

YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**WEB TABANLI SÖZLÜKLER KULLANILARAK EŞ
ANLAMLILAR SÖZLÜĞÜ OLUŞTURMA**

Matematik Müh. Eren BEKÂR

FBE Matematik Müh. Anabilim Dalında

Hazırlanan

YÜKSEK LİSANS TEZİ

Tez Danışmanı : Yrd. Doç. Dr. Nilgün GÜLER BAYAZIT

İSTANBUL, 2010

İÇİNDEKİLER

Sayfa

İÇİNDEKİLER.....	ii
SİMGE LİSTESİ	iv
KISALTMA LİSTESİ.....	v
ŞEKİL LİSTESİ	vi
ÇİZELGE LİSTESİ.....	vii
ÖNSÖZ.....	viii
ÖZET	ix
ABSTRACT	x
1. GİRİŞ.....	1
1.1 Amaç.....	1
2. EŞ ANLAMLILAR SÖZLÜĞÜ.....	3
2.1 WordNet	3
2.2 EuroWordNet.....	5
2.3 BalkaNet	6
3. NEDEN OTOMATİK HAZIRLANMALI?	9
3.1 Wikipedia İle Hazırlama Sebepleri.....	9
3.1.1 Makalelerin Tekil URL'lerden Oluşması.....	10
3.1.2 Çok Sayıda Makale İçermesi	10
3.1.3 Yoğun Bağlantı Yapısı	10
3.1.4 Canlı Güncellemeler	15
3.1.5 Yönlendirme Sayfaları.....	16
3.1.6 Anlam Belirsizliği Sayfaları	16
3.1.7 Kategori Sayfaları	16
3.2 Diğer Hazırlama Yöntemleri	17
3.2.1 WikiRelate	17
3.2.2 PF-IBF	18
3.2.3 Çift İkili Ağaç ve PF-IBF	18
3.2.4 Bağlantıların Birlikte Bulunması	19
3.2.5 Coreference Resolution (Zamir Çözümlemesi).....	20
3.2.6 Kelime Anlam Belirsizliği Giderme	20

3.2.7	Explicit Semantic Analysis	20
3.2.8	Domain Specific Thesauri	21
3.2.9	İki Dilli Sözlük Hazırlama	21
3.2.10	Japonca-İspanyolca Dillerinde Sözlük Hazırlama	22
4.	WIKIPEDIA	24
4.1	Organizasyonun Temel Tanıtımı	24
4.2	Wikipedia İstatistikleri	25
5.	KULLANILAN YÖNTEMLER	26
5.1	PF-IBF	26
5.2	TF-IDF	28
5.3	Bağlantıların Birlikte Bulunması	29
6.	DENEYSEL SONUÇLAR	33
6.1	Kullanılan Yazılımlar	33
6.2	Kullanılan Donanımlar	34
6.3	Test Verisi	34
6.4	Yöntemlerin Karşılaştırılması	35
7.	SONUÇLAR VE ÖNERİLER	51
	KAYNAKLAR	52
	EKLER	54
	EK – 1 Wikipedia veritabanının şeması	55
	EK – 2 Wikipedia makalesinin yapısı	56
	Ek – 3 Yöntemlerin sürelerinin karşılaştırılması	60
	ÖZGEÇMİŞ	67

SİMGE LİSTESİ

$\cos(v_i, v_j)$	Cosine fonksiyonu
ibf	Tersine geri link frekansı
idf	Tersine doküman frekansı
lf	Bağlantı frekansı
$n_{i, j}$	i teriminin j dokümanında geçme adedi
pf	Yol frekansı
tf	Terim frekansı
t_i	İki makale arasındaki i. Yol
v_i	i numaralı makale

KISALTMA LİSTESİ

DBT	Dual Binary Tree
DDİ	Doğal Dil İşleme
IDE	Integrated Development Environment
ESA	Explicit Semantic Analysis
ILI	Inter Lingual Index
LF_IBF	Link Frequency – Inversed Backward Link Frequency
PF-IBF	Path Frequency – Inversed Backward Link Frequency
SQL	Structured Query Language
T-SQL	Transactional - Structured Query Language
TF-IDF	Term Frequency- Inversed Document Frequency
URL	Uniform Resource Locator
XML	Extensible Markup Language

ŞEKİL LİSTESİ

Sayfa

Şekil 1.1 İlişkisel eş anlamlılar sözlüğünün ağ yapısı ile gösterimi.....	1
Şekil 2.1 WordNet’te bulunan temel anlamsal ilişkiler.....	5
Şekil 2.2 EuroWordNet’in veri yapısı	6
Şekil 3.1 Makale için geri bağlantılar ve ileri bağlantılar	11
Şekil 3.2 Türkçe Wikipedia’daki geri bağlantı adetlerinin dağılımları	12
Şekil 3.3 Türkçe Wikipedia’daki ileri bağlantı adetlerinin dağılımı	13
Şekil 3.4 Wikipedia kategorilerinin ağaç yapısı ile gösterimi.....	17
Şekil 3.5 Bitişiklik matrisinin çift ikili ağaca çevrilmesi	19
Şekil 3.6 ESP kullanılarak kavramların arasındaki ilişki hesaplama	21
Şekil 3.7 WordNet ve Wikipedia kullanılarak iki dilli sözlük hazırlama.....	23
Şekil 5.1 PF-IBF metodunda kullanılan makale ve bağlantıları.....	27
Şekil 5.2 Bağlantıların pencere yapısı	30

ÇİZELGE LİSTESİ

Sayfa

Çizelge 2.1 BalkaNet temel istatistikleri	7
Çizelge 2.2 Alt grupların ve kelime tiplerinin dağılımları	8
Çizelge 2.3 Anlamsal ilişkiler	8
Çizelge 3.1 İngilizce ve Türkçe Wikipedia'ların geri bağlantı istatistikleri.....	14
Çizelge 3.2 İngilizce ve Türkçe Wikipedia'ların ileri bağlantı istatistikleri	15
Çizelge 3.3 İngilizce (2006 yılı) ve Türkçe Wikipedia'ların bağlantı istatistikleri.....	15
Çizelge 4.1 7 Kasım 2009 tarihli Wikipedia istatistikleri	25
Çizelge 6.1 Yöntemlerin kullanıldığı bilgisayar yazılımları	34
Çizelge 6.2 Yöntemlerin kullanıldığı bilgisayar donanımları	34
Çizelge 6.3 PF-IBF yönteminin derinlik değerleri için hesaplama değerleri ve süreler	35
Çizelge 6.4 Yöntemlerin test veri setine uygulanması	36
Çizelge 6.5 Yöntemlerin hesaplama süreleri	43
Çizelge 6.6 Spearman's Rank korelasyon hesaplaması sonuçları.....	43
Çizelge 6.7 Ortalama korelasyon değerleri	50

ÖNSÖZ

Eş anlamlılar sözlükleri günümüz internet ortamının hemen her seviyesinde kullanım alanı bulmuştur. Bilgi çıkarımından kelime anlam belirsizliği gidermeye, veri madenciliğinden metin gruplamaya kadar her alanda eş anlamlılar sözlüğü kullanımı yapılan çalışmaların başarısını önemli ölçüde arttırmaktadır. Fakat günümüzde bu sözlükler güncelliğini yitirmekte ve kapsama alanı gittikçe daralmaktadır. Geçmiş senelerde insan eli ile yapılan büyük ölçekli çalışmaları güncel tutmak da bir o kadar zaman almakta ve maliyetli bir iş yükü oluşturmaktadır. Bu nedenle bu tarz sözlüklerin otomatik olarak oluşturulması ve kaynağının mümkün olduğunca güncel tutulması gereği doğmuştur. WEB tabanlı sözlükler bu tarz çalışmalar için oldukça uygundur. Hem farklı dillerde içeriklere sahip olması, hem de insanların istedikleri zaman istedikleri şekilde makalelerini düzenleyebilmeleri ve en önemlisi büyük miktarlarda içeriğe sahip olması Wikipedia'nın en önemli özelliklerinden birkaçı olarak tanımlanabilir. Bu nedenlerden dolayı tez çalışmada veri kaynağı olarak Türkçe Wikipedia kullanılmıştır. Wikipedia üzerinden farklı yöntemlerle eş anlamlılar sözlüğü oluşturulmaya çalışılmış ve aralarındaki farklar gözlemlenmiştir.

Bu çalışmamda bana her türlü desteği gösteren tez danışmanım ve saygıdeğer hocam Yrd. Doç. Dr. Nilgün Güler BAYAZIT'a ve yardımlarından dolayı Yrd. Doç. Dr. Olcay Taner Yıldız hocama teşekkürlerimi sunarım.

ÖZET

Bu tez çalışmasının amacı Türkçe eş anlamlılar sözlüklerini güncel ve geniş kapsamlı olarak otomatik bir şekilde hazırlayabilmektir. Bunun için daha önceden insan eli ile hazırlanan WordNet, EuroWordNet ve içerisinde Türkçe dilinin de bulunduğu BalkaNet projeleri incelenmiş, eksiklikleri gözlemlenmiştir. Bu eksiklikleri gidermek ve amaca yönelik bir eş anlamlılar sözlüğü hazırlamak için Wikipedia'nın Türkçe sürümü kullanılmıştır. Wikipedia'nın kullanılmasının öncelikli sebebi geniş kapsamlı içeriği, her gün binlerce güncelleme yapılması, yeni makale eklenmesi ve birçok dilde içeriğe sahip olmasıdır. Çalışmada eş anlamlılar sözlüğü hazırlanırken PF-IBF, TF-IDF, bağlantıların birlikte bulunması yöntemleri kullanılmıştır. Bu yöntemler detaylı bir şekilde anlatılmış, bir veri seti üzerinde üçünün de sonuçları hesaplanmış ve yorumlanmıştır. Bu örnek veri seti önce bu yöntemler ile işleme konulmuş, daha sonra 17 kişilik bir insan grubu ile de test edilerek yöntemlerin başarısı incelenmiştir.

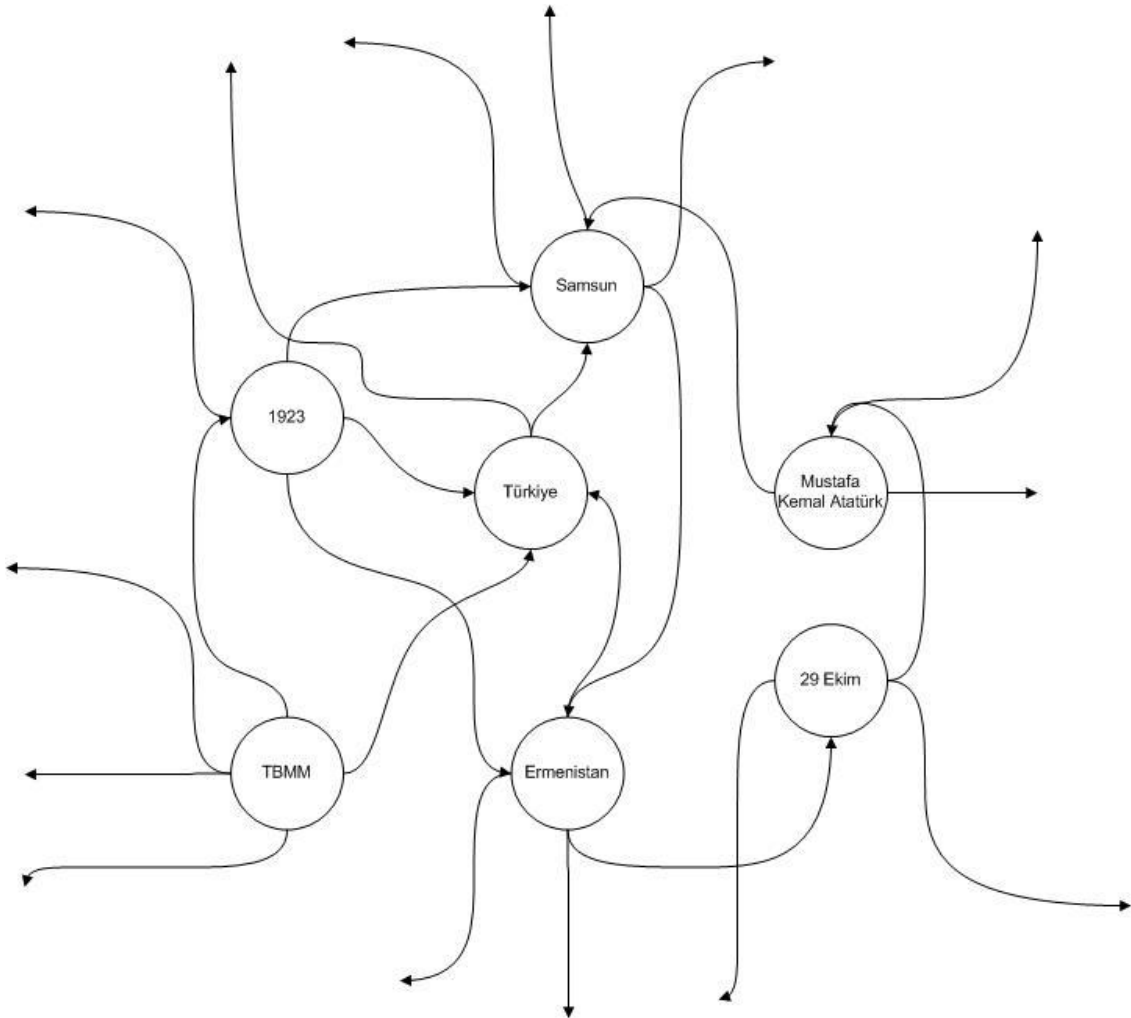
ABSTRACT

The goal of this study is to achieve automated construction of a comprehensive and contemporary Turkish thesaurus. In this aim, manually-constructed WordNet, EuroWordNet and BalkaNet, latter of which includes the Turkish Language in it, have been reviewed and their shortcomings have been noted. The Turkish version of Wikipedia has been used to overcome these shortcomings and develop a new and enhanced Turkish thesaurus. The major reasons for using Wikipedia are its broad index, updates and inclusion of new articles, whose frequency reach thousands within a day period, and finally its availability in several languages. In this study, the method of choice was PF-IBF, TF-IDF and Link Co-occurrence. These methods have been explained in detail and applied to a dataset. In sum, this sample dataset was run by these methods and subsequently tested with responses of 17 human subjects. The efficacy of the methods has been discussed.

1. GİRİŞ

1.1 Amaç

Eş anlamlılar sözlüğü, kelimeleri ve bu kelimelerin arasındaki benzerlik ilişkilerini barındıran bir çeşit sözlüktür. Genel anlamda iki farklı eş anlamlılar sözlüğü vardır. Bunlardan birincisi bağlantısal eş anlamlılar sözlüğüdür. Bağlantısal eş anlamlılar sözlüğü, kelimeler arasındaki "dır/dır", "-in bir parçası" gibi ilişkileri tanımlar. Bir diğer eş anlamlılar sözlüğü tipi ise ilişkisel eş anlamlılar sözlüğüdür. Bu sözlük tipinde bir kelime ile ilişkili diğer kelimeleri belirlemek mümkündür. İlişkisel eş anlamlılar sözlüğünü Şekil 1.1’de gösterildiği gibi bir ağ yapısı ile tanımlayabiliriz. Bu durumda kelimeler ağdaki düğümleri, kelimeler arasındaki ilişkiler de bu düğümler arasındaki bağları oluşturur.



Şekil 1.1 İlişkisel eş anlamlılar sözlüğünün ağ yapısı ile gösterimi

Eş anlamlılar sözlüğü kullanımının doğal dil işleme (DDİ) ve bilgi çıkarımı gibi bir çok alanda etkinliği önceki çalışmalarda kanıtlanmıştır(Strube ve Ponzetto, 2006; Nakayama vd.,

2007a; Nakayama vd., 2007b; Ito vd., 2008; Ponzetto vd., 2006; Mihalcea, 2007; Gabrilovich ve Markovitch, 2007; Milne vd., 2006; Erdmann vd., 2008; Ramírez vd., 2007). Örneğin bilgi çıkarımı alanında, anahtar kelime tabanlı çalışan WEB arama motorları, araması yapılan anahtar sözcükle tamamen uyuşmayan WEB sayfalarını bulamaz. Bu nedenle araştırmalar sorgu genişletme alanında daha da yoğunlaşmıştır. Bu yöntemle sorgulama yapılan kelime ile benzer diğer kelimeler bulunur ve arama sonuçlarına bu kelimeler ile ilgili WEB sayfalarını da katarak sonuçları daha başarılı hale getirmek mümkün olmuştur.

Fakat bu alanlarda yapılan çalışmaların başarılı sonuç vermesi için kullanılan eş anlamlılar sözlüğünün her zaman güncel bilgilere sahip olması ve içeriğinin yeterince geniş olması gerekmektedir. Elle hazırlanan eş anlamlılar sözlüğünün içeriklerini güncel tutmak büyük bir maliyet gerektirir. Bu nedenle güncel bilgileri içeren, bilgisayarların anlayabileceği web tabanlı sözlüklerden faydalanmak yapılması gereken çalışmaları hem otomatikleştirecek, hem de hızlanmasını sağlayacaktır.

2. EŞ ANLAMLILAR SÖZLÜĞÜ

2.1 WordNet

WordNet, 1985 yılında geliştirmeye başlanan ve 2006 yılında da son sürümü olan 3.0'ın yayınlandığı, İngilizce kelimelerden oluşan sözcük veritabanıdır. İçeriğinde 155.287 adet İngilizce kelime barındırır. Bu kelimeler kendi içlerinde eş anlamlılar olarak gruplanmıştır ve bu grupların her birine eş anlamlılar grupları anlamına gelen “synset” ismi verilmiştir. Toplamda 117.659 adet eş anlamlılar grubu içeren bu veritabanının içinde 206.941 adet de kelime anlam çifti bulunmaktadır.

Örneğin İngilizce dilinde genel anlamı “most suitable or right for a particular purpose” olarak verilmiş bir eş anlamlılar grubunun içeriği aşağıdaki gibidir:

- Good: “a good time to plant tomatoes”
- Right: “the right time to act”
- Ripe: “the time is ripe for great sociological changes”

Bu eş anlamlılar grupları da kendi içlerinde birbirlerine anlamsal ilişkiler ile bağlıdır. Bu ilişkiler eş anlamlılar gruplarının tiplerine göre değişiklik göstermektedir. Grup tiplerine göre kullanılmış olan ilişkiler ve açıklamaları aşağıdaki gibidir.

- İsim gruplarında kullanılan ilişkiler:
 - Hipernim (Hypernym): Eğer her X, Y’nin bir çeşidi ise Y, X’in bir hipernimi demektir.
 - Hiponim (Hyponym): Eğer her Y, X’in bir çeşidi ise Y, X’in bir hiponimi demektir.
 - Eş Güdümlü Terim: Eğer X ve Y bir hipernim paylaşıyorsa, bu durumda X ve Y birbirlerinin eş güdümlü terimleridir.
 - Holonim (Holonym): Eğer X, Y’nin bir parçası ise Y, X’in bir holonimidir.
 - Meronim (Meronym): Eğer Y, X’in bir parçası ise Y, X’in bir meronimidir.
- Fiil gruplarında kullanılan ilişkiler:
 - Hipernim (Hypernym): Eğer X aktivitesi Y’nin bir çeşidi ise Y, X’in bir hipernimidir.
 - Troponim (Troponym): Eğer Y aktivitesi, X’in farklı bir biçimde yapılaş şekli ise Y aktivitesi X’in bir troponimidir.
 - Gerektirme: Eğer X’i yaparken Y’yi de yapmanız gerekiyor ise Y fiili X’in bir gerektirmesidir.

- Eş Gdml Terim: Eęer iki fiilin ortak bir hipernimi varsa bu iki fiil birbirlerinin eş gdml terimleridir.
- Hayponim (Hyponym): Aynı iřlevi gren fakat yapılan iři daha detaylı belirten olaydır.
- Sıfat gruplarında kullanılan iliřkiler:
 - İliřkili isimler
 - Benzerlik
 - Fiilin sıfat-fiili
- Zarf gruplarında kullanılan iliřkiler:
 - Kendilerinin tretildięi sıfatlarla baęlantılıdırlar. rneęin “quickly” zarfı “quick” sıfatı ile baęlantılıdır.

Ayrıca bir eş anlamlı grubunun içindeki kelime başka bir eş anlamlı grubundaki dięer bir kelime ile anlamsal olarak baęlı olabilmektedir.

WordNet’de isimler ve fiiller hiyerarřık olarak iliřkilendirilmiřtir. rneęin İngilizce "dog" kelimesi ařaęıdaki hipernim hiyerarřisine sahiptir.

dog, domestic dog, Canis familiaris

=> *canine, canid*

=> *carnivore*

=> *placental, placental mammal, eutherian, eutherian mammal*

=> *mammal*

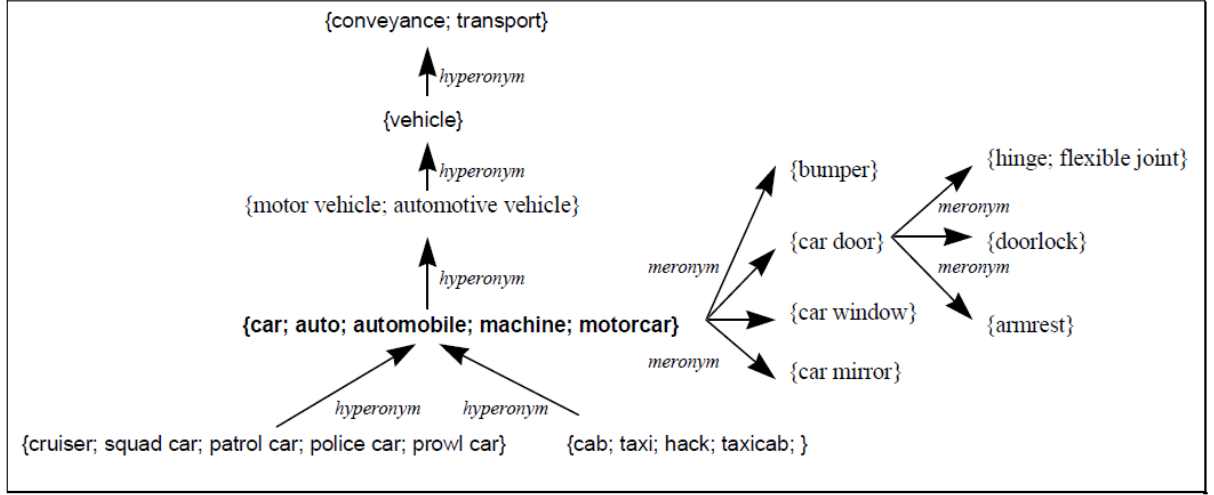
=> *vertebrate, craniate*

=> *chordate*

=> *animal, animate being, beast, ...*

=> ...

Bu hiyerarřide her seviyedeki kelimeler birbirleri ile eş anlamlıdır. WordNet’de hiyerarřinin en st seviyesinde toplam 25 adet isim 15 adet de fiil bulunmaktadır.



Şekil 2.1 Wordnet'te bulunan temel anlamsal ilişkiler.*

2.2 EuroWordNet

EuroWordNet Avrupa dillerinin Wordnet tabanlı olarak geliştirilmiş sözcük veritabanıdır. Fakat Wordnet'ten en önemli farkı, geliştirilmesi yapılan her dilde bulunan kayıtların diğer dillere dizinler (interlingual index, ILI) ile bağlı olmasıdır.

İçeriğinde bulunan diller şunlardır:

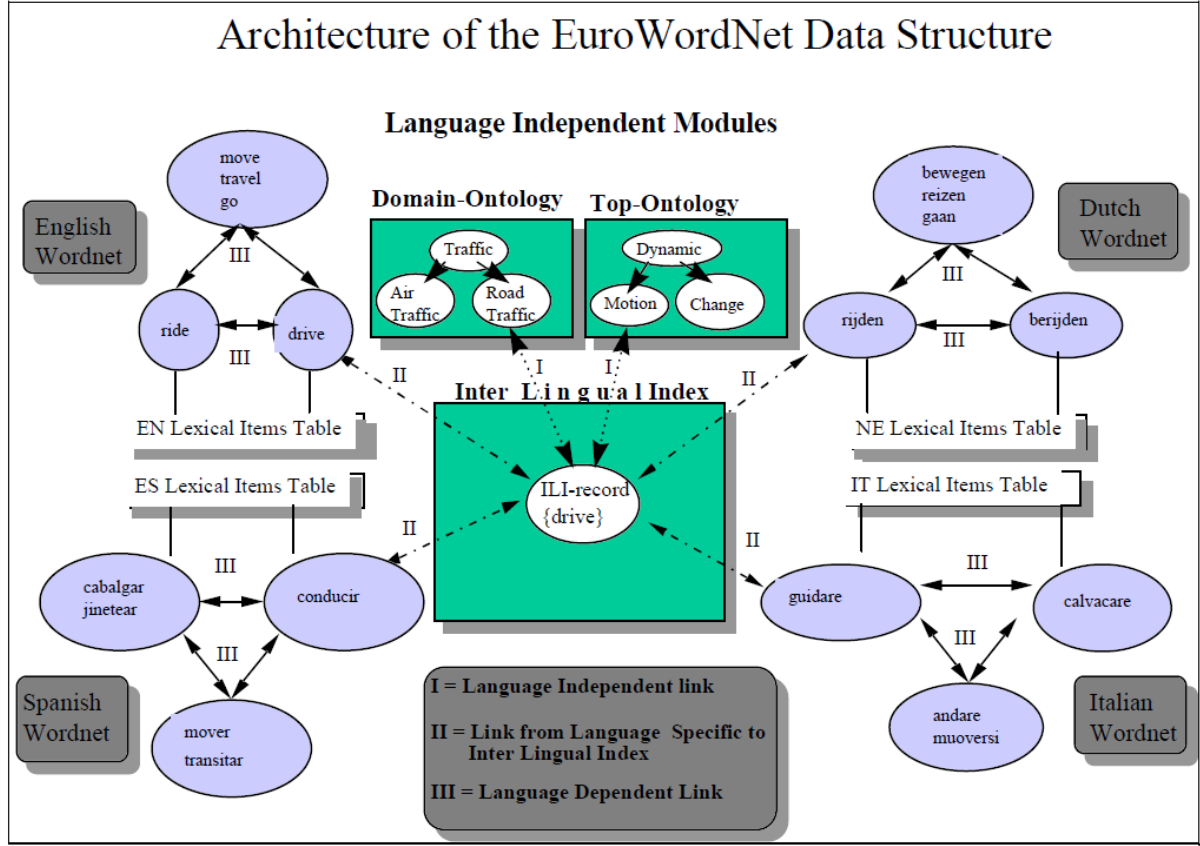
1. Hollandaca
2. İtalyanca
3. İspanyolca
4. İsveççe
5. Almanca
6. Fransızca
7. Çekce
8. Estonyaca.

Çoğu dil WordNet'ten farklı olarak sadece ücretli olarak kullanılabilmektedir, fakat belirli oranlarda ücretsiz olarak kullanımına izin verilen diller de mevcuttur.

Bu dillerde yapılan çalışmalar 1996 yılının Mart ayında başlamış ve 36 ay sürmüştür.

* <http://www.ksl.stanford.edu/onto-std/eurowordnet.pdf>

Veritabanında toplamda 30.000 kavram ve 50.000 kelime anlamı bulunmaktadır. Bu dillerin geliştirilmeleri günümüzde durdurulmuştur.



Şekil 2.2 EuroWordNet'in veri yapısı.**

2.3 BalkaNet

Balkanet projesine öncelikle EuroWordNet'in hiyerarşisinin en üstünde bulunan ve birçok hiponime sahip olan 1.310 temel kavramının Türkçeye çevrilmesi ile başlanmıştır. Bu aşamadan sonra Türkçe dilinde eş anlamlı, zıt anlamlı ve hayponimleri otomatik olarak çıkarmak için bir çalışma yapılmıştır. Bu çalışmalar sonucunda aşağıdaki bilgilere ulaşılmıştır (Bilgin vd., 2004).

Eş Anlamlılar:

hw: $w_1, w_2, \dots, w_n$ formatındaki kayıtlar ile yaklaşık 11.000 eş anlamlı seti çıkarılmıştır. Burada hw asıl kelimeyi, w_i 'ler ise hw ile eş anlamlı bir kelimeyi temsil etmektedir.

** <http://www.ksl.stanford.edu/onto-std/eurowordnet.pdf>

hw: $(w_i)^*, w$ formatındaki kayıtlardan 10.846 adet eş anlamlılar grubu çıkartılmıştır.

Hiperonimler:

"Bir tür" ve "Bir çeşit" gibi kalıplar sayesinde 625 adet hayponim - hiperonim çifti çıkartılmıştır.

"-giller" eki ile 889 adet hayponim - hiperonim çifti çıkartılmıştır.

Zıt anlamlılar:

"karşıtı" kelimesi ile 235 adet zıt anlamlı kelime çifti çıkartılmıştır.

Bundan sonra yapılan çalışmalar ile birlikte aynı yöntemlerle WordNet 1.7 ve 2.0 sürümleri üzerinde çalışmalar yapılmış ve BalkaNet veritabanının genişletilmesi amaçlanmıştır. Bu çalışmalar sonucunda Mart 2004 tarihinde gelinen noktada Türkçe Wordnet'in durumu aşağıdaki gibi olmuştur.

Çizelge 2.1 BalkaNet temel istatistikleri

İstatistik	Adet
Synset	11.628
Synset Üyesi	16.095
Ortalama Synset Boyutu	1,38
Sözcüklerle İlgili Olmayan Konseptler	987
Tanımlar	4.913

Çizelge 2.2 Alt grupların ve kelime tiplerinin dağılımları

Grup Tipi	Adet	Oran
Grup 1	1.218	%100
Grup 2	3.471	%100
Grup 3	3.783	%100
İsim	8.691	%74,7
Fiil	2.556	%22
Sıfat	381	%3,3

Çizelge 2.3 Anlamsal ilişkiler

İlişki Tipi	Adedi
Hiperonim	11.251
Tüm Üye	946
Tüm Parça	1.423
Tüm Kısım	176
Sebep	100
Bir Durumda Olma	577
Zıt Anlamlı Kelime	1.400
Alt Olay	127
Bakınız	270
Fiil Grubu	896
Kategori Alanları	384
Toplam	17.550

3. NEDEN OTOMATİK HAZIRLANMALI?

Bilgisayarların anlayabileceği sözlükleri otomatik olarak oluşturmak en zor alanlardan biridir. Bunun en kolay yolu sözlüklerin insan eli ile hazırlanmasıdır. Geçmişte birçok katılımcı yüksek kalitede eş anlamlılar sözlüğü hazırlamak için uzun zaman harcamıştır. Fakat hazırlanan bu sözlükler geniş ölçekli olmamakla beraber çoğu zaman da yeni bilgileri içermemekte ve güncelliğini yitirmektedir. Bu nedenle sözlükleri otomatik oluşturmak için doğal dil işleme tabanlı birçok çalışma yapılmıştır. Fakat doğal dillerin karmaşıklığı yüzünden, eş anlamlı terimler ile ilgili sorunlar günümüzde hala mevcuttur. Günümüzde bu problemleri çözebilecek bir otomatik eş anlamlılar sözlüğü oluşturma metoduna hala ihtiyaç duyulmaktadır.

Eş anlamlılar sözlüğü, kelimeler arasındaki anlamsal ilişkileri barındıran bir veri yapısıdır. Bu yapıdaki bir sözlüğü oluşturma en kolay yolu insan gücüdür. WordNet bunun en önemli örneklerinden birisidir. Bu sözlük birçok katılımcının katkısı ile hazırlanmıştır. Son sürümünde içinde 155.287 adet kelime ve 206.941 adet hiyerarşik ilişki barındırmaktadır. Fakat şu da aşikardır ki, bu denli geniş bir sözlüğü insan eli ile hazırlamak çok zaman alan bir çalışmadır. Geçmiş senelerdeki binlerce katılımcının hazırladığı bu sözlükler günümüzde güncelliğini kaybetmiştir. Bu nedenle de otomatik sözlük hazırlama amacıyla birçok çalışma yapılmaya devam edilmektedir.

3.1 Wikipedia İle Hazırlama Sebepleri

Web tabanlı sözlüklerin en önemli özelliklerinden birisi de içerdikleri bağlantılardır. Bu bağlantılar WEB yerelliği gibi önemli bilgileri barındırır. WEB yerelliği birbirlerine bağlı olan sayfaların bağlı olmayanlara göre daha fazla benzer olduğunu belirtir. Buradan yola çıkarak bu kuralın WEB tabanlı sözlüklerde de geçerli olduğunu söylemek mümkündür. Hatta web tabanlı sözlüklerde bu kuralın daha güçlü olduğu kabul edilebilir. Çünkü web tabanlı sözlüklerde kullanılan bağlantılar kelimeler arasındaki ilişkilerin kesin tanımlarıdır.

URL tabanlı kelime tanımlama da web tabanlı sözlüklerin önemli özelliklerinden bir tanesidir. Sıradan sözlüklerde dizinler kullanılmasına rağmen bazı durumlarda birden fazla kelime için tek bir dizin kullanılmaktadır. Bu da belirsiz terimlerin aynı makalede tanımlanması anlamına gelmektedir. Bu insanlar için sorun oluşturmaya bile makine tabanlı öğrenme sistemlerinde anlaşılır bir durum olmaktan çıkmaktadır. Örneğin; "Metro Turizm filosuna kattığı 30 yeni Mercedes Travego 2+1 sayesinde yolcularına Suit hizmeti veriyor." cümlesindeki Metro'nun bir turizm firması ya da bir çeşit ulaşım aracı olup olmadığını anlamak insanlar için kolay olsa

da makina tabanlı öğrenme sistemlerinin bunun bir turizm firması ya da bir çeşit ulaşım yolu olup olmadığını anlayabilmek için yeterli bilgisi mevcut değildir.

3.1.1 Makalelerin Tekil URL'lerden Oluşması

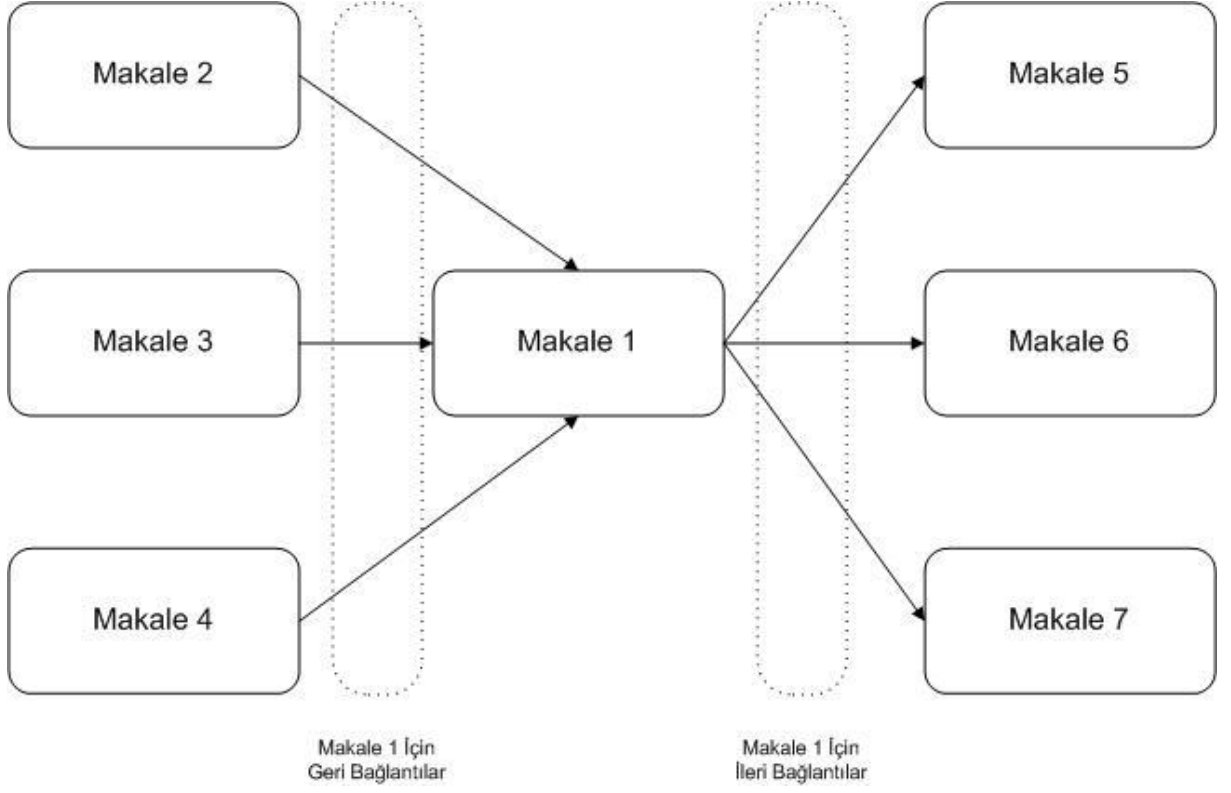
Wikipedia'da her terimin kendi URL'si mevcuttur. Bu da kelimeler arasındaki ilişkileri analiz ederken kelime belirsizliği sorununun oluşmamasını sağlar. İçerdiği yüksek adetteki makale sayının yanında güncellenme sıklığı ve içerdiği yoğun bağlantılı yapısı da Wikipedia'nın önemli özelliklerinden bazılarıdır. Wikipedia; sanat, coğrafya, tarih, bilim, spor, oyunlar ve bunun gibi birçok kategoriyi kapsar.

3.1.2 Çok Sayıda Makale İçermesi

Wikipedia İngilizce dilinde Eylül 2006 tarihi itibarı ile 1,3 milyon makale içermekte idi. Mart 2006'da bu adet 1 milyonken, Mart 2005'te sadece 500.000 adet makale içermekteydi. Wikipedia'nın yüksek hızda güncellenmesi, içeriğinin günden güne artmasına yardımcı olmaktadır. Bu da Wikipedia'nın eş anlamlılar sözlüğü oluşturabilmesindeki önemli avantajlarından bir tanesidir.

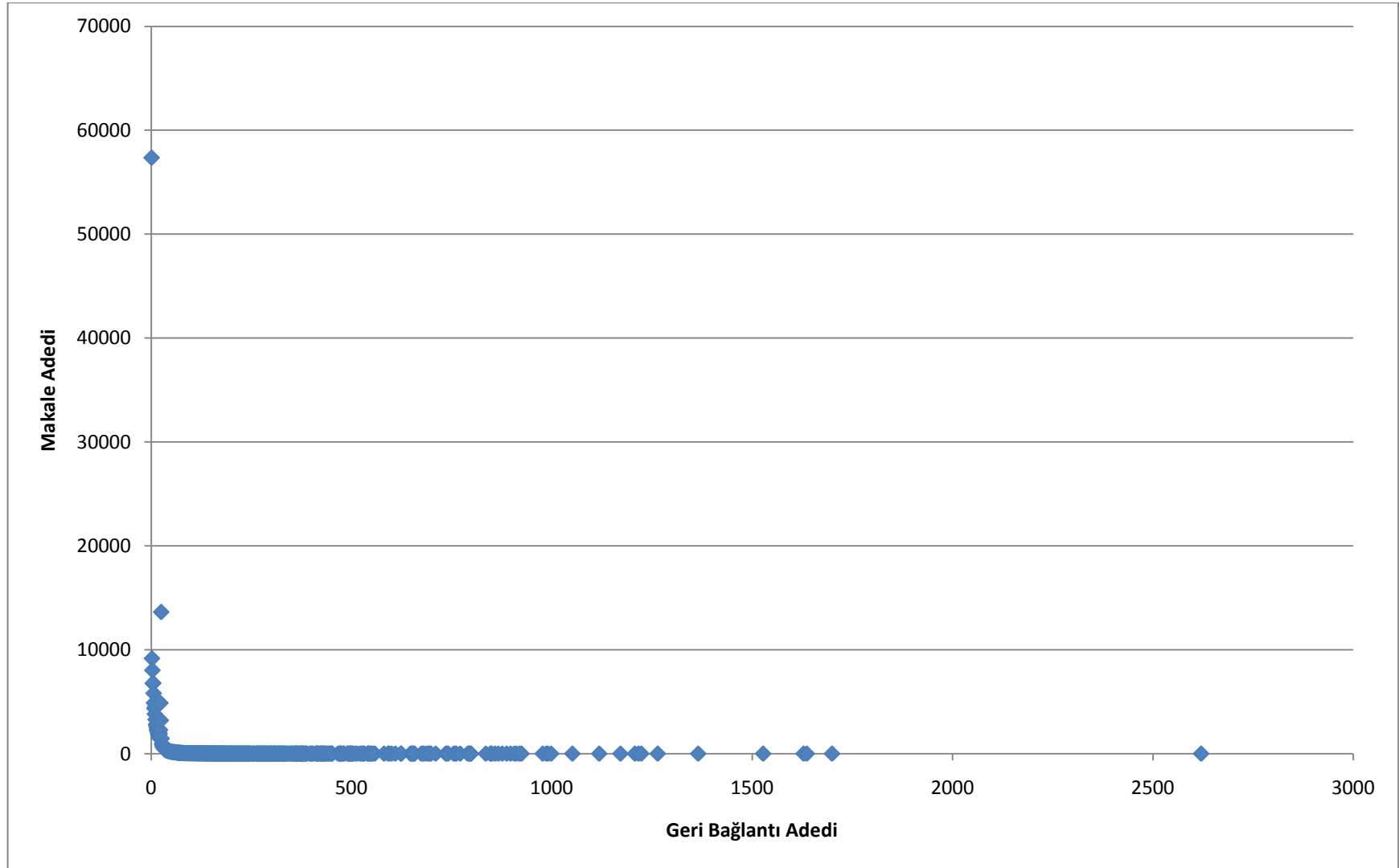
3.1.3 Yoğun Bağlantı Yapısı

Wikipedia'nın ölçeği dışındaki önemli özelliklerinden biri de yoğun bağlantı yapısıdır. Burada yoğun kelimesinin anlamı Wikipedia'nın bir çok dahili bağlantısı olduğunu vurgulamaktadır. Yani Wikipedia makaleleri bu bağlantılar sayesinde birbirlerine sıkı bir şekilde bağlıdır. Bu bağlantı yapısını analiz ederek önemli bilgiler edinmek mümkündür. Wikipedia'daki bağlantıları iki farklı şekilde ifade etmek mümkündür. Bunlardan birincisi, herhangi bir makaleye, diğer makalelerden yapılan bağlantılardır. Bu tarz bağlantılara geri bağlantı denmektedir. Bir diğer bağlantı çeşidi ise bir makaleden diğer makalelere yapılan bağlantılardır. Buna da ileri bağlantı ismi verilmektedir. Bir makalede bulunabilecek iki farklı bağlantı türü Şekil 3.1'de gösterilmiştir.

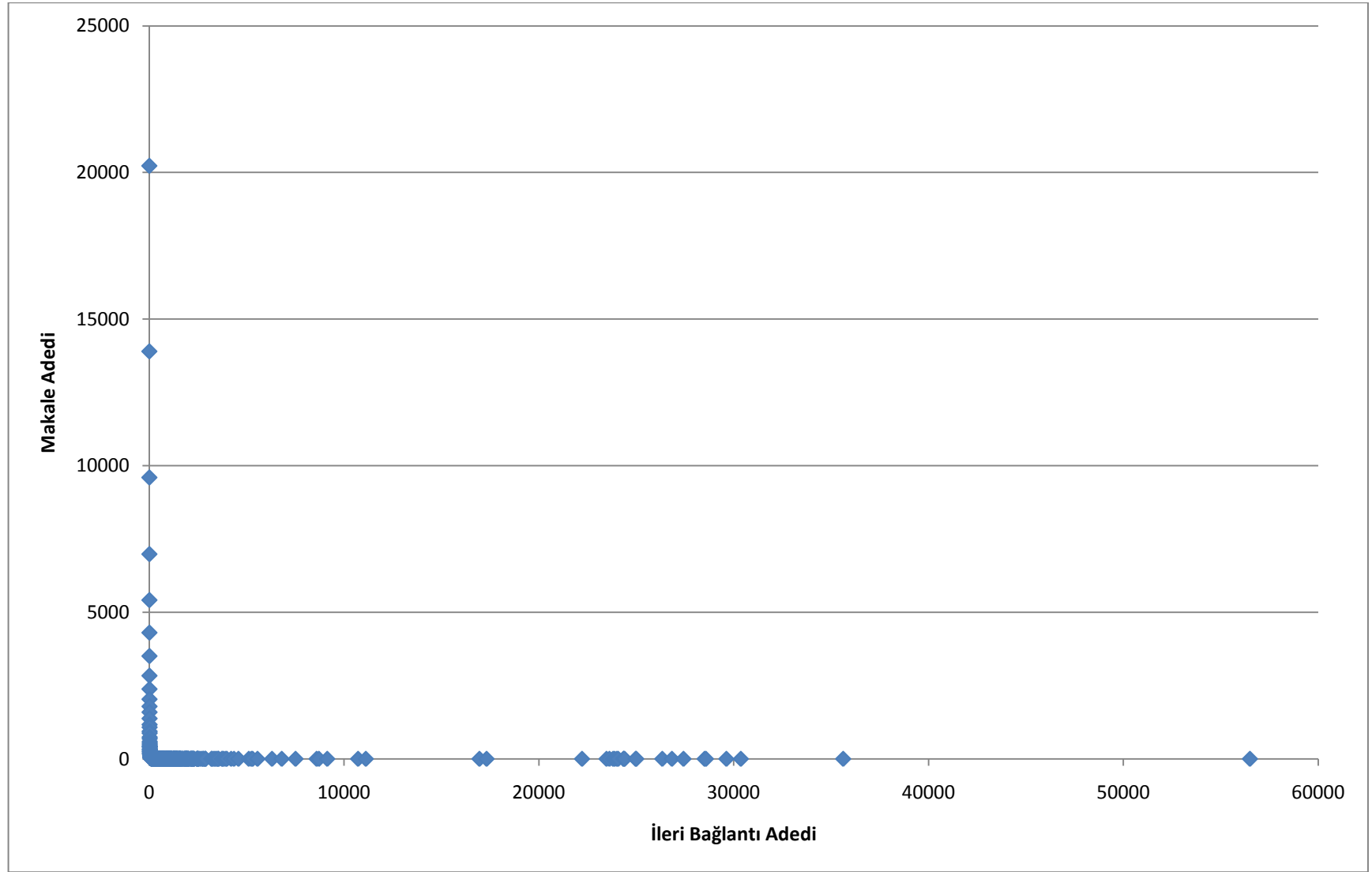


Şekil 3.1 Makale için geri bağlantılar ve ileri bağlantılar

Şekil 3.2’de Türkçe Wikipedia’daki geri bağlantı adetlerinin dağılımları gösterilmektedir. Şekil 3.3’de de Türkçe Wikipedia’daki ileri bağlantı adetlerinin makalelere göre dağılımı gösterilmektedir. Burada zipf dağılımından bahsetmek mümkündür. Yüksek derecede geri bağlantı adedine sahip sadece bir kaç makale mevcutken geri kalan çoğu makalenin geri bağlantı adedi çok azdır.



Şekil 3.2 Türkçe Wikipedia'daki geri bağlantı adetlerinin dağılımları



Şekil 3.3 Türkçe Wikipedia'daki ileri bağlantı adetlerinin makalelere göre dağılımı

Çizelge 3.1’de gösterildiği gibi Türkçe Wikipedia’da sadece 25 makale 10.000’den fazla geri bağlantıya sahipken 196 adet makale 1.000’den fazla geri bağlantıya sahiptir. 100’den fazla geri bağlantıya sahip olan makale adedi ise 2,296’dır.

Türkçe Wikipedia’da yönlendirme bağlantılarını saymadığımızda toplam 2.301.679 adet bağlantı mevcuttur. Bu bağlantılar toplamda 176.388 adet makaleden oluşmaktadır. Yani her makale ortalama 13,04 adet ileri bağlantıya sahiptir. Toplam 87 adet makalede 500’den fazla ileri bağlantı mevcuttur. 1.889 adet makalede de 100’den fazla ileri bağlantı mevcuttur.

Bu bilgiler İngilizce Wikipedia için hesaplandığında adetler şu şekilde bulunmaktadır: 196 makale 10.000’den fazla geri bağlantıya sahipken, 3.198 adet makale 1.000’den fazla geri bağlantıya sahiptir. Ve 67.515 adet makalede 100’den fazla geri bağlantı mevcuttur.

İngilizce Wikipedia’da yönlendirme bağlantılarını saymadığımızda toplamda 49.980.910 adet ileri bağlantı mevcuttur. Bu bağlantılar toplamda 1.686.960 adet makaleden oluşmaktadır. Bu da her bir Wikipedia makalesinin ortalamada 39,62 adet ileri bağlantıya sahip olduğunu göstermektedir. 2.531 adet makalede 500’den fazla ileri bağlantı vardır. 94.931 adet makalede de 100’den fazla ileri bağlantıya sahiptir.

Çizelge 3.1 İngilizce ve Türkçe Wikipedia’ların geri bağlantı istatistikleri

	İngilizce Wikipedia	Türkçe Wikipedia
10,000’den fazla geri bağlantıya sahip makale adedi	196	25
1,000’den fazla geri bağlantıya sahip makale adedi	3.198	196
100’den fazla geri bağlantıya sahip makale adedi	67.515	2.296

Çizelge 3.2 İngilizce ve Türkçe Wikipedia’ların ileri bağlantı istatistikleri

	İngilizce Wikipedia	Türkçe Wikipedia
500’den fazla ileri bağlantıya sahip makale adedi	2.531	87
100’den fazla ileri bağlantıya sahip makale adedi	94.931	1.889

Çizelge 3.3 İngilizce (2006 yılı) ve Türkçe Wikipedia’ların bağlantı istatistikleri

	İngilizce Wikipedia	Türkçe Wikipedia
Toplam bağlantı adedi	49.980.910	2.301.679
Toplam makale adedi	1.686.960	176.388
Makale başına düşen ortalama bağlantı adedi	39,62	13,04

3.1.4 Canlı Güncellemeler

Günümüze kadar yapılan çalışmalarda ilişkisel ve benzerlik ölçümlerinin çoğu WordNet kullanılarak yapılmıştır. WordNet yapısal taksonomi bakımından yeterli bir yapıya sahip olmasına rağmen kapsam olarak bazı soru işaretlerini de beraberinde getirmektedir. Örneğin WordNet 2.1 “Condoleezza Rice”, “Salvador Allende” veya “The Rolling Stones” gibi güncel bilgileri içermemektedir. Fakat Wikipedia bunlar gibi birçok güncel bilgiyi içermektedir. İngilizce sürümü 14 Şubat 2006 itibarı ile 971.518 makaleye sahiptir. Bu makalelerin içerisinde toplamda 18,4 milyon adet bağlantı tanımlanmıştır. Mayıs 2004 itibarı ile de içerdiği kategori bilgileri sayesinde taksonomi bilgisini çıkarmak mümkündür. Bir makale bir ya da daha fazla kategoriye atanabilmektedir. Ocak 2006 itibarı ile Wikipedia makalelerinin %94’ü toplamda 91.502 adet kategori altında gruplanmıştır (Strube ve Ponzetto, 2006).

3.1.5 Yönlendirme Sayfaları

Yönlendirme sayfaları sorgulanan makalenin içeriğinin sorgulandığı şekilde tutulmadığı durumlarda kullanılır. Çoğu zaman eş anlamlı makaleler arasında yönlendirme yapılması için de kullanılmaktadır. Yönlendirme sayfaları benzer olduğu için gereksiz olan makaleleri elemek için de çok uygundur. Bu aynı konuya atıfta bulunan eşdeğer makalelerin varlığını engeller. Şu hallerde kullanılabilir:

- Eş anlamlı sözcükler. (“İktisat” makalesine yönlendiren “Ekonomi” makalesi gibi)
- Bilimsel terimler (“Kabartma Tozu” makalesinin “Sodyum bicarbonat” makalesine yönlendirilmesi gibi)
- Kısaltmalar. (“Amerika Birleşik Devletleri” makalesine yönlendiren “USA” makalesi gibi)
- Hatalı yazımlar. (“Türkiye” makalesinden “Türkiye” makalesine yönlendirmeler gibi)
- Farklı şekillerde yazılan kelimeler. (“Makina” makalesinin “Makine” makalesine yönlendirilmesi gibi)
- Takma isimler. (“Einstein” makalesinin “Albert Einstein” makalesine yönlendirilmesi gibi)

3.1.6 Anlam Belirsizliği Sayfaları

Anlam belirsizliği sayfaları, sorgulaması yapılan şeyin olması muhtemel makalelere bağlantılar içeren bir makalelerdir. Örneğin “Galatasaray” anlam ayrımı makalesi aşağıdaki makalelere bağlantı vermektedir.

- Galatasaray Adası
- Galatasaray Spor Kulübü
- Galatasaray Lisesi
- Galatasaray Üniversitesi
- Galata Kulesi
- Galata Köprüsü

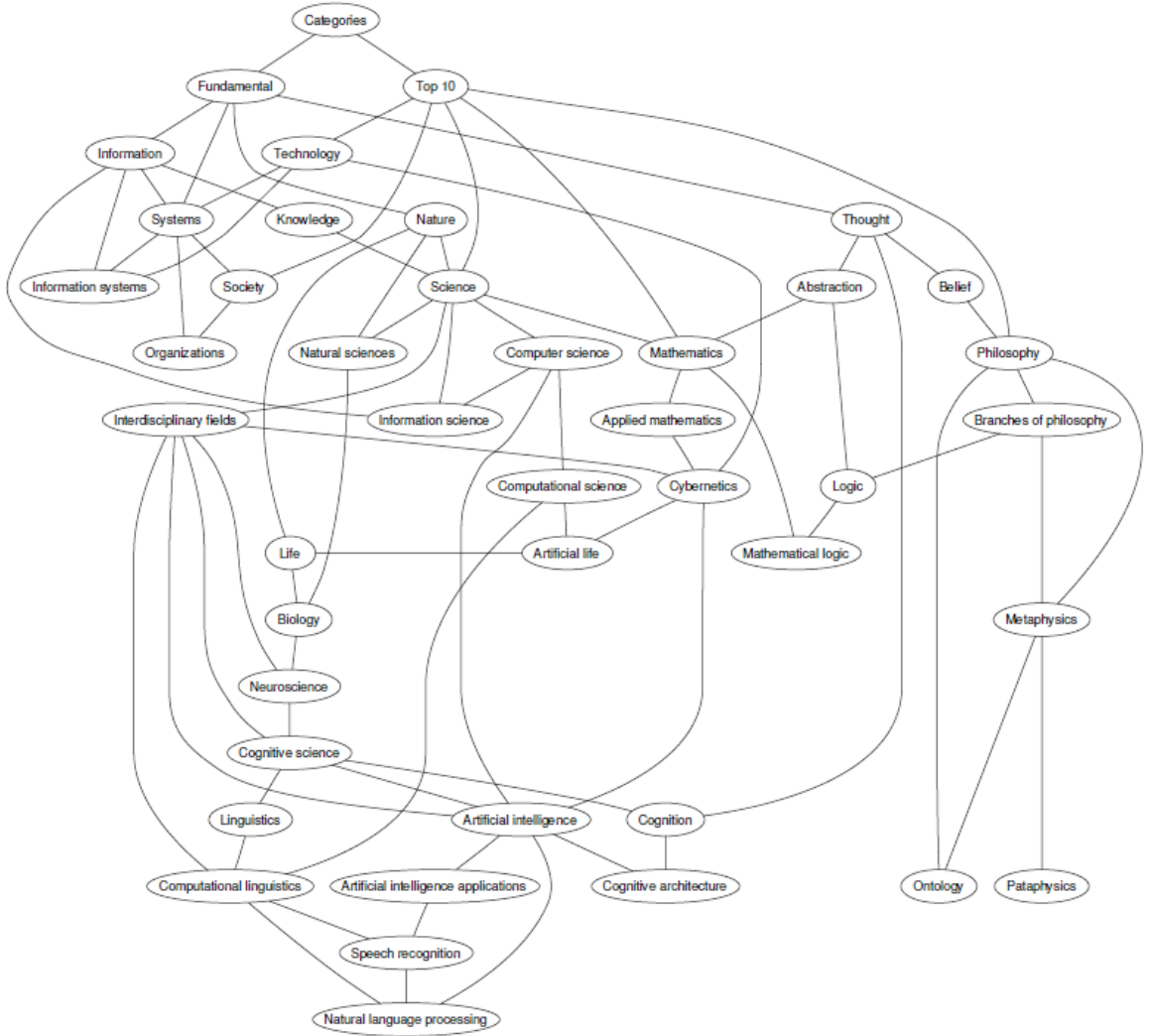
3.1.7 Kategori Sayfaları

Kategori sayfaları o kategoriye ait olan makaleleri ve alt kategorileri içeren ama içeriği olmayan makalelerdir. Bu sayfalar kategori başlığı ile başlar ve alt kategorilerinin isimleri ile devam eder. Wikipedia her makaleyi bir kategoriye atar. Kategori vikipedistler tarafından el

ile atanır, bu nedenle her makalenin bir kategorisi olmak zorunda değildir. Aynı zamanda bazı makaleler birden fazla kategoriye ait olabilirler. Mesela Türkiye makalesi 3 kategoriye aittir:

1. Türkiye Cumhuriyeti tarihi
2. Türkiye
3. Türkçe konuşan bölgeler ve ülkeler

Bu nedenler Türkiye makalesi 3 farklı kategori sayfasında görülür.



Şekil 3.4 Wikipedia kategorilerinin ağaç yapısı ile gösterimi (Ponzetto ve Strube, 2007)

3.2 Diğer Hazırlama Yöntemleri

3.2.1 WikiRelate

Strube ve Ponzetto (2006), Wikipedia'nın folksonomi özelliğini kullanarak iki kelime arasındaki anlamsal benzerliği hesaplamayı amaçlamışlardır. Bunu yaparken de Wikipedia'da

bulunan makalelerin sahip oldukları kategorilerinden faydalanmışlardır. Çalışma sonucunda kelimelerin anlamsal benzerliklerini hem Wikipedia ile hem de WordNet ile hesaplayıp karşılaştırmışlardır.

Bu yöntemde verilen her iki kelime için ilgili Wikipedia makaleleri sorgulanmıştır. Bunu yaparken yönlendirme sayfaları varsa onlar da takip edilmiştir. (Örneğin “TBMM” makalesinin “Türkiye Büyük Millet Meclisi” makalesine yönlendirilmesi gibi) Eğer varılan makale anlam ayrımı makalesi ise buradaki her aday makale ile ilişkilendirme yöntemi tekrarlanarak içlerinden en başarılı sonuç vereni seçilmiştir. Daha sonra bulunan Wikipedia makalelerinin sahip oldukları kategori listeleri çıkartılmıştır. Bu listelerdeki kategoriler arasındaki uzunluklar taksonomi tabanlı ölçümlerle hesaplanmış ve bu hesaplamalar sonucunda en kısa mesafeyi veren makale çifti seçilmiştir.

3.2.2 PF-IBF

Nakayama vd., (2007a), otomatik olarak ilişkisel eş anlamlılar sözlüğü oluşturabilmek için yaptığı çalışmada Wikipedia’nın kendi içindeki bağlantı yapısından faydalanmışlardır. Kullandıkları metod LF-IBF (Link Frequency – Inversed Backward Frequency) daha sonradan PF-IBF (Path Frequency – Inversed Backward Link Frequency) olarak isim değiştirmiştir. Wikipedia’nın bu yapısı sayesinde sadece kelimeler arasındaki anlamsal ilişkileri bulmakla kalmayıp, her bir kelimenin eş anlamlılarını da bulabileceklerini göstermişlerdir. Yapılan çalışma sonucunda 1,7 milyon makaleyi analiz ederek içerisinde yaklaşık 78 milyon ilişki barındıran bir eş anlamlılar sözlüğü oluşturmuşlardır.

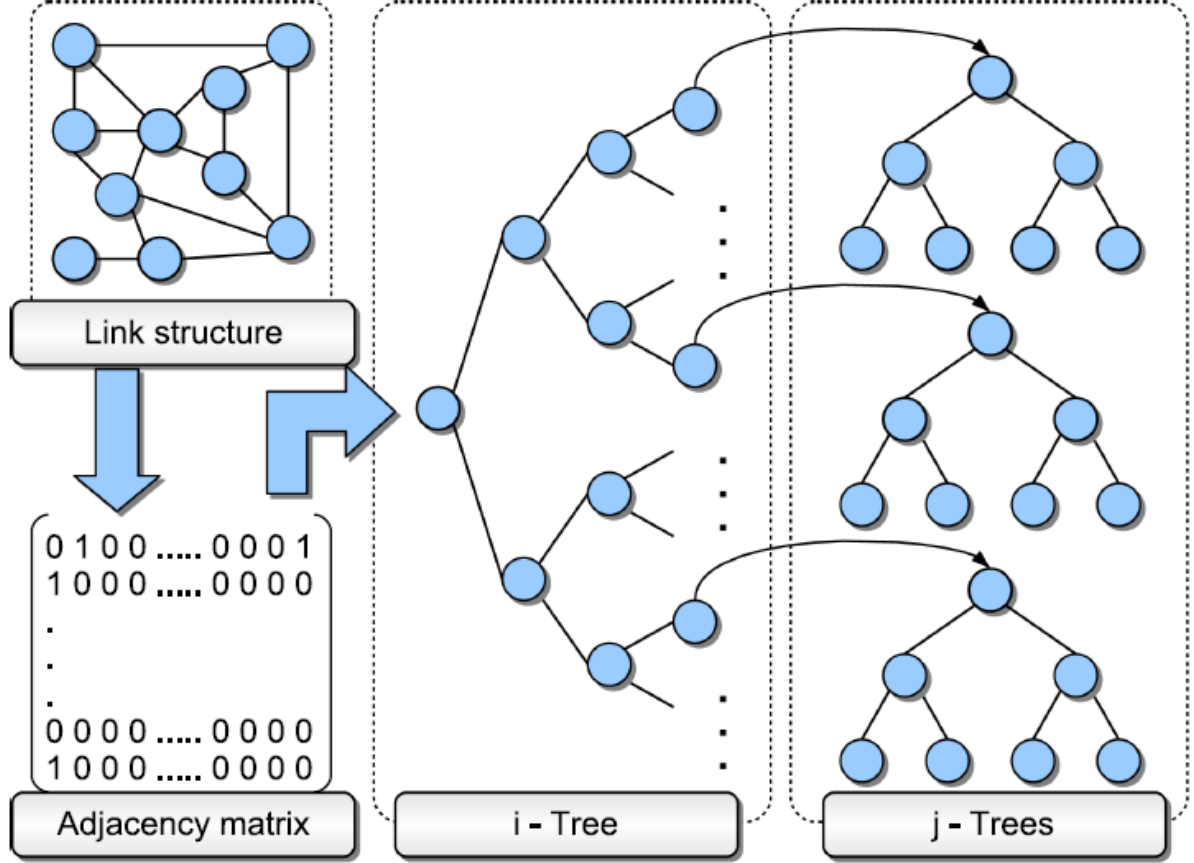
İki Wikipedia makalesi arasındaki ilişki bir makaleden diğer makaleye ulaşılacak yolların sayısı ve her bir yolun uzunluğu ile hesaplanmıştır. Buna ek olarak genel kapsamlı makaleler, diğer bir çok makaleye kısa mesafede bağlantı sağlayabildiğinden bu tarz makaleler de ters orantılı bir şekilde hesaplama dahil edilerek iki makale arasındaki ilişki hesaplanmıştır.

3.2.3 Çift İkili Ağaç ve PF-IBF

Nakayama vd., (2007b) PF-IBF metoduna bir eklenti yaparak bağlantı analizini daha verimli bir şekilde yapıp ilişkisel eş anlamlılar sözlüğü oluşturmayı denemişlerdir. Bu yöntemde DTB (Dual Binary Tree) veri yapısı ve DTB için geliştirilen bir çarpma algoritmasının kullanılmasını önermişlerdir.

Bu yöntemde PF-IBF metodunu daha verimli ve performanslı çalışabilmesi için çift ikili ağaçlar kullanılmıştır. Bu ağaçlar bitişiklik matrisindeki (Adjacency Matrix) verilerin

sıkıştırılmasında kullanılmıştır. Bitişiklik matrisindeki birçok eleman 0 değerine sahip olduğundan hesaplamalarda bunların bir etkisi olmamaktadır. Bu nedenle sadece matrisin içindeki kullanılabilir veriler çift ikili ağaç yapısında tutulduğunda ve bu veri yapısına özel geliştirilen çarpma algoritması ile hesaplamalar yapıldığında ölçümler çok daha performanslı bir şekilde gerçekleştirilebilmektedir.



Şekil 3.5 Bitişiklik matrisinin çift ikili ağaca çevrilmesi (Nakayama vd., 2007b)

3.2.4 Bağlantıların Birlikte Bulunması

Ito vd., (2008) bağlantıların birlikte bulunması analizi ile PF-IBF'den özellikle performans açısından daha ölçeklenebilir bir metot ortaya atmışlardır. Bu yöntemde benzerlik ölçümü için cosine metodunu kullanmışlardır. Bu yöntemde TF-IDF (Term Frequency – Inverse Document Frequency) metodu entegre edilerek normalden daha başarılı hesaplamalar yapıldığını göstermişlerdir.

Burada benzerlik hesaplaması yapılırken bütün Wikipedia makalelerindeki bağlantılardan faydalanılmıştır. Her bir makale için bağlantılar sıralı bir şekilde çıkartılmıştır. Önceden belirlenen pencere uzunluğu içerisindeki her bağlantı ikilisi için gözükme adetleri toplanarak hesaplanmıştır. Bu hesaplama bütün Wikipedia makaleleri için tekrarlandığında her bir

bağlantı ikilisi için kaçar kez aynı pencerede birlikte bulunduğu bilgisi elde edilmiştir. Daha sonra verilen herhangi iki kelime için Wikipedia’da kaç pencere içinde birlikte bulundukları bilgisi ve birlikte bulunma adetleri cosine benzerlik ölçümü fonksiyonu ile iki kelimenin benzerlik değeri hesaplanmıştır.

3.2.5 Coreference Resolution (Zamir Çözümsemesi)

Ponzetto vd., (2006) yaptıkları çalışmada makine öğrenmesi tabanlı coreference resolution problemini Wikipedia ve WordNet gibi farklı kaynaklardan aldıkları ilişkisel özellikler ile çözmeye çalışmışlardır. Bu çalışmada anlamsal benzerlikleri çıkarmak ve ilişkisel ölçüm yapmak için Wikipedia ve WordNet taksonomileri kullanılmıştır.

“Coreference Decision” problemi için maksimum entropi metodu kullanılmıştır. Bunu yaparken MALLET (McCallum A. K., 2002) kütüphanesinden yararlanılmıştır. Verilerdeki aşırı uymanın önlenmesi için öncesinde verileri yumuşatmak hedeflenmiştir. Bunun için de gaussian yöntemi kullanılmıştır.

3.2.6 Kelime Anlam Belirsizliği Giderme

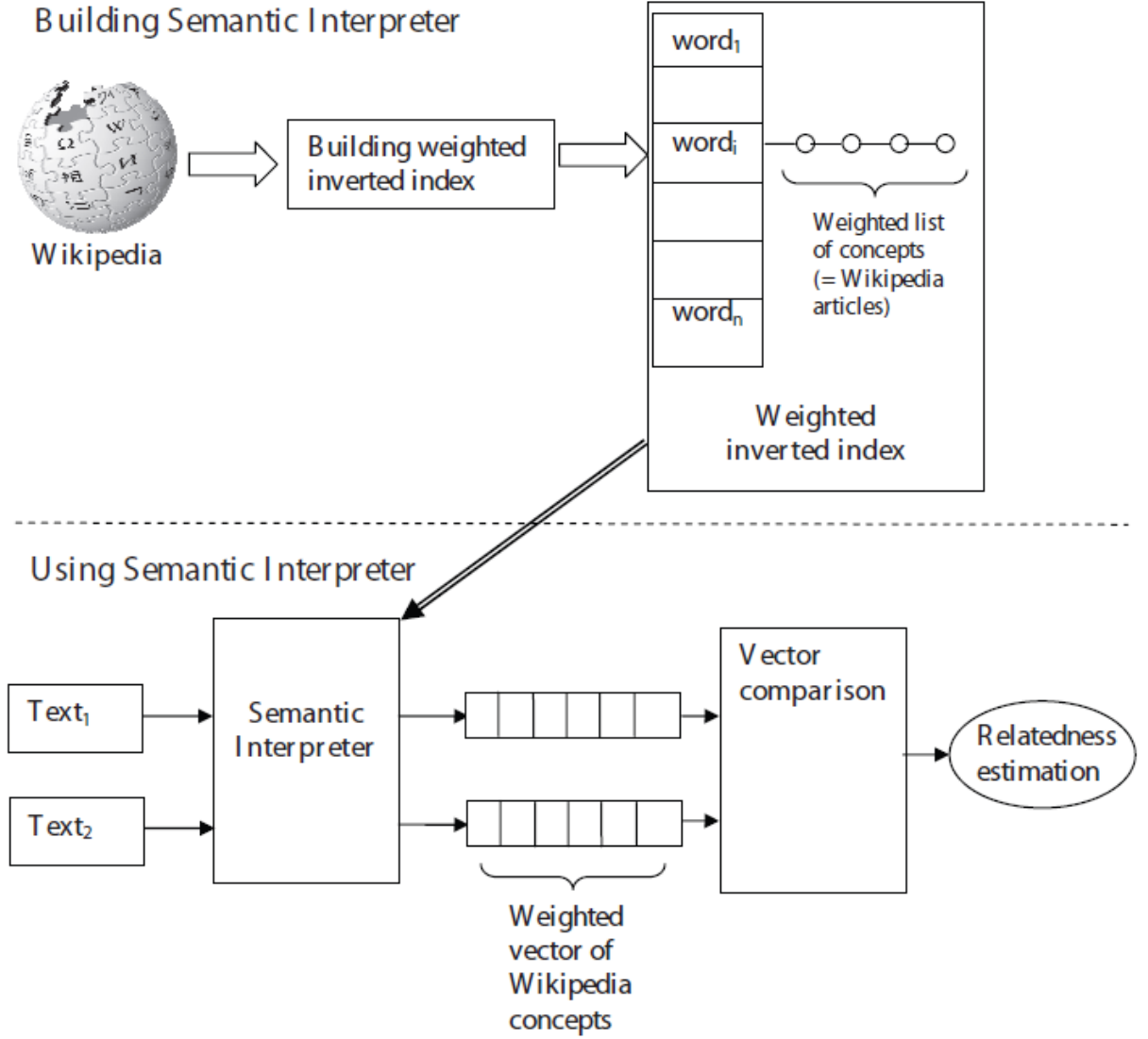
Mihalcea (2007), yaptığı çalışmada kelime anlamı belirginleştirmede Wikipedia’yı bir anlam ayrımı kaynağı olarak kullanmıştır. Önerdiği yöntemde kelimelerin anlamlarını bulmak için Wikipedia’daki bağlantılar kullanılmıştır. SENSEVAL’in alt kümesinde yaptığı deneysel çalışmalarda Wikipedia’daki bilgilerin kelime anlamı belirginleştirme sistemlerinde kullanılabileceğini göstermiştir. Diğer yöntemler ile karşılaştırıldığında hata oranlarında %30 ile %44 arasında bir azalma gözlemlenmiştir.

3.2.7 Explicit Semantic Analysis

Gabrilovich ve Markovitch (2007), yaptıkları çalışmada ESA (Explicit Semantic Analysis) yöntemini kullanarak kelimelerin anlamlarını bulmayı hedeflemişlerdir. Burada anlamı bulunmak istenen kelimeler, Wikipedia makalelerinin ağırlıklandırılmış vektör dizisine çevrilmiştir.

Her Wikipedia makalesi, ilgili makaledeki kelimelerin bir vektörü olarak temsil edilmektedir. Bu vektörlerin girdileri TF-IDF kullanılarak ağırlıklandırılmıştır. Buradaki ağırlıklar da kelimeler ile kavram arasındaki ilişkiyi belirtmektedir. Anlamsal yorumu hızlandırmak için ters çevrilmiş dizinler kullanılmıştır. Ters çevrilmiş dizinler her kelimeyi bir kavram vektörü olarak temsil etmektedir. İki metin parçasının arasındaki ilişkiyi hesaplamak için metin

parçalarına ait anlamsal yorumlayıcıdan geçmiş vektörler cosine metric yöntemi ile karşılaştırılmıştır.



Şekil 3.6 ESP kullanılarak kavramların arasındaki ilişki hesaplama (Gabrilovich ve Markovitch, 2007)

3.2.8 Domain Specific Thesauri

Milne vd. (2006), yaptıkları çalışmada Wikipedia'nın kapsamını ve doğruluğunu incelemek için belli alanlarda eş anlamlılar sözlüklerindeki anlamsal ilişkiler ile Wikipedia'daki kavram ve bağlantıları karşılaştırmışlardır.

3.2.9 İki Dilli Sözlük Hazırlama

Erdmann vd. (2008), yaptıkları çalışmada iki dilli sözlük oluşturmak için Wikipedia bağlantılarını kullanmışlardır. Bunun için sadece diller arasında geçişi sağlayan bağlantılar değil, yönlendirme bağlantıları ve bağlantılardaki metinler de kullanılmıştır. Yöntemin tek

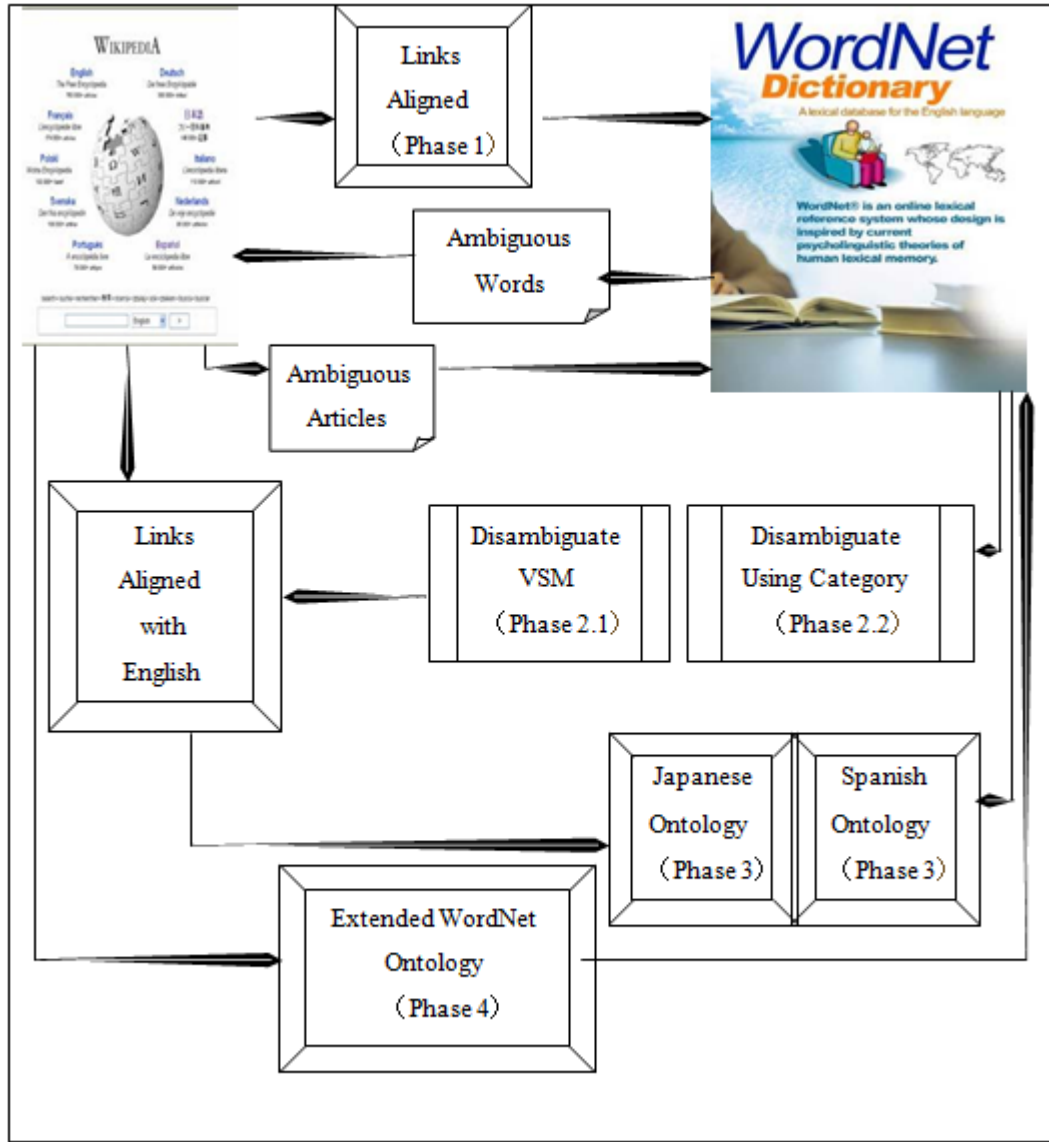
dezavantajı ise kullanılan metotlarda sadece tek bir kelimenin kullanılabilmesidir. Birçok kelimedenden oluşan terimler bu yöntemde kullanılamamaktadır.

Wikipedia makalelerindeki farklı iki dil için yazılmış aynı makalelerin başlıklarının ve içeriklerinin aynı olduğu kabul edilirse, diller arasındaki bağlantılar analiz edilerek terimlerin tercümeleri bulunabilir. Fakat genelde, bir makalede, en fazla bir adet diller arası bağlantı bulunmaktadır. Bu nedenle sadece bu bağlantılar analiz edildiğinde kapsam çok daralmaktadır. Fakat yönlendirme bağlantıları ve bu bağlantıların metni kullanıldığında bir terimin çevirisindeki kullanılabilecek dil adedini çok daha arttırmak mümkün olmaktadır.

3.2.10 Japonca-İspanyolca Dillerinde Sözlük Hazırlama

Ramírez vd. (2007), yaptıkları çalışmada Japonca ve İspanyolca dilleri için bir sözlük oluşturmak amacıyla Wikipedia'yı kullanmışlardır. Bunu yaparken İngilizce dilini de ortak dil olarak kullanmışlardır. Bu sözlük oluşturulurken çıkabilecek kelime anlamı belirsizlikleri için de WordNet'ten faydalanılmıştır. Belirsizlik gideriminde iki yöntem kullanılmıştır. Birincisi vektör uzay modeli, ikincisi ise Wikipedia'daki kategorisel bilgiler. Bu iki metodun birleşimi ise en iyi sonucu vermiştir.

Bu yöntem uygulanırken Wikipedia'daki bağlantıların hepsi kullanılmıştır. Bağlantıların birden fazla anlamı olup olmadığını anlamak için WordNet ile karşılaştırılmıştır. Bir belirsizlik olması durumunda bu belirsizliği gidermek için iki metod kullanılmıştır. Birinci yöntemde, Wikipedia makalelerindeki içerik ile WordNet'teki açıklama cosine benzerlik metodu ile ölçülmüştür. İkinci yöntemde ise, Wikipedia'daki kategori WordNet'in ontolojisindeki ilgili kelime ile karşılaştırılmıştır.



Şekil 3.7 WordNet ve Wikipedia kullanılarak iki dilli sözlük hazırlama (Ramírez vd., 2007)

4. WIKIPEDIA

Wikipedia, içeriğinde geniş çaplı konularda makaleler barındıran, çok dilli, web tabanlı bir ansiklopedidir. Wikipedia, dallarında uzman kişilerce hazırlanmış olan çevirim içi ansiklopedi Nupedia'nın devamı sayılmaktadır. 2001 yılında sadece İngilizce desteği olan ve uzman olmayan kişilere de yazma hakkı tanıyan bir proje olarak ortaya çıkmıştır. Wikipedia bütün dünyaya yayılmış gönüllü makale yazıcıları (vikipedistler) tarafından hazırlanmıştır. Birçok ülkeden vikipedistlerin olmasının sonucu olarak Wikipedia makaleleri birçok dile çevrilmiştir. Her dilde farklı sayıda olmak kaydıyla Wikipedia şu anda 200'den fazla dilde makale barındırmaktadır.

Wikipedia'daki makaleler çok geniş alanlardan oluşmaktadır. Örneğin; bilim alanında biyoloji ve antropoloji; eğlence alanında albüm isimleri, aktörler ve oyuncular ve Harry Potter gibi hayali karakterler alanında birçok içeriğe sahiptir.

Wikipedia, vikipedistlere kullanımı kolay ve sınırlandırılmamış bir ortam sağlar. Sadece yazılan içeriklerin orijinal olması, çalıntı olmaması gerekmektedir.

4.1 Organizasyonun Temel Tanıtımı

Wikipedia, içeriği dünyanın her köşesinden gönüllü insanlar tarafından hazırlanan açık kodlu, kâr amacı gütmeyen ücretsiz sunulan bir ansiklopedidir. İnternet bağlantısı bulunan herhangi bir kişi, tüm makalelerde ekleme, çıkartma, düzenleme yapabilir.

Türkçe Wikipedia, diğer tüm dillerdeki Wikipedialarla birlikte, kâr amacı gütmeyen Wikimedia Vakfı'nın tescilli markasıdır.

Wikimedia kar amacı gütmeyen ve ABD'nin Florida eyaletindeki yasalara tabi bir vakıftır. Vakıf ilk olarak 20 Haziran 2003 tarihinde Wikipedia'nın kurucusu Jimmy Wales tarafından duyurulmuştur.

Wikipedia'nın yazı düzenleme kurallarına uygun olmak ve kabul edilebilir bir standartta olmak kaydıyla, bilgi ve alıntılar eklenebilir, referans verilebilir. Bilgi eklerken ya da makaleyi geliştirirken Wikipedia içeriğinin bozulacağından korkulmasına gerek yoktur, çünkü diğer vikipedistler gerekiyorsa bariz hataları düzeltmektedirler. Ayrıca, Vikipedi Ansiklopedi Yazılımı "MediaWiki", editör kaynaklı hataların kolayca geri alınabilmesini sağlayacak şekilde tasarlanmıştır.

Wikipedia, artık yayın hayatını sona erdirmiş olan Nupedia'nın yan kuruluşu olan bir ansiklopedi yaratma amacıyla kurulmuştur. Nupedia, ayrıntılı bir ön izleme sistemine sahiptir

ve yalnızca yüksek kaliteli katkılar yapılabilirdi, ancak buna bağlı olarak da makalelerin yazımı ve geliştirilmesi çok yavaş olmaktaydı. 2000 yılında, Jimmy Wales (Nupedia'nın kurucusu) ve Larry Sanger Nupedia'yı daha açık ve özgür hale getirmek için çeşitli yönlerde iyileştirmeler yapılmasını uygun gördüler. Wikipedia'nın formatının daha açık, daha az resmi bir ansiklopedi projesi için uygun olduğunu fark ettiler. Sanger, wiki kavramını daha önceden bilen Wales'i kolayca Nupedia için bir Wikipedia kurmaya ikna etmeyi başardı ve Nupedia'nın ilk wikisi 10 Ocak 2000'de internete konuldu.

15 Ocak 2000'de yeni projeye Wikipedia ismi kondu ve kendi alan adı adresi olan wikipedia.com alındı. Bant genişliği ve sunucu ihtiyacı Wales tarafından giderilmiştir.

Mayıs 2001'de, İngilizce olmayan ilk Wikipedia'lar yayın hayatına başladı. Bu diller Katalanca, Çince, Felemenkçe, Almanca, Esperanto, Fransızca, İbranice, İtalyanca, Japonca, Portekizce, Rusça, İspanyolca ve İsveççe idi.

4.2 Wikipedia İstatistikleri

100'den fazla dilde, 3.800.000 makale üzerinde çalışan 48.000 aktif editör vardır. Ağustos 2010 itibariyle Türkçe Wikipedia'da 147.718 makale bulunmaktadır. Wikipedia'nın sahip olduğu bilgi birikimini artırmak için her gün tüm dünyadan yüz binlerce ziyaretçi binlerce değişiklik yapmakta, binlerce yeni makale başlatılmaktadır. Katkıda bulunmak için ziyaretçilerde herhangi bir özel nitelik aranmamaktadır ve her yaştan insan Wikipedia'da makale yazımına yardım etmektedir.

Çizelge 4.1 7 Kasım 2009 tarihli Wikipedia istatistikleri

Wikipedia	Makale Adedi	Bağlantı Adedi	Düzenlenme Adedi
İngilizce	3.370.830	21.073.996	404.654.879
Türkçe	147.823	723.852	8.473.154
Toplam	14.289.139	52.240.962	797.140.803

5. KULLANILAN YÖNTEMLER

5.1 PF-IBF

Wikipedia gibi WEB tabanlı sözlükler içerdikleri makalelerden ve bu makaleleri bağlayan bağlantılardan oluşur. Bu nedenle bu tarz sözlükleri bir G grafi ile ifade edebiliriz.

$$G = \{V, E\} \quad (5.1)$$

V : makale kümesi,

E : bağlantı kümesi

PF-IBF yönteminde iki makale arasındaki benzerlik ilişkisini hesaplamak için iki farklı hesaplama kullanılmaktadır.

1. V_i makalesinden V_j makalesine giden bütün yolların adedi
2. V_i makalesinden V_j makalesine giden her yolun uzunluğu

İki makalede bulunan ortak bağlantı sayısı ne kadar çok ise bu iki makale arasındaki benzerliğin o kadar fazla olduğu söylenebilir. Buna ek olarak ortak içerdikleri bağlantıların arasındaki uzunluklara bağlı olarak iki makale arasındaki ilişkinin de daha güçlü veya daha zayıf olması beklenmektedir. Diğer bir deyişle iki makale G grafiğinde ne kadar yakın konumlanmışsa ve ne kadar çok ortak bağlantı paylaşıyorlarsa bu iki makale arasındaki benzerliğin diğer makalelere oranla daha fazla olduğu söylenebilir.

V_i ve V_j arasındaki bütün yolları $T = \{t_1, t_2, \dots, t_n\}$ olarak tanımlansın. Bu durumda, PF (yol frekansı) şu şekilde tanımlanmaktadır:

$$pf(v_i, v_j) = \sum_{k=1}^n \frac{1}{d(t_k)} \quad (5.2)$$

$d()$ fonksiyonu, t_k yolunun uzunluğuna bağlı olarak artan bir fonksiyon olarak tanımlanmaktadır. $d()$ için logaritma fonksiyonu gibi monoton artan bir fonksiyon kullanılabilir.

Buna ek olarak iki farklı makale arasındaki bağlantı sayısı da benzerlik oranına etki eden bir faktördür. Örneğin bir makalenin diğer makalelere çok sayıda bağlantı verdiğini kabul edelim. Bu bağlantılar nedeni ile bu makale birçok makale arasında çok fazla adette kısa yollar bulunmaktadır. Bu nedenle sadece PF fonksiyonu kullanıldığında bu tarz makaleler ile birçok makale arasında güçlü benzerlik oranı hesaplanabilmektedir. Fakat bu tarz makaleler genel

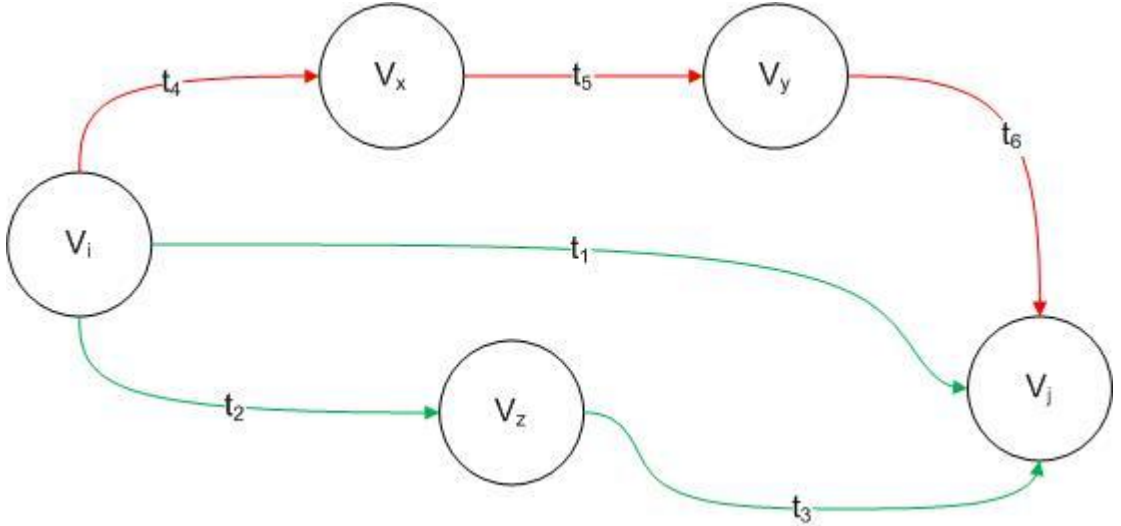
yapıda makaleler olarak tanımlanmalıdır. Ve çoğu zaman da genel yapıdaki makalelerin önemi yüksek değildir. Bu sorunu aşabilmek için *PF* fonksiyonunu ters çevrilmiş geri bağlantı frekansı (inversed backward link frequency - *IBF*) fonksiyonu ile birlikte kullanmak gerekmektedir. Bu durumda benzerlik oranı aşağıdaki şekilde hesaplanmalıdır:

$$pfibf(v_i, v_j) = pf(v_i, v_j).ibf(v_j) \quad (5.3)$$

$$ibf(v_j) = \log \frac{N}{bf(v_j)} \quad (5.4)$$

N ; Wikipedia'daki toplam makale sayısını belirtmektedir. $bf(v_j)$ ise v_j makalesinin sahip olduğu toplam geri bağlantı adedini temsil etmektedir. Bu yöntemle sadece belli başlı makaleler ile bağlantı paylaşan bir makale kendisinden çok bağlantı paylaşan bir makaleye göre daha yüksek bir *PF-IBF* değerine sahip olacaktır.

Bu yöntemde hesaplama yapılırken v_i makalesinden v_j makalesine giden bağlantılar aranırken grafikte iki makalenin arasında kalan ortak makaleler de kullanılmaktadır. Fakat hesaplamalarda ortak makaleleri ararken derine inildikçe hesaplama performansının da düştüğü gözlemlenmiştir. Bu nedenle hesaplamalarda maksimum derinlik için 2 değeri kullanılmıştır.



Şekil 5.1 PF-IBF metodunda kullanılan makale ve bağlantıları

Şekil 5.1’de de gösterildiği gibi v_i makalesinden v_j makalesine giden yollardan t_1 ve t_2 - t_3 ’den oluşan iki yol kullanılırken t_4 - t_5 - t_6 ’dan oluşan yol 2 uzunluğundaki maksimum derinliği aştığı için kullanılmamaktadır.

5.2 TF-IDF

TF-IDF ağırlıklandırma yöntemi bilgi çıkarımı ve metin madenciliğinde sık kullanılan bir yöntemdir. Yöntemin amacı bir kelimenin belli bir veri setindeki dokümanda ne kadar öneme sahip olduğunu hesaplamaktır. Bu yöntemin varsayımına göre bir kelime o dokümanda ne kadar çok geçmiş ise dokümandaki önemi de o kadar fazladır. Yöntemin çeşitli varyasyonları arama motorlarında sıkça kullanılmaktadır.

Örneğin elimizde “bir şarkısın sen” şeklinde bir sorgu ve bunu arayacağımız bir dokümanlar kümesi olsun. Öncelikle bu kelimelerin hiç birinin geçmediği dokümanları veri setinden çıkartmalıyız. Daha sonra bu kelimelerin her birinin bütün dokümanlarda geçme adetlerini hesaplamalıyız. Bu adet terim frekansı (term frequency) olarak geçmektedir. Fakat “bir” kelimesi çok genel bir kelime olduğundan ve asıl anlamlı kelimeler olan “şarkısın” ve “sen” kelimelerinin frekansından çok daha fazla çıkacağından hesaplarımızın hatalı çıkmasına sebep olacaktır. Bunun için tersine doküman frekansı (inverse document frequency) kullanılmaktadır. Böylece veri setinde çok sık geçen kelimelerin ağırlıkları, daha nadiren geçenlere göre azaltılmış olur.

Verilen bir dokümandaki bir kelimenin terim adedi, basitçe o dokümanda o kelimenin kaç kere geçtiği şeklinde tanımlanabilmektedir. Diğerlerine göre daha uzun dokümanlarda frekansı aranan kelime diğerlerine göre daha az öneme sahip olsa bile terim frekansı olarak daha büyük bir adet verecektir. Bu da hesaplamalarda sapmaya neden olacaktır. Bu nedenle terim frekansı hesaplanırken bir normalizasyon işlemi uygulanır.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (5.5)$$

$n_{i,j}$: i . terimin j . dokümanda geçme adedidir.

$\sum_k n_{k,j}$: terimi ise j . dokümandaki bütün kelimelerin adetlerinin toplamıdır.

Tersine doküman frekansı (*IDF*) ise bir terimin tüm veri setinin içindeki önemini hesaplar. Doküman adedi, o terimi içeren doküman adedine bölünerek *idf* hesaplanabilmektedir.

$$idf_i = \log \frac{|D|}{|\{d : t_i \in d\}|} \quad (5.6)$$

$|D|$: veri setindeki toplam doküman adedi

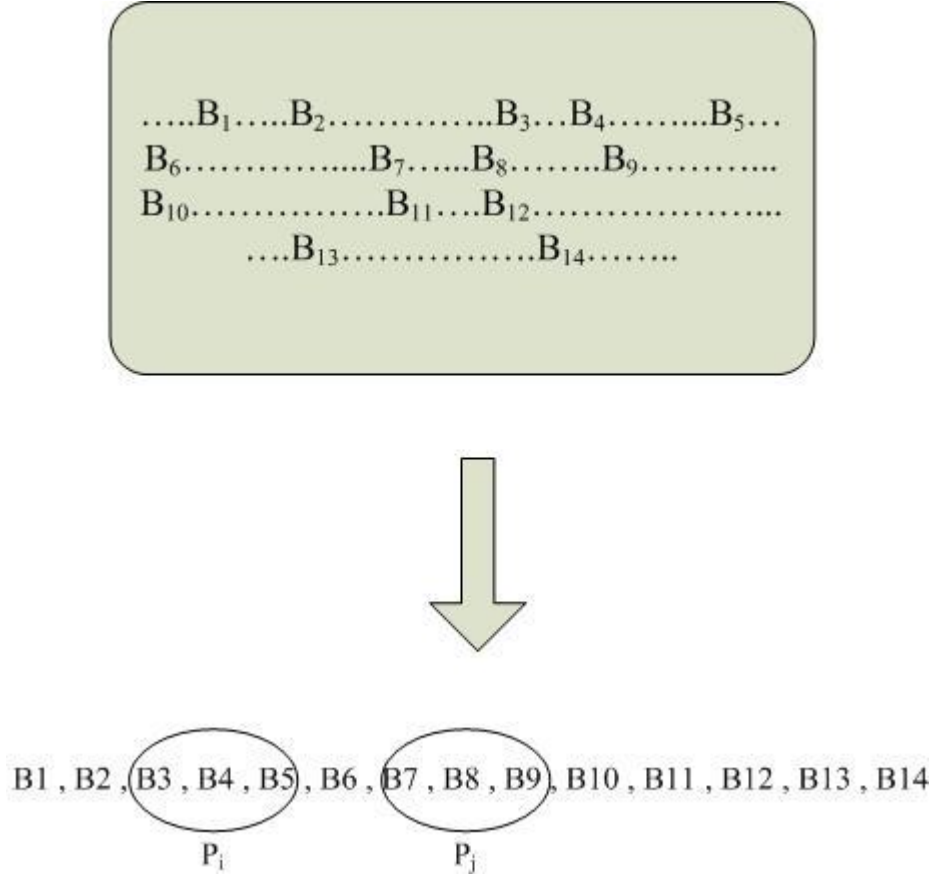
$|\{d : t_i \in d\}|$: i. terimin geçtiği doküman adedi.

5.3 Bağlantıların Birlikte Bulunması

Bağlantıların birlikte bulunması yönteminin TF-IDF yöntemine göre avantajı, hesaplamayı yerel bilgiler ile değil genel bilgiler ile yapmasıdır. Analizin bütün Wikipedia makalelerinin birlikte kullanılarak yapılması, sonucun tek bir makalenin içerik kalitesine bağlı olmamasını sağlamıştır. TF-IDF yönteminde sadece ilgili makalenin içeriğindeki bilgiler kullanılıyorken bağlantıların birlikte bulunması yönteminde ilgili makalenin karakteristiklerini bulmak için bütün Wikipedia'daki bağlantılardan faydalanılmaktadır. Buna ek olarak bağlantıların birlikte bulunması yöntemi ile analiz yapıldığında hesaplamalardaki performans veri miktarı ile doğru orantılıdır. PF-IBF metodundaki gibi büyük miktarda ve uzun hesaplamalar gerektirmez.

Bu yöntem kullanılırken içeriklerdeki bağlantılar farklı metinler ile isimlendirilmiş olsa bile, eğer aynı makalenin URL'sine bağlantı sağlıyorsa bu iki bağlantı aynı bağlantılar olarak tanımlanır. Örneğin “MS” ve “Microsoft” isminde iki adet bağlantımız olsun. Bunların ikisi de “Microsoft” makalesine yönlendirildiği için aynı bağlantı olarak tanımlanmalıdır.

Eğer bir makaledeki iki farklı bağlantı belirli bir aralıkta bulunuyorsa bu iki bağlantı birlikte bulunur şeklinde tanımlanabilir. Genelde bu aralık n adet bağlantının içinde olduğu bir pencere olarak belirlenir. Makaledeki bağlantı adedi arttığında birlikte bulunan bağlantı kombinasyonu da artacaktır. Fakat bu şekilde bir limit belirlendiğinde hesaplamalar içinde kullanılan kombinasyon adedi daha kolay hesaplanabilir seviyede kalacaktır. Özellikle makaledeki bağlantı adedi arttığında, birbirine yakın bulunan iki bağlantının benzerliği, aynı makaledeki birbirine yakın olmayan herhangi iki bağlantının benzerliğinden çok daha fazla olduğu gözlemlenebilir.



Şekil 5.2 Bağlantıların birlikte bulunması metodunda kullanılan bağlantı ve pencere yapısı

Şekil 5.2’de birlikte bulunma sıklığının hesaplanmasının genel bir görünümü gösterilmektedir. Buradaki pencere genişliği 3 olarak alınmıştır. Şekil 5.2’de B_i terimleri makaledeki bağlantıları temsil etmektedir. Bu hesaplama modelinin akışı şu şekildedir:

1. Bir makaledeki bütün bağlantılar makalede gözükmeye sıralarına göre çıkarılır.
2. Belirlenen pencere en soldan itibaren birer birer sağa kaydırılır.
3. Pencere içerisinde kalan her iki bağlantının birlikte bulunma sıklığı sayılır.

Bu akış Wikipedia makalelerinin hepsine uygulanmıştır. Ve tüm Wikipedia’daki bağlantı ikililerinin birlikte bulunma adetleri üst üste toplanarak hesaplanmıştır. Genel olarak birlikte bulunma frekansını hesaplamak için kullanılan dört farklı yöntem mevcuttur:

1. Cosine:

$$\text{Cosine}(x, y) = \frac{f_{xy}}{\sqrt{f_x \cdot f_y}} \quad (5.7)$$

2. Mutual Information:

$$MI(x, y) = \log \frac{P(x, y)}{P(x)P(y)} \quad (5.8)$$

3. Dice Coefficient:

$$Dice(x, y) = \frac{2 \cdot f_{xy}}{f_x + f_y},$$

$$(0 < Dice(x, y) \leq 1) \quad (5.9)$$

4. Improved Dice Coefficient:

$$IDice(x, y) = w(f_{xy}) \frac{2 \cdot f_{xy}}{f_x + f_y}, \quad (5.10)$$

$$w(f_{xy}) = \begin{cases} f_{xy} \\ \log f_{xy} \end{cases}$$

Burada f_x ve f_y , sırası ile x ve y bağlantılarının bulunduğu pencere adedidir. f_{xy} ise x ve y bağlantılarının birlikte bulunduğu pencere adedidir. Bu metotlarla hesaplanan birlikte bulunma metotlarına birinci dereceden birlikte bulunma metodu denilmektedir. Çünkü bu metotlarda eğer iki bağlantı direkt olarak birlikte bulunuyor ise benzer olarak tanımlanabilir. Fakat Schütze ve Pedersen (1997) yaptıkları çalışmada birinci dereceden birlikte bulunma metodunun eksikliklerine dikkat çekmiş ve ikinci dereceden birlikte bulunma metodunu önermiştir. Bazı veri setlerinde eğer içerikler yeterli değil ise aslında benzer olabilecek durumdaki iki bağlantı direkt olarak aynı pencere içinde bulunmayabilir. Bu iki bağlantı aynı pencerede birlikte bulunmasalar da eğer aynı birlikte bulunma desenine sahip ise ikinci dereceden birlikte bulunma metodu bunların benzerliğini ortaya çıkarabilir. Ito vd., (2008) daha önceki deneysel çalışmalarında birkaç yüz bağlantı çifti ile yapılan testlerde bağlantı çiftlerinin %50'si birinci dereceden birlikte bulunma metodu ile benzerlikleri ortaya çıkmaya bile ikinci derecen birlikte bulunma metodu ile benzerlikleri hesaplanabilmiştir. Bunun sebebi içeriklerin yetersizliği nedeni ile iki bağlantı aynı pencerede bulunamıyor, fakat aynı birlikte bulunma desenine sahip olduğu için benzerlikleri ikinci dereceden birlikte bulunma metodu sayesinde ortaya çıkabiliyor.

Bu yöntemde ikinci dereceden birlikte bulunma metodunu kullanmak için öncelikle her bir bağlantının vektörü oluşturulur ve buna bağlantı vektörü denir. Bir bağlantı vektörü birinci derecen birlikte bulunma desenlerini ifade eder. i . bağlantının bağlantı vektörünü ifade eden v_i aşağıdaki şekilde tanımlanır:

$$v_i = \{l_{i1}, l_{i2}, l_{i3}, \dots, l_{in}\}$$

Burada l_{ij} , i . ve j . bağlantıların birinci dereceden birlikte bulunmasını ifade eder. Bütün bağlantı vektörleri oluşturulduktan sonra iki bağlantı arasındaki benzerliği hesaplamak için aşağıdaki gibi cosine metodu kullanılmaktadır:

$$\begin{aligned} \cos(v_i, v_j) &= \frac{v_i \cdot v_j}{|v_i| |v_j|} \\ &= \frac{\sum_{k=1}^n l_{ik} l_{jk}}{\sqrt{\sum_{k=1}^n l_{ik}^2} \sqrt{\sum_{k=1}^n l_{jk}^2}} \end{aligned} \quad (5.11)$$

6. DENEYSEL SONUÇLAR

6.1 Kullanılan Yazılımlar

Kelime benzerliklerinin Wikipedia kullanılarak hesaplanmasında öncelikle Wikipedia içeriğinin indirilmesi gerekmektedir. Bunun için <http://download.wikimedia.org/trwiki/latest/> adresinden Wikipedia'nın SQL dosyaları indirilmiştir. Bu dosyaların içeriğinde veritabanında bulunması gereken tabloları oluşturmak için gerekli SQL cümlecikleri ve bu tabloların içeriğinde bulunan verileri tablolara eklemek için gerekli olan SQL cümlecikleri bulunmaktadır. MySQL veritabanı kurulumu yapıldıktan sonra bu dosyaların içeriğinde bulunan veriler veritabanına eklenmiştir. Yöntemlerde kullanılan veritabanı sorgularının performansı ve T-SQL dilinin MS SQL Server'da yapılabilecek çalışmalara etkisi düşünüldüğünden MySQL veritabanında bulunan tablolar içerikleri ile birlikte MS SQL Server 2008 veritabanına geçirilmiştir. Bu işlem tamamlandıktan sonra <http://download.wikimedia.org/trwiki/latest/> adresinden Türkçe Wikipedia içeriklerinin bulunduğu XML formatındaki dosyalar indirilmiştir. Bu dosyaların içeriği EK-1 de bulunabilir. Her makalenin içeriğinde bulunan bağlantıların formatı aşağıdaki şekildedir:

[[Yıldız Teknik Üniversitesi Fen Edebiyat Fakültesi|Fen-Edebiyat Fakültesi]]

Burada “[[” ve “]]” karakterleri arasında kalan kısım bağlantının yapılacağı makalenin ismini ve bağlantısını makale içerisinde hangi metinle gözükeceğini belirtmektedir. Örneğin bu bağlantıda bağlantı verilen Wikipedia makalesinin ismi “Yıldız Teknik Üniversitesi Fen Edebiyat Fakültesi” olarak geçmektedir. Fakat makale üzerinde bunun yerine “Fen-Edebiyat Fakültesi” metni gözükecektir.

Bu ve bunun gibi bütün bağlantıların sıraları, makalede gözükeceği metinleri ve yönlendirme yapılacak makalelerin isimleri her makale için ayrı ayrı çıkartılıp veritabanında saklanmıştır. Bunun için Java programlama dilinde yazılan önışleme kodu kullanılmıştır. Bu aşamadan sonra kullanılacak yöntemler Java programlama dili ile yazılmış ve oluşturulan veritabanı kullanılarak kelime benzerlikleri hesaplanmıştır.

Kullanılan yazılım ve sürümleri Çizelge 6.1'deki gibidir.

Çizelge 6.1 Yöntemlerin kullanıldığı bilgisayar yazılımları

İşletim Sistemi	Windows 7 Ultimate Version 6.1
MySql	5.1.49
MS SQL Server	Microsoft SQL Server 2008 (SP1) - 10.0.2531.0
Java	1.6 Update 16
IDE	IntelliJ IDEA Ultimate 9.0

6.2 Kullanılan Donanımlar

Bu geliştirmelerin yapıldığı ve uygulamanın kullanıldığı bilgisayarın donanımsal özellikleri aşağıdaki gibidir.

Çizelge 6.2 Yöntemlerin kullanıldığı bilgisayar donanımları

İşlemci	Intel® Core™2 Duo işlemci T7700, 2.40 GHz
Bellek	2048 MB
Dahili Sabit Disk Sürücüsü	250 GB
Sabit Disk Denetleyicisi	SATA
Sabit Disk Sürücüsü Hızı	5400 rpm

6.3 Test Verisi

Yapılan çalışmaları değerlendirebilmek için İngilizce eş anlam sözlüğü oluşturma amacıyla yapılan çalışmalarda kullanılan WordSimilarity-353* test derlemindeki kelime çiftlerinden Türkçe Wikipedia’da bulunan 139 kelime çiftinden oluşan bir kelime seti hazırlanmıştır.

*<http://www.cs.technion.ac.il/~gabr/resources/data/wordsim353/>

Bu kelime çiftleri 17 kişilik bir test grubuna puanlandırılmıştır. Test grubundan her kelime çifti için kelimelerin anlam ilişkilerine göre 0 ile 10 arasında puan verilmesi istenmiştir. Kullanılan yöntemlerin sonuçları ile insanların verdiği puanlar Spearman’s Rank korelasyon

hesaplaması ile karşılaştırılmıştır. Bu hesaplamada eğer iki yöntemin sonuçları ne kadar benzer ise korelasyon değeri de o kadar yüksek çıkmaktadır. Bu değer en fazla 1, en düşük -1 çıkabilir. Eğer korelasyon değeri negatif çıkmışsa bu iki yöntemin sonuçları arasında negatif yönde bir ilişki olduğunu gösterir.

6.4 Yöntemlerin Karşılaştırılması

PF-IBF yöntemi ile yapılan hesaplamalarda en yüksek derinlik 2 olarak belirlenmiştir. Bu değer belirlenirken değişik derinlikler ile testler yapılmıştır. Derinlik için 1 değeri kullanıldığında benzerliğin yüksek olması gereken makaleler arasında hesaplama sonucunda çok düşük ya da sıfır değerli benzerlik oranları bulunmuştur. 3 ve 4 derinlik değerleri ile yapılan hesaplamalarda da benzerlik oranlarının değerlerinde artış gözlemlense de sıralama ve doğruluklarında yeterli bir değişim gözlenmemesinin yanında hesaplama performansının da düştüğü gözlemlenmiştir. Ortalama hesaplama değerleri ve toplam süreler Çizelge 6.3’de gösterilmiştir. Bu hesaplamalar sonucunda en uygun derinlik değeri olarak 2 seçilmesi uygun görülmüştür.

Çizelge 6.3 PF-IBF yönteminin derinlik değerleri için hesaplama değerleri ve süreler

Derinlik Değeri	Toplam Süre (milisaniye)
1	9
2	84
3	3.811
4	153.121

Hem 1. dereceden hem de 2. dereceden bağlantıların birlikte bulunması yönteminde yapılan hesaplamalarda pencere boyutu 2 ile 7 arasında değiştirilerek sonuçlar gözlemlenmiştir. Pencere boyutunu 2 ve 3 aldığımızda sonuçların hesaplanmasında performans yüksek değerlere ulaşırken hesaplama sonuçlarındaki başarı diğer pencere boyutlarına göre daha az çıkmıştır. 5 ve yukarıdaki pencere boyutunda da başarı 4’e göre çok artmamış fakat yapılan hesaplamaların çoğalması nedeniyle hesaplama süresi artmıştır. Bu nedenle hesaplamalarda pencere boyutunun 4 olarak kullanılmasına karar verilmiştir.

Yapılan çalışmalar sonucunda PF-IBF, TF-IDF, 1. dereceden bağlantıların birlikte bulunması ve 2. dereceden bağlantıların birlikte bulunması metotlarının sonuçları Çizelge 6.4'deki gibidir.

Çizelge 6.4 Yöntemlerin test veri setine uygulanması

Makale 1	Makale 2	PF-IDF	TF-IDF	1. Dereceden Bağlantıların Birlikte Bulunması	2. Dereceden Bağlantıların Birlikte Bulunması
Afet	Sel	39,83102	0	0	0,01396
Afet	Kuraklık	24,29817	0	0	0,05384
Afet	Doğa	21,11101	0	0	0,01216
Afet	Su	19,11235	451.281,88	0	0,04897
Afet	Meksika	16,79512	0	0	0,04202
Apple	Apple_Computer	143,89738	0	0,5	0,2758
Apple	Pixar	112,79777	0	0	0,08352
Apple	IPhone	110,24011	126.010,22	0	0
Apple	Ipod	106,55784	2.138,45	0	0
Apple	CEO	103,60521	0	0	0,07471
Basketbol	Brezilya	63,7997	237.821,62	2,43133	0,17421
Basketbol	Futbol	27,80364	749.447,39	6,06118	0,30858
Basketbol	Beyzbol	25,31566	0	5,09325	0,40667
Basketbol	Meksika	22,3935	201.209,14	1,80907	0,18027
Basketbol	Tenis	20,8971	77.891,40	4,46006	0,38739

Bilgisayar	Yazılım	58,31656	102.657,30	3,43982	0,39709
Bilgisayar	Klavye	52,20734	0	2,61538	0,12435
Bilgisayar	İnternet	42,95463	1.288.683,84	3,40805	0,09748
Bilgisayar	Kimya	27,10091	0	1,91485	0,15372
Bilgisayar	Televizyon	23,53804	179.779,35	3,10341	0,19606
Bilim	Fizik	275,7375	17.982.083,82	2,8172	0,3417
Bilim	Kimya	239,43554	1.805.367,38	2,18025	0,27602
Bilim	Enerji	147,44137	308.038,10	2,7398	0,14713
Bilim	Güneş	124,61239	33.157.010,44	1,85312	0,04462
Bilim	Canlı	106,61595	142.649,57	2,52982	0,12984
Bilişim_bilimi	Bilgisayar	85,84623	4.060.060,69	0	0,2154
Bilişim_bilimi	Yazılım	54,14112	200.566,27	0	0,12966
Bilişim_bilimi	Fizik	42,05986	314.981,43	0	0,22678
Bilişim_bilimi	Bilim	27,50438	17.296.404,10	0	0,1869
Bilişim_bilimi	Kimya	27,10091	0	1	0,07039
Boks	Futbol	16,68218	91.507,64	5,82636	0,28472
Boks	Maraton	8,68732	0	0	0,07978
Boks	Beyzbol	8,43855	0	3,9092	0,53497
Boks	Doktor	7,67869	0	0	0,02277
Boks	Tenis	6,9657	0	4,56	0,28919
Caz	Virtüöz	18,3934	0	2	0,2392
Caz	Müzik	12,89035	0	4,25763	0,35468

Caz	Meksika	5,59837	0	0,8	0,26722
Caz	Telefon	0	0	0	0,00013
Caz	Tren	0	0	0,5	0,11225
Döviz	Dolar	30,23287	0	1,75	0,14661
Döviz	Para	22,72016	0	1,44338	0,16988
Döviz	Banka	12,6889	0	0	0,24723
Döviz	İsrail	10,86542	0	0	0,06029
Döviz	Fiyat	9,58882	0	0	0,09939
Faiz	Para	34,7237	0	3	0,16282
Faiz	Banka	28,70051	3.100,93	2	0,19549
Faiz	Telefon	0	0	0	0,00002
Faiz	Tren	0	0	0	0,02137
Faiz	Bilgisayar	0	0	0	0,03422
Gezegen	Güneş	188,23896	1.849.340,43	3,82974	0,43953
Gezegen	Atmosfer	47,80354	223.418,89	2,58873	0,36051
Gezegen	Yıldız	27,84714	5.003.763,09	4,37798	0,60427
Gezegen	Futbol	22,24291	0	0,7303	0,02301
Gezegen	Su	19,11235	138.782,81	0,55556	0,09167
Gökbilimci	Takımyıldız	25,08797	0	0	0,01878
Gökbilimci	Atmosfer	21,7785	0	0	0,06408
Gökbilimci	Güneş	19,08797	446.487,76	1,60357	0,04215
Gökbilimci	Anne	17,65526	0	0	0,00317

Güneş	Gezegen	271,52854	1.849.340,43	3,82974	0,43953
Güneş	Atmosfer	268,60152	3.452.330,25	4,49073	0,32638
Güneş	Yıldız	200,18533	96.649.409,14	6,54935	0,36656
Güneş	Su	165,64038	2.680.637,90	2,80141	0,12639
Güneş	Enerji	104,44855	113.778,87	2,79129	0,10511
Hastane	Doktor	24,34087	28.246,31	1,11803	0,16297
Hastane	Çocuk	19,71957	0	0	0,07401
Hastane	Millet	17,91577	23.332,88	0	0,09071
Hastane	Hemşire	13,72843	0	1,5	0,19829
Hastane	Sigorta	12,94416	0	0	0,09101
Hayvanlar	Yumurta	51,92465	44.797,47	0,77152	0,04508
Hayvanlar	Kuşlar	33,06674	2.573.019,79	7,61687	0,62125
Hayvanlar	Memeliler	32,9966	2.092.916,78	17,92242	0,45541
Hayvanlar	Canlı	31,39234	27.136,09	3,23077	0,10889
Hayvanlar	Bilim	13,75219	9.961.202,60	0,5	0,01716
Istakoz	Deniz	10,86493	0	0	0,0462
Istakoz	Hayvanlar	7,37487	0	0	0,01155
Istakoz	Radyo	6,86198	0	0	0,00107
Istakoz	Su	6,37078	65.013,81	0	0,03135
Istakoz	Bilgisayar	6,31921	0	0	0,00736
İçecekler	Alkol	69,47879	0	0	0
İçecekler	Şarap	56,32181	147.494,74	0	0

İçecekler	Kahve	32,72593	245.571,85	0	0
İçecekler	Su	31,85392	151.428,94	0	0
İçecekler	Sıvı	26,96379	14.273,68	0	0
Jaguar	Brezilya	414,29833	4.217.324,88	2,05922	0,13161
Jaguar	Kaplan	184,608	17.083.533,39	2,70875	0,37774
Jaguar	Meksika	171,75645	892.017,61	2,99054	0,10676
Jaguar	Memeliler	152,35361	2.064.839,59	2	0,06379
Jaguar	Başkan	104,07546	8.246,28	0	0,01239
Kaplan	Jaguar	85,41518	17.083.533,39	2,70875	0,37774
Kaplan	Orman	78,6087	1.447.732,35	1,20268	0,05266
Kaplan	Meksika	61,58212	892.517,82	2,58199	0,03563
Kaplan	Memeliler	57,5736	3.718.795,46	2,7	0,02086
Kaplan	Başkan	56,04063	8.250,90	0	0
Kedi	Memeliler	18,6032	222.901,54	0,66815	0,11776
Kedi	Kaplan	16,65391	806.830,88	1,52554	0,12766
Kedi	Jaguar	16,13897	1.036.772,60	1,76777	0,13548
Kedi	Hayvanlar	14,74975	259.474,17	1,35225	0,02757
Kedi	Göz	7,13795	30.536,84	1,8	0,2279
Konser	Müzik	23,10571	0	2,66501	0,23419
Konser	Telefon	0	0	0,40825	0,00012
Konser	Tren	0	0	0	0,08682
Konser	Bilgisayar	0	0	0	0,07023

Konser	Kitap	0	0	0,70711	0,07791
Memeliler	Hayvanlar	41,1884	2.092.916,78	17,92242	0,45541
Memeliler	Orman	26,2029	0	2,64211	0,10997
Memeliler	Kuşlar	21,636	1.227.975,10	4,5	0,17702
Memeliler	Millet	17,91577	0	0	0,00359
Memeliler	Bebek	17,08659	0	0	0,0202
Microsoft	Windows	182,49807	2.671,17	6,3424	0,38531
Microsoft	Microsoft_Windows	134,7953	1.910.690,62	14,70277	0,28376
Microsoft	İşletim_sistemi	116,50296	98.335,43	10,89476	0,53269
Microsoft	Paul_Allen	108,10618	0	2,13977	0,08814
Microsoft	ABD	103,90354	0	5,84145	0,20483
Proton	Fizik	6,38726	52.243,00	1,91881	0,04461
Proton	Telefon	0	0	0	0,01233
Proton	Tren	0	0	0,5	0,01418
Proton	Bilgisayar	0	0	0,33333	0,01434
Proton	Kitap	0	0	0	0,01044
Sel	Deniz	38,28496	39.200,48	0,77152	0,08623
Sel	Su	22,83902	162.569,71	0,7303	0,03886
Sel	Afet	15,72843	0	0	0,01396
Sel	Kuraklık	9,39981	0	1,8	0,20748
Sel	Yaşam	8,54329	0	0	0,03082
Takımyıldız	Güneş	38,17594	37.642,25	1,34374	0,19212

Takımyıldız	Yıldız	11,03417	0	1,21356	0,10944
Takımyıldız	Yaşam	8,54329	0	0,5	0,06177
Takımyıldız	Proton	7,68732	0	0	0,00562
Takımyıldız	Uzay	7,53592	0	0,46291	0,07507
Yazılım	Bilgisayar	22,65413	102.657,30	3,43982	0,39709
Yazılım	Klavye	14,56287	0	0,54772	0,09725
Yazılım	Ofis	9,66167	0	0	0,08759
Yazılım	Yaşam	8,54329	0	0	0,01632
Yazılım	Otomobil	6,78162	0	0	0,11878
Yıldız	Güneş	703,13576	96.649.409,14	6,54935	0,36656
Yıldız	Gezegen	304,67234	5.003.763,09	4,37798	0,60427
Yıldız	Enerji	288,18085	239.440,17	2,06559	0,17573
Yıldız	Atmosfer	174,22801	3.632.601,38	2,24537	0,4058
Yıldız	Uzay	162,66264	465.267,28	2,13615	0,32624

Hesaplamalar yapılırken geçen toplam süre ve ortalama süreler Çizelge 6.5’de milisaniye cinsinden gösterilmiştir. Ek 3’de ise her yöntem için kelime çiftlerinin hesaplanma süreleri gösterilmiştir.

Çizelge 6.5 Yöntemlerin hesaplama süreleri

	PF-IBF	TF-IDF	1. Dereceden Bağlantıların Birlikte Bulunması	2. Dereceden Bağlantıların Birlikte Bulunması
Toplam Süre (ms)	80	44.814	4.409	12.135
Ortalama Süre (ms)	0,575539568	322,4	31,71942446	87,30215827

Test veri setindeki bütün kelime çiftlerinin tüm yöntemler ile hesaplanmaları test grubunun puanlamaları Spearman's Rank korelasyon hesaplama yöntemi ile hesaplandığında korelasyon sonuçları Çizelge 6.6'da gösterilmiştir. Bu değerlere göre ortalama korelasyonlar hesaplandığında (Çizelge 6.7) en başarılı yöntemin PF-IBF olduğu gözlemlenmiştir.

Çizelge 6.6 Spearman's Rank korelasyon hesaplaması sonuçları

Makale 1	Makale 2	PF-IDF	TF-IDF	1. Dereceden Bağlantıların Birlikte Bulunması	2. Dereceden Bağlantıların Birlikte Bulunması
Afet	Sel	0,9	0,00000	-0,5	-0,5
Afet	Kuraklık	0,9	0,00000	-0,5	-0,5
Afet	Doğa	0,9	0,00000	-0,5	-0,5
Afet	Su	0,9	0,00000	-0,5	-0,5
Afet	Meksika	0,9	0,00000	-0,5	-0,5
Apple	Apple_Computer	0,35	0,70000	0,25	-0,1
Apple	Pixar	0,35	0,70000	0,25	-0,1
Apple	IPhone	0,35	0,70000	0,25	-0,1
Apple	Ipod	0,35	0,70000	0,25	-0,1

Apple	CEO	0,35	0,70000	0,25	-0,1
Basketbol	Brezilya	-0,05	-0,05000	0,95	0,65
Basketbol	Futbol	-0,05	-0,05000	0,95	0,65
Basketbol	Beyzbol	-0,05	-0,05000	0,95	0,65
Basketbol	Meksika	-0,05	-0,05000	0,95	0,65
Basketbol	Tenis	-0,05	-0,05000	0,95	0,65
Bilgisayar	Yazılım	0,5	0,55000	0,5	-0,6
Bilgisayar	Klavye	0,5	0,55000	0,5	-0,6
Bilgisayar	İnternet	0,5	0,55000	0,5	-0,6
Bilgisayar	Kimya	0,5	0,55000	0,5	-0,6
Bilgisayar	Televizyon	0,5	0,55000	0,5	-0,6
Bilim	Fizik	0,7	-0,10000	0,6	0,9
Bilim	Kimya	0,7	-0,10000	0,6	0,9
Bilim	Enerji	0,7	-0,10000	0,6	0,9
Bilim	Güneş	0,7	-0,10000	0,6	0,9
Bilim	Canlı	0,7	-0,10000	0,6	0,9
Bilişim_bilimi	Bilgisayar	0,8	0,30000	-0,1	0,1
Bilişim_bilimi	Yazılım	0,8	0,30000	-0,1	0,1
Bilişim_bilimi	Fizik	0,8	0,30000	-0,1	0,1
Bilişim_bilimi	Bilim	0,8	0,30000	-0,1	0,1
Bilişim_bilimi	Kimya	0,8	0,30000	-0,1	0,1
Boks	Futbol	0,5	0,30000	0,85	0,7

Boks	Maraton	0,5	0,30000	0,85	0,7
Boks	Beyzbol	0,5	0,30000	0,85	0,7
Boks	Doktor	0,5	0,30000	0,85	0,7
Boks	Tenis	0,5	0,30000	0,85	0,7
Caz	Virtüöz	0,9	-0,15000	0,95	0,85
Caz	Müzik	0,9	-0,15000	0,95	0,85
Caz	Meksika	0,9	-0,15000	0,95	0,85
Caz	Telefon	0,9	-0,15000	0,95	0,85
Caz	Tren	0,9	-0,15000	0,95	0,85
Döviz	Dolar	0,8	-0,50000	0,65	0,7
Döviz	Para	0,8	-0,50000	0,65	0,7
Döviz	Banka	0,8	-0,50000	0,65	0,7
Döviz	İsrail	0,8	-0,50000	0,65	0,7
Döviz	Fiyat	0,8	-0,50000	0,65	0,7
Faiz	Para	0,65	0,30000	0,65	0,7
Faiz	Banka	0,65	0,30000	0,65	0,7
Faiz	Telefon	0,65	0,30000	0,65	0,7
Faiz	Tren	0,65	0,30000	0,65	0,7
Faiz	Bilgisayar	0,65	0,30000	0,65	0,7
Gezegen	Güneş	0,6	1,00000	0,9	1
Gezegen	Atmosfer	0,6	1,00000	0,9	1
Gezegen	Yıldız	0,6	1,00000	0,9	1

Gezegen	Futbol	0,6	1,00000	0,9	1
Gezegen	Su	0,6	1,00000	0,9	1
Gökbilimci	Takımyıldız	0,6	0,75000	0,75	0,9
Gökbilimci	Atmosfer	0,6	0,75000	0,75	0,9
Gökbilimci	Güneş	0,6	0,75000	0,75	0,9
Gökbilimci	Anne	0,6	0,75000	0,75	0,9
Güneş	Gezegen	-0,3	-0,30000	-0,1	-0,1
Güneş	Atmosfer	-0,3	-0,30000	-0,1	-0,1
Güneş	Yıldız	-0,3	-0,30000	-0,1	-0,1
Güneş	Su	-0,3	-0,30000	-0,1	-0,1
Güneş	Enerji	-0,3	-0,30000	-0,1	-0,1
Hastane	Doktor	0,2	0,45000	0,65	0,8
Hastane	Çocuk	0,2	0,45000	0,65	0,8
Hastane	Millet	0,2	0,45000	0,65	0,8
Hastane	Hemşire	0,2	0,45000	0,65	0,8
Hastane	Sigorta	0,2	0,45000	0,65	0,8
Hayvanlar	Yumurta	-0,4	-0,40000	0,6	0,5
Hayvanlar	Kuşlar	-0,4	-0,40000	0,6	0,5
Hayvanlar	Memeliler	-0,4	-0,40000	0,6	0,5
Hayvanlar	Canlı	-0,4	-0,40000	0,6	0,5
Hayvanlar	Bilim	-0,4	-0,40000	0,6	0,5
Istakoz	Deniz	0,85	0,35000	-0,15	0,85

Istakoz	Hayvanlar	0,85	0,35000	-0,15	0,85
Istakoz	Radyo	0,85	0,35000	-0,15	0,85
Istakoz	Su	0,85	0,35000	-0,15	0,85
Istakoz	Bilgisayar	0,85	0,35000	-0,15	0,85
İçecekler	Alkol	0,4	0,00000	-0,5	-0,5
İçecekler	Şarap	0,4	0,00000	-0,5	-0,5
İçecekler	Kahve	0,4	0,00000	-0,5	-0,5
İçecekler	Su	0,4	0,00000	-0,5	-0,5
İçecekler	Sıvı	0,4	0,00000	-0,5	-0,5
Jaguar	Brezilya	0,5	0,90000	0,3	0,7
Jaguar	Kaplan	0,5	0,90000	0,3	0,7
Jaguar	Meksika	0,5	0,90000	0,3	0,7
Jaguar	Memeliler	0,5	0,90000	0,3	0,7
Jaguar	Başkan	0,5	0,90000	0,3	0,7
Kaplan	Jaguar	0,6	0,60000	0,2	0,6
Kaplan	Orman	0,6	0,60000	0,2	0,6
Kaplan	Meksika	0,6	0,60000	0,2	0,6
Kaplan	Memeliler	0,6	0,60000	0,2	0,6
Kaplan	Başkan	0,6	0,60000	0,2	0,6
Kedi	Memeliler	0,3	0,00000	-0,5	-0,6
Kedi	Kaplan	0,3	0,00000	-0,5	-0,6
Kedi	Jaguar	0,3	0,00000	-0,5	-0,6

Kedi	Hayvanlar	0,3	0,00000	-0,5	-0,6
Kedi	Göz	0,3	0,00000	-0,5	-0,6
Konser	Müzik	0,3	-0,50000	0,95	0,3
Konser	Telefon	0,3	-0,50000	0,95	0,3
Konser	Tren	0,3	-0,50000	0,95	0,3
Konser	Bilgisayar	0,3	-0,50000	0,95	0,3
Konser	Kitap	0,3	-0,50000	0,95	0,3
Memeliler	Hayvanlar	0,9	0,65000	0,85	0,9
Memeliler	Orman	0,9	0,65000	0,85	0,9
Memeliler	Kuşlar	0,9	0,65000	0,85	0,9
Memeliler	Millet	0,9	0,65000	0,85	0,9
Memeliler	Bebek	0,9	0,65000	0,85	0,9
Microsoft	Windows	0,8	0,85000	0,9	0,6
Microsoft	Microsoft_Windows	0,8	0,85000	0,9	0,6
Microsoft	İşletim_sistemi	0,8	0,85000	0,9	0,6
Microsoft	Paul_Allen	0,8	0,85000	0,9	0,6
Microsoft	ABD	0,8	0,85000	0,9	0,6
Proton	Fizik	0,55	0,55000	0,6	0,65
Proton	Telefon	0,55	0,55000	0,6	0,65
Proton	Tren	0,55	0,55000	0,6	0,65
Proton	Bilgisayar	0,55	0,55000	0,6	0,65
Proton	Kitap	0,55	0,55000	0,6	0,65

Sel	Deniz	-0,1	0,25000	-0,15	-0,5
Sel	Su	-0,1	0,25000	-0,15	-0,5
Sel	Afet	-0,1	0,25000	-0,15	-0,5
Sel	Kuraklık	-0,1	0,25000	-0,15	-0,5
Sel	Yaşam	-0,1	0,25000	-0,15	-0,5
Takımyıldız	Güneş	0,2	0,10000	0,5	0,7
Takımyıldız	Yıldız	0,2	0,10000	0,5	0,7
Takımyıldız	Yaşam	0,2	0,10000	0,5	0,7
Takımyıldız	Proton	0,2	0,10000	0,5	0,7
Takımyıldız	Uzay	0,2	0,10000	0,5	0,7
Yazılım	Bilgisayar	1	0,30000	0,75	0,4
Yazılım	Klavye	1	0,30000	0,75	0,4
Yazılım	Ofis	1	0,30000	0,75	0,4
Yazılım	Yaşam	1	0,30000	0,75	0,4
Yazılım	Otomobil	1	0,30000	0,75	0,4
Yıldız	Güneş	0,3	0,70000	0,7	0,2
Yıldız	Gezegen	0,3	0,70000	0,7	0,2
Yıldız	Enerji	0,3	0,70000	0,7	0,2
Yıldız	Atmosfer	0,3	0,70000	0,7	0,2
Yıldız	Uzay	0,3	0,70000	0,7	0,2

Çizelge 6.7 Ortalama korelasyon değerleri

	PF-IBF	TF-IDF	1. Dereceden Bağlantıların Birlikte Bulunması	2. Dereceden Bağlantıların Birlikte Bulunması
Ortalama Korelasyon	0,475899281	0,268	0,428057554	0,382014388

7. SONUÇLAR VE ÖNERİLER

Tez çalışmasında otomatik olarak eş anlamlılar sözlüğü oluşturmak için Türkçe Wikipedia veritabanı kullanılmıştır. Bu veritabanı içerisinde bulunan bağlantılar kullanılarak kelimeler arasındaki benzerlik oranları 4 farklı yöntemle hesaplanılmıştır.

PF-IBF yönteminde maksimum kullanılacak yol uzunluğunun yapılan hesaplamalar ve gözlemler sonucunda 2 olarak alınması uygun görülmüştür. Bunun dışındaki değerlerde hesaplama performansı ve doğruluğunun sapmalar gösterdiği gözlemlenmiştir.

Bağlantıların birlikte bulunması yönteminde iki farklı yöntem uygulanmıştır. Bunlardan birincisi birinci dereceden birlikte bulunma yöntemidir. Bu yöntemde pencere genişliği parametresi nedeni ile bazı ilişkili kelime çiftlerinin arasındaki benzerlik oranları istenildiği kadar yüksek çıkmadığı gözlemlenmiştir. Bunun için ikinci dereceden birlikte bulunma yöntemi kullanılmış ve bu tarz kelime çiftlerinin arasındaki benzerlik oranlarının hesaplanabilmesi sağlanmıştır. Her iki metotta da pencere genişliğinin yapılan gözlemler sonucunda 4 olarak alınması uygun görülmüştür. 4 değerinin altında kalan değerlerde başarı istenilen düzeye ulaşmadığı, 4'ün üzerinde kalan değerler ile de başarı yeterli miktarda artmadığı için pencere genişliği parametresinin en uygun değerinin 4 olacağı görülmüştür.

Yapılan çalışmaların insanların kararları ile ne kadar benzer olduğunu görebilmek için Türkçe Wikipedia'da bulunan 139 kelime çiftinden oluşan bir kelime seti hazırlanmıştır. Bu kelime çiftleri hazırlanırken 28 adet kelime seçilmiştir. Bunların eşleştirileceği kelimeler ise uygulamanın PF-IBF yöntemi ile yapılan hesaplamalar sonucunda en benzer bulunan 5 kelime olarak seçilmiştir. Bu kelime çiftleri 17 kişilik bir test grubuna puanlandırılmıştır. Her kelime çifti için kelimelerin ilişkilerine göre 0 ile 10 arasında puan verilmesi istenmiştir.

PF-IBF yöntemi ile yapılan hesaplamaların sonuçları ile test grubunun verdiği puanlar Spearman's Rank korelasyon hesaplaması ile karşılaştırılmıştır. Yapılan hesaplamalar sonucunda negatif korelasyona sahip toplam 4 adet makale olduğu gözlemlenmiştir. Bu makaleler Basketbol, Güneş, Hayvanlar ve Sel'dir. Bu makalelerin içeriklerinde bulunan bağlantı adetleri diğer makalelere oranla daha az olduğu için hesaplamaların sonuçları yeterince başarılı çıkmamaktadır. Bir diğer sebep ise makalelerin içeriklerinde bulunan bağlantıların diğer bağlantılar ile yeterince ilişkili olmamasıdır.

Korelasyon sonuçlarında 139 adet kelime çiftinin 119 tanesinin korelasyonu pozitif hesaplanmıştır. Bunların 84 tanesinin korelasyonu 0,5 ve üzerindedir. Bütün kelime çiftleri için korelasyon ortalaması ise 0,476'dır.

KAYNAKLAR

- Erdmann, M., Nakayama, K., Hara, T. ve Nishio, S., (2008a), "An Approach for Extracting Bilingual Terminology from Wikipedia", Proc. of International Conference on Database Systems for Advanced Applications (DASFAA), 380-392.
- Erdmann, M., Nakayama, K., Hara, T. ve Nishio, S., (2008b), "A Bilingual Dictionary Extracted from the Wikipedia Link Structure", Proc. of International Conference on Database Systems for Advanced Applications (DASFAA), 686-689.
- Gabrilovich, E. ve Markovitch, S., (2007), "Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis", In Proceedings of the 20th International Joint Conference on Artificial Intelligence, 1606-1611.
- Ito, M., Nakayama, K., Hara, T. ve Nishio, S., (2008), "Association Thesaurus Construction Methods based on Link Co-occurrence Analysis for Wikipedia", Proceeding of the 17th ACM conference on Information and knowledge management.
- Kassner, L., Nastase, V. ve Strube, M., (2008), "Acquiring a taxonomy from the German Wikipedia", In Proceedings of the 6th International Conference on Language Resources and Evaluation, Marrakech, Morocco.
- Liu, J. ve Birnbaum, L., (2007), "Measuring Semantic Similarity between Named Entities by Searching the Web Directory", Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, 461-465.
- Mihalcea, R., (2007), "Using Wikipedia for Automatic Word Sense Disambiguation", In Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL 2007), Rochester.
- Milne, D., Medelyan, O. ve Witten, I. H., (2006), "Mining Domain-Specific Thesauri from Wikipedia: A case study", Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence.
- Nakayama, K., Hara, T. ve Nishio, S., (2007a), "Wikipedia Mining for an Association Web Thesaurus Construction", Proc. of International Conference on Web Information Systems Engineering (WISE), 322-334.
- Nakayama, K., Hara, T. ve Nishio, S., (2007b), "A Thesaurus Construction Method from Large Scale Web Dictionaries", Proc. of International Conference on Advanced Information Networking and Applications (IEEE AINA), 932-939.
- Ponzetto, S. P. ve Strube, M., (2006a), "WikiRelate! Computing semantic relatedness using Wikipedia", In Proceedings of the 21st National Conference on Artificial Intelligence, Boston 1419-1424.
- Ponzetto, S. P. ve Strube, M., (2006b), "Exploiting semantic role labeling, WordNet and Wikipedia for coreference resolution", In Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, New York, N.Y., 4-9 June 2006, 192-199.
- Ponzetto, S. P. ve Strube, M., (2007a), "Knowledge derived from Wikipedia for computing semantic relatedness", Journal of Artificial Intelligence Research, 181-212.

- Ponzetto, S. P. ve Strube, M., (2007b), "Deriving a large scale taxonomy from Wikipedia", In Proceedings of the 22nd Conference on the Advancement of Artificial Intelligence, Vancouver, B.C., Canada, 1440-1445.
- Ponzetto, S. P. ve Strube, M., (2007c), "An API for measuring the relatedness of words in Wikipedia", In Companion Volume to the Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic, 49-52.
- Ponzetto, S. P. ve Strube, M., (2008), "WikiTaxonomy: A Large Scale Knowledge Resource", In Proceedings of the 18th European Conference on Artificial Intelligence, Patras, 751-752.
- Ramírez, J., Asahara, M. ve Matsumoto, Y., (2008), "Japanese-Spanish Thesaurus Construction Using English as a Pivot"
- Zirn, C., Nastase, V. ve Strube, M., (2008), "Distinguishing between instances and classes in the Wikipedia taxonomy", In Proceedings of the 5th European Semantic Web Conference, Tenerife, Spain, 1-5 June 2008, 376-387.
- Bilgin, O., Çetinoğlu, Ö. ve Oflazer, K., (2004), "Building a Wordnet for Turkish", Romanian Journal of Information Science and Technology, Volume 7, Numbers 1-2, 163 – 172.
- McCallum, A. K., (2002), "MALLET: A Machine Learning for Language Toolkit"

EKLER

- EK – 1 Wikipedia veritabanının şeması
- EK – 2 Wikipedia makalesinin yapısı
- Ek – 3 Yöntemlerin sürelerini karşılaştırılması

EK – 1 Wikipedia veritabanının şeması



EK – 2 Wikipedia makalesinin yapısı

```

<page>
<title>Yıldız Teknik Üniversitesi</title>
<id>21678</id>
<revision>
<id>4610147</id>
<timestamp>2008-12-18T16:02:59Z</timestamp>
<contributor>
<ip>88.245.8.248</ip>
</contributor>
<comment>/* Ünlü Mezunları */</comment>
<text xml:space="preserve">{ {Üniversite_Künyesi |adı          =Yıldız Teknik
Üniversitesi |amblemi      =[[Resim:YTÜ_Amblem.jpg|150px|Yıldız Teknik Üniversitesi
Logosu ]] |slogan          =Her Zaman Parlayan Yıldız |eğitim dili    =[[Türkçe]] (%30
[[İngilizce]]) |kuruluş     =[[1911]] |türü          =[[Devlet üniversitesi|Devlet]] |rektör
=Prof. Dr. [[İsmail Yüksek]] |şehir          =[[İstanbul]] |ülke      =[[Türkiye]] |akademik
birimler =9 fakülte&lt;br /&gt;2 enstitü &lt;br /&gt; Meslek Yüksek Okulu &lt;br /&gt;
Yabancı Dil Yüksek Okulu|akademik personel =1384( 178 profesör, 86 Doçent, 222 Yardımcı
Doçent, 135 Öğretim Görevlisi, 107 Okutman, 580 Araştırma Görevlisi, 46 Uzman) |yayın
sayısı    = - |öğrenci sayısı  = 21.473 |yerleşke      = [[Yıldız, İstanbul|Yıldız]] (Merkez),
[[Şişli]], [[Ayazağa]], [[Davutpaşa]]|yerleşke yüzölçümü= Merkez Kampüs:500.000m²
|bağlantılar    = [[IAESTE]]|web_sitesi    =[http://www.yildiz.edu.tr
yildiz.edu.tr]} }"Yıldız Teknik Üniversitesi" ("YTÜ"), [[İstanbul]]'da bulunan yedi devlet
üniversitesinden biridir.==Tarihçe====Kondüktör Mekteb-i Alisi Dönemi (1911-
1922)===Vilayet nafia idarelerinin "fen memuru" (eski adıyla kondüktör, yeni adıyla
[[tekniker]]) gereksinimlerini karşılamak amacıyla 1911'de Kondüktör Mekteb-i Alisi adıyla,
Paris'teki "[[Ecol de Conducteur]]"ün müfredat programı esas alınarak [[Bayındırlık
Bakanlığı]]'na bağlı bir okul kurulmuş ve okula öğrenci kaydına 22 Ağustos 1911'de
başlanmıştır.===Nafia Fen Mektebi Dönemi (1922-1937)===1922'de okulun adı Nafia Fen
Mektebi'ne dönüştürülmüş, öğrenim süresi 1926'da 2.5 yıla ve 1931'de 3 yıla
çıkarılmıştır.===İstanbul Teknik Okulu Dönemi (1937-1969)===[[Resim:Ytu
08.jpg|thumb|200px|left|Yıldız Teknik Üniversitesi]]Türkiye'de imar işlerinin ve teknik
hizmetlerin artması nedeniyle Fen memurları ile yüksek mühendisler arasında oluşan
Türkiye'de imar işlerinin ve teknik hizmetlerin artması nedeniyle oluşan boşluğu doldurmak
amacıyla 19 Aralık 1936 tarihinde yayımlanan ve 1 Haziran 1937 tarihinde yürürlüğe giren
3074 sayılı Yasa ile Nafia Fen Mektebi lağvedilerek yerine Teknik Okul kurulmuştur. 2 yıllık
Fen memuru ve 4 yıllık mühendislik bölümleri olan okula [[Yıldız Sarayı]] müstemilatından,
bugün de kullanılmakta olan binalar tahsis edilmiş ve buraya taşınılmıştır.İlk kuruluşta fen
memuru ve mühendislik dalında öğrenci yetiştiren inşaat ve makina bölümleri varken 1942 ve
1943 ders yılından itibaren mühendislik kısmında elektrik ve mimarlık bölümleri açılmıştır.
Okul, 26 Eylül 1941 tarihinde yayımlanan İstanbul Yüksek Mühendis Okulu ve Teknik
Okulu'nun Maarif Vekaleti'ne devri Hakkında Kanun uyarınca Nafia Vekaleti'nden alınarak
Maarif Vekaletine bağlanmıştır.[[Milli Eğitim Bakanlığı]]'nın 07 Haziran 1949 tarihli
kararıyla Harita ve Kadastro Mühendisliği kurulmuş ve Türkiye'de bu dalda mühendis
yetiştiren ilk kuruluş olarak 1949-1950 ders yılında öğretime başlamıştır. 1951-1952 ders
yılından itibaren Teknikerlik kısmı kapatılmıştır.1959-1960 ders yılında İstanbul Teknik
Okulu içinde bir ihtisas bölümü açılarak bir yıllık öğrenim sonunda yüksek mühendis ve
yüksek mimar ünvanları verilmeye başlanmıştır.===İstanbul Devlet Mühendislik ve Mimarlık
Akademisi Dönemi (1969-1982)===03 Haziran 1969 tarihinde yayımlanan 1184 sayılı Devlet

```

Mühendislik ve Mimarlık Akademileri Yasası ile, özerkliği olan yüksek dereceli bir öğretim ve araştırma kurumu olarak kurulmuştur. 1971'de özel yüksekokullar 1472 sayılı yasa ile kapatılmış, bunlardan mühendislikle ilgili olanları İstanbul Devlet Mühendislik ve Mimarlık Akademisi'ne bağlanmıştır.===Yıldız Üniversitesi Dönemi (1982-1992)===İstanbul Devlet Mühendislik ve Mimarlık Akademisi ile buna bağlanmış olan mühendislik yüksekokulları, Kocaeli Devlet Mühendislik ve Mimarlık Akademisi ve Kocaeli Meslek Yüksekokulu'nun ilgili fakülte ve bölümleri, 20 Temmuz 1982 tarihli 41 Sayılı Kanun Hükmünde Kararname ve bu Kararnamenin değiştirilerek kabulüne dair 30 Mart 1983 tarihli 2809 sayılı Yasa ile Yıldız Üniversitesi adı ile kurulmuştur. Yeni kurulan üniversite Fen-Edebiyat, Mühendislik, Kocaeli'nde bulunan Meslek Yüksekokulu, Fen Bilimleri Enstitüsü, Sosyal Bilimler Enstitüsü ve Rektörlüğe bağlı Yabancı Diller, Atatürk İlkeleri ve İnkılap tarihi, Türk Dili, Beden Eğitimi ve Güzel Sanatlar bölümlerinden oluşmuştur.===Yıldız Teknik Üniversitesi Dönemi (1992-)===[[Resim:Ytu 09.jpg|thumb|200px|Yıldız Teknik Üniversitesi]]3 Temmuz 1992 tarih ve 3837 sayılı Yasa ile üniversitenin adı Yıldız Teknik Üniversitesi olarak değiştirilmiş; Mühendislik Fakültesi; Elektrik-Elektronik, İnşaat, Makina ve Kimya-Metalurji Fakülteleri olarak dört fakülteye ayrılmış, ayrıca İktisadi ve İdari Bilimler Fakültesi kurulmuş; Kocaeli Mühendislik Fakültesi ile Kocaeli Meslek Yüksekokulu ayrılarak Kocaeli Üniversitesi adı altında örgütlenmiştir. Günümüzde üniversite 9 fakültesi, 2 enstitüsü, 2 meslek yüksekokulu, yabancı diller yüksekokulu ve 20.000'i aşan öğrencisi ile eğitim-öğretim ve araştırma faaliyetlerini sürdürmektedir.

<ref>http://www.yildiz.edu.tr/tarihce.php</ref>==Akademik

Birimler====Fakülteler====[[Yıldız Teknik Üniversitesi Elektrik-Elektronik Fakültesi|Elektrik-Elektronik Fakültesi]]====Önceden Yıldız Üniversitesi Mühendislik Fakültesi dahilindeki Elektrik Mühendisliği, Elektronik ve Haberleşme Mühendisliği ve Bilgisayar Bilimleri ve Mühendisliği bölümlerinin 1992 yılında Yıldız Teknik Üniversitesi, Elektrik-Elektronik Fakültesi adı altında toplanmasıyla oluşmuştur. Üniversitenin [[Yıldız Yerleşkesi]]'nde yer alır.*Bilgisayar Mühendisliği Bölümü*Elektrik Mühendisliği Bölümü*Elektronik ve Haberleşme Mühendisliği Bölümü====[[Yıldız Teknik Üniversitesi Fen Edebiyat Fakültesi|Fen-Edebiyat Fakültesi]]====Sekiz bölümden oluşan Fakültenin ilk kuruluşundan bu yana hizmet veren bölümler Fizik, Kimya ve Matematik bölümleridir. Fakülte Davutpaşa Yerleşkesi'ndeki tarihi taş binada faaliyetlerini sürdürmektedir.*Batı Dilleri ve Edebiyatları Bölümü*Biyoloji Bölümü*Fizik Bölümü*İnsan ve Toplum Bilimleri Bölümü*İstatistik Bölümü*Kimya Bölümü*Matematik Bölümü*Türk Dili ve Edebiyatı Bölümü====[[Yıldız Teknik Üniversitesi Makina Fakültesi|Makina Fakültesi]]====Makina Fakültesi, Mühendislik Fakültesinden 1992 yılında ayrılmıştır. Makina Fakültesi,1993 yılında kurulmuş olup, Fakülte bölümlerinden, Makina Mühendisliği 1938-1939, Endüstri Mühendisliği 1993 ve Gemi İnşaatı ve Gemi Makineleri Mühendisliği 1982 yıllarında eğitime başlamıştır. [[Yıldız Yerleşkesi]]'ndedir.*Makine Mühendisliği Bölümü*Mekatronik Mühendisliği Bölümü*Endüstri Mühendisliği Bölümü*Gemi İnşaatı ve Gemi Makineleri Mühendisliği Bölümü*Gemi Makineleri İşletme Mühendisliği Bölümü====[[Yıldız Teknik Üniversitesi Mimarlık Fakültesi|Mimarlık Fakültesi]]====1942 yılında kurulan Mimarlık Fakültesi, lisans düzeyinde Mimarlık ve Şehir ve Bölge Planlama bölümleri, 11 uzmanlık programı (bilgisayar ortamında mimarlık, bina araştırma ve planlama, mimari tasarım, mimarlık tarihi ve kuramı, röle ve restorasyon, yapı, yapı fiziği, kentsel koruma ve planlama, kentsel mekan organizasyonu ve tasarım, kentsel dönüşüm ve planlama, peyzaj planlama) ile 80 öğretim üyesi, 5 öğretim görevlisi, 43 araştırma görevlisi, 1300'ü aşkın lisans, 350'yi aşkın yüksek lisans, 150 doktora öğrencisi ve 10000'den fazla mezunuyla faaliyetlerini sürdürmektedir. Fakülte [[Yıldız Yerleşkesi]]'nde 1880'lerde yapılmış tarihi binasında faaliyetlerini sürdürmektedir.*Mimarlık Bölümü*Şehir ve Bölge Planlama Bölümü====[[Yıldız Teknik Üniversitesi İktisadi ve İdari Bilimler Fakültesi|İktisadi ve İdari Bilimler Fakültesi]]====İktisadi ve İdari Bilimler Fakültesi 1992 yılında [[Çukursaray]]

binasında faaliyetine başlamıştır. Fakülte, 1993 yılında kurulan İktisat Bölümü ile faaliyetlerine başlamıştır. 1996 yılında İşletme Bölümü ve 1999 yılında Siyaset Bilimi ve Uluslararası İlişkiler Bölümü kurulmuştur. Fakülte 2000'den beri, Yıldız Sarayının bir parçası olan [[Yaverler Dairesi]]'nde hizmet vermektedir. Fakülte binası aynı zamanda 1925 senesinden 1936 senesine kadar hizmet veren ve 1936 yılında Atatürk'ün isteği ile Ankara'ya taşınan bugünkü adı ile Ankara Üniversitesi Siyasal Bilgiler Fakültesi'ne (Mekteb-i Mülkiye) ev sahipliği yapmıştır. *İktisat Bölümü*İşletme Bölümü*Siyaset Bilimi ve Uluslararası İlişkiler Bölümü=====[[Yıldız Teknik Üniversitesi İnşaat Fakültesi|İnşaat Fakültesi]]=====1911 yılında kurulmuş olan, üniversite'nin en köklü bölümlerinden İnşaat Mühendisliği Bölümünü'nün de bağlı bulunduğu fakülte, Çevre Mühendisliği ve Jeodezi ve Fotogrametri Mühendisliği eğitimi de vermektedir. Üniversitenin [[Davutpaşa Yerleşkesi]]'nde bulunmaktadır. *İnşaat Mühendisliği Bölümü*Çevre Mühendisliği Bölümü*Jeodezi ve Fotogrametri Mühendisliği Bölümü=====[[Yıldız Teknik Üniversitesi Kimya-Metalurji Fakültesi|Kimya-Metalurji Fakültesi]]=====Kimya-Metalurji Fakültesi, Kimya Mühendisliği, Metalurji ve Malzeme Mühendisliği, Matematik Mühendisliği ve Biyomühendislik bölümlerinden oluşmaktadır. [[Davutpaşa Yerleşkesi]]'ndedir. *Biyomühendislik Bölümü*Kimya Mühendisliği Bölümü*Matematik Mühendisliği Bölümü*Metalurji ve Malzeme Mühendisliği Bölümü=====[[Yıldız Teknik Üniversitesi Eğitim Fakültesi|Eğitim Fakültesi]]=====Fakülte [[Davutpaşa Yerleşkesi]]'nde yer almaktadır. *Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü*Yabancı Diller Eğitimi Bölümü=====[[Yıldız Teknik Üniversitesi Sanat ve Tasarım Fakültesi|Sanat ve Tasarım Fakültesi]]=====1997 yılında kurulan Sanat ve Tasarım Fakültesinde 7 lisans, 5 lisansüstü programı mevcuttur. Geçmiş 1989 yılında kurulan Müzecilik Yüksek Lisans Programı'na kadar gitmektedir. [[Yıldız Yerleşkesi]]'ndedir. *İletişim Tasarımı Bölümü*Müzik ve Sahne Sanatları Bölümü:Müzik Toplulukları Programı:Duysal Sanatlar Tasarımı Programı:Dans Programı*Sanat Bölümü:Bileşik Sanatlar Programı:Sanat Yönetimi Programı:Fotoğraf ve Video Programı=====Enstitüler=====[[Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü|Fen Bilimleri Enstitüsü]]=====Üniversitede yüksek lisans eğitimine 1959 yılında (o zamanki adı İstanbul Teknik Okulu) "İhtisas Bölümü" adıyla başlanmıştır.Enstitünün 21 anabilim dalı ve 99 programı (49 tezli, 4 tezsiz, 4 ikinci öğretimde tezsiz yüksek lisans ve 42 doktora) bulunmaktadır. Programlardan 57'si mühendislik, 22'i mimarlık ve 20'si temel bilimler alanları ile ilgilidir.=====[[Yıldız Teknik Üniversitesi Sosyal Bilimler Enstitüsü|Sosyal Bilimler Enstitüsü]]=====1983 tarihinde kurulan Sosyal Bilimler Enstitüsü'sünde 8 anabilim dalına bağlı 19 programda çalışmalar yapılmaktadır. Enstitü [[Yıldız Yerleşkesi]] tarihi [[Çukursaray]] binasındadır.=====Yüksekokullar=====[[Yıldız Teknik Üniversitesi Meslek Yüksekokulu|Meslek Yüksekokulu]]=====Meslek Yüksekokulu [[Maslak Yerleşkesi]]'nde yer almaktadır. *İktisadi ve İdari Programlar Bölümü:Bankacılık ve Sigortacılık:Emlak ve Emlak Yönetimi:İşçi Sağlığı ve İş Güvenliği:Muhasebe:Turizm ve Otel İşletmeciliği*Teknik Programlar Bölümü:Bilgisayar Teknolojisi ve Programlama:Harita ve Kadastro:İnşaat:Makine:Restorasyon:Seramik:Tasarım ve Basım Yayımcılık=====[[Milli Saraylar ve Tarihi Yapılar Meslek Yüksekokulu]]=====Milli Saraylar ve Tarihi Yapılar Meslek Yüksekokulu'nun kurulmasına ilişkin protokol, Eski TBMM Başkanı Bülent Arınç ve YTÜ Rektörü Prof. Dr. Durul Ören tarafından imzalanmıştır.=====[[Yıldız Teknik Üniversitesi Yabancı Diller Yüksekokulu|Yabancı Diller Yüksekokulu]]=====Yabancı Diller Yüksekokulu, Temel İngilizce ve Modern Diller olmak üzere iki bölümden oluşmaktadır. Temel İngilizce Bölümü'nde yeterlik sınavında başarılı olamayan öğrencilere İngilizce Hazırlık eğitimi verilir. İleri düzeyde İngilizce dersleri ile diğer yabancı dillerin dersleri Modern Diller Bölümü tarafından verilmektedir. Yabancı Diller Yüksekokulu [[Davutpaşa Yerleşkesi]]'nde yer almaktadır. *Temel İngilizce*Modern Diller Bölümü=====Rektörlüğe Bağlı Birimler=====*Atatürk İlkeleri ve İnkılap Tarihi Bölümü*Beden Eğitimi Bölümü*Enformatik Bölümü*Türk Dili Bölümü

<ref><http://www.yildiz.edu.tr/akademik.php></ref>==Öğrenci Kulüpleri<ref><http://www.yildiz.edu.tr/tarihce.php></ref>==Mayıs 2007 itibariyle Yıldız Teknik Üniversitesi'nde aşağıdaki kulüpler aktiftir:[<http://www.kulupler.yildiz.edu.tr/index.php?a=4>]*Atatürkçü Düşünce Topluluğu Kulübü*Atıcılık ve Paintball*Basın ve Yayın Kulübü*Bilimsel Düşünce Kulübü*Bilişim Kulübü*Çevre ve Teknolojileri Kulübü*[[Yıldız Teknik Üniversitesi Dağcılık Kulübü|Dağcılık Kulübü]] ([[YTÜDAK]])*Dans Kulübü*Denizcilik Kulübü*Ekoloji Kulübü*Ekonomi Kulübü*Elektrik ve Elektronik Müh.Enst. Kulübü*Etik Kulübü*Felsefe ve Düşünce Kulübü*Fotoğraf Kulübü (YTÜFÖK)*Güneş Enerji Sistemleri Kulübü [<http://www.gesk.yildiz.edu.tr> (GESK)]*Havacılık Kulübü (YTÜHAVK)*Kalite ve Verimlilik Kulübü (KVK)*Kimya Teknolojileri Kulübü*Kulüpler Birliği*Makina Teknolojileri Kulübü [www.makinateknolojileri.org]*Müzik Kulübü*Plastik Sanatlar Kulübü*Radyo Kulübü*Robotik ve Otomasyon Kulübü*Satranç Kulübü*Sinema Kulübü*Siyaset ve Uluslararası İlişkiler Kulübü*Socrates Kulübü*Sosyal Sorumluluk Kulübü*Spor Kulübü*Tiyatro Kulübü [www.yildizoyunculari.blogspot.com]*Uluslararası Staj Değişimi Kulübü*Yapı ve Gelişim Kulübü*İşletme Kulübü==Ünlü Mezunları==*[[Tarık Akan]]*[[Ali Coşkun]]*[[Erkut Aslanoğlu]]*[[Ayşegül Abadan]]*[[Kenan İmirzalıoğlu]]*[[Hamdi Alkan]]*[[Hasan Doğan]]*[[Ahu Türkpençe]]*[[Oğuzhan Karataş]]*[[Cenk Arkış]]==Kaynak==<div class="references-small"><ref><http://www.yildiz.edu.tr> Yıldız Teknik Üniversitesi İnternet Sayfası*<ref><http://www.jfm.yildiz.edu.tr> Yıldız Teknik Üniversitesi Jeodezi Ve Fotogrametri Mühendisliği Bölümü*<ref><http://www.ce.yildiz.edu.tr> Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği Bölümü*<ref><http://www.inm.yildiz.edu.tr> YTÜ İnşaat Müh. İnternet Sayfası*<ref><http://www.bilisim.yildiz.edu.tr> YTÜ Bilişim Kulübü*<ref><http://www.a305teyim.com> Yıldız Teknik Üniversitesi Makine Fakültesi Öğrencileri Platformu*<ref><http://matematikmuhendisleri.blogspot.com/> Matematik Mühendisleri*<ref><http://www.yildizz.com> Yıldız Teknik Üniversitesi Gayriresmi Öğrenci Platformu*<ref><http://www.ytuforum.com> Yıldız Teknik Üniversitesi Öğrenci Forumu}} {YTÜ birim}} {İstanbul'daki Üniversiteler}} {Türkiye'deki Teknik Üniversiteler}} [[Kategori:Yıldız Teknik Üniversitesi|*]][[Kategori:Beşiktaş, İstanbul]][[Kategori:1911'de kurulan kuruluşlar]][[en:Yıldız Technical University]][[simple:Yıldız Technical University]]</text></revision></page>

Ek – 3 Yöntemlerin sürelerinin karşılaştırılması

Makale 1	Makale 2	PF- IDF	TF- IDF	1. Dereceden Bağlantıların Birlikte Bulunması	2. Dereceden Bağlantıların Birlikte Bulunması
Afet	Sel	1	2	8	3
Afet	Kuraklık	0	1	45	153
Afet	Doğa	1	2	54	626
Afet	Su	0	25	27	1636
Afet	Meksika	1	14	19	1772
Apple	Apple_Computer	1	0	54	147
Apple	Pixar	2	0	16	24
Apple	IPhone	2	0	17	25
Apple	Ipod	1	0	13	4
Apple	CEO	1	1121	21	139
Basketbol	Brezilya	1	24	80	1400
Basketbol	Futbol	0	19	100	346
Basketbol	Beyzbol	0	2	27	59
Basketbol	Meksika	1	13	66	36
Basketbol	Tenis	1	1	73	45
Bilgisayar	Yazılım	1	2	104	213
Bilgisayar	Klavye	1	3	41	30
Bilgisayar	İnternet	1	3	51	260

Bilgisayar	Kimya	1	7	43	79
Bilgisayar	Televizyon	1	5	52	65
Bilim	Fizik	2	2	40	163
Bilim	Kimya	2	3	31	19
Bilim	Enerji	1	2	25	35
Bilim	Güneş	2	2	17	147
Bilim	Canlı	2	1	28	20
Bilişim_bilimi	Bilgisayar	0	3	17	14
Bilişim_bilimi	Yazılım	1	0	1	5
Bilişim_bilimi	Fizik	1	1	3	12
Bilişim_bilimi	Bilim	0	1	2	8
Bilişim_bilimi	Kimya	0	0	2	7
Boks	Futbol	0	5	40	20
Boks	Maraton	0	0	58	9
Boks	Beyzbol	0	1	4	9
Boks	Doktor	0	0	20	50
Boks	Tenis	0	1	15	17
Caz	Virtüöz	0	1	34	32
Caz	Müzik	0	3	53	50
Caz	Meksika	0	4	31	30
Caz	Telefon	0	10659	674	1447
Caz	Tren	0	1	22	16

Döviz	Dolar	0	0	18	6
Döviz	Para	0	0	25	7
Döviz	Banka	0	0	6	4
Döviz	İsrail	0	10	16	49
Döviz	Fiyat	0	0	12	2
Faiz	Para	0	0	14	4
Faiz	Banka	0	0	1	3
Faiz	Telefon	0	10615	80	235
Faiz	Tren	0	0	1	3
Faiz	Bilgisayar	0	1	2	9
Gezegen	Güneş	0	1	54	14
Gezegen	Atmosfer	1	0	8	8
Gezegen	Yıldız	0	1	29	10
Gezegen	Futbol	0	5	18	20
Gezegen	Su	1	1	8	13
Gökbilimci	Takımyıldız	0	1	26	5
Gökbilimci	Atmosfer	0	1	3	11
Gökbilimci	Güneş	1	2	10	14
Gökbilimci	Anne	0	0	2	4
Güneş	Gezegen	1	1	30	14
Güneş	Atmosfer	1	1	27	13
Güneş	Yıldız	1	2	39	15

Güneş	Su	1	2	30	19
Güneş	Enerji	1	1	29	15
Hastane	Doktor	0	0	20	7
Hastane	Çocuk	1	0	3	5
Hastane	Millet	0	0	9	4
Hastane	Hemşire	0	0	14	3
Hastane	Sigorta	0	0	14	3
Hayvanlar	Yumurta	1	1	25	8
Hayvanlar	Kuşlar	1	0	15	6
Hayvanlar	Memeliler	1	0	13	7
Hayvanlar	Canlı	0	0	4	5
Hayvanlar	Bilim	1	1	5	10
Istakoz	Deniz	0	0	16	7
Istakoz	Hayvanlar	0	0	0	3
Istakoz	Radyo	0	0	12	6
Istakoz	Su	0	1	1	10
Istakoz	Bilgisayar	0	1	2	8
İçecekler	Alkol	0	1	30	5
İçecekler	Şarap	0	0	13	11
İçecekler	Kahve	0	1	14	7
İçecekler	Su	1	2	1	19
İçecekler	Sıvı	0	0	16	1102

Jaguar	Brezilya	1	8	35	37
Jaguar	Kaplan	0	0	11	4
Jaguar	Meksika	1	4	14	34
Jaguar	Memeliler	0	1	4	5
Jaguar	Başkan	1	0	12	8
Kaplan	Jaguar	1	0	2	4
Kaplan	Orman	1	1	13	24
Kaplan	Meksika	2	7	9	28
Kaplan	Memeliler	1	1	3	4
Kaplan	Başkan	1	0	1	4
Kedi	Memeliler	0	0	33	15
Kedi	Kaplan	0	1	6	8
Kedi	Jaguar	0	0	4	7
Kedi	Hayvanlar	0	1	5	7
Kedi	Göz	0	1	17	6
Konser	Müzik	0	0	16	7
Konser	Telefon	0	10856	128	216
Konser	Tren	0	0	1	4
Konser	Bilgisayar	0	1	3	9
Konser	Kitap	0	0	17	4
Memeliler	Hayvanlar	1	0	13	6
Memeliler	Orman	1	1	11	10

Memeliler	Kuşlar	0	1	8	7
Memeliler	Millet	0	1	2	4
Memeliler	Bebek	0	0	2	5
Microsoft	Windows	1	1	65	9
Microsoft	Microsoft_Windows	0	2	83	11
Microsoft	İşletim_sistemi	1	0	47	8
Microsoft	Paul_Allen	1	1	13	6
Microsoft	ABD	0	549	620	352
Proton	Fizik	0	1	16	9
Proton	Telefon	0	10775	256	203
Proton	Tren	0	1	1	4
Proton	Bilgisayar	0	1	6	9
Proton	Kitap	0	0	2	5
Sel	Deniz	0	1	2	8
Sel	Su	0	1	3	10
Sel	Afet	0	0	1	2
Sel	Kuraklık	0	0	1	2
Sel	Yaşam	0	0	13	3
Takımyıldız	Güneş	0	1	5	11
Takımyıldız	Yıldız	0	1	3	6
Takımyıldız	Yaşam	0	1	1	2
Takımyıldız	Proton	0	0	2	3

Takımyıldız	Uzay	0	0	11	5
Yazılım	Bilgisayar	0	1	20	12
Yazılım	Klavye	0	0	8	6
Yazılım	Ofis	1	0	6	4
Yazılım	Yaşam	0	0	2	4
Yazılım	Otomobil	0	1	24	7
Yıldız	Güneş	16	1	34	20
Yıldız	Gezegen	2	0	13	12
Yıldız	Enerji	2	1	18	12
Yıldız	Atmosfer	1	1	11	10
Yıldız	Uzay	1	1	9	8

ÖZGEÇMİŞ

Doğum Tarihi 21.02.1984

Doğum Yeri İstanbul

Lise 1999 - 2002

Beşiktaş Atatürk Anadolu Lisesi

Lisans 2002 – 2006

Yıldız Teknik Üniversitesi Kimya Metalurji
Fakültesi Matematik Mühendisliği Anabilim Dalı

Yüksek Lisans 2007 - 2010

Yıldız Teknik Üniversitesi Fen Bilimleri
Enstitüsü Matematik Mühendisliği Anabilim Dalı

Çalıştığı Kurumlar

2006 – devam

Pozitron Bilgisayar ve İnt. Hiz. A.Ş.
Yazılım Mühendisi