektir. Figür 4’te “Bir çiftlikte 25 inek vardır. Çiftliğe 4 inek girdi. Kaç inek oldu?” probleminin öğretilme ekran çıktısı gösterilmektedir.



Şekil 4. Çözümlemeyen bir problemin sisteme öğretilmesi

III. SONUÇ VE GELECEK ÇALIŞMALAR

İlköğretim seviyesindeki Türkçe aritmetik problemler incelendiğinde bu problemlerin çeşitli tiplerde olduğu anlaşılmaktadır. Geliştirilen uygulamada kullanıcı tarafından girilen problemin hangi tipte olduğunun anlaşılması ve bu tipe göre problemin çözülmesi sağlanmıştır. Bu sebeple Türkçe aritmetik problemler için dört temel soru kalıbı belirlenmiş –toplama, çıkarma, çarpma ve bölme- olup soru tiplerinin çözülmesi için anahtar kelimeler oluşturulmuştur. Anahtar kelimeler ile soru tipinin belirlenmesi için XML formatında veri dosyası hazırlanmıştır. Bu dosyada kelimeler ve ait oldukları gruplar tutulmaktadır. Yapılan çalışmanın sonucunda dört tipteki aritmetik problemlerin çözülme başarısı, soru kalıbı kurallarına ve Türkçe dilbilgisi kurallarına uyulduğu takdirde %92 olarak gözlemlenmiştir. Bunun nedeni bölme ve çarpma soruları çok çeşitli olup gruplamanın yapılmasının zorluğundan kaynaklanmaktadır.

İlerleyen aşamalarda; soru tipi sayısının artırılmasıyla daha çok sayıda problemin çözülmesi, bilinmeyen bir soru tipi geldiğinde uygulamanın bu soru tipini ve çözümünü otomatik öğrenmesi, soru kalıbı belirlenirken kullanılan nitelik sayılarının artırılması ve bu sayede soru kalıbının belirlenme başarısının artırılması hedeflenmiştir. Ayrıca Türkçe aritmetik problem çözümlerinin yanında Türkçe aritmetik problem üretmekte hedeflerimiz arasındadır.

IV. KAYNAKLAR

- [1] P. M. Nadkarni, L. Ohno-Machado, W. W. Chapman, *Natural Language Processing: An Introduction*, Journal Of The American Medical Informatics Association: JAMIA [J Am Med Inform Assoc], vol. 18, no. 5, pp. 544-51, 2011.
- [2] V. V. Nabiyev, *Yapay Zeka-Problemler-Yöntemler-Algoritmalar*, 2. Baskı, Seçkin Yayıncılık, (2005).
- [3] D.G. Bobrow, A. Collins, *Representation and Understanding*, Academic Pres, Newyork, (1975).
- [4] E. Charniak, 1969. *Computer solution of calculus word problems*, **Proceedings of the 1st international joint conference on Artificial intelligence**, Washington, (1975) 303- 316.
- [5] A. Bundy, L. Byrd, G. Luger, C. Mellish, R. Milne, M. Palmer, *Solving mechanics problems using meta-level inference*, Department of Artificial Intelligence, Edinburgh University, Edinburgh Research Archive, (1979) 1017-1027.
- [6] T. H. Huang, Y. H. Liu, H. C. Chang, *Educational Technology & Society*, 15(1) (2012) 248-259.
- [7] G. Polya, *How to solve it*, Princeton University Press, Princeton, New Jersey, (1945).

- [8] H. E. Chang, Y. T. Sung, S. F. Lin, *Computers & Education*, 46(2) (2006) 140-151.
- [9] B. S. Ilany, B. Margolin, *Creative Education, Scientific Research*, 1(3) (2010) 138-148.
- [10] A.C.Cem Say (2001) **DOI: 10.1142/S021800140100089**
- [11] Ü. Çakırođlu, *World Applied Sciences Journal*, **6(2)** (2009) 196-202.
- [12] M.D. Akın, A.A. Akın, *Türk Dilleri için Açık Kaynaklı Doğal Dil İşleme Kütüphanesi: Zemberek*, *Elektrik Mühendisliği*, **431** (2007) 38-44.
- [13] S. Kaba, *Zemberek Projesi*, Yıldız Teknik Üniversitesi Açık Kaynak Günleri '09, (2009).