

Hayat Bilgisi Veri tabanları CSdb ve CSoyun

M.Fatih AMASYALI
BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

Sunum içeriği

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



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

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, yakındaki oduncuların seslerini 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)
<http://www-formal.stanford.edu/leora/commonsense/>

Kemik

Hayat Bilgisi (Commonsense) Nedir?

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



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

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

Kemik

Bağlam tabanlı hesaplama Context-aware computing

- Bağlam:
 - “bulunulan zamana, yere, etraftaki nesneler ve insanlara dair bilgi”
 - Schilit and Theimer and Ryan



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

Bağlam tabanlı hesaplama

- Bağlamı anlamak için mevcut sensörler/kaynaklar:
 - Mikrofon, GPS, Kamera, sistem saati, hız, hareket, ivme algılayıcılar, kişisel bilgiler (takvim, arkadaşlar, telefon rehberi), trafik bilgisi, ?
- Gelecek sensörler:
 - RGB-Depth kameralar, Isı ölçerler, basınç ölçerler, tıbbi sensörler, ?



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

Bağlam tabanlı Hesaplamada Fikirler

- Cep telefonunuz, takviminiz ve sistem saatinden bilgi alarak bir toplantıda olduğunuzda zil sesini otomatik kısar.
- Cep telefonunuz bulunduğunuz ortamın ses seviyesine göre zil sesini ayarlar.
- Örnek: oyun Pou, bankacılık (iGaranti), başka ?



Kemik

Bağlam ve hayat bilgisine sahip bir takvim uygulaması

- “Özurfa Kebapçısı’nda Hilmi ile öğlen yemeği” bilgi girişi için, Sisteme daha önce Hilmi’nin vejetaryen olduğu girilmişse sistem sizi uyarır, onay ister.
- “Gece 3’te kahvaltı” bilgi girişi için uyarı
- Şu anda bulunduğunuz yerden saatinde varamayacağınız bir yerdeki randevu için uyarı
- Açık olmayan bir günde/saatte müze ziyareti için uyarı



Kemik

Mevcut durum

- Sensörler gelişmeye devam ediyor
- Bağlam bilgisini kullanan uygulamalar artıyor.
- Bağlamı anlamak ve (etkin) kullanabilmek için hayat bilgisine ihtiyaç var.



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

Düşünme zamanı

- Gruplar oluşturun.
- Bağlam tabanlı hesaplama için uygulama fikirleri geliştirip listeleyin.
- 10 dakika



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

Hayat Bilgisi Veri Tabanları nasıl oluşturulur?

2 yaklaşım, 2 örnek

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



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

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

Kemik

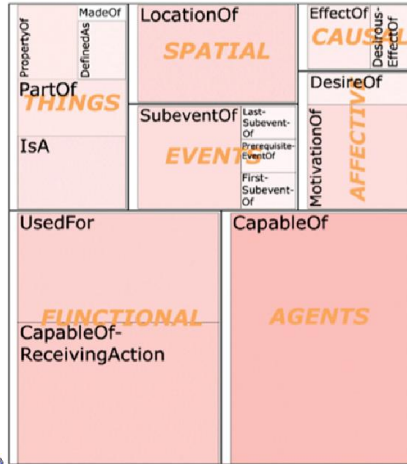
ConceptNet'in oluşturulması

- 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



Kemik

ConceptNet'in Yapısı

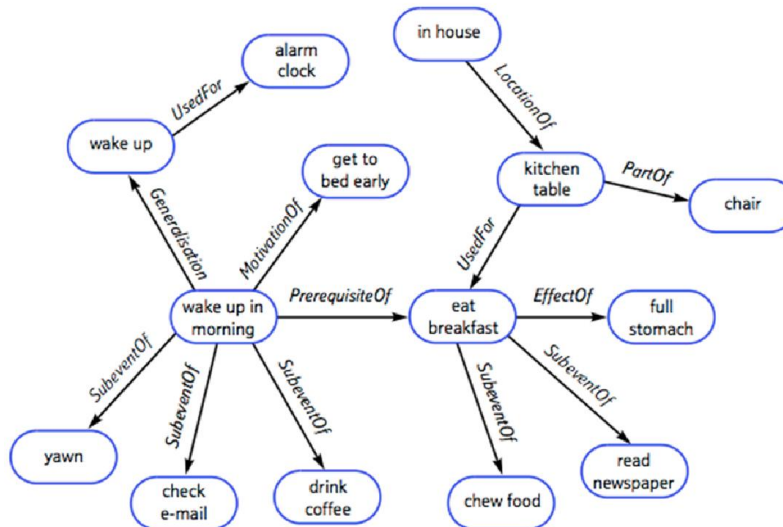


- 1.6 millyon (kavram-ilişki-kavram) üçlüsü
- 20 ilişki türü



Kemik

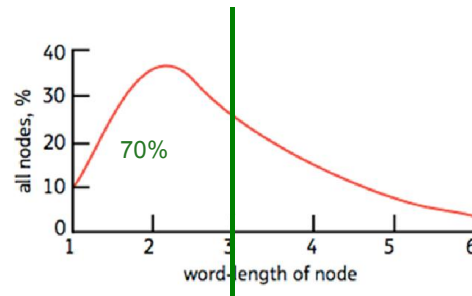
ConceptNet örneği



Kemik

ConceptNet'e dair

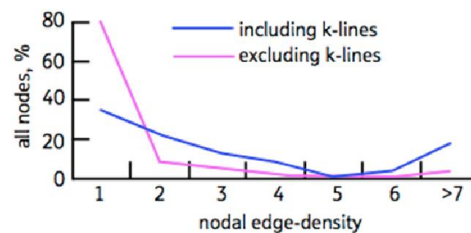
- Kavramlardaki kelime sayılarının histogramı



Kemik

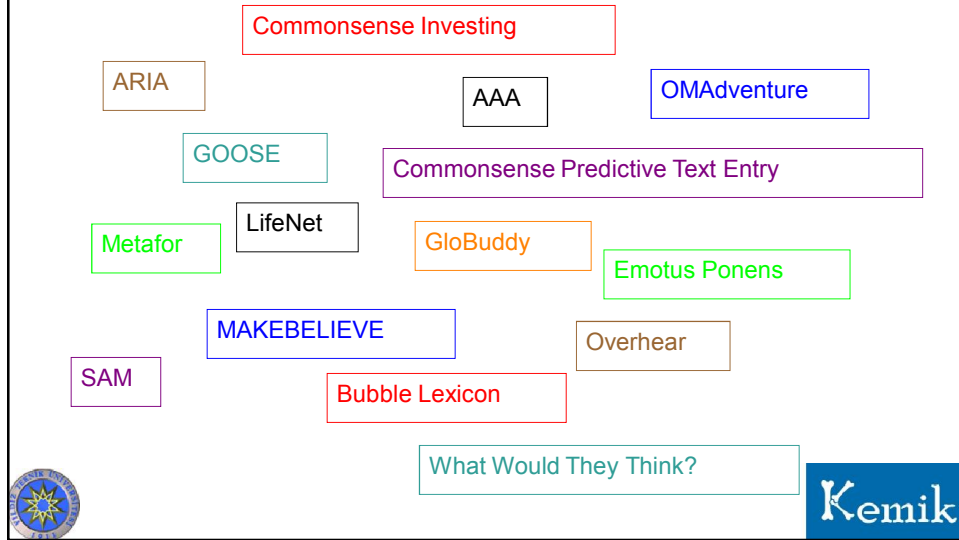
ConceptNet'e dair

- Kavramlara giren ve kavramlardan çıkan ilişki sayılarının histogramı



Kemik

ConceptNet Uygulamaları

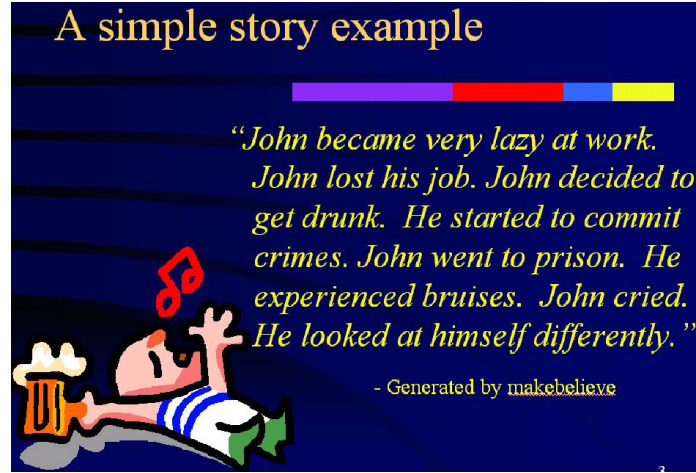


Commonsense ARIA

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



MakeBelieve (Hikaye Üretici)



Kemik

OMAdventure (Oyun)



- OMAdventure, kullanıcının hedef bir objeyi bulmak için dünyada gezindiği bir oyun
- Dünya ve objeler Openmind veritabanı ve Altavista'nın resim arama motoruyla oluşturuluyor.
- OMAdventure, aslında hayat bilgisi veri tabanına bilgi girişi için bir yol



Kemik

OMAdventure Nasıl çalışıyor?

- Kullanıcı yeni bir yer girdiğinde etrafı ve oradaki objeler otomatik olarak oluşturuluyor.
- Oyun başladığında kullanıcıya hedef bir obje veriliyor. Kullanıcı dünyada gezinerek onu bulmaya çalışıyor.
- Oyunda herhangi bir anda kullanıcı sihir yapabilir. Sihirle mevcut odaya/yere bir obje ekleyebilir ya da istediği yeni bir yere gidebilir. Tabi ki hedef objeyi bulunduğu odaya ekleyemez.
- Kullanıcı sihir yapmak istediğinde OpenMind veritabanına (üretilmiş bir soruya cevap vererek) bilgi eklemek zorunda.



Kemik



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

Düşünme Zamanı

- **Yeni** gruplar oluşturun.
- Hayat bilgisi toplamak için ve kullanmak için fikirler geliştirip listeleyin.
- 10 dakika



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

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.



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

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

Kemik

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
zerah_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

Kemik

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
allens	0	26-oct-2010	(Seed) 100.0
an_american_in_paris	0	26-oct-2010	(Seed) 100.0
beowulf	161	09-nov-2010	100.0
big_lebowski	120	19-jun-2010	(Seed) 100.0
boogie_nights	5	12-jan-2010	100.0
close_encounters	1	12-jan-2010	100.0
collateral	161	09-nov-2010	100.0
dick_tracy	4	12-jan-2010	100.0
die_hard	120	19-jun-2010	(Seed) 100.0
doctor_zhivago	161	09-nov-2010	100.0
empire_strikes_back	120	19-jun-2010	(Seed) 100.0
feris_bueller_s_day_off	255	19-may-2011	(Seed) 100.0
fight_club	120	19-jun-2010	(Seed) 100.0
forrest_gump	120	19-jun-2010	(Seed) 100.0
goblet_of_fire	161	09-nov-2010	100.0

Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

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.nl.cmu.edu

8,146 FOLLOWERS
1,863 TWEETS

NELL ile iletişimi sürdür
Bugün Twitter'a kıld

Ad ve soyad
E-posta
Şifre
Kaydol

Tweetler

NELL (@cmunell) I think "Halazepam" is a #Drug (bit.ly/ybtHn) 7 dk

NELL (@cmunell) I think "rockwell kont" is a #VisualArtist (bit.ly/revVxsl) 1 sa

NELL (@cmunell) I think "santa_dal_united_party" is a #PoliticalOrganization that is not part of the government (bit.ly/z7K1S9) 5 sa

NELL (@cmunell) 4 sa

Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

- 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



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

Türkçe Hayat Bilgisi Veritabanı

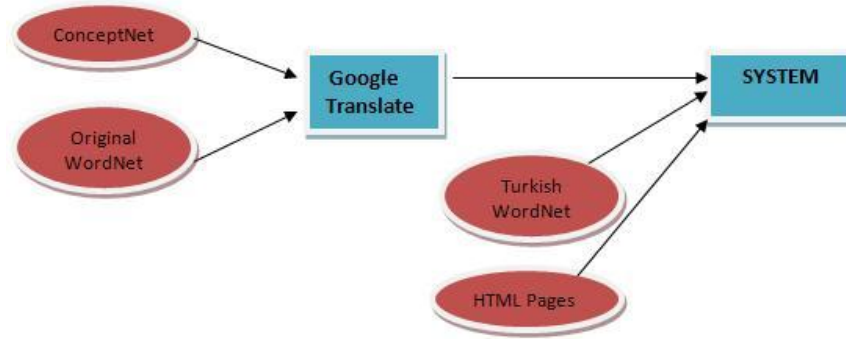
- Literatürdeki çalışmalar genelde ingilizce için
- Türkçe için bir ilk
- Hangi yaklaşım kullanılmalı:
 - Cyc, OpenMind?
 - **Mevcutları tercüme etmek / kullanmak**
 - **Frekans tabanlı bilgi çıkarımı**
- Veriler nasıl tutulmalı?
 - (Kavram - ilişki türü – kavram) üçlülere



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

Veri Kaynakları



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

Veri tabanının oluşturulması

- Kaynaklardaki 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)



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

Veri tabanından örnekler

<u>Bunun için ne gerekir?</u>	<u>Neye sebep olur?</u>	<u>Bundan neler yapılır?</u>
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
<u>Ne için kullanılır?</u>	<u>Bu ne yapabilir?</u>	<u>Nerede bulunur?</u>
asker-savaş	kuş-uçmak	oda-bina
çatal-yemek	kişi-yürümek	kişi-oda
top-oyunmak	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

Kemik

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
Eşanlımlı	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

Kemik

Problemler

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



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

www.kemikoyun.yildiz.edu.tr/commonsense

CSOYUN Gelişim: 40% SEVİYE: 11 TOPLAM PUAN: 23208 ZAFER: 113 MUKABİL: 11 BOZGUN: 6 Çıkış Ayarlar

Hoşgeldin Hızlı fatih amasyalı

PILAN CANAVARLARI

- Serkan Özcan 650752
- Caner Bey 400977
- fatih amasyalı 23000
- Abbas MEMİŞ 18015
- Sinem Topcu 4079

Sistem İstatistikleri

İlgi Sayısı: 1094552

Kelime Sayısı: 640646

Kalancı Sayısı: 32

Listeyi Güncelle

Bildiklerim | Araştır | Tahminlerim | Bana Öğret | Birlikte Eğitim

Tahminlerim

Özellikler kazanma hesaplamalı
İşte etatle peykoşmak
istediğim bir tahmin var
kangızda. Tabiki bu sadece
bir tahmin ve bana göre doğru
ya size göre? Benimle aynı
düşüncedeyseniz Doğru
butonuna, farklı bir
düşüncedeyseniz Yanlış
butonuna, kendi hakkında
fikiriniz yoksa Ben Bilmem
butonuna tıklayabilirsiniz.

hayal
Bu neyin ilk adımıdır?
yaşamak

Doğru **Yanlış** **Ben Bilmem**

Davetlisiniz!

Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

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.



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

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.

SEVİYE	TOPLAM PUAN	ZAFER	MUKABİL	BOZGUN
21	121101	113	15	1

PUAN CANKARILARI

Kullanıcı	Puan
Serkan Özcan	650943
Gençer Bey	401025
fatih amasyalı	121020

Bildiklerim | Araştır | Tahminlerim | Bana Öğret | Birlikte Eğitim

Diğer

Eşanlamlı

adapte olmak

uyumak

5.0

Bildiklerim

Sistem istiyor listede ve çok önüne Eger doğru 5 yıldız atmayı Bana k



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

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'da

- **Tahminlerim:** Bu oyun CSoyun'u benzerler 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.

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)



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

- **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.



CSoyun'da


- **Birlikte Eğitim:** Bana öğret oyununun iki kişinin karşılıklı oynadığı versiyonudur. Sistemdeki kullanıcılar ikiye 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






Kemik

Birlikte Eğitim

CSOYUN  **Hosgeldin Yenilikçi fatih amasyalı**

Bölümün %43	SÖZÜŞ	TÖPLAM PUAN	ZAFER MUKABİL ROZGUN	114 17 1.	DİKUS Ayanlar
21	121981				

PUAN CANAVARLARI

	Serkan Özcan
	Cihan Bey
	Fatih Amasyalı
	Abbas MEMİŞ
	Benker Nard

Sistem İstatistikleri
Tilgi Sayısı: 1234784
Kelime Sayısı: 70077

Bildiklerim Araştır Tahminlerim Bana Öğret Benle Eğitir

Oyuna Başla **Oyuna Dışla**

Puanınız: 1 Rakip puanı: 0 Tebrikler kazandınız!

SORU: karbon Bu neyin hammaddesidir?

Kelime Giriniz:

Girmiş Olduğunuz Kelimeler

- almıs
- kurgun kalem
- kalem ucu
- kümü

Birlikde Eğitim

İşte Haliç ardayosı arın buradı. Bilginiz gücünü göstermek için Haliç'e kadar gitmeye gerek kalmadı. Nasreddin Haliç'i alt etmeye teklifiyi Oyuna Başla butonuyla basarak sorularınıza doğru cevaplar verebileceğin geliyor. Rakibiniz belirlenip soru sorulduktan sonra soldaki elona cevaplarınızı yazıp enter tuşuna basabilirsiniz.



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

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 oyunlarıyla üretilmekte, yeni kavramlar bana öğret ve birlikte eğitim oyunlarıyla 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-iliski türü-kavram) üçlülerine ait toplam oy ve toplam puan bilgileri tutulmaktadır. Hesaplanan ortalama puanlarla bilgi üçlülerinin doğruluk değerleri hesaplanmaktadır.



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

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.



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

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
Eşanlamı	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 Kavramıdır	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
Zıtanlamı	840	4.641	Durumundadır	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 Zıtanlamı	440	4.332	Fiil - Fiil	7	5.000



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

Gelecek Çalışmalar

- Bilgi miktarının arttırılması
 - Cümlelerin öğelerinin kullanımıyla nesne-yer, eylem-yer, özne-eylem gibi ilişki türlerine ait bilgi ikililerinin toplanması
 - Bilgi çıkarımı yöntemlerinin iyileştirilmesi ve yeni kaynaklarda çalıştırılması
- Bilgilerin kalitesinin arttırılması
 - CSoyun'un yeni platformlarda çalıştırılması (kullanıcı sayısını arttırmak için)
- Bilgileri kullanan uygulamaların gerçekleştirilmesi
 - Akıllı arama motoru, paragraf anlama



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

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

Kemik

Hayat bilgisi Veri tabanlarının kalitesi nasıl ölçülür?

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



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

Hayat Bilgisi Veritabanlarının Uygulama alanları

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



Kemik

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



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik

Kemik Doğal Dil İşleme Grubu:
www.kemik.yildiz.edu.tr



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Kemik