

Hesaplamalı Anlambilim Ders Notları

# Hayat Bilgisi Veri tabanları

Mehmet Fatih AMASYALI  
Hesaplamalı Anlambilim  
Ders Notları  
BÖLÜM 2



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Bölüm 2

- Hayat Bilgisi (Commonsense) Nedir?
- Neden ihtiyaç var?
- Benzer çalışmalar
- Türkçe Hayat Bilgisi Veritabanı
  - Kaynaklar
  - Veri tabanının oluşturulması
  - Kavramlar arası ilişki türleri
- CSoyun
- Otomatik Bilgi Çıkarımı



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Hayat Bilgisi (Commonsense) Nedir?

- Herkesin bildiğini varsaydığımız bilgiler
  - İnsanlar her gün yemek yer.
  - İnsanlar geceleri uyur.
  - Balıklar denizde yaşar.
  - Elma yuvarlıktır.
  - Kediler miyavlar.



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Hayat Bilgisi Veri tabanlarına Neden ihtiyaç var?

- Bilgisayarlara herşeyi söylemek gerekir. Söylediklerimizden söylemediklerimizi anlayamazlar / tahmin edemezler.
- Bunu yapabilmenin yolu, bilgisayarlara Hayat Bilgisini vermek.
- Bu sayede bilgisayarlar da insanlar gibi “Ali lokantada yemek yedi” cümlesinden “Ali para ödemiştir” bilgisini çıkarabileceklerdir.



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Bir hikayemiz var: Kurt ve Kız

- Küçük bir kız babaannesini ziyaret etmek için ormanda yürümeye başlıyor. Çalılıkların arkasındaki aç bir kurt kızı görüyor. Onu yemeğe karar veriyor. Küçük kıza tam saldıracakken, oduncuların şarkılarını duyuyor. Çalılıkların ardında saklanıp kararından vazgeçiyor.
- Problem: Aşağıdakileri açıklayınız.
  - Kurt kararından neden vazgeçti?
  - Küçük kız korktu mu?
- Contributed by [Pat Hayes](#) and [Lokendra Shastri](#) (9th July, 1997)
- More examples at <http://www-formal.stanford.edu/leora/commonsense/>



NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Duygudurum analizi

- Yemeğim 5 dakikada geldi. Çok teşekkürler XXX.
- Yemeğim 55 dakikada geldi. Çok teşekkürler XXX.
- XXX sayesinde 1 saatlik filmi 3 saatte izledim.
- XXX sayesinde 1 saatte 5 film indirdim.



NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

Hayat Bilgisi Veri Tabanları nasıl oluşturulur?

## 2 temel yaklaşım

- Bilgileri, az sayıda bilgi mühendisi özenle girsin yaklaşımı (Cyc, 20 yılda 1.5 milyon kaliteli bilgi, [www.cycfoundation.org/concepts](http://www.cycfoundation.org/concepts) )
- Bilgileri, çok sayıda web kullanıcısı girsin yaklaşımı (OpenMind, 8000 kişi, 3 yılda 1 milyon bilgi, [www.openmind.org](http://www.openmind.org) )



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Popüler Çalışmalar

- Cyc (1984, anlamsal ağ, uzman kişiler, elle)
- Wordnet (1985, anlamsal ağ, uzman kişiler, elle)
- OpenMind / ConceptNet (2002, gönüllü kullanıcılar, otomatik)
- NELL (2010, hayat bilgisi, yarı otomatik)



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

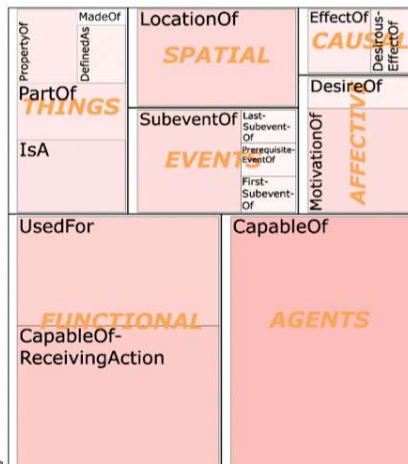
NOVA Research Lab

## ConceptNet'in oluşturulması

- Çıkarım aşaması
  - Gönüllü kullanıcıları serbest formda girdikleri cümlelerden (OMCS corpus)
  - Cümle → bilgi üçlüsü
  - Tüm bilgilerin %90'ı
- Zenginleştirme aşaması
  - Mevcut bilgilerden yeni bilgiler üretmek
  - Tüm bilgilerin %10'u



## ConceptNet'in Yapısı

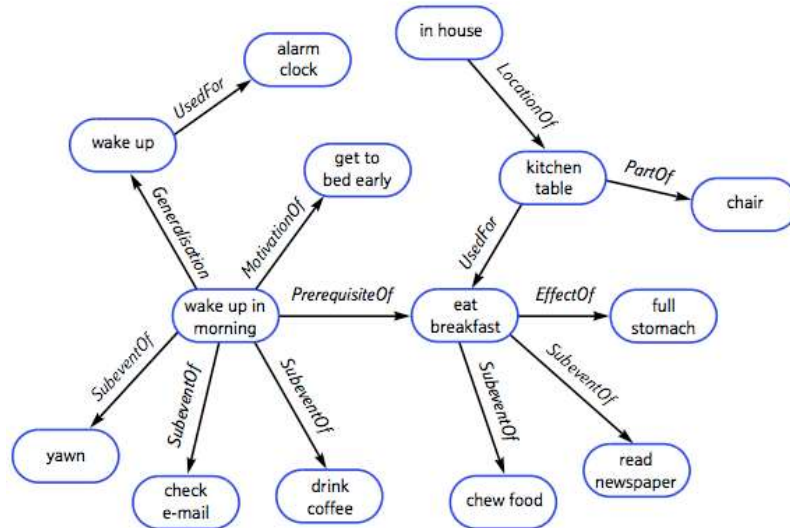


- 1.6 milyon (kavram-ilişki-kavram) üçlüsü
- 20 ilişki türü



Hesaplamalı Anlambilim Ders Notları

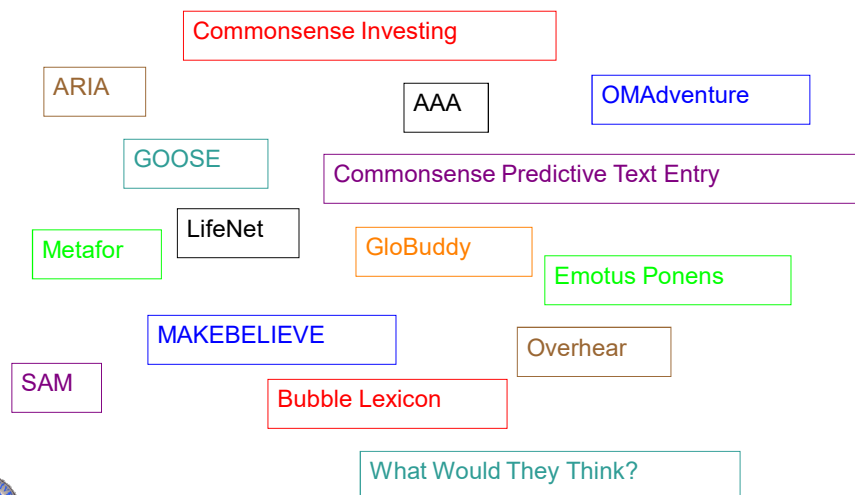
## ConceptNet örneği



NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## ConceptNet Uygulamaları



NOVA Research Lab

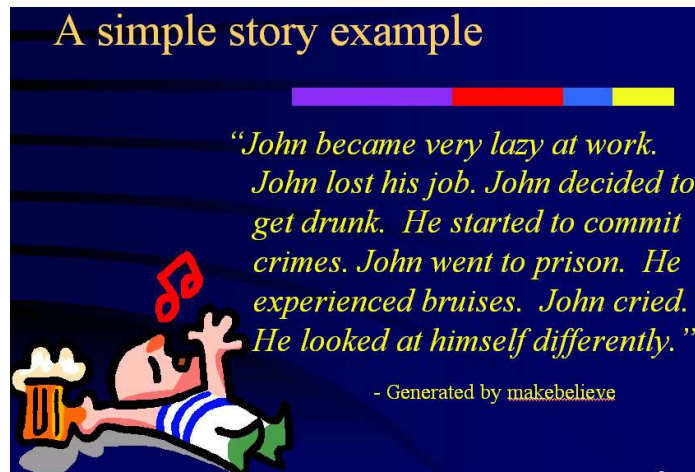
## Commonsense ARIA

- Kullanıcısının yazdığı email'i inceleyerek içeriğine uygun fotoğraflar önerir / ekler.



NOVA Research Lab

## MakeBelieve (Hikaye Üretici)



NOVA Research Lab

## Düşünme Zamanı

- X kişilik gruplar oluşturun.
- Hayat bilgisi toplamak, toplanmış iyileştirmek ve olası uygulamalar için yeni fikirler geliştirip listeleyin.
- 20 dakika



## Never-Ending Language Learning - NELL

<http://rtw.ml.cmu.edu/rtw/>

- 2010 yılında çalışmaya başlayan, web sayfaları üzerinde iteratif olarak çalışan, yarı otomatik, sürekli yeni ilişkiler öğrenen bir sistemdir.
- Sistem morfolojik çözümleme şablonları, HTML kodu şablonları ve metin şablonları kullanarak bilgi toplamaktadır.
- Sisteme başlangıçta her ilişki türü için doğruluğu bilinen kavram ikilileri verilmekte, sistem buradan çeşitli şablonları öğrenmekte ve bu şablonlardan yeni kavram ikilileri bulmaktadır. Bu yeni kavram ikililerinden yeni şablonlar çıkarmakta ve bu çevrim sürmektedir.





# NELL

- Bu çevrim arttıkça üretilen şablonların ve bunlardan üretilen ilişkilerin doğruluklarının keskin bir şekilde düştüğünü gören tasarımcılar her şablonun bir kısıtı olması gerektiğine karar vermişlerdir. Örneğin “bu nerede yapılır” ilişkisi için bulunan kavram ikililerinden ilkinin bir eylem, ikincisinin bir mekan olması kısıtını koyup bulunan kavram ikilisi bu kısıtlara uyuyorsa veri tabanına dahil etmişlerdir.
- Sistem tasarımcıları ayrıca, bu kısıtlara uyan ve uymayan kavram ikililerinin sayılarını kullanarak şablonlara ait güvenilirlik katsayıları hesaplamışlardır. Tasarımcılar, sistemin bu yeni haliyle veri kümesinin boyutu büyüdükçe güvenilirliğinin keskin biçimde azalmasının önüne geçmişlerdir.



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

# NELL

So far, NELL has accumulated over 15 million candidate beliefs by reading the web, and it is considering these at different levels of confidence. NELL has high confidence in 1,027,444 of these beliefs — these are displayed on this website. It is not perfect, but NELL is learning. You can track NELL's progress below or @cmunell on Twitter, browse and download its knowledge base, read more about our technical approach, or join the discussion group.

**Recently-Learned Facts** [twitter](#) [Refresh](#)

instance	iteration	date learned	confidence
zarah_fiske_road is a street	526	08-mar-2012	99.3
ksnf is a company	526	08-mar-2012	100.0
ford_super_duty_f_53_motorhome is a model of automobile	524	05-mar-2012	100.0
squishiness is a physical characteristic	524	05-mar-2012	93.8
bill_bradley is a U.S. politician	524	05-mar-2012	100.0
professionals is a profession that is a kind of doctors	529	12-mar-2012	100.0
jeffrey_r_immelt is the CEO of general electric	529	12-mar-2012	100.0
jerry_stackhouse coaches the team dallas_mavericks	529	12-mar-2012	93.8
guns is a type of weapon produced in u_s	529	12-mar-2012	99.8
eastern_european_countries is a country located in the geopolitical location bulgaria	529	12-mar-2012	93.8

Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

## Hesaplamalı Anlambilim Ders Notları

## NELL

**NELL Knowledge Base Browser**

log in | preferences | help/instructions | feedback

**movie**  
(category)

See [metadata](#) for movie  
4,976 instances, 1 page

A release of a visual narrative composed of a series of moving images that is considered a complete presentation and is intended to be classified as a film. A film can be of any length of running time and presented in a theatrical, television, internet-streaming and direct-to-home-video presentation. (FreeBase)

instance	iteration	date learned	confidence
<a href="#">aliens</a>	0	28-oct-2010	(Seed) 100.0
<a href="#">an_american_in_paris</a>	0	28-oct-2010	(Seed) 100.0
<a href="#">beowulf</a>	161	09-nov-2010	100.0
<a href="#">big_lebowski</a>	120	19-jun-2010	(Seed) 100.0
<a href="#">boogie_nights</a>	5	12-jan-2010	100.0
<a href="#">close_encounters</a>	1	12-jan-2010	100.0
<a href="#">collateral</a>	161	09-nov-2010	100.0
<a href="#">click_tracy</a>	4	12-jan-2010	100.0
<a href="#">die_hard</a>	120	19-jun-2010	(Seed) 100.0
<a href="#">doctor_zhivago</a>	161	09-nov-2010	100.0
<a href="#">empire_strikes_back</a>	120	19-jun-2010	(Seed) 100.0
<a href="#">ferris_bueller_s_day_off</a>	255	19-may-2011	(Seed) 100.0
<a href="#">fight_club</a>	120	19-jun-2010	(Seed) 100.0
<a href="#">forrest_gump</a>	120	19-jun-2010	(Seed) 100.0
<a href="#">goblet_of_fire</a>	161	09-nov-2010	100.0

Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

## Hesaplamalı Anlambilim Ders Notları

## NELL

**NELL**  
@cmunell

I am a machine reading research project at Carnegie Mellon, periodically tweeting facts I read. Please follow me, and reply with corrections so I can improve! Pittsburgh PA - http://tw.mt.cmu.edu

8,146 TWEETS  
1 TAKIP ETKİLERİ  
1,863 TAKİPÇİLERİ

**NELL ile iletişimi sürdür**  
Bugün Twitter'a katıl  
Ad ve soyad  
E-posta  
Şifre  
Kaydol

**Tweetler**

- NELL (@cmunell) I think "Hatazepam" is a #Drug (bit.ly/ybtpHn) 7 dk
- NELL (@cmunell) I think "rockwell kent" is a #VisualArtist (bit.ly/vvVxst) 1 sa
- NELL (@cmunell) I think "Janata\_dal\_united\_party" is a #PoliticalOrganization that is not part of the government (bit.ly/z7K1Sb) 3 sa
- NELL (@cmunell) 4 sa

Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

- 2010'dan bugüne 50 milyon aday ilişki öğrendi.
- Bunların 1.8 milyonunun güvenilirliği yüksek.

## NELL

- İnanç/güvenilirlik nasıl ölçülür?
  - Şablon güvenilirliği
    - Ürettiği ikililerin doğruluk oranı
  - İlişki güvenilirliği
    - Üretildiği şablonun güvenilirliği
    - Kaç farklı şablondan üretildiği
    - Kullanıcı geri beslemeleri

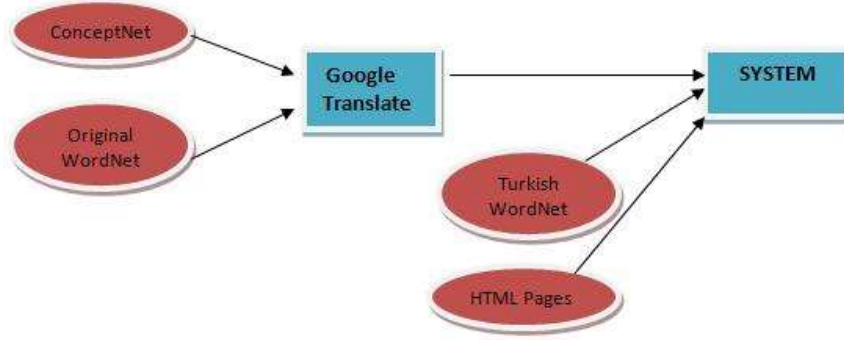


## Türkçe Hayat Bilgisi Veritabanı

- Literatürdeki çalışmalar genelde İngilizce için
- Türkçe için bir ilk
- Hangi yaklaşım kullanılmalı:
  - Cyc, OpenMind?
  - **Mevcutları tercüme etmek / kullanmak**
  - **Frekans tabanlı bilgi çıkarımı**
- Veriler nasıl tutulmalı?
  - (Kavram - ilişki türü – kavram) üçlüleri



## Veri Kaynakları



## Veri tabanının oluşturulması

- Kaynaklardaki (ConceptNet, Wordnet) format farklılıkları, basit bir veri formatı kullanılarak çözüldü.
- Otomatik çeviri (problemlili ancak en az maliyetlisi bu)
- Frekans tabanlı bilgi çıkarımı (400 bin web sitesinin html kodu üzerinde Zemberek)



## Hesaplamalı Anlambilim Ders Notları

## Veri tabanından örnekler

<i>Bunun için ne gerekir?</i>	<i>Neye sebep olur?</i>	<i>Bundan neler yapılır?</i>
yazmak-araştırmak	öldürmek-ceza	taş-köprü
denemek-para	doğurmak-hayat	çelik-makine
uyumak-yatmak	sevmek-umut	su-bulut
seyahat etmek-enerji	sevmek-acı	kağıt-gazete
öğrenmek-okumak	ateş-acı	yün-kumaş
yaşam-yiyecek	öldürmek-üzüntü	kumaş-gömlek
<i>Ne için kullanılır?</i>	<i>Bu ne yapabilir?</i>	<i>Nerede bulunur?</i>
asker-savaş	kuş-uçmak	oda-bina
çatal-yemek	kişi-yürümek	kişi-oda
top-oynamak	bilgisayar-düşünmek	elbise-mağaza
ördek-yemek	çocuk-düşmek	kemik-kişi
hastalık-öldürmek	bıçak-kesmek	asker-savaş
baş-düşünmek	gemi-batmak	öğrenci-okul



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

## Hesaplamalı Anlambilim Ders Notları

İlişki Türü	Concept Net	Orijinal Wordnet	Türkçe Wordnet	Web	İlişki Türü	Concept Net	Orijinal Wordnet	Türkçe Wordnet	Web
Ne için kullanılır?	36864	0	0	0	Yaklaşık Zıtanlamalı	0	0	1678	0
Bu ne yapabilir?	51549	0	0	0	Durumundadır	0	0	1546	0
Nerede bulunur?	30778	0	0	0	Bölümün Bütünü	0	27842	2385	0
Ne arzu eder?	5989	0	0	0	Üyenin Bütünü	0	57717	2907	0
Bunun için ne gerekir?	17822	0	0	0	Benzer Anlam	0	21999	504	0
Bunun ne özellikleri var?	11214	0	0	0	Parçanın Bütünü	0	0	230	0
Neyden yapılmış?	1000	0	0	0	Zıtanlamalı	0	3463	0	0
Neyin bir parçası?	8105	0	0	0	Sıfatın Eylemi	0	115	0	0
İçerdiği olaylar nelerdir?	20330	0	294	0	Birlikte geçmek	0	433	0	0
Bunun tanımı nedir?	2721	0	0	0	Bu neyi gerektirir?	0	1990	0	0
Neye sebep olur?	13010	907	237	0	Bunun içeriği nedir?	0	2349	0	0
Neyi ister?	7777	0	0	0	Sıfatın İsmi	0	1885	0	0
Hangi hedef için bu yapılır?	5297	0	0	0	İsim Hali	0	6087	0	0
Bunun için ilk önce ne yaparsın?	3147	0	0	0	Fiil - Fiil	0	0	0	10255
Bu ne tarafından oluşturulur?	107	0	0	0	İsim - Fiil	0	0	0	200542
Buna neler yapılır/uygulanır?	145	0	0	0	İsim Tamlaması	0	0	0	3370
Bu hangi olayla biter?	2839	0	0	0	Sıfat - Fiil	0	0	0	16312
Eşanlamalı	0	124320	6999	0	Sıfat - Sıfat	0	0	0	3735
Üst Kavramıdır	34566	282137	24141	0	Sıfat - Tamlaması	0	0	0	25250
Benzer Fiiller	0	2807	758	0	Toplam ilişki sayısı	253260	534051	42455	259464
Alan adı nedir?	0	0	776	0					



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

# Problemler

- Otomatik çeviri hataları
  - Yanlış çeviri
  - Çeviri yapılamama
  - Çözüm ? Elle çeviri
- Bilgi kalitesi
  - Çözüm (çok sayıda kişiden geri besleme almak-CSoyun)



Gelişme 4%  
 Hoşgeldin Hızlı fatih amasyalı

SEVİYE	TOPLAM PUAN	ZAFER	113
11	23208	MUKABEL	11
		BOZGUN	0

Çıkış  
 Ayarlar

PUAN CANAVARLARI  
 650752  
 400977  
 23000  
 18015  
 4079

Bildiklerim Araştır Tahminlerim Bana Öğret Birlikte Eğitim

Tahminlerim  
 Dedikodu kazanına hoşgeldiniz!  
 İşte sizinle paylaşmak  
 istediğim bir tahmin var  
 karşınızda. Tabiki bu sadece  
 bir tahmin ve bana göre doğru  
 ya size göre? Benimle aynı  
 düşünüyorsanız Doğru  
 butonuna, farklı bir  
 düşüncedeyse Yanlış  
 butonuna, konu hakkında  
 fikriniz yoksa Ben Bilmem  
 butonuna tıklayabilirsiniz.

hayal  
 Bu neyin ilk adımıdır?  
 yaşamak

Doğru Yanlış Ben Bilmem



*Davetlisiniz!*

## CSoyun

- Amaç: CSdb'deki bilgi sayısını ve kalitesini arttırmak
- Kullanıcıları teşvik etmek için puan aldıkları 5 oyun tasarlandı.
- Kullanıcılar sisteme verdikleri her geri besleme için puan alıyor.
- Kullanıcılar yeni ilişkiler ve kavramlar girebiliyor ki bu sayede bilgi sayısı artıyor.
- Geri beslemelerle ilişkilerin güvenilirlikleri / kalitesi artıyor.
- Sistem veri tabanında olmayan yeni ilişkiler üretip geri besleme alabiliyor.



## CSoyun'da

- **Bildiklerim:** Kullanıcı öncelikle bir ilişki türü seçer. Sistem CSdb'den o ilişkiye sahip kavram ikililerini rastgele kullanıcıya gösterir. Kullanıcı bu bilgilere oy verir.



## CSoyun'da

- **Araştır:** Kullanıcı öncelikle bir kavram girer ve bir ilişki türü seçer. Sistem CSdb'den o kavramla o ilişkiye sahip kavramları rastgele kullanıcıya gösterir. Kullanıcı bu bilgilere oy verir.

**CSOYUN** Gelişme: 30% SEVİYE: 21 TOPLAM PUAN: 121101 ZAFER: 113 MUKABİL: 15 BOZGUN: 1

Hoşgeldin Yenilikçi fatih amasyalı

**PUAN CANAVARLARI**

Ad	Puan
Serkan Özcan	650943
Caner Bey	401025
fatih amasyalı	121020
Abbas MEMİŞ	106703
Berker Nural	43760

**Sistem İstatistikleri**

İliski Sayısı: 1234689  
Kelime Sayısı: 700075

**Bildiklerim**

Kelime	İliski	Not
kedi	Nerede bulunur?	
ahır		0.0
Connecticut		0.0
temiz çamaşır		0.0
New Jersey		0.0
çarba		0.0
kedi göster		1.0
kum kutusu		0.0
restoran		0.0
kanarya kafesi		0.0

**Araştır**

Çok g...  
mi ist...  
heme...  
saldı...  
neyim...  
Eğer...  
mera...  
bir ta...  
bunu...  
Sizni...  
bilgile...  
yardı...  
diğer...  
çok bi...  
eluru...

## CSoyun'da

- **Tahminlerim:** Bu oyun Csoyun'u benzerli sistemlerden ayıran en önemli özelliğidir. Sistemde, CSdb'de yer alan relationlar kullanılarak non-monotonic reasoning yapılarak CSdb'de yer almayan hipotezler üretilir ve hipotezler tablosuna kaydedilir. Hipotezlerin 4 farklı üretilme çeşidi Tablo 'da verilmiştir. Hipotezler onaylandıklarında relations tablosuna kaydedilmektedirler.
- **Monotonic çıkarım nedir?**
- **non-monotonic çıkarım nedir?**

Çıkarım türü	Örnek çıkarımlar
(X-R1-Y) and (Y-R2-Z) → (X-R3-Z)	(martı - altkavramıdır - kuş) ve (martı - bu ne yapabilir - uçmak) → (kuş - bu ne yapabilir - uçmak)
(X-R1-Y) → (X-R2-Y)	(sinemaya gitmek - içerdiği olaylar nelerdir - mısır yemek) → (sinemaya gitmek - bunun için ilk önce ne yaparsın - mısır yemek)
(X-R1-T) and (Z-R1-T) and (X-R2-Y) → (Z-R2-Y)	(Kedi - altkavramıdır - memeli) ve (köpek - altkavramıdır - memeli) ve (kedi - sahiptir - kürk) → (köpek - sahiptir - kürk)
(X-R1-Y) and (X-R2-Z) → (Z-R3-Y)	(bıçak - nerede bulunur - mutfak) , (bıçak - ne için kullanılır - kesmek) → (kesmek - nerede yapılır - mutfak)





## Tahminlerim

CSoyun

Gelişme 4%

SEVİYE 11

TOPLAM PUAN 23208

ZAFER 113

MUKABİL 11

BOZGUN 0

Çıkış

Ayarlar

Bildiklerim Araştır Tahminlerim Bana Öğret Birlikte Eğitim

PUAN CANAVARLARI

Serkan Özcan 650752

Caner Bey 400977

fatih amasyalı 23000

Abbas MEMİŞ 18015

Sinem Topcu 4079

Sistem İstisnaları

Diğer Sayısı: 1094552

Kelime Sayısı: 640646

Kullanıcı Sayısı: 32

Listeyi Güncelle

hayal

Bu neyin ilk adımdır?

yaşamak

Doğru Yanlış

Ben Bilmem

Tahminlerim

Dedikodu kazanma heggeldiniz!  
İzne sizlerle paylaşmak  
istediğim bir tahmin var:  
karpuzda. Tabiki bu sadece  
bir tahmin ve bana göre doğru  
ya size göre? Benimle aynı  
düşüncedeyseniz Doğru  
butonuna, farklı bir  
düşüncedeyseniz Yanlış  
butonuna, kuma hakikide  
fikriniz yoksa Ben Bilmem  
butonuna tıklayabilirsiniz.



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

## CSoyun'da

- **Bana Öğret:** Kullanıcıya bir kavram ve bir ilişki türü verilir. Bu kavramla bu ilişki türüne sahip kavramlar istenir. Kullanıcının verdiği bilgi eğer CSdb'de bulunuyorsa kullanıcı extra puan alır.

Windows Internet Explorer

CS http://www.kemikayun.yildiz.edu.tr/commonsense/csoyun.php#

CSoyun

Gelişme 30%

SEVİYE 21

TOPLAM PUAN 121133

ZAFER 113

MUKABİL 16

BOZGUN 1

Çıkış

Ayarlar

Bildiklerim Araştır Tahminlerim Bana Öğret Birlikte Eğitim

PUAN CANAVARLARI

Serkan Özcan 650943

Caner Bey 400977

fatih amasyalı 23000

Abbas MEMİŞ 18015

Sinem Topcu 4079

Sistem İstisnaları

Diğer Sayısı: 1234690

Oyna Bula Oyna Bula

Oyunu Bula

Oyun başladı...

SORU: su Neyden yapılmış?

Kelime Giriniz:

Giriş Oluştur Kelimeler

değin

hığın

Bana Öğret

Hegeldiniz öğretilen Bana bir harf öğretilen 40 yıl aynı malımla temizlenir. Kaldığı her günün dedi mi? Harf ileylese hemer Oyuna Bula bulaşma basarak bulaş bulaşlar öğretilir. Siz veriyorum dediği aile aile dileyeyiz. Sordular amiralar ceng vererek gın bulduki olanı cengleri yapıp enter tapu basabiliriz.

İletişim Merkezi  
İnternet arşivi

18.03.2012

09.03.2012



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

## CSoyun'da

- **Birlikte Eğitim:** Bana öğret oyununun iki kişinin karşılıklı oynadığı versiyonudur. Sistemdeki kullanıcılar ikişerli eşleştirilir. Eşleşen kullanıcılara aynı soru (bu kavramla bu ilişki türüne sahip kavramlar nelerdir?) sorulur. Verilen aynı cevaplar için kullanıcılar extra puan alır.



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

## Birlikte Eğitim

The screenshot shows the CSoyun web application interface. The header includes the CSoyun logo, a progress bar at 34%, and a sidebar with user avatars and names. The main content area displays a question about carbon and a list of related terms. The right sidebar contains a 'Birlikte Eğitim' dropdown menu and a description of the game.



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

## CSoyun

- Tasarlanan 5 oyunla sistem müdahale gerektirmeyen, kendi kendine gelişebilir bir yapıya kavuşmuştur. Bilgilerin güvenilirlikleri kullanıcı oylarıyla belirlenmekte, yeni ilişkiler tahminlerim, bana öğret ve birlikte eğitim oyunları ile üretilmekte, yeni kavramlar bana öğret ve birlikte eğitim oyunları ile CSdb'ye girilmektedir. Bu sayede zaman içerisinde (oyunlar oynandıkça) sistemin içerdiği bilgi miktarı ve bilgilerinin güvenilirliği artacaktır.
- Sistemde ayrıca oyunların parametrelerinin (kullanılacak ilişki türleri, oyun süreleri, oyun puanları vs.) ayarlanması için bir de admin ekranları tasarlanmıştır.
- Sistemde (kavram-ilişki türü-kavram) üçlülerine ait toplam oy ve toplam puan bilgileri tutulmaktadır. Hesaplanan ortalama puanlarla bilgi üçlülerinin doğruluk değerleri hesaplanmaktadır.



## CSoyun istatistikleri

Sistemin açılışı: Ocak 2010

	1.rapor	2.Rapor
Rapor Zamanı	18 Nisan 2010	23 Şubat 2012
Oy Sayısı	34 bin	155 bin
Kullanıcı Sayısı	65	182
İlişki türü sayısı	40	45
İlişki sayısı	1.09 milyon	1.2 milyon
Kavram sayısı	475 bin	700 bin

Bir ilişkinin doğruluk oyu en düşük 0, en yüksek 5 olarak verilmektedir. Verilen oyların ortalaması 5 üzerinden yaklaşık 3.1'dir. Kullanıcılar tarafından oylanan tekil ilişki sayısı yaklaşık 99 bin'dir. Buna göre oyunun kullanıma açıldığı 2 yıl boyunca veri tabanındaki ilişkilerin %8.3'ü oylanmıştır.



Hesaplamalı Anlambilim Ders Notları

## İlişki türlerine göre ortalama puanlar

İlişki Türü	Oy Sayısı	Ort. Puan	İlişki Türü	Oy Sayısı	Ort. Puan
Nerede bulunur?	4450	3.059	Bunun için ilk önce ne yaparsın?	437	4.797
Eşanlımlı	4163	1.557	Üyenin Bütünü	419	2.138
Kullanım amacı nedir? Ne için kullanılır?	2843	3.222	Buna neler yapılır/uygulanır?	408	4.883
Bunun için ne gerekir?	2583	3.293	Sıfat - Sıfat	374	2.210
Neye sebep olur?	2184	2.839	Bunun ne özellikleri var?	355	4.919
Bu ne yapabilir?	1728	3.028	İsim Tamlaması	338	3.641
Ne arzu eder?	1413	2.654	Parçanın Bütünü	292	4.523
İçerdiği olaylar nelerdir?	1339	3.645	Benzer Anlam	290	2.864
Bölümün Bütünü	1181	1.997	Bunun içeriği nedir?	283	3.638
Üst Kavramdır	1168	2.755	Sıfat - Fiil	259	2.813
Neyin bir parçası?	871	4.439	İsim - Fiil	213	2.298
Benzer Fiiller	857	1.503	Sıfat - Tamlaması	201	4.209
Zıtanlamalı	840	4.641	Durumdadır	84	4.357
Hangi hedef için bu yapılır?	815	3.885	Bunun tanımı nedir?	58	5.000
Neyden yapılmış?	813	4.618	Birlikte geçmek	53	5.000
Bu neyi gerektirir?	614	3.441	Sıfatın İsmi	30	5.000
Bu hangi olayla biter?	569	4.224	Alan adı nedir?	24	5.000
Kimin/Neyin tarafından yaratılır/oluşturulur?	521	4.796	İsim Hali	20	4.300
Neyi ister?	456	4.309	Sıfatın Eylemi	15	5.000
Yaklaşık Zıtanlamalı	440	4.332	Fiil - Fiil	7	5.000

Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Hayat bilgisi Veri tabanlarının kalitesi nasıl ölçülür?

- Doğruluk: Sistemden rastgele seçilecek 100 ilişkiden %X'inin bir insan tarafından doğru olarak nitelendirilmesi
- Kapsayıcılık: Gereken tahmini sayı milyonlar.
  - Kavram sayısının yeterliliği
  - İlişki sayısının yeterliliği
- Bu ölçümlerle, sürekli gelişen bir sistemin kapsayıcılık ve doğruluk performanslarının zaman içindeki değişimi izlenebilir.

Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Hayat bilgisi Veri tabanlarının kalitesi nasıl ölçülür?

- Kavram sayısının yeterliliği:
  - Büyük metin kütüphanelerindeki kavramların % kaçını mevcut?
- İlişki sayısının yeterliliği:
  - Bir insanın rastgele verilen bir kavram hakkında 1 dakikada üreteceği ilişkilerin % kaçını mevcut?
  - Verilen bir ilişki türü hakkında bir insanın üreteceği ilişkilerin % kaçını mevcut?
  - İnsanlı ölçümlerde tek bir kişi yetmez



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Hayat Bilgisi Veritabanlarının Uygulama alanları

- Sorgu zenginleştirme
- Metin sınıflandırma
- Soru cevaplama
- Metin Anlama
- Otomatik çıkarım (Textual Entailment)
- Uzman sistemler



NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Bilgi Çıkarımı Nedir? Doğal Dil İşleme Uygulamalarının Neresinde Yer Alır?

- Sayfa Arama: Sorguya karşılık gelen Web sayfalarını bulma
- Paragraf Arama: Sorguya karşılık gelen paragrafları bulma
- **Bilgi Çıkarımı:** Önceden tanımlanmış şablonlara uygun bilgileri bulma
- Soru Cevaplama: Kullanıcı sorusunun cevabını bulma
- Metin Anlama: Metinleri insanların anlayabildiği gibi anlama



NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Bilgi Çıkarımı - örnek

- 21 yaşındaki inşaat işçisi Kazım Yaprak, evine dönerken para meselesi yüzünden tartıştığı arkadaşı Hilmi Baker tarafından bıçaklanarak öldürüldü.
- **Katil:** Hilmi Baker
- **Kurban:** Kazım Yaprak
- **Sebepe:** Para
- **Suç Aleti:** Bıçak



NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Otomatik Bilgi Çıkarımı - Temel Metotlar

- Düzenli ifadeleri (Regex) kullanmak
- Yazım dilindeki şablonları kullanmak
- Öğelerine ayrılmış metinleri kullanmak
- Dinamik web sayfalarındaki tekrarlanan HTML tag'lerini kullanmak
- Morfolojik çözümlemelerdeki şablonları kullanmak



NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Düzenli ifadeleri (Regex) kullanmak

DOCUMENT:	EXTRACTED DATA:
Capitol Hill- 1 br twnhme.	<Bedrooms: 1
D/W W/D. Pkg incl \$675.	Price: 675>
3BR upper flr no gar. \$995.	<Bedrooms: 3
(206) 999-9999  	Price: 995>
Extraction rule:	* (<Digit>) 'BR' * '\$' (<Nmb>)
Output:	Rental {Bedrooms @1} {Price @2}

Bedroom ::= ( br || brs || bdrm || bedrooms || bedroom )



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Yazım Dilindeki Şablonlar

- İstenen ilişki türü için şablonlar bulup büyük metinlerde bu şablonlara uyan çiftleri çıkarmak

[memurlar.NET.FORUM](http://memurlar.NET.FORUM)

Tüm Forumlar Mühendisler İnşaat, Çevre, Gıda ve Diğer Mühendislikler, Yönetici: Hiçkimse Sadece forum sahipleri yazabilir. ...  
forum.memurlar.net/forum.asp?F\_ID=49 - 27k - [Önbellek](#) - [Benzer sayfalar](#)

[Greenpeace ses kayıtları ve diğer multimedya içeriği](#)

... Greenpeace Uluslararası > Ana Sayfa > görüntü/ses/oyun > Ses kayıtları ve diğer multimedya içeriği. Ses kayıtları ve diğer multimedya türleri. ...  
www.greenpeace.org/turkey\_tr/ multimedia/media-view?type=gp\_other - 37k - [Önbellek](#) - [Beni](#)

[List 2000 : Kidz 2000 / Spor ve Diğer Aktiviteler /](#)

... On-line Alışveriş. Spor. Seyahat ve Turizm. List2000 : Reklam, List2000 : Reklam. 24 Kasım 2004, Çarşamba, Kidz 2000, Spor ve Diğer Aktiviteler / ...  
www.list2000.com.tr/tr/kidz\_2000/ sports\_other\_activities/index.asp - 24k - [Önbellek](#) - [Beni](#)

[DERGİLER VE DİĞER PERİYODİKLER](#)

DERGİLER VE DİĞER PERİYODİKLER. AKSİYON. Kuruluş Tarihi: 10 Aralık 1994. Niteliği: Haber-Haftalık. Adres:..... Çobançeşme Mah. Kalender Sok. ...  
www.byegm.gov.tr/TURKBASINI/dergiler-turkce.htm - 67k - [Önbellek](#) - [Benzer sayfalar](#)

[Birleşmiş Milletler Teşkilatı ve Organları](#)

... Bilim, güzel sanatlar ve diğer konular ile ilgili bilgi artırıcı yayımlar hazırlamak, başarılı doktora öğrencilerine Albert Einstein ödülü ...  
www.icep.org.tr/turkish/bm/index.asp - 25k - 7 Ara 2004 - [Önbellek](#) - [Benzer sayfalar](#)

[Türkiye İş Bankası | www.isbank.com.tr](#)

Yeni Türk Lirası. Sözleşmeler ve Diğer Hukuki Belgeler. ...  
www.isbank.com.tr/yti/yti-islem-sozlesme.html - 12k - [Önbellek](#) - [Benzer sayfalar](#)

[Ekol Bilgisayar Canon ve Diğer Lazer Toner Fiyat Listesi](#)

Ekol Bilgisayar Canon ve Diğer Lazer Kartuşları Fiyatları. Canon ve Diğer Lazer Tonerler Fiyat Listesi. CANON LASER PRINTERLER İÇİN. ...



NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Öğelerine ayrılmış metinler

- “ye” fiilinin nesneleri yiyecek olarak sınıflandırılabilir.

[geyiq.com/forum](http://geyiq.com/forum) - Taksim borsada bira ile yanm döner **yerken**

geyiq.com/forum > geyiq alanı > kıl oluyorum, dumur oldum > Taksim borsada bira ile yanm döner **yerken**. Orjinalini görmek için ...  
www.geyiq.com/forum/archive/index.php/t-10219.html - 3k - [Önbellek](#) - [Benzer sayfalar](#)

[Yerken Family Grave Search](#)

... It's like we've always known each other! - Pam from CA. Advertisement. Click Here. Search Page for Surname: **Yerken**. Name: First, Middle, **Yerken** Last. ...  
www.findagrave.com/surnames/y/**yerken**.html - 13k - [Önbellek](#) - [Benzer sayfalar](#)

[Hurriyetim](#)

... Kelebek, 25.05.2004. Ödülü, Bush kraker **yerken** söylemeyin, ... Umanım kimse ona bu ödülü kazandığını, o kraker **yerken** söylemez' diye yanıtladı. ...  
www.hurriyetim.com.tr/ haber0/,sid=436@mvid=416896.00.asp - 42k - [Önbellek](#) - [Benzer sayfalar](#)

[MILLİYET INTERNET - BUSINESS](#)

... Zeytin **yerken** alzheimer oluyorum haberimiz yok. Zeytini, zehirli tekstil boyası ile veya demir sülfat gübresi ile karatıp satıyorlar. ...  
www.milliyet.com.tr/2003/12/12/business/bus07.html - 30k - [Önbellek](#) - [Benzer sayfalar](#)

[TürkiyeOnline.com - Haber](#)

... sağlık. Mantar **yerken** dikkat Havalanın ısınmasıyla birlikte doğada ortaya çıkan mantarların bilimsiz olarak tüketilmesinin, zehirlenmelere neden ...  
www.turkiyeonline.com/haber/saglik/ haber.php?story=2004\_04\_02\_mantar - 20k - [Önbellek](#) - [Benzer sayfalar](#)



NOVA Research Lab



## Hesaplamalı Anlambilim Ders Notları

## “iç” fiilinin nesneleri içecek olarak sınıflandırılabilir.

### Web

#### [Su içerken okunacak dua](#)

Su **içerken** okunacak dua. Suyu Besmele ile üç nefesde içmeli, her nefesde ağzını bardaktan çekmelidir. Peygamber efendimiz de böyle yaparlardı. ...

[www.guzelislam.com/dinbilgiler/yn/dua/dua\\_su\\_icerken\\_okunacak\\_dua.htm](http://www.guzelislam.com/dinbilgiler/yn/dua/dua_su_icerken_okunacak_dua.htm) - 5k - [Önbellek](#) - ]

#### [Zemzem içerken okunacak dua](#)

Zemzem **içerken** okunacak dua. Zemzem suyu ikram edildiğinde, güzel kokuda olduğu gibi geri çevirmemeli, almalıdır. Su içtikten ...

[www.guzelislam.com/dinbilgiler/yn/dua/dua\\_zemzem\\_icerken\\_okunacak\\_dua.htm](http://www.guzelislam.com/dinbilgiler/yn/dua/dua_zemzem_icerken_okunacak_dua.htm) - 6k

[ [www.guzelislam.com](http://www.guzelislam.com) alanından daha fazla sonuç göster ]

#### [SIGARA WEB SİTESİ - FORSNET](#)

SIGARAYI BIRAKIN. ETRAFINIZDAKİ HERKES SIGARA **İÇERKEN** SİZ BIRAKABİLİRSİNİZ. Bu çok genel bir problemidir. Etrafınızda ...

[www.sigara.gen.tr/sigarayi\\_birakin/etraftakiler.html](http://www.sigara.gen.tr/sigarayi_birakin/etraftakiler.html) - 10k - [Önbellek](#) - [Benzer sayfalar](#)

#### [HaberX - 'Polisi, Peker ile çay içerken görürseniz ...](#)

'Polisi, Peker ile çay **içerken** görürseniz şaşırmayacaksınız' 10.10.2004, 03:06. Sedat Peker'in önceki gece tutuksuz yargılanmak ...

[www.haberx.com/n/155332/polisi-peker-ile-cay-icerken.htm](http://www.haberx.com/n/155332/polisi-peker-ile-cay-icerken.htm) - 17k - [Önbellek](#) - [Benzer sayfalar](#)

#### [geyiq.com/forum - Sigara İçerken Yorum Yapanlara ..](#)

geyiq.com/forum > Eğlence > kıl oluyorum, dumur oldum > Sigara **içerken** Yorum Yapanlara.. PDA. Orjinalini görmek için tıklayın ...

[www.geyiq.com/forum/archive/index.php/t-6883.html](http://www.geyiq.com/forum/archive/index.php/t-6883.html) - 11k - [Önbellek](#) - [Benzer sayfalar](#)

#### [Yeni Safak Online - Gülşen Ortacı : Kahve içerken iyi ...](#)

... Kahve **içerken** iyi düşünün. Kahve ve çay özellikle yetişkinler tarafından çokça ve bilinçsizce tüketiliyor. Oysa uyanıcı ...

[www.yenisafak.com.tr/arsiv/2002/haziran/27/gortac.html](http://www.yenisafak.com.tr/arsiv/2002/haziran/27/gortac.html) - 17k - [Önbellek](#) - [Benzer sayfalar](#)



NOVA Research Lab

## Hesaplamalı Anlambilim Ders Notları

## Şablonlar

- Üretkenlik: ürettiği bilgi miktarı
- Güvenilirlik: ürettiği bilgilerin doğruluk oranı
- İkisi birbiriyle çelişir



NOVA Research Lab

## Open domain yaklaşımı

- Belirli fiillerin en çok görülen şablonları çıkarılmış ve bu şablonlarla bilgi çıkarımı yapılmaktadır. Bu şablonlar rasgele metinlerden değil, konuya özel metinlerden elde edilmektedirler.

{ Company | Person } **controls** Company

{ Company | Person } **earns** Money { for | from } Goods-or-Services

{ Company | Country } **exports** Goods to Country

Coperorg **invests** Money in { Financial-Instrument | Market | Country | Company }

noun BE INDUCED BY noun **activation of these PROTEIN was induced by PROTEIN**  
 noun INDUCE noun **PROTEIN induced the tyrosine phosphorylation**  
 noun BIND TO noun **the drugs bind to two different PROTEIN**  
 noun BIND noun **motifs previously found to bind the cellular factors**  
 noun BINDING noun **the TATA-box binding protein**  
 the BINDING of noun **the binding of PROTEIN**



## Dinamik Web Sayfalarından Bilgi Çıkarımı

Amazon Kitap Sayfası HTML kodu

Elde edilen bilgi

```
....
</td></tr>
</table>
<b class="sans">The Age of Spiritual Machines : When Computers
Exceed Human Intelligence</b><br>
<font face=verdana,arial,Helvetica size=-1>
by <a href="/exec/obidos/search-handle-url/index=books&field-author=
Kurzweil%2C%20Ray/002-6235079-4593641">
Ray Kurzweil</a><br>
</font>
<br>
<a
href="http://images.amazon.com/images/P/0140282025.01.LZZZZZZZ.jpg
">
</a>
<font face=verdana,arial,Helvetica size=-1>
<span class="small">
<span class="small">
<b>List Price:</b> <span class="listprice">$14.95</span><br>
<b>Our Price:</b> <font color=#990000>$11.96</font><br>
<b>You Save:</b> <font color=#990000><b>$2.99 </b>
(20%)</font><br>
</span>
<p> <br>...
```

Title: **The Age of Spiritual Machines**  
 :  
 : **When Computers Exceed**  
**Human Intelligence**  
 Author: **Ray Kurzweil**  
 List-Price: **\$14.95**  
 Price: **\$11.96**  
 :  
 :



## Temel Fikir

- Birçok web sayfası veritabanlarından dinamik olarak oluşturuluyor.
- Dinamik web sayfalarında html tag'leri tekrar eder.
- Tekrar eden kalıplar arasında aynı tür bilgiler yer alır.



## Alışveriş Robotları

- Tekrarlı HTML taglerinin kullanılarak bilgi çıkarılan sistemlere örnek olarak çeşitli web sitelerinde satılan ürünlerin bilgilerini tek bir sayfada toplayan sistemler verilebilir.
- Ticari Siteler:
  - MySimon
  - Cnet
  - BookFinder



Hesaplamalı Anlambilim Ders Notları

## Alışveriş/Haber Toplama Robotlarının Çalışma Adımları

- 1- Her kaynak site için bilgi çıkarım mekanizmasını kur.
- 2- Kullanıcıdan sorgusunu al. (tür, fiyat vs.)
- 3- Her site için:
  - Kullanıcı sorgusunu siteye gönder
  - Sonuç sayfasını al.
  - Sonuç sayfasını o sayfanın bilgi çıkarım mekanizmasıyla işle. Sonuçlarını kendi veritabanına kaydet.
- 4- Sonuçları fiyatlara/tarihe vb. kritere göre sırala.
- 5- Sonuçları HTML formatına çevir. Kullanıcıya döndür.



NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Morfolojik çözümlemelerdeki şablonları kullanmak, Amasyalı, 2005

- “Bütün X’ler Y’dir” ilişkisine uyan ikililerin otomatik olarak bulunması
- ADIMLAR
  - Şablonların bulunması
  - Şablonlara uygun ikililerin bulunması
  - İkililerin eklerine ayrılması
  - Verilerden örnekler
  - Sınıflandırma Metodu
  - Sınıflandırma Sonuçları



NOVA Research Lab

## Şablonların bulunması

- Keşfetmek istediğimiz ikililerin aralarındaki ilişki türü belirlenir. Ör: “Tüm X’ler Y’dir”.
- Bilinen X,Y ikilileri Google’da aratılır.
- X ve Y arasındaki şablonlar ve frekansları belirlenir.
- En yüksek frekanslı olan şablonlar bu ilişki türünün şablonları olurlar.



## Bulunan Şablonlardan Örnekler Tüm X’ler Y’dir için

- |                 |                      |
|-----------------|----------------------|
| • ve diğer      | ⇒ ve her türlü       |
| • ler ve diğer  | ⇒ lerden biri olan   |
| • ve benzeri    | ⇒ leri ve diğer      |
| • veya diğer    | ⇒ larından biri olan |
| • türü olan     | ⇒ lerinden biri olan |
| • ları ve diğer | ⇒ lardan biri olan   |
| • lar ve diğer  | ⇒ adı olan           |
|                 | ⇒ ve her tür         |



Hesaplamalı Anlambilim Ders Notları

## Bulunan Şablonlardan Örnekler

### X'in yeri Y'dir için

- |                                       |                   |
|---------------------------------------|-------------------|
| • y deki x                            | x, y ye zz km     |
| • y de bulunan x                      | x (y ye zz km     |
| • y de x                              | x, y              |
| • x y de                              | x - y             |
| • x y ili sınırları içerisinde        | x bulunduğu yer:y |
| • y ili sınırlarında kalan x          | y-x               |
| • y ili sınırları içinde bulunan x    | x(y               |
| • y ilçesi sınırları içinde bulunan x | x(y)              |
| • x y nin sınırları içerisinde        |                   |
| • x/y                                 |                   |
| • x / y                               |                   |
| • x-y                                 |                   |
| • x y ye zz km                        |                   |



NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Şablonlara uygun ikililerin bulunması

- Google'da bulunan şablonlar aratılır.
- Sonuç sayfalarındaki şablonların sağ ve sollarındaki kelimeler alınır ve bir dosyaya kaydedilir.



NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Şablonlara uygun ikililerden örnekler

- kontrolör personel
  - teçhizat malzeme
  - kemer teçhizat
  - protein gıda
  - Azerbaycan bölge
  - Ceyda yardımcı
  - komünizm ideoloji
  - delta Gediz
  - kurum Kocaelispor
  - fotoğrafçı Robert
- ⇒ tür flamingo
  - ⇒ ünite aksesuar
  - ⇒ bedel masraf
  - ⇒ din azınlık
  - ⇒ çelik yapı
  - ⇒ yem araç
  - ⇒ kız sıfat
  - ⇒ yapı sorun
  - ⇒ ölçü şart



NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## İkililerin elle sınıflandırılması

- Bulunan ikililerden hangilerinin “Tüm X’ler Y’dir” ilişkisine sahip olup olmadığı elle işaretlenir.



NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Eklerine Ayrılmış Kelimelerden Örnekler

- **adamlarından biri olan OZAN**
  - Noun+ A3pl+ P3sg+ Abl
  - Noun+ A3sg+ Pnon+ Nom
- **dualar ve her türlü ibadet**
  - Noun+ A3pl+ Pnon+ Nom
  - Noun+ A3sg+ Pnon+ Nom



NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## Verilerin Modellenmesi

- Farklı ek türü sayısı 53
- Farklı şablon türü sayısı 16
- 2 kelime  $(53 \times 2) +$  şablon türü = 107 özellik
- Artık, her bir örnek 107 boyutlu bir uzayda bir noktadır.



NOVA Research Lab

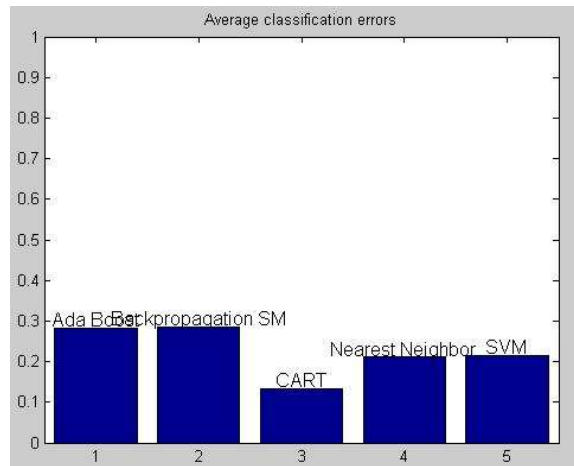


## Eğitim ve Test verileri

- Eğitici sınıflandırma metotlarında eğitim ve test verilerine ihtiyaç var.
- 365 eğitim, 365 test verisi oluşturuldu.
- Her iki veri setinde de 103'er tane istediğimiz, 262'şer tane istemediğimiz türden veri var.



## Farklı sınıflandırıcılar için sonuçlar



## Frekansların Kullanımı

- Çok daha fazla Internet sayfasından binlerce ikili elde edilir.
- Bulunan ikililerin tekrar sayıları (frekansları) bulunur.
- En çok tekrar eden ikililer veritabanına kaydedilir.



## Metot Hakkında

- Sınıflandırma başarısı %100 olmadığından frekans kullanımı zorunlu. Dolayısıyla çok fazla sayfa işlemek gerekli.
- Birden fazla kelimedenden oluşan kelime grupları arasındaki ilişkiler bulunamıyor.
- Her tür ilişkiye uygulanabilir değil.  
Ör: eş anlamlı, zıt anlamlı vs.
- **~%85'lik bir başarı** ile sınıflandırma yapılabildiği görülmüştür.



- Son büyük girişim: 2018
- Allen Institute for Artificial Intelligence-AI2 – Alexandria
- Birçok cs bilgisini dil modellerine katma girişimi?



## Commonsense / reasoning kaynaklar

- Choice of Plausible Alternatives (COPA): An evaluation of commonsense causal reasoning <http://people.ict.usc.edu/~gordon/copa.html>
- RACE: Large-scale ReAding Comprehension Dataset From Examinations <https://arxiv.org/abs/1704.04683>
- Story Cloze Test and ROCStories Corpora <https://www.cs.rochester.edu/nlp/rocstories/>
- ATOMIC: An Atlas of Machine Commonsense for If-Then Reasoning <https://arxiv.org/abs/1811.00146>
- CommonsenseQA: A Question Answering Challenge Targeting Commonsense Knowledge <https://www.tau-nlp.org/commonsenseqa>
- Pretrained Encyclopedia: Weakly Supervised Knowledge-Pretrained Language Model <https://arxiv.org/abs/1912.09637>
- Oren Etzioni - Learning Common Sense: a Grand Challenge for Academic AI Research <https://www.youtube.com/watch?v=m7PVqEPRcxU&feature=youtu.be>
- ERNIEs: [arxiv.org/abs/1904.09223](https://arxiv.org/abs/1904.09223) [arxiv.org/abs/1905.07129](https://arxiv.org/abs/1905.07129)
- If we want to train AI to do what humans want, we need to study humans <https://distill.pub/2019/safety-needs-social-scientists/>



## Referanslar

- Lenat, D. B., Ramanathan V. G., Karen P., Dexter P., ve Shepherd M., "CYC: Toward programs with common sense", The Communications of the ACM, 33(8):31-49 (1990).
- Push Singh, Thomas Lin, Erik T. Mueller, Grace Lim, Travell Perkins ve Wan Li Zhu, "Open Mind Common Sense: Knowledge acquisition from the general public", Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems, Irvine, CA, 2002.
- Liu, H. ve Singh, P., "ConceptNet: A Practical Commonsense Reasoning Toolkit", BT Technology Journal, Volume 22. Kluwer Academic Publishers, 2004.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. ve Miller, K., "Introduction to WordNet: An On-line Lexical Database", 1993.
- Bilgin, O., Çetinoğlu, Ö. ve Oflazer, K., "Building a WordNet for Turkish", Romanian Journal of Information Science and Technology, 7(1-2), 163-172, (2004).
- <http://code.google.com/p/zemberek/>
- Context awareness in Ubiquitous Computing, Man Lok (Simon) Yau
- Generating a Dynamic Gaming Environment Using OMCS, David Gottlieb, Josh Juster
- A. Carlson, J. Betteridge, R.C. Wang, E.R. Hruschka Jr. and T.M. Mitchell, "Coupled Semi-Supervised Learning for Information Extraction", In Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM), 2010

