# Kelimeler Arası Anlamsal İlişkilerin Bulunmasında Word2vec ile Şablonların Karşılaştırılması

# Comparison of Templates with Word2vec in Finding Semantic Relations Between Words

Kaan ANT[1] Uğur SOĞUKPINAR[2] Mehmet Fatih AMASYALI[3]

[1,2,3]Bilgisayar Mühendisliği Bölümü
Elektrik-Elektronik Fakültesi
Yıldız Teknik Üniversitesi, İSTANBUL

Email: kaanantt@gmail.com sogukpinar.ugur@gmail.com mfatih@ce.yildiz.edu.tr

***Özetçe*— Doğal dil işleme çalışmalarının daha etkili yapılabilmesi için kelimeler arası anlamsal ilişkileri içeren veri tabanlarının kullanımı giderek yaygınlaşmaktadır. Kelime torbası yaklaşımı yerine önerilen anlamsal uzaylar kelimeler arası ilişkilerin büyüklüklerini vermekte ancak ilişki türünü ifade etmemektedir. Bu çalışmada anlamsal uzayların ilişki türü bulmada nasıl kullanılabileceği gösterilmiş ve şablonlar yöntemiyle karşılaştırılması yapılmıştır. Oldukça büyük (1 GB) bir derlem üzerinde elde edilen sonuçlara göre "üst kavramıdır" ve "zıt anlamlıdır", ilişkileri için anlamsal uzaylar daha başarılı olurken, "nerede bulunur", "neyden yapılmıştır" ilişki türlerinde ve ilişkisizliğin belirlenmesinde şablonlar yaklaşımı daha başarılı olmuştur.**

***Anahtar Kelimeler—doğal dil işleme; hayat bilgisi veritabanları; anlamsal uzaylar; ilişki şablonları.***

***Abstract*—The use of databases those containing semantic relationships between words is becoming increasingly widespread in order to make natural language processing work more effective. Instead of the word-bag approach, the suggested semantic spaces give the distances between words, but they do not express the relation types. In this study, it is shown how semantic spaces can be used to find the type of relationship and it is compared with the template method. According to the results obtained on a very large scale, while "is_a" and "opposite" are more successful for semantic spaces for relations, the approach of templates is more successful in the relation types "at_location", "made_of" and " non relational ".**

***Keywords—natural language processing; commonsense databases; semantic spaces; relation templates.***

## I. GİRİŞ

Kelimeler ya da kelime grupları arasındaki ilişki türlerini içeren veri tabanları çeşitli doğal dil işleme görevlerinde (dil çevirisi, diyalog yönetim sistemleri, eşanlamlı ifade oluşturma, soru cevaplama, duygu durum analizi vb.) yaygın olarak kullanılmaya başlanmıştır. En çok bilinen veri tabanlarına örnek olarak Wordnet [1], Cyc [2], NELL [3], ConceptNet [4] , CSDB [5] verilebilir.

Metin temsilinde kullanılan kelime torbası (bag of words) yaklaşımında kelimeler arası anlamsal benzerlikler dikkate alınmamaktadır. Bu durumu iyileştiren anlamsal uzaylar yaklaşımında ise kelimeler d boyutlu vektörlere dönüştürülür. Bu d boyutlu uzaydaki vektörler arası yakınlık anlamsal yakınlıkla doğru orantılıdır. Bu sayede her bir kelimeye anlamsal olarak en yakın kelimeler bulunabilir. Anlamsal uzayların oluşturulmasında bugüne kadar pek çok çalışma yapılmıştır [6,7,8]. Bununla birlikte hepsinin temel varsayımı Harris'in önerdiği [9], belirli bir kelime penceresi içinde sıklıkla geçen kelimelerin birbirlerine anlamsal olarak yakın olmasıdır. Hızı ve dilden bağımsız yöntemi sebebiyle anlamsal uzay oluşturmada en yaygın kullanılan araç Word2vec'dir [8].

Anlamsal uzaylarda kelimeler vektörlerle temsil edildiğinden iki kelime arasındaki anlamsal benzerliğin büyüklüğünü bulmaya imkan vermektedir. Ancak bu anlamsal ilişkinin türü hakkında bir bilgi vermemektedir.

Bu çalışmada anlamsal uzaylarda kelimeler arası ilişki türünün belirlenmesi için bir yöntem önerilmiş, yöntemin başarısı çeşitli ilişki türleri üzerinde şablonlar yöntemiyle karşılaştırılmıştır.

Bu çalışma ASYU-2017 Konferansında sunulmuş ve sadece özet olarak
Bildiri Özetleri Kitabı (ISBN: 978-605-84722-8-0)'nda yayımlanmıştır.

Bildirinin 2. bölümünde Word2vec'in detayları verilmiştir. 3. bölümde karşılaştırmada kullanılan şablonlar yöntemi anlatılmıştır. 4. bölümde ise önerilen ilişki türü bulma yöntemi sunulmuştur. 5. bölümde kullanılan metin derlemi ve deneysel sonuçlar verilmiştir. Son bölümde ise sonuçlar tartışılmış ve gelecek planları açıklanmıştır.

## II. ANLAMSAL UZAYLAR VE WORD2VEC

Kelimelerin d boyutlu sayısal vektörlere dönüştürüldüğü anlamsal uzaylardaki temel amaç anlamsal olarak yakın kelimelerin vektörlerinin de birbirlerine yakın olmasıdır. Bu amacı gerçekleştirmek için bugüne kadar yapılmış birçok çalışma bulunmaktadır. Bununla birlikte Mikolov'un önerdiği Word2vec hızı ve başarılı sonuçları sayesinde en çok kullanılan araç haline gelmiştir [8]. Ayrıca Word2vec kütüphanesinin içinde analogy isimli bir fonksiyonla (kelime1→kelime2), (kelime3 → kelimeX) sorusuna cevap verilebilmektedir. Bu ifade "kelime3 ile hangi kelime arasında kelime1 ile kelime2 arasındaki ilişki vardır" şeklinde okunabilir. Bununla birlikte analogy'nin performansının ölçüldüğü ilişki türleri oldukça sınırlıdır. Para birimi, ülke başkenti, fiillerin geçmiş zamanı gibi ilişki türleri kullanılmıştır. Kelimeler arasındaki önemli anlamsal ilişki türleri olan üst kavramıdır, eşanlamlıdır, nerede bulunur vb. içinse bu ilişkilerin otomatik bulunmasıyla ilgili Word2vec'in üzerinde herhangi bir çalışma yapılmamıştır. Yapılan bir çalışmada [10], bir kelimeye ait en yakın kelimelerin listesi alınmış ve Wordnet ile bu ilişkilerin türleri çıkarılmıştır. Bu çalışmada ve bizim kendi deneylerimizde gördüğümüz üzere bir kelimeye en yakın kelimelerin listesinde bu önemli anlamsal ilişkilere sahip kelimeler bulunmaktadır Ancak hangi kelimeyle ne tür bir ilişkiye sahip olunduğu otomatik olarak bulunamamaktadır. Tablo 1'de deneylerimizde elde ettiğimiz "okul" kelimesine en yakın kelimeler gösterilmiştir.

| Kelime | Benzerlik |
|---|---|
| ilköğretim | 0,781 |
| anaokul | 0,767 |
| öğretmen | 0,766 |
| öğrenci | 0,757 |
| dershane | 0,745 |
| yatılı | 0,739 |
| ilkokul | 0,725 |
| dersane | 0,684 |
| lise | 0,664 |
| anasınıfı | 0,660 |

**Tablo I**. "okul" kelimesine en yakın kelimeler ve benzerlik oranları

Yukarıdaki değerler derlem üzerinde word2vec ile cosinüs benzerlikleri hesaplanması sonucu elde edilmiştir. Tablo 1 incelendiğinde Word2vec ile "okul" kelimesiyle anlamsal ilişki içinde olan kelimelerin başarıyla bulunduğu buna karşılık, farklı türde ilişkilerin bir arada bulunduğu görülmektedir.

### *A. Şablonlar Yöntemi*

Kelimeler arası ilişki türlerinin otomatik olarak bulunmasında en yaygın kullanılan yöntemlerden birisidir [11]. Bu yöntemde öncelikle her bir ilişki türünü ifade eden şablonlar tanımlanmaktadır. Ardından bu şablonlar kelimelerle birlikte büyük metin derlemlerinde aratılmakta ve çıkan sonuçlara göre ilişki türleri belirlenmektedir. Örneğin "üst kavramıdır" ilişki türü için "bir tür Kelime1 olan Kelime2" şablonu kullanılabilir. Bu şablon kullanılarak aralarında "üst kavramıdır" ilişkisi olduğu düşünülen 2 kelime (Ör: kelime1=hayvan kelime2=kedi) bu şablon içine yerleştirilerek oluşan "bir tür hayvan olan kedi" karakter dizisinin derlemde kaç kez geçtiği aratılır. Geçme sayısına göre kelime1 ve kelime2 arasında bu ilişkinin olup olmadığına karar verilir. Şablonlar yönteminin bir başka kullanımında ise kelimelerden birisi şablonun içine konulur. Derlemde şablonun geçtiği yerlerde diğer kelimeye karşılık gelen kelimeler alınır.

### *B. Anlamsal Uzaylarda İlişki Türü Bulma*

Daha önce tanıtılan analogy fonksiyonunun içyapısı incelendiğinde Eşitlik 1'i kullandığı görülmüştür.

$$Kelime3 - KelimeX = Kelime1 - Kelime2 \quad (1)$$

Eşitlik 1 kullanılarak VektörX bulunur.

$$VektörX = Kelime3 - Kelime1 + Kelime2 \quad (2)$$

KelimeX, VektörX'e en yakın kelime olarak belirlenir.

$$k = {}^{argmin}_{i \in N}( dist\ (VektörX, Kelime_i)) \quad (3)$$

$$KelimeX = Voc(k) \quad (4)$$

Eşitlik 3 ve 4'te *N*, derlemdeki tekil kelime sayısını, *Voc* ise koordinatlar matrisini göstermektedir. *dist* fonksiyonu ise kosinüs benzerliğidir.

Kullanılan bu yaklaşımda aynı ilişki türüne sahip kelime ikilileri arasındaki fark vektörlerinin birbirine benzer olduğu varsayılmıştır. Bu varsayımdan hareketle önerdiğimiz yöntemde önce aralarındaki ilişki türlerini bildiğimiz kelime ikilileri ve vektörleri toplanmıştır. Girişlerin bu 2 vektörden elde edildiği, çıkışın ise 2 kelime arasındaki ilişki türü olduğu bir yapay öğrenme veri kümesi oluşturulmuştur. Örneğin C adet ilişki türünün her biri için aralarındaki ilişki türü bilinen 50'şer kelime ikilisi kullanılırsa problem C adet sınıfa sahip, 50*C örnekli bir sınıflandırma problemine dönüşmektedir. Bu adımın ardından istenilen yapay öğrenme algoritması ile ilişki türü tahmini yapılabilmiştir.

Bu çalışma ASYU-2017 Konferansında sunulmuş ve sadece özet olarak
Bildiri Özetleri Kitabı (ISBN: 978-605-84722-8-0)'nda yayımlanmıştır.

## III. DENEYSEL YÖNTEMLER

Bu bölümde karşılaştırılan her iki yöntem için yapılan işlemler verilmiştir.

### *A. Kullanılan Metin Derlemi*

Kullanılan her 2 yöntem de büyük boyutlu bir metin derlemine ihtiyaç duymaktadır. Bunun için Boğaziçi Üniversitesinin haber metinlerinden oluşturduğu derlem kullanılmıştır [12].

Bu derlemin kelime morfolojik çözümlenmesi yapılmış ve kelimelerin gövdeleri alınmış hali kullanılmıştır [13]. Ek olarak sayı içeren tüm kelimeler silinmiştir. Bu işlemler sonucunda derlemde 28632 tekil kelime, toplamda yaklaşık 143 milyon kelime kalmıştır.

### *B. Karşılaştırma İçin Kullanılan Kelime İkilileri*

Yaptığımız çalışmada anlamsal ilişkilerin otomatik belirlenebilmesi ve yöntemlerin performanslarının ölçülebilmesi için gerekli olan kelime ikilileri ve ilişki türleri belirlenmiştir. Tablo 2'de seçilen ilişki türleri ve her bir ilişki türüne ait kelime ikilisi sayıları görülmektedir. Oluşturulan ikililerdeki kelimelerin tümü derlem içinde geçen kelimelerden seçilmiştir.

| İlişki Türü | Örnek ikililer | İkili Sayısı |
|---|---|---|
| Nerede Bulunur | ağaç-orman, altın-maden, çiçek-bahçe | 169 |
| Üst Kavramıdır | kaktüs-bitki, kalp-organ, kedi-hayvan | 168 |
| Zıt Anlamıdır | sert-yumuşak, hızlı-yavaş, güzel-çirkin | 170 |
| Neyden Yapılır | çamur-toprak, ev-tuğla, cümle-sözcük | 150 |
| İlişkisiz | sınıf-pilav, sezgi-bina, sorgu-muz | 586 |

**Tablo II**. İlişki türleri, örnek ikililer ve bu türlere ait kelime ikilileri sayıları

Seçilen ilişki türlerine ek olarak, ilişkili olmamanın da tahmin edilebilmesi için "ilişkisiz" türü eklenmiştir.

### *C. Şablonlar Yöntemi Tasarımı*

Şablonlar yönteminde "ilişkisiz" türü haricindeki ilişki türleri için elle şablonlar oluşturulmuştur. Oluşturulan şablonlara ait düzenli ifadeler Tablo 3'te verilmiştir.

| **İlişki türü** | **Düzenli ifade** |
|---|---|
| Nerede Bulunur | (?P<w>[^]+?)[td][ea]ki (?P<w>[^]+?) |
| Nerede Bulunur | (?P<w>[^]+?)[td][ea] (?P<w>[^]+?) bul |
| Nerede Bulunur | (?P<w>[^]+?)[td][ea] (?P<w>[^]+?) var |
| Nerede Bulunur | (?P<w>[^]+?) (?P<w>[^]+?)[iuıü][nm] parça |
| Nerede Bulunur | (?P<w>[^]+?) (?P<w>[^]+?)[dt][ae][n] al |
| Neyden Yapılır | (?P<w>[^]+?)(?P<w>[^]+?)[dt][ae][n] (yapılmış\|oluşmuş\|üretilmiş) |
| Neyden Yapılır | (?P<w>[^]+?)[dt][ae][n] (yapılmış\|oluşmuş\|üretilmiş) (?P<w>[^]+?) |
| Neyden Yapılır | (?P<w>[^]+?)(oluşan\|yapılan\|üretilen) (?P<w>[^]+?) |
| Üst Kavramıdır | (?P<w>[^]+? bir (?P<w>[^]+?)[dt][iuıü][r] |
| Üst Kavramıdır | (?P<w>[^]+?) kapsar (?P<w>[^]+?)[i,u,ı,ü] |
| Üst Kavramıdır | (?P<w>[^]+?) (?P<w>[^]+?)[iuıü] kapsar |
| Üst Kavramıdır | (?P<w>[^]+?) [iuıü][n] (tümü\|hepsi\|tamamı) (?P<w>[^]+?)[dt][iuıü][r] |
| Üst Kavramıdır | (?P<w>[^]+?) bir (çeşit\|tür) (?P<w>[^]+?)[dt][iuıü][r] |

**Tablo III**. İlişki türü- düzenli ifade şablonları

Tablo 3'de görüldüğü üzere aynı ilişki türü için birçok şablon tanımlanmıştır. Aralarındaki ilişki tahmin edilmek istenen kelime ikilisi bu şablonların hepsine yerleştirilerek derlemde aratılmıştır ve geçme sayıları bulunmuştur. Bir kelime ikilisi en çok hangi ilişki türüne ait şablonlarda geçiyorsa o ilişki türüne dahil edilmiştir. Eğer hiçbir şablonda geçmiyorsa "ilişkisiz" sınıfına atanmıştır. "Zıt anlamıdır" ilişki türü için genelleştirilebilir şablonlar bulunamamıştır. Bu sebeple şablonlar yöntemi bu ilişki türü için başarısız kabul edilmiştir.

Derlemin indekslenmesinde ve şablonların aranmasında bir arama motoru kütüphanesi olan Apache Spark kullanılmıştır [14].

### *D. Anlamsal Uzaylar Yöntemi Tasarımı*

Bölüm III.A'da anlatılan metin derlemi Word2vec ile işlenmiştir. Eğitimde C-BOW yöntemi, kelime vektörlerinin boyutu olarak 200, kelime penceresi boyutu olarak ise 5 kullanılmıştır. C-BOW yönteminde her bir kelimenin vektör uzayında bir vektör olarak kabul edilmesi esasına dayanır. Vektör bileşenleri her kelimenin ağırlık veya önemini temsil eder. İki vektör arasındaki benzerlik, kosinüs benzerlik ölçüsü kullanılarak hesaplanır.

Yapay öğrenme veri kümesinin oluşturulmasında 201 özellik çıkarılmıştır. Kelimeler 200 boyuta sahiptir. 200 boyutlu fark vektörü her kelime ikilisi için çıkarılarak kullanılmıştır. Fark vektörüne ek olarak, iki vektör arasındaki kosinüs benzerliği de kullanılmıştır. Bu özelliğin özellikle "ilişkisiz" sınıfının belirlenmesinde etkin olduğu görülmüştür. Elde edilen 201 özellikli veri kümesi arff formatına çevrilerek WEKA kütüphanesinde kullanılmıştır [15].

## IV. DENEYSEL SONUÇLAR

Denemelerde ilk olarak anlamsal uzaydan elde edilen arff dosyası üzerinde çeşitli sınıflandırıcılar çalıştırılmıştır. Bu denemelerde sınıflandırma işlemi yapılırken 3 kez 10 katlı çapraz geçerlemeyle elde edilen 30 test sonucunun ortalamaları kullanılmıştır. Sonuçlar Tablo 4'te verilmiştir. Sınıflar arası örnek dağılımları eşit olmadığında performans karşılaştırmasında sınıf dağılımlarıyla ağırlıklandırılmış F-ölçütü kullanılmıştır. Denemelerde orijinal 201 boyutlu veri kümesi, CFS [16] yöntemiyle özellik seçimi yapılarak elde edilmiş 6 boyutlu veri kümesi ve PCA ile varyansın %95'ini açıklayan 173

Bu çalışma ASYU-2017 Konferansında sunulmuş ve sadece özet olarak
Bildiri Özetleri Kitabı (ISBN: 978-605-84722-8-0)'nda yayımlanmıştır.

boyutlu veri kümesi kullanılmıştır. CFS (Korelasyona Dayalı Özellik Seçimi), değerlendirme formülüne uygun bir korelasyon ölçüsü ve buluşsal arama stratejisini birleştiren bir algoritmadır. Bu sayede özellik indirgemenin ve uzay dönüşümünün performans üzerindeki etkileri de görülebilmiştir.

| | F-ölçütü (std. Sapma) % | | |
|---|---|---|---|
| Yöntem | Orijinal | Özellik seçimli | PCA |
| Naive Bayes | 34,4(4) | 53,4(3) | 36,3(5) |
| Destek Vektör Makineleri | 59,3(3) | 52,2(3) | 56,5(3) |
| En yakın 1 Komşu | 59,5(4) | 48,4(3) | 43,1(3) |
| Rastsal Ormanlar | 41,6(4) | 52,5(4) | 38,1(3) |
| Karar Ağacı(J48) | 50,9(4) | 49,7(4) | 38,2(3) |
| Basit Lojistik Regresyon | 60,1(4) | 54,2(4) | 56,7(5) |

**Tablo IV**. Anlamsal uzaydaki deneme sonuçları

Tablo 4 incelendiğinde en başarılı sonuçların basit lojistik regresyon, en yakın 1 komşu ve destek vektör makineleri yöntemleriyle alındığı görülmektedir. Özellik seçimi ile 6 boyuta indirgemenin sadece Naive Bayes ve Rastsal ormanlar için daha iyi olduğu görülmektedir. Orijinal ve PCA ile dönüştürülmüş yeni uzaydaki sonuçlar genellikle birbirine yakındır.

Orijinal özelliklere göre en başarılı 3 yöntemin aralarında istatistiksel anlamlı bir farkın olup olmadığının testi için eşli t-test kullanılmış ve %95 olasılıkla fark olmadığı görülmüştür.

Sınıf bazlı analiz için en yüksek başarıya sahip basit lojistik regresyona ait analizler Tablo 5'te, en yakın 1 komşuya ait analizler ise Tablo 6'da verilmiştir.

| Sınıflar | F-ölçütü | Kesinlik | Duyarlılık |
|---|---|---|---|
| Nerede Bulunur | 43,8 | 43,4 | 44,2 |
| Neyden Yapılmış | 36,9 | 43,4 | 32,2 |
| Zıt Anlamıdır | 45,4 | 51,5 | 40,6 |
| Üst Kavramıdır | 38,5 | 39,5 | 37,5 |
| İlişkisiz | 81,9 | 77,1 | 87,4 |

**Tablo V**. Basit lojistik regresyonun sonuçlarının sınıf bazlı analizi

| Sınıflar | F-ölçütü | Kesinlik | Duyarlılık |
|---|---|---|---|
| Nerede Bulunur | 41,9 | 58,9 | 32,5 |
| Neyden Yapılmış | 47,1 | 53,6 | 42 |
| Zıt Anlamıdır | 42,3 | 37,6 | 48,2 |
| Üst Kavramıdır | 60,2 | 60,4 | 60 |
| İlişkisiz | 72,9 | 69,7 | 76,5 |

**Tablo VI**. En yakın 1 komşunun sonuçlarının sınıf bazlı analizi

Tablo 5 ve 6 birlikte incelendiğinde en yüksek başarı ile tahmin edilen sınıfın her iki yöntem içinde "ilişkisiz" sınıfı olduğu görülmektedir. Bununla birlikte oldukça önemli bir ilişki türü olan "üst kavramıdır" ın en yakın 1 komşu ile çok daha iyi belirlenebildiği görülmektedir. Bu iki yöntemden "üst sınıfıdır" ilişki türünü daha iyi tahmin ediyor olması sebebiyle en yakın 1 komşuyu kullanmak daha etkili olabilir.

Kelimeler arası ilişkileri tahmin etmek için kullandığımız ikinci yöntem olan şablonlara ait sonuçlar ise Tablo 7'de verilmiştir. "Zıt anlamıdır" ilişki türü şablonlar yöntemiyle belirlenemediğinden tabloda yer almamaktadır.

| Sınıflar | F-ölçütü | Kesinlik | Duyarlılık |
|---|---|---|---|
| Nerede Bulunur | 74.1 | 68.7 | 81.3 |
| Neyden Yapılmış | 67.2 | 60.1 | 78.7 |
| Üst Kavramıdır | 36.2 | 95.1 | 22.6 |
| İlişkisiz | 86.8 | 84.3 | 87.4 |

**Tablo VII**. Şablonlar yöntemiyle elde edilen arama sonuçlar (%)

Tablo 7'de görüldüğü üzere "üst kavramıdır" ve "zıt anlamıdır" haricindeki tüm ilişki türlerinde şablonlar yöntemi anlamsal uzaylardan çok daha başarılı olmuştur.

## V. SONUÇ VE GELECEK ÇALIŞMALAR

Kelimeler arası anlamsal ilişkilerin otomatik olarak bulunmasına yönelik yaptığımız bu çalışmada anlamsal uzaylardaki fark vektörleri ve şablonlar yöntemi karşılaştırılmıştır. Sonuçlara göre ilişki türüne bağlı sonuçlar elde edilmiştir. "üst kavramıdır" ve "zıt anlamıdır" ilişki türleri için anlamsal uzaylar yöntemi, "nerede bulunur", "neyden yapılmıştır" ve "ilişkisiz" ilişki türleri için şablonlar yöntemi daha başarılı olmuştur.

İlişki türlerinin belirlenmesinde kullanılan derlemin boyutu ve türü önem göstermektedir [11]. Gelecek çalışmalar olarak, çeşitli derlemler üzerinde çalışılması, kullanılan 2 yöntemin sonuçlarının birleştirilmesi, tekil kelimelere ek olarak kelime grupları üzerinde çalışılması düşünülmektedir.

## KAYNAKÇA


[1] Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. ve Miller, K., "Introduction to WordNet: An On-line Lexical Database", 1993.

[2] Lenat, D. B., "Cyc: A Large-Scale Investment in Knowledge Infrastructure", The Communications of the ACM, 38(11):33-38, 1995..

[3] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr., Tom M. Mitchell, "Toward an Architecture for Never-Ending Language Learning", AAAI Publications, Twenty-Fourth AAAI Conference on Artificial Intelligence, 2010.

[4] Liu, H. ve Singh, P., "ConceptNet: A Practical Commonsense Reasoning Toolkit", BT Technology Journal, (22), Kluwer Academic Publishers, 2004

[5] Amasyalı, M. F., İnak, B. ve Ersen, M. Z., "Türkçe Hayat Bilgisi Veri Tabanının Oluşturulması", Akademik Bilişim Konferansı, 2010.

[6] M.Fatih Amasyalı, Aytunç Beken, "Measurement of Turkish Word Semantic Similarity and Text Categorization Application", SIU 2009

[7] Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "Glove: Global Vectors for Word Representation." EMNLP. Vol. 14. 2014.


Bu çalışma ASYU-2017 Konferansında sunulmuş ve sadece özet olarak
Bildiri Özetleri Kitabı (ISBN: 978-605-84722-8-0)'nda yayımlanmıştır.

[8] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems pp. 3111-3119.

[9] Z.S., Haris, "Mathematical structures of language", Wiley, s.12, 1968.

[10] Handler, Abram. An empirical study of semantic similarity in WordNet and Word2Vec. 2014. PhD Thesis. Columbia University.

[11] Gürkan Şahin, M.Fatih AMASYALI, "Iterative Information Extraction from Large Text Collections", EMO Bilimsel Dergi, Cilt 4, Sayı 7, 13-20, 2014

[12] Sak, Haşim, Tunga Güngör, and Murat Saraçlar. "Turkish language resources: Morphological parser, morphological disambiguator and web corpus." Advances in natural language processing. Springer Berlin Heidelberg, 2008. 417-427.

[13] Tuğba Yıldız, Savaş Yıldırım, Banu Diri, (2013). "Extraction of Part-Whole Relations from Turkish Corpora", CICLing 2013: Computational Linguistics and Intelligent Text Processing, LNCS 7816, 126-138.

[14] Apache spark docs. http://spark.apache.org/.

[15] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1), 10-18.

[16] Hall, Mark A. Correlation-based feature selection for machine learning. 1999. PhD Thesis. The University of Waikato.

Bu çalışma ASYU-2017 Konferansında sunulmuş ve sadece özet olarak
Bildiri Özetleri Kitabı (ISBN: 978-605-84722-8-0)'nda yayımlanmıştır.