

Kelime Anlamları

(Word Semantics)



Prof.Dr. Banu Diri



İçerik

- Kelime Anlam Benzerliği
- WORDNET
 - Rakamlarla Wordnet
 - İlişki türleri
 - Wordnet'te anlamsal benzerlik ölçümleri
 - Bağ sayma
 - Leacock & Chodorow (1998)
 - Wu & Palmer (1994)
 - Ortak / Müşterek bilgi
 - Jiang-Conrath (1997)
 - Lin (1998)



Kelime Anlam Belirsizliđi

Bir kelimenin ne anlama geldiđini nasıl biliriz?

Kalem



kale m ¹	yazma, çizme vb. işlerde kullanılan araç	Kalem masanın üstünde.
------------------------	---	---------------------------

ACABA ?

- Basit bir kelime için bile tek bir anlam yoktur
- Hangi anlamını seçeceğimize karar vermek için kelimenin içinde bulunduğu bağlamı bilmemiz gerekir
- Bağlamda olan ve kelimenin gerçek anlamıyla olan bağıntısını belirleyen kavrama *sense* denir



”Kalem kağıda ihtiyacım var.”

“İki kalem erzak aldım.”

kalem ¹	yazma, çizme vb. işlerde kullanılan araç	Kalem masanın üstünde.
kalem ²	Resmî kuruluşlarda yazı işlerinin görüldüğü yer.	Kalemimiz bugün hizmet vermeyecektir.
kalem ³	yontma işlerinde kullanılan ucu sivri veya keskin araç.	Oymacı kalemi.
kalem ⁴	çeşit, tür	Beş kalem ilaç.
kalem ⁵	bazı deyimlerde yazı	Kaleme almak.
kalem ⁶	yazar	Peyami Safa, edebiyatımızın usta kalemlerindendir.

Anlam belirsizliğini gidermek üzerine yarışmalar yapılmaktadır

Senseval/Semeval 1998 yılında başlamıştır



Kelime Anlam Belirsizliğini gidermek için iki yaklaşım vardır:

❑ Sözcük Seçim

Seçilmiş Sözcük Yaklaşımı

Tüm Sözcükler Yaklaşımı

Kelime Anlam Belirsizliğini gidermek için kullanılan yöntemler:

❑ Yararlanılan kaynaklara göre

- Bilgi Tabanlı Yöntemler
- Derlem Tabanlı Yöntemler
- Melez Yöntemler

❑ Denetimli ve Denetimsiz Yöntemler



❑ Sözcük Seçim Yöntemi (Lexical Sample) Seçilmiş Sözcük Yaklaşımı

- Anlamlarıyla birlikte küçük bir kelime grubu seçilir
- Bu kelimeler içerisinde geçtiği cümleler göz önünde bulundurularak bir insan tarafından etiketlenir
- Avantajı, kelimelerin mümkün olan bütün anlamları tablodaki gibi listelenerek (kalem örneği) etiketleme yapacak kişilerin kelimelerin bütün anlamlarını atlamadan etiketleme yapabilmesi sağlanır
- Böylece, elde sınırlı sayıdaki kelimelerin tüm anlamları veya seçilmiş sayıdaki anlamları için etiketleme gerçekleştirilmiş olur



❑ Sözcük Seçim Yöntemi (Lexical Sample) Tüm Sözcükler Yaklaşımı

- Verideki bütün kelimeler aynı anda sisteme yüklenir
- İşaretleyiciler cümlelerdeki bütün kelimeleri anlamlarına göre işaretler
- Kelimenin kökü, büyük harfle yazılıp yazılmadığı, cümlenin hangi ögesi olarak kullanıldığı gibi özellikler kelime anlamlarını birbirinden ayırmak için kullanılır
- Kelimeler anlamlarına göre sınıflandırıldığında karar ağacı gibi sınıflandırma tabanlı yöntemler sistemi, yeni gelen kelimenin anlamını otomatik olarak etiketlemek için eğitilir



□ Belirsizlik Giderme Yöntemleri

Bilgi Tabanlı Belirsizlik Giderme Yöntemleri

Sözcük anlamlarının belirtildiği kaynakları kullanır

- ❖ Sözlük
- ❖ Kelime Ağı (WordNet)

❖ Sözlük Anlamının Bağlam ile Örtüşmesini Temel Alan Yöntemler

- LESK Algoritması
- Sözcükler ve anlamlarını, yakın anlamlı sözcükleri kullanır
- *Oxford Advanced Learner's Dictionary* %50-70 arası başarı



Lesk Algoritması

1. Verilen tümce içerisindeki belirsizliği giderilecek olan sözcük seçilir
2. Sözcüğün, sözlükteki açıklaması ve verilen örnek tümceler incelenir
3. Eşleşen sözcükler tespit edilir
4. En çok örtüşme hangi anlamdan elde edilmiş ise,
belirsizliği belirlenmek istenen sözcüğe o anlam atanır
5. Tümcedeki anlam taşıyan kelimeler kullanılır, işlevsel kelimeler çıkarılır
6. Türkçe için çalışılıyorsa sözcüklerin kök veya gövde biçimleri kullanılır

Örnek : Eren'nin *konak* gibi büyük evi var.

konak	
1) Büyük ve gösterişli ev. Örnek: <i>Konaktan okula dönerken çok yoruldum.</i>	→ çakışan 3 sözcük
2) Vali, kaymakam gibi yüksek dereceli devlet görevlilerinin resmî konutu. Örnek: <i>Vali konağının önünde korumalar vardı.</i>	→ çakışan 2 sözcük
3) Konakçı. Örnek: <i>Yağmur nedeniyle konaklar yola çıkamadı.</i>	
4) Araba veya hayvanla bir günde alınan yol. Örnek: <i>Buradan orası beş konaktır.</i>	
5) Yolculukta geceyi geçirmek için inilen, konaklanan yer. Örnek: <i>Kervan konağa bu akşam ulaşacak.</i>	
6) Kundak çocuklarının başlarında görülen kepek tabakası. Örnek: <i>Çağla'nın başında konak oluştu.</i>	
7) Gözde oluşan ince tabaka. Örnek: <i>Erdem'in gözünde konak var.</i>	



LESK Algoritmasının *zayıf* yönleri:

1. Sözlükteki açıklamalarda, örnek tümcedeki sözcükler geçmeyebilir
2. Birden fazla açıklama ve örnek tümcede aynı sayıda örtüşme olabilir

Lesk → *Corpus Lesk* (sözcük örtüşmelerine ağırlık değeri verir)

Lesk → *Uyarlanmış Lesk* (sadece seçilen sözcük değil, diğer sözcükler ile ilişkili diğer sözcük tanımları da dikkate alınır)

❖ Anlamsal Kelime Ağı Üzerinde Benzerlik Ölçen Yöntemler

İki sözcüğün anlam açısından birbirine ne kadar benzediği ve aralarındaki ilişkinin çıkarılmasına dayanır

- **Sözcük benzerliği** : İki sözcüğün eş veya yakın anlamlı olmasının ölçüsüdür
- **Sözcük ilişkisi** : Sözcükler arasındaki ilişkiyi gösterir. Zıt anlamlı iki sözcük arasında ilişki derecesi (akıllı-aptal) yüksektir ancak, aralarında benzerlik olmayabilir. Buna karşılık iki sözcük birbiri ile ilişkili olabilir ama benzerlik olmayabilir (benzin-araba)



❑ Belirsizlik Giderme Yöntemleri

Derlem Tabanlı Yöntemler

❖ Denetimli Yöntemler

- ✓ Makine öğrenmesi yöntemlerine (Karar Ağaçları, Naive Bayes, Destek Vektör Makineleri, ...) anlamları etiketlenmiş bir eğitim kümesi verilerek sistem modellenir
- ✓ Etiketli veri kümesi üzerinden çıkarılan özellikler ve etiket bilgisini kullanır

❖ Yarı Denetimli Yöntemler

❖ Denetimsiz Yöntemler

- ✓ Bir sözcüğü anlamı ile etiketlemeden önce, sözcüğün hangi anlam kümesine üye yapılacağına karar verir
- ✓ *Bağlam kümeleme* kullanılan yöntemlerden biridir
- ✓ Sözcükler bir vektör olarak gösterilir. Aynı ve yakın anlamdaki sözcük vektörleri aynı bağlam kümesi içerisinde yer alır
- ✓ İki sözcük vektörünün benzerliği hesaplanır



Yöntemlerin Karşılaştırılması

Yöntem	Üstünlük	Eksiklik
Bilgi Tabanlı	Daha yüksek doğrulukta sonuç üretir	Algoritmalar sözcük örtüşmesine dayalı, örtüşme sıklığı sorunu gözlemlenebilir. Başarım sözcük tanınamalarına bağlıdır
Denetimli	Gerçekleştirme açısından diğer yöntemlere üstündür	Kaynakları kısıtlı olan dillerde elde edilen sonuçlar yetersizdir
Denetimsiz	İşaretlenmiş derlem gerekmiyor	Gerçekleştirilmesi zor ve daha düşük başarımlı sonucu verir



WORDNET

- WordNet (kelime ağı), bir dilin tüm kelimelerini, bu kelimelerin tanımlarını ve aralarındaki anlambilimsel ilişkileri içeren bir veritabanıdır
- İlişkiler, eş kümeler (synset) arasında tanımlanmıştır
- 1985 yılında, Princeton Üniversitesinde elle 10 yıla yakın bir sürede **İNGİLİZCE** için oluşturulmuştur
- Günümüzde 200'den fazla dili kapsar

WordNet Hangi Alanlarda Kullanılır

- Doğal Dil İşleme (metin analizi, makine çevirisi, anlam belirsizliği giderme)



WordNet ile Sözlük Arasındaki Fark Nedir?

- Sözlüğe benzer, isimler, sıfatlar, fiiller ve zarflar ayrı olarak gruplanır
- Eş anlamlı kelimeler *synset* (117 bin adet) adı verilen kümeler olarak tutulur
- WordNet içerisinde kelimeler hiyerarşik bir yapı altında dizilir
- *Synset*'ler eş anlamlı öbekler oluşturulduktan sonra anlamsal kapsamlarına göre sıralanırlar
- Anlamsal kapsamı daha geniş olan terimler *hypernym*, daha spesifik olanlar ise *hyponym* olarak adlandırılır
- Bir sözcük, anlamı daha geniş olan bir terimin *hyponym*'i iken, aynı zamanda daha dar anlamlı bir kelimenin *hypernym*'i olabilir



Örnek

“hayvan” sözcüğü “canlı” teriminin altında listelenirken, kendi altında “kedi” sözcüğünü içersin.

“kedi” sözcüğü, “hayvan” sözcüğünün **hyponym**’i, “canlı” sözcüğü de “hayvan” sözcüğünün **hypernym**’i olur.

“canlı” terimi, altında “insan” ve “bitki” gruplarında olabilir

Türkçe WordNet Çalışmaları Nelerdir?

Türkçe üzerine yapılmış en büyük çalışmalarından biri KeNet’tir.

(Starlang Yazılım olarak Işık üniversitesi bünyesinde geliştirilmiş Türkçe WordNet KeNet’e açık kaynak ve ücretsiz olarak ulaşılabilir)

KeNet 80 binden fazla *synset* içerir

Sabancı Üniversitesinin yürüttüğü bir başka proje **BalkaNet**



WordNet deki ilişkiler

Synsets arasında en sık kodlanan ilişki süper-bağımlı ilişkidir
(hyperonymy, hyponymy veya ISA ilişkisi olarak geçer)

{Furniture}, {piece_of_furniture} en genel hali
{bed} ve {bunkbed} daha spesifik

Böylece,
WordNet, mobilyaların ranza, yatak içerdiğini;
yatak ve ranza gibi kavramların mobilya kategorisini oluşturduğunu anlar



Kavram		Kavram		Kavram	
Türkçe	İngilizce	Türkçe	İngilizce	Türkçe	İngilizce
Üst Kavram	Hyperonym	Parçanın Bütünü	Holo Portion	Sonucudur	Is Caused By
Alt Kavram	Hyponym	Yaklaşık Karşıt Anlam	Near Antonym	Durumundadır	Be in State
Bölümün Bütünü	Holo Part	Bütünün Parçası	Mero Portion	Durumudur	State of
Bütünün Bölümü	Mero Part	Alt Olay	Sub Event	Yaklaşık Eş Anlamlı	Near Synonym
Üyenin Bütünü	Holo Member	Olaydır	Is Event of		
Bütünün üyesi	Mero Member	Nedenidir	Causes		



Alt Kavram (Alt Anlamlılık) - (Hyponym)

- ❖ Sözcükler birbirleriyle aynı cins ve türden olma açısından benzerlikler taşırlar

Örnek

‘kuş’ sözcüğü canlılar ve hayvanlar üst kavramlarının bir alt türüdür. Serçe, güvercin, çalıkuşu, bülbül gibi sözcükler de ‘kuş’ üst kavramının alt örnekleridir.



Eş Anlamlılık - (Hyponym)

- ❖ Farklı okunan ve yazılan ancak aynı kavramsal içeriği ve anlamsal özellikleri olan iki sözcük arasındaki ilişkidir

Örnek

‘ak ve beyaz’

‘yürek ve kalp’

Sözcükler belli bir bağlamda birbirlerinin yerine kullanılabilirler mi?

- ❖ **Eşdizimsellik** anlam eş anlamlılığını etkileyen ve sözcüklerin birbiri yerine geçmesini engelleyen özelliklerdendir

Örnek

‘anamın ak sütü’

‘kalpsiz adam’



Gerçek Eş Anlamlılık Çok Azdır

Eş anlamlılığın üç derecesi vardır:

- ❖ **Tam eş anlamlılık:** Anamların tam benzerliğidir
Sözcükler birbirlerinin yerine kullanılabilir
Üzüntü, elem, keder, dert
- ❖ **Önermesel eşanlamlılık:** Gerektirme ile tanımlanır.
Eşanamlı sözcükler her zaman birbirinin yerine kullanılamaz
Ali kaplumbağayı gördü
Ali tosbağayı gördü
Ali araştırmacıları kaplumbağaları incelerken gördü
Ali araştırmacıları tosbağaları incelerken gördü
- ❖ **Yarı eşanlamlılık:** Anamları birbirleriyle tamamen aynı değildir
yüzünden
-den dolayı
nedeniyle
sebebiyle
dolayısıyla



Eş anlamlılık

İlgili sözcük çiftlerinin farklı yönleri olabilir

- Dilin farklı lehçelerine ait olabilirler
günebakan, ayçiçeği, dövranber
- Bazı sözcükler benzer durumlarda farklı biçimlerde kullanılırlar
adam, herif
- Bazı sözcüklerin duygusal anlamları vardır
politikacı, devlet adamı
- Bazı sözcükler sadece belli bir bağlamda kullanılırlar
(alnım) ak, beyaz ten



Karşıt Anlamlılık- (Antonym)

- ❖ İki sözcük arasında anlamsal olarak birbirinin karşıtı olma durumudur
kısa-uzun
ters-düz
üst-alt
- ❖ Temel olarak iki tür karşıtlık vardır
 - Derecelendirilebilen/derecelendirilemeyen
 - Ters/bakışımli

Derecelendirilebilen karşıtlık İki sözcük arasında belli bir ölçütün iki uç noktasında olma ilişkisi vardır

Uzun/kısa, genç/yaşlı (“daha ” ve “çok” sözcükleriyle kullanılabilir)

- Olumsuzları karşıtlık taşımazlar
 - o «kısa değil», «uzun»
 - o «genç değil», «yaşlı» anlamına gelmez



Derecelendirilemeyen karşıtlık İki kavram arasındaki karşıtlığın, bölgesel bir sınır gibi, tam olduğu durumda ortaya çıkar

ölü/canlı, bekar/evli

- Olumsuz biçimleri birbirleri ile eşit anlam taşır
«ölü değil», «canlı»
«bekar değil», «evli» anlamına gelir

Ters karşıtlık (yön gösteren) Sözcüklerden biri bir yönde diğeri ise onun karşıtı yönünde bir hareketi gösteriyorsa
sağ/sol, itmek/çekmek

Bakışimli karşıtlık İki sözcük arasında birbirinin karşıtını içermeye ilişkisi vardır
öğretmen/öğrenci, doktor/hasta
Ahmet Ali'ye borç verdi . Ali Ahmet'ten borç aldı



Eş Soslilik - (Homophony)

- ❖ Kimi sözcükler aynı ses ve yazı biçimde olmalarına rağmen birbirlerinden tümüyle bağımsız kavramsal içerik ve anlamsal özelliklere sahiptirler

Örnek

‘yüz’

yüz (I) a. 1. Doksan dokuzdan sonra gelen sayının adı

yüz (II) a. 1. Başta, alın, göz, burun, ağız, yanak ve çenenin
bulunduğu ön bölüm, sima, çehre, surat

yüz (III) e. 1. Kol, bacak, yüzgeç vb. organların özel hareketleriyle su
yüzeyinde veya su içinde ilerlemek, durmak

yüz (IV) e. (-i) 1. Derisini çıkarmak, derisini soymak



Eş Anlamlılık - (Polysemy)

- ❖ Bir göstergenin yansıttığı temel anlamın yanı sıra yeni yeni kavramları da anlatır durumda olması
- ❖ Eğer bir sözcükten benzetme yoluyla başka anlamlar türetilirse ortaya çıkan sözcükler arasındaki ilişki çok anlamlılık ilişkisidir

Örnek

YÜZ insan yüzü, yorgan yüzü

Baş

- ✓ İnsan ve hayvanlarda beyin, göz, kulak, burun, ağız gibi organları kapsayan vücudun üst bölümünde bulunan organ (Salona girerken başının üzerine çiçekleri attı)
- ✓ Bir topluluğu yöneten kimse (devletin başı)
- ✓ Başlangıç (satır başı)
- ✓ Temel (her şeyin başı sağlıktır)
- ✓ Arazide en yüksek nokta, tepe (yokuşun başı)
- ✓ Yiyeceklerde tane (üç baş soğan)



Eğretileme - (Metaphor)

❖ Bir kavramı diğer bir kavram üzerinden anlamak/anlatmak için kullanılır

- Eğer bir kavram diğeri üzerinden benzetme ile ifade edilirse iki kavram arasındaki anlamsal ilişki eğretilemedir

Örnek

Borç içinde yüzmek

Borç => deniz

Borçlu olmanın miktarı => denizin derinliği

Denizin içinde yüzmek => çok fazla borcun içinde olmak

Eğretileme

Kan beynine sıçramak

İçinde öfke biriktirmek

Öfke kusmak

Burnundan ateşler saçmak

Tepesinin taşı atmak

Öfkeyle dolmak



Parça-bütün İlişkisi (Part-whole)

- ❖ Bir bütünün parçası ile bütün arasındaki ilişkiyi anlatır
 - Sözcükler bir bütün ile onun parçalarını gösteriyorsa, kurdukları anlamsal ilişkidir

Örnek

masa-çekmece

kitap-sayfa

ev-çatı



Rakamlarla Wordnet

Category	Unique Forms	# of Senses
Noun	117,097	145,104
Verb	11,488	24,890
Adjective	22,141	31,302
Adverb	4,601	5,720



WordNet Noun Relations

Relation	Definition	Example
Hypernym	From concepts to superordinates	<i>breakfast</i> → <i>meal</i>
Hyponym	From concepts to subtypes	<i>meal</i> → <i>lunch</i>
Has-Member	From groups to their members	<i>faculty</i> → <i>professor</i>
Member-Of	From members to their groups	<i>copilot</i> → <i>crew</i>
Has-Part	From wholes to parts	<i>table</i> → <i>leg</i>
Part-Of	From parts to wholes	<i>course</i> → <i>meal</i>
Antonym	Opposites	<i>leader</i> → <i>follower</i>

Hyponym (alt kavram)/Hypernym (üst kavram) (Is-A relationships)
Bütün ilişkilerin %60'ı



Verb Relations in WordNet

Relation	Definition	Example
Hypernym	From events to superordinate events	<i>fly</i> → <i>travel</i>
Troponym	From events to their subtypes	<i>walk</i> → <i>stroll</i>
Entails	From events to the events they entail	<i>snore</i> → <i>sleep</i>
Antonym	Opposites	<i>increase</i> ⇔ <i>decrease</i>

Adjective & Adverb Relations in WordNet

Relation	Definition	Example
Antonym	Opposite	<i>heavy</i> ⇔ <i>light</i>
Adverb	Opposite	<i>quickly</i> ⇔ <i>slowly</i>

Troponym: bir şeyin nasıl yapıldığı çeşidini gösterir

Entails: bir şeyin nasıl yapıldığını gösterir



Wordnet'le anlamsal benzerlik bulma

- İki temel yaklaşım
 - Bağ sayma
 - Taksonomi yeterli
 - Ortak / Müşterek bilgi (Mutual Information)
 - Taksonomi ve corpus kullanır



Leacock & Chodorow (1998)

$$sim_{LC}(c_1, c_2) = -\log \frac{len(c_1, c_2)}{2L}$$

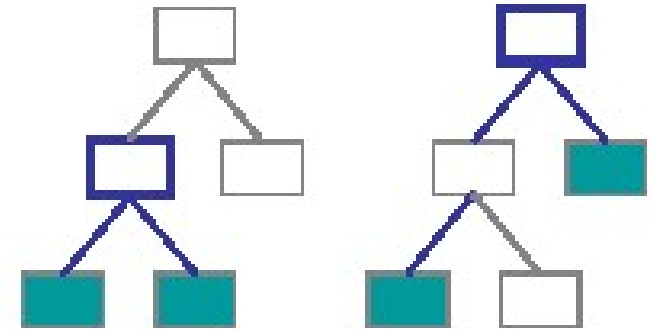
- *len(c1,c2)* iki synset arasındaki en kısa yolun uzunluğu (*benzerlik değeriyle ters orantılı*)
- *L*, tüm taksonominin derinliği



Wu & Palmer (1994)

$$\text{sim}_{\text{Wu\&Palmer}}(c_1, c_2) = \frac{2 \times N_3}{N_1 + N_2 + 2 \times N_3}$$

- N_1 ve N_2 , en yakın ortak üst synset'lerine iki synsetin IS-A bağlarıyla uzaklıkları (*benzerlik değeriyle ters orantılı*)
- N_3 , en yakın ortak üst synset'in kök synset'e IS-A bağlarıyla uzaklığı (*büyüklüğü ortak synset'in spesifikliğini gösterir*)



Jiang-Conrath (1997)- Lin (1998)

Diagram illustrating the relationship between two similarity measures, sim_K and sim_L , using the analogy of a fruit basket.

The basket is divided into two sections: A (green) and B (purple).

Section A contains 2 apples and 1 banana.

Section B contains 1 apple and 1 banana.

The formula for $sim_K(c_1, c_2)$ is shown as:

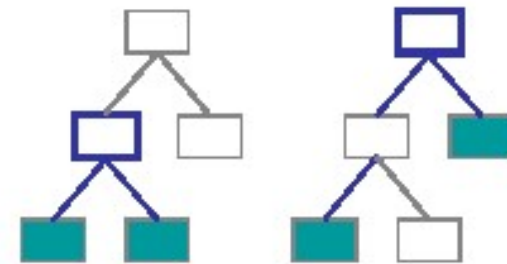
$$sim_K(c_1, c_2) = \frac{2 \log(p(lcs(c_1, c_2)))}{\log(p(c_1)) + \log(p(c_2))}$$

The formula for $sim_L(c_1, c_2)$ is shown as:

$$sim_L(c_1, c_2) = \frac{2 * \log(p(lcs(c_1, c_2)))}{\log(p(c_1)) + \log(p(c_2))}$$

The diagram shows that sim_K is a simplified version of sim_L , where the denominator is not part of the fraction.

$lcs(c1,c2)$ en yakın ortak üst synset



- A, her iki kavramı da içeren en spesifik kavramı kullanır
(İki kavramın beraber geçtiği doküman sayısına benzer)
- B, iki kavramdan herhangi birini içeren doküman sayısına benzer



Benzerlik Metotlarının Karşılaştırılması

- Bütün metotlar İngilizce 38 kelime çiftine uygulanmış
- Bulunan benzerlik değerlerinin, insan yargılarıyla olan korelasyonları yandaki tabloda verilmiştir

<i>Method</i>	<i>Type</i>	<i>Correlation</i>
Wu & Palmer 1994	Edge Counting	0.74
Li 2003	Edge Counting	0.82
Leacock & Chodorow 1998	Edge Counting	0.82
Resnik 1999	Info. Content	0.79
Lin 1998	Info. Content	0.82
Lord 2003	Info. Content	0.79
Jiang & Conrath 1998	Info. Content	0.83
Tversky 1977	Feature Based	0.73
Adapted Lesk 2002	Feature Based	0.37*
Rodriguez 2003	Hybrid	0.71



Kaynaklar

- Steve Vincent, Text Extraction, Similarity and WordNet
- http://www.ise.gmu.edu/~carlotta/teaching/INFS-795-s06/readings/Similarity_in_WordNet.pdf
- <http://www.cs.utah.edu/~sidd/documents/mstthesis03ppt.pdf>
- <https://starlangyazilim.com/kelime-anlam-belirsizligi-giderme-nedir/>
- ‘Türkçe Doğal Dil İşleme’, Eşref Adalı, Akçağ, 2020
- <https://acikders.ankara.edu.tr/>



Son ☹

