

EGE ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

(DOKTORA TEZİ)

**ONTOLOJİ TABANLI BİR ANLAMSA
İLİNTİLİLİK YÖNTEMİ GELİŞTİRİLMESİ VE
YÖNTEMİN METİN BAĞDAŞIKLIĞININ
OTOMATİK OLARAK ÖLÇÜMÜNDE
KULLANILMASI**

Görkem GİRAY

Tez Danışmanı : Yrd. Doç. Dr. Murat Osman ÜNALIR

Bilgisayar Mühendisliği Anabilim Dalı

Bilim Dalı Kodu : 619.01.00

Sunuş Tarihi : 14.03.2011

Bornova-İZMİR

2011

Görkem GİRAY tarafından doktora tezi olarak sunulan “Ontoloji Tabanlı Bir Anlamsal İntililik Yöntemi Geliştirilmesi ve Yöntemin Metin Bağdaşıklığının Otomatik Olarak Ölçümünde Kullanılması” başlıklı bu çalışma E.Ü. Lisansüstü Eğitim ve Öğretim Yönetmeliği ile E.Ü. Fen Bilimleri Enstitüsü Eğitim ve Öğretim Yönergesi’nin ilgili hükümleri uyarınca tarafımızdan değerlendirilerek savunmaya değer bulunmuş ve **14 Mart 2011** tarihinde yapılan tez savunma sınavında aday oybirliği/oyçokluğu ile başarılı bulunmuştur.

Jüri Üyeleri:

İmza

Jüri Başkanı : Yrd. Doç. Dr. Murat Osman ÜNALIR

Raportör Üye: Prof. Dr. Oğuz DİKENELLİ

Üye : Prof. Dr. Yasemin TOPALOĞLU

Üye : Yrd. Doç. Dr. Tuğkan TUĞLULAR

Üye : Yrd. Doç. Dr. Murat KOMESLİ

ÖZET

ONTOLOJİ TABANLI BİR ANLAMSAL İLİNTİLİLİK YÖNTEMİ GELİŞTİRİLMESİ VE YÖNTEMİN METİN BAĞDAŞIKLIĞININ OTOMATİK OLARAK ÖLÇÜMÜNDE KULLANILMASI

GİRAY, Görkem

Doktora Tezi, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Yrd. Doç. Dr. Murat Osman ÜNALIR
Mart 2011, 165 sayfa

Bilim ve teknolojideki gelişmeler bilgisayarları, hayatımızın ayrılmaz ve önemli bir parçası haline getirmiştir. Bilgisayarların her alanda kullanımının yaygınlaşmasıyla birlikte elektronik ortamda üretilen ve saklanan verinin hacmi ciddi boyutlara ulaşmıştır. Elektronik ortamda saklanan verinin yönetimi ve bu veri üzerinde bilgisayarların zekice işlemler yapabilmesi gittikçe önem kazanmaktadır. Anlamsal web vizyonu çerçevesinde, bilgisayarların veri üzerinde zekice işlemler yapabilmesi için bilgisayarlara, verinin anlamı üzerinde yorum yapabileceği verinin biçimsel olarak ifade edilmiş şekilde sağlanması gerekmektedir. Anlamsal web çalışma alanı biçimsel olarak ifade edilmiş veri olarak ontolojilerin kullanımını önermektedir. Bilgisayarlar, ontolojileri kullanarak hem kavramlar arasındaki ilişkileri bulabilme hem de kendilerine sağlanan veriden ayrı olarak ifade edilmemiş veriye de çıkarsama yaparak ulaşma olanağı bulacaktır.

Elektronik ortamda saklanan verinin önemli bir bölümünü de doğal dille yazılmış belgeler oluşturmaktadır. Bu belgelerin, birçok şirketin sahip olduğu elektronik veri hacminin yaklaşık %80-85'ini oluşturduğu tahmin edilmektedir (Uren et al., 2006). Bu belgelere örnek olarak web sayfaları, kelime işleme programlarıyla üretilmiş belgeler verilebilir. Bu belgelerin anlaşılabilir bir biçimde yazılmış olması, bu belgelerin başkaları tarafından kavranmasını kolaylaştıracaktır ve belgelerin kalitesini yükseltecektir. Doğal dille yazılmış bir metnin anlaşılır olması için, metindeki dil öğelerinin açıkladıkları özellik ve durumlar arasında anlam bağıntıları bulunması, yani metnin bağdaşık olması gerekmektedir. Bu

koşulun otomatik olarak bilgisayarlar tarafından değerlendirilebilmesi, elektronik ortamda saklanan veri üzerinde yapılabilecek zekice işlemlere iyi bir örnektir. Bunun yapılabilmesi için doğal dille yazılmış metinlerin anlamsal olarak etiketlenmesi, yani metindeki kelimelerin ontolojilerdeki kavramlarla ilişkilendirilmesi gerekmektedir. Böylece bilgisayarlar, içeriğini bilmediği doğal dille yazılmış bir metnin anlamı hakkında işlenebilir bir veriye sahip olacaktır.

Metin bağdaşıklığının otomatik olarak değerlendirilmesi için çeşitli yöntemler önerilmiştir. Yapılan deneyler sonucunda, bu yöntemlerden bazılarının ölçeklenebilir olmadığı, bazılarının ise güvenilir sonuçlar vermediği belirlenmiştir. Hem ölçeklenebilir hem de güvenilir sonuçlar veren bir yöntem olarak, metindeki ard arda gelen cümleler arasındaki anlamsal ilintililikler ya da benzerlikler ölçülerek metnin bağdaşıklığının değerlendirilmesi önerilmiştir (Foltz et al., 1998; Lapata and Barzilay, 2005). Bu tez kapsamında, metin bağdaşıklığının otomatik olarak değerlendirilmesi için, metindeki cümleler arasındaki benzerlik yerine ilintililiğin ölçülmesi önerilmiştir. Bunun için ontoloji tabanlı bir anlamsal ilintililik ölçüm yöntemi geliştirilmiştir. Bu yöntemin başarısı, insanların ilintililik yargıları ile karşılaştırılarak sınılandıktan sonra metnin bağdaşıklığının otomatik olarak değerlendirilmesinde kullanılmıştır.

Anahtar sözcükler: Anlamsal etiketleme, ontoloji, metin kavrama, bağdaşıklık, OWL, anlamsal ilintililik ölçümü

ABSTRACT**DEVELOPMENT OF AN ONTOLOGY-BASED SEMANTIC
RELATEDNESS MEASUREMENT METHOD AND ITS
APPLICATION TO AUTOMATIC MEASUREMENT OF TEXT
COHERENCE**

GİRAY, Gökem

Ph.D. in Computer Engineering

Supervisor: Yrd. Doç. Dr. Murat Osman ÜNALIR

March 2011, 165 pages

Advances in science and technology have made computers an integral part of our lives. With the widespread use of computers, the volume of electronic data produced and stored has grown exponentially; data management and intelligent data process capabilities are becoming increasingly important. Computers should be provided with formally described data in order to process data intelligently. Semantic web research recommends the use of ontologies for representing the meaning of data formally. By using ontologies, computers can explore relationships between concepts and discover implicitly stated facts by making inferences.

An important part of electronically stored data is made up of documents written in natural language, such as web pages and products of text processing programs. It is estimated that such documents account for 80-85% of the data stored by many companies (Uren et al., 2006). The level of understandability these documents possess will directly affect readers' comprehension levels and document quality. For the sake of the understandability of a text written in natural language, there must be semantic connections among the language elements in the text, i.e. the text must be coherent. Being able to evaluate text coherence automatically is a good example of intelligent data processing, and in order to do this, texts written in natural language must be annotated with ontological concepts. Computers will then be able to process a text written in natural language at the semantic level.

Various methods have been proposed for the automatic evaluation of text coherence. As a result of experiments, some of these methods are not scalable and

some of them do not give reliable results. However, a method, which is both scalable and provides reliable results, evaluated text coherence by measuring relatednesses or similarities between the consecutive sentences (Foltz et al., 1998; Lapata and Barzilay, 2005). In this thesis, measuring relatedness between the consecutive sentences instead of similarity has been proposed in order to assess text coherence. An ontology-based semantic relatedness measurement method has been developed for this purpose. After the success of this method has been evaluated against human judgment, the method has been used for the automatic assessment of text coherence.

Keywords: Semantic annotation, ontology, text comprehension, coherence, OWL, semantic relatedness measurement

TEŞEKKÜR

Ege Üniversitesi Bilgisayar Mühendisliği bölümündeki lisans eğitimim süresince olduğu gibi, doktora eğitimim sırasında da beni yönlendiren, beni anlamsal web çalışma alanıyla tanıştıran değerli hocam Yrd. Doç. Dr. Murat Osman Ünalır'a teşekkürü bir borç bilirim. Tezin biçimlenmesinde değerli katkılarını aldığım, tez izleme komitemdeki değerli hocalarım Prof. Dr. Oğuz Dikenelli'ye ve Yrd. Doç. Dr. Tuğkan Tuğlular'a çok teşekkür ederim. Yapıcı eleştirileri ve yorumları için değerli jüri üyelerim Prof. Dr. Yasemin Topaloğlu'na ve Yrd. Doç. Dr. Murat Komesli'ye teşekkürlerimi sunarım.

Beni daha önce hiç tanımamasına rağmen benimle doktora programı giriş sınavı öncesinde görüşmeyi ve sonrasında danışmanım olmayı kabul ederek Bilgisayar Mühendisliği doktora programına başlamama vesile olan İTÜ Bilgisayar Mühendisliği öğretim üyesi Yrd. Doç. Dr. D. Turgay Altılar'a çok teşekkür ederim.

Önerdiğimiz yöntemin başarısının ölçülmesi için hazırladığımız deney ortamının sağlıklı olması için benimle tecrübelerini paylaşan Pamukkale Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü öğretim üyesi Yrd. Doç. Dr. Sevgi Özgüngör'e çok teşekkür ederim. Metin bağdaşıklığı otomatik ölçüm sonuçlarını değerlendirmek için kullandığım veri kümesini benimle paylaşan Massachusetts Institute of Technology'deki Bilgisayar Bilimleri ve Yapay Zeka Laboratuvar'ından Doç. Dr. Regina Barzilay'a ve University of Edinburgh'daki Bilişim Enstitüsü'nden Mirella Lapata'ya teşekkürlerimi sunarım.

Çalışma süresince gerekli verilerin sağlanmasında bana yardımcı olan Ahu Öztürk'e, Savaş Ceylan'a, Mustafa Güneş'e, Emine Ünalır'a, İbrahim Çınar'a ve dağıttığım anketi dolduran tüm katılımcılara çok teşekkür ederim. Matematiksel gösterimler konusunda bana destek olan Koray Özer'e teşekkür ederim.

İş hayatıyla birlikte yürütülmesi çok kolay olmayan doktora eğitimi sürecini benim için mümkün olduğu kadar kolaylaştıran değerli yöneticim Özkan Dönmez'e çok teşekkür ederim.

Tez yazma sürecinde yakın ilgisini esirgemeyen, bana sürekli manevi destek veren Ünal Er'e ve İbrahim Çınar'a çok teşekkür ederim.

Beni bu günlere getiren, iyi bir eğitim hayatım olması için ellerinden geleni yapan annem Suzan Giray’a ve babam Ahmet Erdem Giray’a, hayatımın ayrılmaz bir parçası olan kardeşim Gökhan Giray’a ve anneannem Sabahat Demirdirek’e teşekkür ederim. Tez sürecinde benden desteğini esirgemeyen ve istatistiksel çalışmalarda bana yardımcı olan eşim Dr. Muazzez Deniz Giray’a çok teşekkür ederim.

Son olarak, bu araştırmanın yapılması için gerekli temeli hazırlayan, bilim okyanusuna bir damla bırakan tüm bilim insanlarına ve araştırmacılara teşekkür ederim.

İÇİNDEKİLER

Sayfa

ÖZET	v
ABSTRACT	vii
TEŞEKKÜR	ix
İÇİNDEKİLER	xi
ŞEKİLLER DİZİNİ	xvii
ÇİZELGELER DİZİNİ	xxi
KISALTMALAR DİZİNİ	xxv
1 GİRİŞ	1
1.1 Tezin Katkıları	2
1.2 Tezin İçeriği	3
2 ONTOLOJİ VE ANLAMSAL ETİKETLEME	5
2.1 Ontoloji	5
2.2 Anlamsal Etiketleme	8
2.3 Özet ve Değerlendirme	10
3 METİN KAVRAMA VE BAĞDAŞIKLIK	12
3.1 Metin Kavrama Modelleri	13
3.1.1 Kintsch'in kavrama modeli	14

İÇİNDEKİLER (devam)

Sayfa

3.1.2 Just ve Carpenter’ın kavrama modeli.....	15
3.1.3 Perfetti’nin kavrama modeli.....	17
3.2 Metnin Zihinsel Temsili.....	18
3.3 Bağdaşıklık.....	21
3.4 Başarılı Metin Kavrama ve Bağdaşıklık	23
3.5 Bağdaşıklığın Ölçülmesi	23
3.6 Özet ve Değerlendirme	26
4 İLİNTİLİLİK VE BENZERLİK.....	30
4.1 İlintililik ve Benzerlik Ölçümü	30
4.1.1 Patika tabanlı ilintililik ve benzerlik ölçüm yöntemleri.....	33
4.1.2 Bilgi içeriği tabanlı ilintililik ve benzerlik ölçüm yöntemleri.....	44
4.1.3 Özellik tabanlı ilintililik ve benzerlik ölçüm yöntemleri.....	45
4.1.4 Geometrik model tabanlı ilintililik ve benzerlik ölçüm yöntemleri.....	49
4.1.5 Karma ilintililik ve benzerlik ölçüm yöntemleri.....	52
4.2 İlintililik ve Benzerlik Ölçümünde Kullanılan Kaynaklar	56
4.3 Özet ve Değerlendirme	59

İÇİNDEKİLER (devam)

Sayfa

5 ONTOLOJİ TABANLI BİR ANLAMSAL İLİNTİLİLİK ÖLÇÜM YÖNTEMİ	66
5.1 Bilişsel Temel	66
5.2 Zihinsel Temsil ve Ontoloji	71
5.3 Ontoloji Tabanlı Anlamsal İlintililik Ölçüm Yönteminin Geliştirilmesi	72
5.3.1 Ontolojik patika	73
5.3.2 Ontolojik patika ile ilintililik ölçümü	75
5.3.3 Anlamsal ilintililik ölçümü için OWL dil öğelerinin değerlendirilmesi	77
5.3.4 OWL dil öğelerinin anlamsal ilintililik katsayılarının belirlenmesi	82
5.4 Ontoloji Tabanlı Anlamsal İlintililik Ölçüm Yönteminin Başarısının Değerlendirilmesi	90
5.5 Uygulama.....	95
5.5.1 Uygulamada kullanılan araçlar	95
5.5.2 Uygulamanın ayrıntıları.....	96
5.6 Özet ve Değerlendirme	99
6 ONTOLOJİ TABANLI ANLAMSAL İLİNTİLİLİK ÖLÇÜM YÖNTEMİ KULLANILARAK METİN BAĞDAŞIKLIĞININ OTOMATİK OLARAK DEĞERLENDİRİLMESİ.....	101

İÇİNDEKİLER (devam)

Sayfa

6.1 Metin Bağdaşıklığının Otomatik Olarak Değerlendirilmesi İçin Bir Yöntem	102
6.1.1 Yarı otomatik eşleştirme	103
6.1.2 Eşleştirme sonuçlarının değerlendirilmesi	107
6.1.3 Metinlerin ve ontolojilerin gözden geçirilmesi	107
6.1.4 Metin bağdaşıklığının otomatik olarak değerlendirilmesi	108
6.2 Uygulama	110
6.2.1 Uygulamada kullanılan araçlar	111
6.2.2 Uygulamanın ayrıntıları	114
6.3 Yöntemin Başarısının Değerlendirilmesi	126
6.4 İlgili Çalışmalar	133
6.5 Özet ve Değerlendirme	135
7 SONUÇ VE TARTIŞMA	137
7.1 Ontoloji Tabanlı Anlamsal İlintililik Ölçüm Yönteminin Kullanılabileceği Alanlar	143
7.2 Metin Bağdaşıklığının Otomatik Olarak Ölçümü ve Gereksinim Mühendisliği	144
7.3 Gelecek Çalışmalar	145

İÇİNDEKİLER (devam)

Sayfa

KAYNAKLAR DİZİNİ.....	149
ÖZGEÇMİŞ.....	165
EKLER	
Ek 1. Türkçe – İngilizce Kavramlar Sözlüğü	
Ek 2. Web Ontology Language (OWL)	
Ek 3. OWL Öğelerinin Anlamsal Ağırlıklarını Belirlemek için Kullanılan Kavram Çiftleri.....	
Ek 4. Anketin Açıklama Yazısı.....	
Ek 5. Katılımcıların İki Kavram Arasındaki İntililik için Verdiği Değerlerin OWL Dil Öğeleri Bazında Dağılımı.....	
Ek 6. Ontoloji Tabanlı Anlamsal İntililik Yönteminin Başarısını Ölçmek için Kullanılan Kavram Çiftleri.....	
Ek 7. Uygulamalarda Kullanılan Araçların ve Kütüphanelerin Sürümleri ve Web Sayfaları	
Ek 8. Bir Ontolojideki İki Kavram Arasındaki Anlamsal İntililik Değerini Hesaplamak için Kullanılan Sınıflar, Değişkenler ve Yordamlar	
Ek 9. Metin Bağdaşıklığını Hesaplamak için Kullanılan Sınıf ve Yordam	
Ek 10. Latin Kare Deseni	
Ek 11. Deneyde Kullanılan Metinler.....	

İÇİNDEKİLER (devam)

Sayfa

Ek 12. Deneyde Elde Edilen Metin Bağdaşıklık Değerleri	
--	--

ŞEKİLLER DİZİNİ

<u>Şekil</u>	<u>Sayfa</u>
Şekil 2.1: Anlamsal etiketleme için bir örnek (Kiryakov et al., 2004).....	9
Şekil 3.1: Waugh ve Norman'ın değiştirilmiş bilişsel modeli (Waugh, N. C. and Norman, D. A., 1965; Solso et al.'den, 2005).	13
Şekil 3.2: Okuduğunu kavramada temel yapıların ve süreçlerin şematik bir gösterimi (Just and Carpenter, 1980, 1987; Solso et al.'den, 2005).....	16
Şekil 3.3: Perfetti'nin metin kavrama modeli (Verhoeven and Perfetti, 2008).	18
Şekil 3.4: Metnin zihinsel temsili için bir örnek a) Metin, b) Metin tabanı, c) Durum modeli (Kintsch, 1994).....	20
Şekil 4.1: Wu ve Palmer'ın örnek kavram hiyerarşisi (Wu and Palmer, 1994). ...	34
Şekil 4.2: Patikaların ağırlıkları.....	39
Şekil 4.3: Gen ontolojisindeki “Hücre içi membran bağımlı organel” kavramının yönlü döngüsüz çizgesi.....	42
Şekil 4.4: İki özellik kümesi arasındaki ilişki (Tversky, 1977).....	45
Şekil 4.5: Örnek bir ontoloji (Tous and Delgado, 2006).	49
Şekil 4.6: Bilgi-tabanlı anlamsal yorumlayıcı (Gabrilovich and Markovitch, 2009).	51
Şekil 4.7: Anlamsal temsil gücü yelpazesi (Daconta et al., 2003).	57
Şekil 4.8: Köpeğin biyolojik taksonomideki yeri.	58
Şekil 5.1: Örnek bir anlamsal ağ (Collins and Quillian, 1969; Solso et al.'den, 2005).	67

ŞEKİLLER DİZİNİ (devam)

<u>Şekil</u>	<u>Sayfa</u>
Şekil 5.2: Anlamsal bilginin işlenmesinde etkinleştirmenin yayılması teorisi (Collins and Loftus, 1975; Solso et al.'den, 2005).	69
Şekil 5.3: “Mavi” kavramından yola çıkılarak “bulut” kavramının etkinleştirmenin yayılmasıyla hatırlanması (Solso et al., 2005).	69
Şekil 5.4: Bellek mimarisinin Nesne-Özellik-İlişki modeli (Wang, 2007).....	70
Şekil 5.5: İnsanın ilintililik yargısı ve bilgisayar ortamında ilintililik ölçümü.	71
Şekil 5.6: OWL ile geliştirilmiş “üniversite” konu alanı ile ilgili bir ontolojinin kesiti.	73
Şekil 5.7: İlintililik ölçümü için anlamsız bir patika örneği.	74
Şekil 5.8: “Dil bilimsel örnek” ve “ontolojik örnek” ilişkilerinin gösterildiği örnek bir bilgi modeli kesiti.	74
Şekil 5.9: K_1 ve K_3 kavramları arasındaki patika.....	75
Şekil 5.10: OWL ile geliştirilmiş “kütüphane” konu alanı ile ilgili bir ontolojinin kesiti.	76
Şekil 5.11: OWL ile geliştirilmiş “ulaşım” konu alanı ile ilgili bir ontolojinin kesiti.	79
Şekil 5.12: OWL ile geliştirilmiş “içecek” konu alanı ile ilgili bir ontolojinin kesiti.	81
Şekil 5.13: Arasında <code>rdfs:subClassOf</code> (özelden genele) dil ögesi ile ilişki kurulmuş iki kavram arasındaki ilintililik için katılımcıların verdiği değerlerin dağılımı.	89

ŞEKİLLER DİZİNİ (devam)

<u>Şekil</u>	<u>Sayfa</u>
Şekil 5.14: Katılımcıların ilintililik değerleri ile otomatik olarak hesaplanan ilintililik değerleri.	94
Şekil 5.15: OWL ile oluşturulmuş bir ontolojideki iki kavram arasındaki anlamsal ilintililiğin hesaplayan uygulamanın ana adımları ve bu uygulamada kullanılan araçlar.	97
Şekil 6.1: Metin bağdaşıklığının otomatik ölçüm yöntemi için çözüm yaklaşımı.	101
Şekil 6.2: Metin bağdaşıklığının otomatik olarak değerlendirilmesi için önerilen yöntemin ana adımları.	103
Şekil 6.3: Doğal dil anlamının aşamaları (Allen, 1995).	104
Şekil 6.4: Morfemin yapısı.	105
Şekil 6.5: POS etiketleme.	105
Şekil 6.6: Bir cümnenin kelime ve kavram vektörü olarak temsili.	106
Şekil 6.7: GATE kullanıcı arayüzü.	112
Şekil 6.8: Ön hazırlık aşaması için geliştirilen uygulamanın ana adımları ve bu uygulamada kullanılan araçlar.	115
Şekil 6.9: Ön hazırlık aşaması için geliştirilen uygulamadaki bileşenlerin GATE kullanıcı arayüzündeki görünümü.	116
Şekil 6.10: Ön hazırlık aşaması için geliştirilen uygulamadaki işlem kaynakları.	117
Şekil 6.11: Metnin ontolojilerle eşleştirilmeden önce bellekte temsil edildiği veri yapısı.	119

ŞEKİLLER DİZİNİ (devam)

<u>Şekil</u>	<u>Sayfa</u>
Şekil 6.12: Anlamsal etiketleme için gerekli bileşenlerin GATE kullanıcı arayüzündeki görünümü.....	120
Şekil 6.13: Anlamsal etiketleme için gerekli işlem kaynakları ve bu kaynakların çalışma sırası.	120
Şekil 6.14: Örnek bir metnin ontolojideki kavramlarla eşleştirilmesi.	122
Şekil 6.15: Anlamsal çözümleme aşaması için geliştirilen uygulamanın ana adımları ve bu uygulamada kullanılan araçlar.	124
Şekil 6.16: Örnek bir metin için Gizli Anlam Çözümleme yöntemi ile yapılan bağdaşıklık ölçümü sonucu.	131
Şekil 6.17: Kavram ilişkilerinin ağırlıkları (Serrano et. al., 2009).	134
Şekil 6.18: Örnek bir metin parçasının Bilişsel Okuma Dizinleme Modeli (Serrano et. al., 2009).....	135
Şekil 7.1: OWL dil öğeleri bazında ontolojik patika boyunca anlamsal ilintililik değerinin değişimi.	147

ÇİZELGELER DİZİNİ

<u>Çizelge</u>	<u>Sayfa</u>
Çizelge 3.1: Bir cümleinin önermelere ayrıştırılmış biçimi (Kintsch, 1998; Rouet'den, 2006).	15
Çizelge 3.2: Bağdaşıklık ilişkileri (Wolf and Gibson, 2004).	23
Çizelge 3.3: Benzerlik ve ilintililik ölçümlerinin bağdaşıklık tahmin başarısı için iki senaryo.....	29
Çizelge 4.1: Düğüm rota çizelgesi.....	40
Çizelge 4.2: “Hücre içi membran bağımlı organel” kavramının yönlü döngüsüz çizgesindeki kavramların anlamsal katkı değerleri.	43
Çizelge 4.3: “Hücre içi organel” kavramının yönlü döngüsüz çizgesindeki kavramların anlamsal katkı değerleri.....	43
Çizelge 4.4: Örnek bir bilgi tabanının içlemsel ve kaplamsal bilgi bileşenleri.	48
Çizelge 4.5: Kavramlar dizinindeki anlamsal ilişkiler (Daconta et al., 2003).....	58
Çizelge 4.6: Patika tabanlı ilintililik ve benzerlik ölçüm yöntemleri.	60
Çizelge 4.7: Bilgi içeriği tabanlı ilintililik ve benzerlik ölçüm yöntemleri.....	62
Çizelge 4.8: Özellik tabanlı ilintililik ve benzerlik ölçüm yöntemleri.	63
Çizelge 4.9: Geometrik tabanlı ilintililik ve benzerlik ölçüm yöntemleri.....	64
Çizelge 4.10: Karma ilintililik ve benzerlik ölçüm yöntemleri.	65
Çizelge 5.1: OWL dil öğeleri.....	77
Çizelge 5.2: Seçilen 201 ontolojideki OWL dil öğelerinin dağılımı.....	84

ÇİZELGELER DİZİNİ (devam)

<u>Çizelge</u>	<u>Sayfa</u>
Çizelge 5.3: Oluşturulan kavram çiftlerinin, aralarında ilişki kuran OWL dil öğelerine göre dağılımı.	85
Çizelge 5.4: 143 kavram çiftinin Türkçe ve İngilizce sürümlerinin değerlendirme sonuçları için yapılan bağımsız gruplar t-testi sonuçları.	87
Çizelge 5.5: OWL dil öğelerinin anlamsal ilintililik katsayıları.	88
Çizelge 5.6: Simetrik olmayan OWL dil öğelerinin anlamsal ilintililik katsayıları.	90
Çizelge 5.7: 24 kavram çiftinin Türkçe ve İngilizce sürümlerinin değerlendirme sonuçları için yapılan bağımsız gruplar t-testi sonuçları.	92
Çizelge 5.8: Benzerlik ölçüm yöntemlerinin ürettiği sonuçların insanların verdiği benzerlik değerleriyle korelasyonları.	94
Çizelge 6.1: Ontolojilerdeki kavramlarla eşleştirilen kelime kategorileri.	118
Çizelge 6.2: Lapata ve Barzilay'ın deneyinde katılımcıların bağdaşıklık değerlendirmeleri ile otomatik bağdaşıklık değerlendirmeleri arasındaki korelasyon değerleri.	128
Çizelge 6.3: Deneyde kullanılan metinlerin anlamsal etiketleme işlemi sonuçlarına ait değerler.	129
Çizelge 6.4: Metin bazında eşleşme oranlarının dağılımı.	129
Çizelge 6.5: Seçilen 201 ontolojideki terim tiplerinin dağılımları.	130
Çizelge 6.6: Bu tez kapsamında ontolojiler kullanılarak yapılan deneyde katılımcıların bağdaşıklık değerlendirmeleri ile otomatik bağdaşıklık değerlendirmeleri arasındaki korelasyon değerleri.	131

ÇİZELGELER DİZİNİ (devam)

Çizelge

Sayfa

Çizelge 6.7: Bu tez kapsamında WordNet kavramlar dizini ve derlemler kullanılarak yapılan deneyde katılımcıların bağdaşıklık değerlendirmeleri ile otomatik bağdaşıklık değerlendirmeleri arasındaki korelasyon değerleri. 132

Çizelge 7.1: Katılımcıların bağdaşıklık değerlendirmeleri ile Gizli Anlam Çözümleme yöntemi ve değişik derlemler kullanılarak hesaplanan otomatik bağdaşıklık değerlendirmeleri arasındaki korelasyon değerleri. 141

Çizelge 7.2: Metinlerin Gizli Anlam Çözümleme yöntemi ile hesaplanan en düşük ve en yüksek bağdaşıklık değerleri. 142

KISALTMALAR DİZİNİ

Kısaltmalar

ARDA	Advanced Research and Development Activity.
DUC	Document Understanding Conferences.
GATE	General Architecture for Text Engineering.
LSA	Latent Semantic Analysis.
NASA	National Aeronautics and Space Administration.
NIST	National Institute of Standards and Technology.
OMG	Object Management Group.
OWL	Web Ontology Language.
OWL DL	Web Ontology Language Description Logics.
OWL Full	Web Ontology Language Full.
OWL Lite	Web Ontology Language Lite.
OWL-S	Web Ontology Language for Services.
POS	Part Of Speech.
RDF	Resource Description Framework.
RDFS	RDF Schema.
SWEET	Semantic Web for Earth and Environmental Terminology.
TAC	Text Analysis Conference.

KISALTMALAR DİZİNİ (devam)**Kısaltmalar**

UML	Unified Modeling Language.
UMLS [®]	Unified Medical Language System [®] .
URI	Uniform Resource Identifier.
W3C	The World Wide Web Consortium.
WWW	World Wide Web.
XML	eXtensible Markup Language.

1 GİRİŞ

Bilgisayarların kullanım alanlarının genişlemesiyle birlikte elektronik ortamda üretilen ve saklanan verinin hacmi ciddi boyutlara ulaşmıştır. Elektronik ortamda saklanan verinin önemli bir bölümünü de doğal dille yazılmış belgeler oluşturmaktadır. Bu belgelerin, birçok şirketin sahip olduğu elektronik veri hacminin yaklaşık %80-85'ini oluşturduğu tahmin edilmektedir (Uren et al., 2006). Elektronik ortamda saklanan belgelerin yönetimi ve bu belgeler üzerinde bilgisayarların zekice işlemler yapabilmesi gittikçe önem kazanmaktadır.

Anlamsal web çalışma alanındaki gelişmeler, bilgisayar bilimleri için yeni olmayan ontolojilerin, farklı alanlarda kullanımını tetiklemiştir. Özellikle, bir konu alanındaki bilgiyi biçimsel olarak temsil etmesi nedeniyle, ontolojilerin anlamsal etiketleme için kullanımı önerilmektedir. Bu tez kapsamında da, doğal dil ile yazılmış metinlerin bağdaşıklıklarının, ontolojilerin yardımıyla otomatik olarak değerlendirilmesi, yani metinler üzerinde bilgisayarların zekice bir işlem yapması hedeflenmiştir.

Bir metnin anlaşılabilir olması için, okuyucu o metni okuduğunda zihninde bağdaşık bir temsilin oluşması gerekmektedir. Bağdaşık bir zihinsel temsilin oluşabilmesi için, metindeki dil öğelerinin açıkladıkları özellik ve durumlar arasında anlam bağıntıları bulunması, yani metnin bağdaşık olması gerekmektedir.

Metin bağdaşıklığının otomatik olarak değerlendirilmesi için çeşitli yöntemler önerilmiştir. Yapılan deneyler sonucunda, bu yöntemlerden bazılarının ölçeklenebilir olmadığı, bazılarının ise güvenilir sonuçlar vermediği belirlenmiştir. Hem ölçeklenebilir hem de güvenilir sonuçlar veren bir yöntem olarak, metindeki ard arda gelen cümleler arasındaki anlamsal ilintililikler ya da benzerlikler ölçülerek metnin bağdaşıklığının değerlendirilmesi önerilmiştir (Foltz et al., 1998; Lapata and Barzilay, 2005).

Anlamsal benzerlik, anlamsal ilintililiğin özel bir durumudur. Anlamsal olarak birbirine benzer kavramlar, birbirleriyle ortak yönleri ile ilintilidir. Fakat iki kavram arasında farklı ilişkilerden kaynaklanan bağlantılar da bulunabilir. Aralarında parça-bütün ya da zıtlık ilişkisi bulunan iki kavram, anlamsal olarak benzer olmamasına rağmen anlamsal olarak ilintili olabilir. Bir okuyucu bir metni okurken, kavramlar arasında sadece benzerlik ilişkisini kurduğunu söylemek yanlış olacaktır. Okuyucunun kavramlar arasındaki çok çeşitli ilişkileri kurarak bir

zihinsel temsil oluřturması, mevcut okuduđunu kavrama modellerine daha uygun dūřmektedir. Bundan dolayı bu alıřmada, Foltz et al.'nin de nerdiđi gibi (Foltz et al., 1998), bir metnin bađdařıklıđını deđerlendirirken ard arda gelen cūmler arasındaki anlamsal benzerliđin lۆlmesi yerine anlamsal ilintililiđin lۆlmesi tercih edilmiřtir. Foltz et al. anlamsal ilintililiđi lۆmek iin istatistik tabanlı olan Gizli Anlam zۆmleme yntemini (Landauer and Dumais, 1997) kullanmaktadır (Foltz et al., 1998). Bu alıřmada tercih edilen yntem ise anlamsal ilintililik lۆm iin ontolojileri kullanmaktadır.

Bu tez kapsamında, bir metinde ard arda gelen cūmler arasındaki anlamsal ilintililiđi lۆmek iin ontoloji tabanlı bir anlamsal ilintililik lۆm yntemi geliřtirilmiřtir. Bu yntemin bařarısı, insanların ilintililik yargıları ile karřılařtırılarak sınanmıřtır. Yntemin anlamsal ilintililik deđerleri ile insanların verdiđi anlamsal ilintililik deđerleri arasındaki korelasyonun¹ ($r = 0,685$; $p < 0,01$) anlamlı ve pozitif ynde olduđu grlmřtir.

Geliřtirilen anlamsal ilintililik lۆm yntemi, dođal dil ile yazılmıř metinlerin bađdařıklıđının otomatik olarak deđerlendirilmesinde kullanılmıřtır. Bunun iin, metinlerdeki kavramlar, ontolojilerdeki kavramlarla eřleřtirilmiřtir. Daha sonra, ontolojiler yardımıyla ard arda gelen cūmler arasındaki anlamsal ilintililikler deđerlendirilmiřtir. Cūmler arasındaki ilintililiđi lۆmek iin ise, o cūmleyi oluřturan kavramlar arasındaki anlamsal ilintililikler hesaplanmıřtır. Bu hesaplama iin tez kapsamında geliřtirilen ontoloji tabanlı anlamsal ilintililik yntemi kullanılmıřtır.

1.1 Tezin Katkıları

Tezin katkıları bařlıca  madde altında zetlenebilir:

1. OWL dil gelerinin anlamları dikkate alınarak, OWL ile oluřturulmuř bir ontolojideki iki kavram arasındaki anlamsal ilintililiđi lۆen bir yntem geliřtirilmiřtir. Bu yntem, biliřsel psikolojideki teoriler zerine

¹ Korelasyon katsayısı iki deđerřken arasındaki dođrusal iliřkinin lۆlerinden biridir. İki deđerřkenin birlikte deđerřim lۆs olan korelasyon katsayısı -1 ile +1 arasında deđer almaktadır. Pozitif korelasyon katsayısı deđerřkenlerden birinin deđerı artarken diđerinin deđerinin arttıđını, negatif korelasyon katsayısı ise deđerřkenlerden birinin deđerı artarken diđerinin deđerinin azaldıđını gstermektedir. Korelasyon katsayısının 0 olması durumunda ise iki deđerřken arasında dođrusal bir iliřki olmadıđı anlařılmaktadır (İkiz vd., 1996).

temellendirilmiştir. Yöntemde, ontolojik patika kavramı geliştirilmiştir. OWL dil öğelerinin, anlamsal ilintililik ölçümünde kullanılmak üzere, iki kavram arasındaki ilintililiğe katkıları sayısal olarak ifade edilmiştir. Bu sayısal değerler, insanların ilintililik yargıları ölçülerek ve sonuçları istatistiksel olarak değerlendirilerek hesaplanmıştır.

2. Anlamsal ilintililik yönteminin başarısının değerlendirilmesi için 24 kavram çiftinden oluşan Türkçe bir veri kümesi hazırlanmıştır. Bu kavram çiftlerinin arasındaki ilintililik, 66 katılımcı tarafından puanlanmıştır. Bu veri kümesi, başka ilintililik yöntemlerinin başarısının değerlendirilmesi için de kullanılabilir. Aynı veri kümesinin İngilizce sürümü de bulunmaktadır.
3. Bir metnin bağdaşıklığının otomatik olarak değerlendirilmesi için, metindeki cümleler arasındaki ilintililiğin, bu tez kapsamında geliştirilen ontoloji tabanlı anlamsal ilintililik yöntemi ile ölçülmesi önerilmiştir. Geliştirilen ontoloji tabanlı anlamsal ilintililik ölçüm yönteminin, metin bağdaşıklığının otomatik olarak değerlendirilmesindeki başarısı deney ile ölçülmüştür.

1.2 Tezin İçeriği

2. bölümde, anlamsal etiketleme ve ontoloji kavramı tanıtılmıştır. Anlamsal etiketleme ile elektronik ortamda saklanan çeşitli kaynakların nasıl zenginleştirilebileceği ve bu sayede bu kaynaklar üzerinde nasıl işlemler yapılabileceğine dair örnekler verilmiştir.

3. bölümde, metin kavrama süreci için önerilen çeşitli modeller, kavrama sonucu zihinde oluşan temsiller, kavramanın başarısını etkileyen bağdaşıklık kavramı ve bağdaşıklığın ölçüm yöntemleri anlatılmıştır.

4. bölümde, ilintililik ve benzerlik kavramları açıklanmıştır. Literatürdeki ilintililik ve benzerlik ölçüm yöntemleri tanıtılmıştır.

5. bölümde, tez kapsamında geliştirilen ontoloji tabanlı anlamsal ilintililik ölçüm yöntemi ve bu yöntemin geliştirilme süreci anlatılmıştır.

6. bölümde, ontoloji tabanlı anlamsal ilintililik ölçüm yöntemi kullanılarak, doğal dil ile yazılmış metinlerin anlaşılabilirliğinin, yani bağdaşıklığının, nasıl değerlendirileceği anlatılmıştır.

7. bölüm, tez çalışması kapsamında ulaşılan önemli sonuçları ve bu çalışmanın geliştirilebilecek yönlerini özetlemektedir.

Ek 2’de, OWL dilinin öğeleri ve kullanım alanları anlatılmıştır.

2 ONTOLOJİ VE ANLAMSAL ETİKETLEME

Bilgisayarların hayatımızın ayrılmaz ve önemli bir parçası haline gelmesiyle birlikte, elektronik ortamda saklanan bilginin hacmi de ciddi oranda artmıştır. Bu hacim artışına karşılık sadece bilginin depolanacağı araçları geliştirmek yeterli değildir. Elektronik ortamdaki bilginin saklandıktan sonra bu bilgiye tekrar erişimin sağlıklı ve kolay biçimde sağlanması ya da elektronik bilginin bilgisayarlar tarafından işlenerek insanlara bazı sonuçlar verilmesi gereksinimlerinin de karşılanması gerekmektedir. Dolayısıyla elektronik ortamda saklanan bilginin yönetilmesi çok önemli bir gereksinim olarak karşımıza çıkmaktadır.

Anlamsal web çalışma alanındaki gelişmeler, bilgisayar bilimleri için yeni olmayan ontolojilerin, farklı alanlarda kullanımını tetiklemiştir. Ontolojiler, doğal dil ile yazılmış metinlerin anlamsal etiketlemesi için de kullanılabilir. Özellikle, bir konu alanındaki bilgiyi biçimsel olarak temsil etmesi ve yeniden kullanımı desteklemesi nedeniyle, ontolojilerin anlamsal etiketleme işlemlerinde kullanımı önerilmektedir. Doğal dil ile yazılmış metinlerin ontolojilerdeki kavramlarla ilişkilendirilmesiyle birlikte, bilgisayarlar bu metinlerdeki kavramlar arasındaki ilişkiler hakkında işlenebilir bir bilgiye sahip olabilirler ve böylece bu metinler üzerinde daha zekice işlemler yaparak insanlara farklı ya da daha kaliteli hizmetler sunabilirler.

Bu bölümde, ontoloji kavramı ve anlamsal etiketleme tanıtılacaktır. Ayrıca, anlamsal etiketleme kullanılarak yapılabilecek işlemlere çeşitli örnekler verilecektir.

2.1 Ontoloji

Anlamsal web çalışma alanının gerek akademik gerekse iş dünyasında ilgi çekmesiyle birlikte, uzun zamandır yapay zeka çalışma alanında kullanılan ontolojiler, daha geniş bir kitlenin dikkatini çekmiştir. Ontolojiler, verinin anlamının bilgisayarların işleyebileceği biçimde temsil edilmesini sağlayarak, verinin daha zekice işlenebilmesine olanak vermektedir. Buradaki “zekice işleme” kavramının insan davranışlarını taklit etme anlamına gelmediği anlamsal web’in fikir babası olan Tim Berners-Lee tarafından açıkça belirtilmiştir (Berners-Lee, 1998). Dolayısıyla ontolojilerin kullanımı beraberinde, bilgisayara daha fazla işlenebilir verinin uygun biçimde sunulmasını gerektirmektedir. Buna karşılık,

bilgisayarın uygun biçimde tanımlanmış veri üzerinde yapacağı çıkarsama ile saklı (yani açık olarak ifade edilmeyen) bilgiye ulaşılması mümkün olacaktır.

Bilgisayar dünyasında en yaygın kullanılan ontoloji tanımı, “paylaşılan bir kavramsallaştırmanın biçimsel ve net bir belirtimi”dir (Gruber, 1993). Uschold ve Gruninger bu tanımda geçen her kelimeyi açıklayarak bu kısa tanımın ifade ettiği uzun ontoloji tanımını açıklamışlardır (Uschold and Gruninger, 2004). “Kavramsallaştırma”, insanların dünyadaki varlıklar üzerine nasıl düşündüklerinin soyut bir modelini ifade eder. Bu soyut model genellikle özel bir konu alanı ile sınırlandırılmıştır. “Net bir belirtim” ise soyut modeldeki kavramlara ve ilişkilere net isimler verildiği ve net tanımlar yapıldığı anlamına gelmektedir. Bir kavramın ya da ilişkinin tanımı, o terimin anlamının ifade edilmesidir. Bir başka deyişle, bir terimin diğer terimlerle ilişkisinin nasıl olacağını belirtir. “Biçimsel” ifadesi, anlam tanımının biçimsel bir dille temsil edildiğini ve böylece tanım üzerindeki belirsizliklerin, farklı anlam çıkarma olasılıklarının ortadan kaldırıldığını ifade etmektedir. Bundan dolayı biçimsel temsil, otomatik çıkarsama yapma olanağını sağlamaktadır. “Paylaşılan” kelimesi ise ontolojilerin, farklı uygulamalar ve topluluklar arasında yeniden kullanımı amaçladıklarını ve desteklediklerini ifade etmektedir. Sowa’ya göre ontolojinin konusu, bir alanda var olan ya da var olabilecek şeylerin kategorilerinin incelenmesidir. Böyle bir çalışmanın sonucu, L dilini kullanan bir kişinin görüş açısından, D alanı hakkında konuşmak amacıyla, D alanında bulunacağı varsayılan şeylerin tiplerinin bir katalogudur, yani bir ontolojidir (Sowa, 1999).

Ontolojileri, anlamsal web için vazgeçilmez kılan ve diğer –kavramsal– modellerden ayıran, geniş kesimler tarafından kabul edilmiş özellikleri vardır. Bu özellikler aşağıdaki listedeki başlıklar altında incelenebilir:

- **Anlamsal temsil gücü:** Anlamsal web’de kullanılabileceği düşünülen ontolojilerin temsil edildiği dillerin anlamsal temsil gücü, sistem geliştirmede kullanılan kavramsal modelleme dillerine göre daha yüksektir.
- **Hedeflenen kullanıcı kitlesinin büyüklüğü:** Ontolojiler üzerinde genelde daha büyük kullanıcı grupları uzlaşmaya varır. Sistem geliştirmede kullanılan kavramsal modeller ise göreceli olarak daha küçük kullanıcı grupları tarafından geliştirilir ve kullanılır.

- **Paylaşılabilir:** Ontolojilerin anlamsal web çalışma alanındaki gelişmelerle birlikte ilgi çekmesinin nedenlerinden birisi paylaşılabilir olmasıdır. Anlamsal web, geniş kesimlerin üzerinde uzlaşmaya vardığı ontolojilerin, web'deki bilginin temsilinde kullanılması üzerine temellendirilmiştir.
- **Yeniden kullanılabilirlik ve genişleme:** Ontolojiler yeniden kullanımı destekler ve ontolojilerin genişletilmesi mümkündür.
- **Açık dünya yaklaşımı:** Açık dünya yaklaşımında bir önermenin doğruluk değeri doğru, yanlış ya da bilinmiyor olabilir. Kapalı dünya yaklaşımında ise doğruluğu bilinmeyen ya da çıkarsanamayan bir önerme yanlıştır. Anlamsal web vizyonunu gerçekleştirmek için geliştirilen ontoloji dilleri açık dünya yaklaşımını benimsemektedir. Bu dillerin, web gibi yoğun ve sürekli değişen veri barındıran, dağıtık ve birbiriyle çelişen verinin bulunabileceği bir ortamda kullanılmak için tasarlandığı düşünüldüğünde, açık dünya yaklaşımının bu koşullara çok daha uygun olduğu anlaşılabacaktır. Diğer taraftan yazılım mühendisliğinde kullanılan UML gibi modelleme dilleri ise kapalı dünya yaklaşımını benimsemektedir. Bir dilin, açık ya da kapalı dünya yaklaşımını benimsemesi, üretilen bir model üzerinde yapılacak bir çıkarsamada farklılıklar oluşturacaktır (Baclawski, 2002). Örneğin, bir modelde her kişinin bir ismi olmasının zorunlu olduğu bir ortamda, UML ile oluşturulmuş bir modelde, bir kişinin isminin olmaması, her kişinin bir ismi olması gerektiği kuralını bozduğu için modelde bir hata oluşacaktır. Fakat aynı durum açık dünya yaklaşımını uygulayan bir dille temsil edilen bir modelde oluştuğunda bir kişinin isminin olmaması durumu, o kişinin isminin bilinmediği şeklinde yorumlanacaktır.
- **Betimleyicilik:** Ontolojiler, doğaları gereği genellikle betimleyici² modeller sınıfına girer. Sistem geliştirmek için kullanılan modeller kuralcıdır. Sistem, bu modeller doğrultusunda geliştirilir. Ontolojiler ise

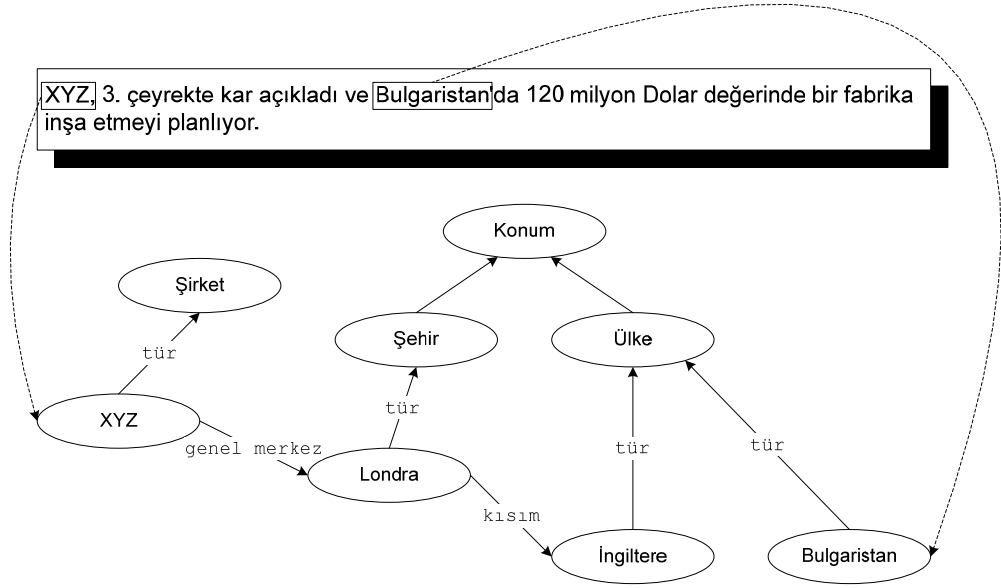
² Modeller, değişik özelliklerine göre sınıflandırılabilir. Bir model, var olan özgün bir şeyi temsil edebilir (örneğin bir fotoğraf) ya da yapılacak bir şeyi betimleyebilir (örneğin bir bina planı). Birinci durumdaki model “betimleyici model”, ikinci durumda ise “kuralcı model” olarak adlandırılabilir (Ludewig, 2003). Bir modelin betimleyici ya da kuralcı olmasını belirleyecek olan, model ile özgün olan arasındaki ilişkidir. Örneğin, bir bina planı var olan bir bina için çizilmişse, bu plan betimleyici, aynı plan inşa edilecek bir bina için çizilmişse kuralcı olur.

daha çok var olanı tanımlar. Açık dünya yaklaşımı da ontolojilerin betimleyici doğasını destekler niteliktedir.

2.2 Anlamsal Etiketleme

Anlamsal etiketleme, elektronik ortamda saklanan bir kaynağın biçimsel olarak ifade edilmiş bir anlamsal model ile ilişkilendirilmesi işlemi olarak tanımlanabilir. Anlamsal etiketleme, elektronik ortamda saklanan, doğal dil ile yazılmış metinler, veritabanlarındaki metin alanları, video ve ses dosyaları gibi her türlü kaynağa uygulanabilir. Kaynakların anlamsal olarak etiketlenmesiyle birlikte bu kaynaklar üzerinde çeşitli işlemler yapılabilir. Bu işlemlere örnek olarak anlamsal arama, sınıflandırma, daha ileri seviyede üst veri oluşturma gibi işlemler verilebilir.

Doğal dil ile yazılmış bir metnin anlamı bilgisayar tarafından anlaşılamaz. Fakat bu metindeki kelimeler ya da kelime öbekleri bir ya da birden çok ontolojideki kavramlar ile eşleştirilirse, bu durumda bilgisayar bu metnin anlamına dair bazı işlemleri bir insana ihtiyaç duymadan yapabilir (Malik et al., 2010). Anlamsal etiketler, metinlerdeki kavramları ve bu kavramlar arasındaki ilişkileri biçimsel olarak tanımlar (Uren et al., 2006). Bir ontoloji kullanılarak anlamsal olarak etiketlenmiş bir metin Şekil 2.1’de gösterilmiştir. Bu örneğe göre, metinde geçen “XYZ” kelimesinin ontolojideki bir kavramla ilişkilendirilmesi ile birlikte, “XYZ”nin bir “Şirket” olduğu ve genel merkezinin “Londra” “Şehri”nde bulunduğu sonucuna bir bilgisayar çıkarsama yaparak varabilir. Benzer şekilde, “Bulgaristan” kelimesinin anlamsal olarak etiketlenmesiyle birlikte, bu kavramın bir “Ülke” olduğu ve bir “Konum” belirttiği sonucuna yine bir bilgisayar ulaşabilir.



Şekil 2.1: Anlamsal etiketleme için bir örnek (Kiryakov et al., 2004).

Doğal dil ile yazılmış metinlerin ontolojilerle etiketlenmesiyle birlikte bu metinler üzerinde anlamsal arama yapma olanağı doğmaktadır. Etiketleme olmadan bu metinler üzerinde sadece sözdizimsel seviyede arama yapılabilirken, metinlerin anlamsal olarak etiketlenmesiyle birlikte bu metinler üzerinde anlamsal seviyede de arama yapılabilir. Örneğin, metinler içinde “İngiltere” kelimesi arandığında sözdizimsel seviyede arama yapabilen bir yazılım, sonuç olarak sadece bu kelimeyi içeren metinleri getirecektir. Fakat metinlerin anlamsal olarak etiketlenmiş olması durumunda, ontolojideki “Londra şehri İngiltere’nin bir kısmıdır” (Bkz. Şekil 2.1) ilişkisinden yola çıkılarak sonuç metinleri olarak “Londra” kelimesini içeren metinler de kullanıcıya gösterilir. Ayrıca birden fazla anlama sahip olabilecek kelimelerle arama yapıldığında, sonuçların içinde sadece kullanıcı tarafından istenen anlamda kullanılan kelimelerin bulunduğu metinlerle sınırlandırılabilmesi için anlamsal etiketler kullanılabilir. Metindeki kelimelerin ontolojilerdeki kavramlarla ilişkilendirilmesiyle birlikte kelime anlamı belirginleştirme işlemi yapılarak kullanıcıya aranan kelimenin hangi anlamıyla arandığı sorularak elde edilecek sonuçların daha doğru olması sağlanabilmektedir. Böylece anlamsal seviyede yapılan aramalar hem kesin isabet hem de erişim isabeti oranlarını arttırmaktadır (Lacavik et al., 2007).

Bu çalışmada, bir metnin anlamsal olarak etiketlenmesi ile bu metnin bağdaşıklığının daha sağlıklı ölçülebileceği savunulmaktadır. Bir metnin bağdaşıklığının otomatik olarak ölçülmesi için en çok kullanılan yöntem, ardışık cümlelerde kullanılan kelimelerin benzerliğinin ya da ilintililiğin hesaplanması

üzerine temellendirilmiştir. Buna göre anlamsal olarak etiketlenmiş metinlerde, ardışık cümlelerde kullanılan kelimeler arasındaki ontolojik ilişkilerin bulunması ve bu ilişkilerin bir anlamsal ilintililik değerine dönüştürülmesi, metin bağdaşıklığını otomatik olarak ölçmek için bir yöntem olarak bu çalışma kapsamında önerilmektedir.

Doğal dil ile yazılmış metinler üzerinde anlamsal seviyede çeşitli işlemler yapılabilmesi için öncelikle anlamsal etiketleme işleminin yapılması gerekmektedir. Bundan dolayı anlamsal etiketleme işleminde kullanılmak üzere çeşitli araçlar geliştirilmiştir. El ile anlamsal etiketleme için kullanılan araçlara örnek olarak Annotea (Kahan and Koivunen, 2001), OntoMat-Annotizer (Hands Schuh et al., 2001); yarı otomatik ve otomatik araçlara örnek olarak ise KIM (Popov et al., 2003), SemTag (Dill et al., 2003) verilebilir. Bu çalışma kapsamında geliştirilen örnek uygulamada da, KIM platformunda olduğu gibi, doğal dil işleme işlevleri GATE aracının uygulama programlama arayüzü kullanılarak geliştirilmiştir. Akademik dünyada üretilen etiketleme araçlarına ek olarak Microsoft, metinlerdeki kelimelerin ya da kelime öbeklerinin ontolojilerdeki kavramlarla etiketlenmesini sağlayan, Word 2007 yazılımı için bir eklenti duyurmuştur. Bu eklentinin temel amacı, ontolojilerin kolayca geliştirilmesini ve geçerlenmesini mümkün kılarak ontolojilerin daha geniş kitleler tarafından kullanılabilmesini ve bir metin yazarken yazarın metnin içeriğiyle ilgili anlamsal bilgiyi metne ekleyebilmesini sağlamaktır (Word Add-in For Ontology Recognition, 2011).

2.3 Özet ve Değerlendirme

Anlamsal etiketleme, elektronik ortamda saklanan bir kaynağın (doğal dil ile yazılmış bir metin, bir video ya da ses dosyası, vs.) anlamsal bir model ile ilişkilendirilmesiyle, bu kaynağın zenginleştirilmesi olarak tanımlanabilir. Doğal dil ile yazılmış bir metnin, biçimsel bir dille ifade edilmiş bir anlamsal model ile ilişkilendirilmesiyle birlikte, bilgisayar metnin anlamı hakkında işlenebilir bir veriye sahip olur. Bilgisayar, metnin ilişkilendirildiği anlamsal model üzerinde bazı işlemler ve çıkarsamalar yaparak o metne yönelik çeşitli işlemler yapabilir. Bu işlemlere örnek olarak, bu çalışma kapsamında önerilen, ontoloji tabanlı bir anlamsal ilintililik ölçüm yöntemi kullanarak metin bağdaşıklığının otomatik olarak ölçülmesi verilebilir.

Ontolojiler, anlamsal web çalışma alanında genel kabul görmüş şekliyle, bir alandaki kavramları, bu kavramlar arasındaki ilişkileri modeller. Genellikle ontolojiler geniş kesimler tarafından kabul görürler; dolayısıyla olgunlaşmış modeller olarak kabul edilebilirler. Bunun yanında yeniden kullanıma ve genişlemeye açıktırlar. Anlamsal web uygulamalarında kullanılması hedeflenen ontolojiler biçimsel olarak ifade edilir. Bu özellikleri göz önüne alındığında, ontolojilerin anlamsal etiketleme sürecinin bir girdisi olarak kullanılması fikri araştırmaya değer bulunmaktadır (Bontcheva et al., 2006). Olgunlaşmış ve biçimsel olarak ifade edilmiş bir ontoloji, elektronik ortamda saklanan kaynaklarla ilişkilendirildiğinde, bu kaynaklar üzerinde bilgisayarların anlamsal seviyede çeşitli işlemler yapabilmesini mümkün kılan bir araç olarak düşünülebilir.

Bu çalışma kapsamında, doğal dil ile yazılmış metinlerin bağdaşıklığını otomatik olarak değerlendirmek için ontolojilerden faydalanılmıştır. Bir metnin bağdaşıklığının nasıl değerlendirilebileceğini ortaya koymak için, öncelikle insanın okuduğunu nasıl kavradığını açıklayan bir modelin sunulması gerekmektedir. Bir sonraki bölümde, insanın okuduğunu kavrama süreci irdelenerek bağdaşık bir metnin hangi özelliklere sahip olması gerektiği üzerinde durulmuştur.

3 METİN KAVRAMA VE BAĞDAŞIKLIK

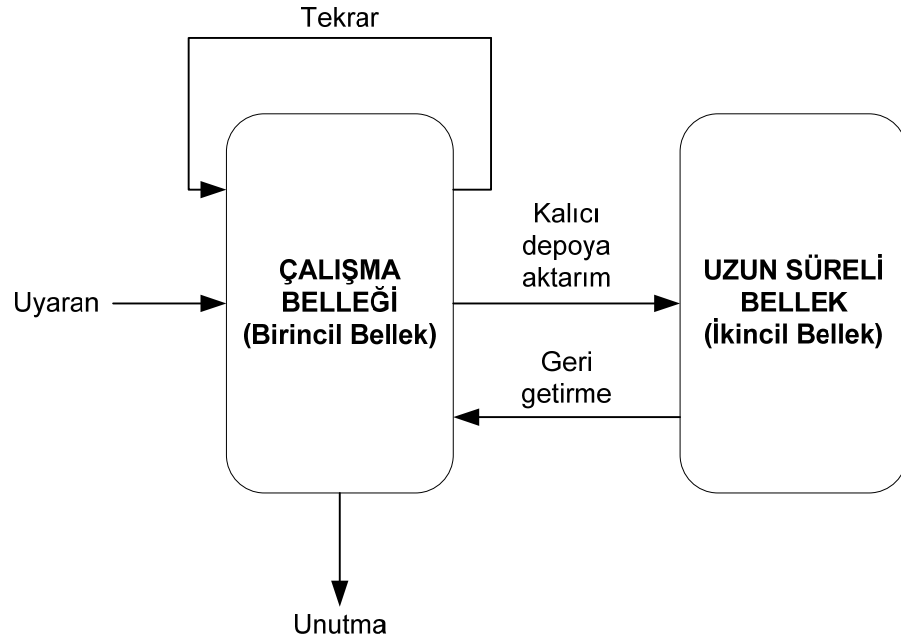
Elektronik ortamda üretilen ve saklanan metinlerin anlaşılabilir olması bu metinlerin kalitesini arttırmaktadır ve metinlerin başka kişiler tarafından kullanımını daha verimli kılmaktadır. Bundan dolayı doğal dil ile yazılmış metinlerin otomatik olarak değerlendirilmesi ve sonuçların yazarlarıyla paylaşılması, metinlerin daha anlaşılır yazılması için bilgisayarların insanlara sunduğu önemli bir hizmet olacaktır. Bir metnin insanlar tarafından ne derecede anlaşılıp anlaşılamayacağını otomatik olarak değerlendiren bir yöntem önermeden önce bu bölümde, insanın okuduğunu kavrama süreci açıklanmıştır ve bir metnin anlaşılabilir olması için hangi özelliklere sahip olması gerektiği anlatılmıştır.

İnsanların okumalarının nedeni genellikle, basılı formda temsil edilen materyalin anlamını çıkarmaktır. Yazılı materyalin anlamını anlama olayı kısaca okuduğunu kavrama olarak tanımlanmaktadır (Solso et al., 2005).

Kavrama sürecinde elde edilen bilgi, geçici ya da kalıcı olarak, bellek adı verilen bilişsel yapılarda depolanır. Çalışma belleği, kavrama, öğrenme, çıkarsama gibi karmaşık bilişsel görevler yerine getirilirken bilginin geçici olarak depolanmasını ve işlenmesini sağlayan bellek sistemi olarak tanımlanabilir (Baddeley, 1992). Kavrama sırasında bir dizi sembol algılanır ve işlenir. Çalışma belleği, kavrama sırasında birbirini izleyen kelimelerin işlenerek fikirler oluşturulması sürecinde ortaya çıkan ara ve son yapıları depolar. Kısaca çalışma belleği, kısa bir süre içinde erişilecek bilginin depolanmasına odaklanır (Just and Carpenter, 1992). Çalışma belleğinin bilgi depolama kapasitesi sınırlıdır. Uzun süreli belleğin en önemli ayırt edici özelliği, kodlama, bilgiyi soyutlama, yapı, kapasite ve süreklilikteki çeşitliliğidir (Solso et al., 2005). Uzun süreli bellek, çok fazla bilgiyi sınırsız süre ile (bir insanın yaşamı boyunca) depolayabilir. Çalışma belleğinin aksine uzun süreli bellek, bilgiyi soyutlar ve karmaşık yapılar içinde depolar.

Şekil 3.1’de, Waugh ve Norman’ın Solso et al. tarafından değiştirilmiş bilişsel modeli gösterilmiştir (Waugh, N. C. and Norman, D. A., 1965; Solso et al.’den, 2005). Bu modelde çalışma belleği ve uzun süreli bellek arasındaki ilişki açıklanmaktadır. Bu modele göre, uyaranlar öncelikle çalışma belleğine gelir. Çalışma belleğinin kapasitesi sınırlıdır. Bundan dolayı çalışma belleğine yeni uyaranlar geldikçe eski öğeler çalışma belleğinden atılır ve unutulur. Öğeler tekrarlandıkça çalışma belleğinde tutulur ya da uzun süreli belleğe aktarılabilir.

Uzun süreli bellekteki bilgiler ise uyarınları işlemek için çalışma belleğine geri çağırılır (Waugh, N. C. and Norman, D. A., 1965).



Şekil 3.1: Waugh ve Norman'ın değiştirilmiş bilişsel modeli (Waugh, N. C. and Norman, D. A., 1965; Solso et al.'den, 2005).

Bu bölümde, metin kavrama süreci için önerilen çeşitli modeller, kavrama sonucu zihinde oluşan temsiller, kavramanın başarısını etkileyen bağdaşıklık kavramı ve bağdaşıklığın ölçüm yöntemleri anlatılmaktadır.

3.1 Metin Kavrama Modelleri

Metin kavramayı açıklamaya çalışan teoriler, okumanın kelimeyi görmekle başlayan, okuduğu cümleleri anlamaya kadar birbirini takip eden süreçlerden oluştuğunu düşünerek bu süreci bölümlere ayırmıştır (Solso et al., 2005). Ayrıca, kavrama sırasındaki psikolojik süreçlerin incelenebilmesi için, metnin içeriğinin anlamının nasıl temsil edileceğinin belirlenmesi, hem teorik hem de yöntemsel açıdan önem taşır. Teorik açıdan, bir insanın, okuduğundan ne kavradığının anlaşılabilmesi için, metinden kavranması gerekenlerin bilinmesi gerekir. Yöntemsel açıdan, kavrama seviyesinin ölçümü, metindeki bilginin kullanılmasının sağlandığı, özetleme, kısa soru cevaplama gibi deneyler ile yapılır. Okuyucunun verdiği cevapların karşılaştırılabileceği bir metin içeriği temsili, deney sonuçlarının değerlendirilebilmesi için gereklidir (Rouet, 2006).

Kintsch'in kavrama modeli, hem metin kavrama süreci için bir model önerir hem de metnin anlamının önermeler ile temsil edilebileceğini savunur. Just ve Carpenter'ın geliştirdiği kavrama modeli, harflerin fiziksel özelliklerinin çıkartılması, kelimelerin kaydedilip kelime dağarcığına ulaşılması, duruma göre işlevlerin belirlenmesi gibi işlemleri içeren farklı basamakları ve çalışma belleği ile uzun süreli bellek gibi daha kalıcı bilişsel yapıları içerir. Perfetti'nin kavrama modeli ise görsel girdiyi dilsel bir temsile dönüştüren süreçleri (kelime tanıma) ve oluşturulan temsil üzerinde işlem yapan süreçleri (kelime metin bütünleştirme) içerir.

3.1.1 Kintsch'in kavrama modeli

Kintsch'in kavrama modeli, kelimelerin algılanmasından metnin anlamının bir temsiline oluşturulmasına kadar tüm okuma sürecini kapsar. Model, özellikle metnin anlamının kavranması üzerine vurgu yapmaktadır.

Kintsch'in kavrama modeli, “önerme”ler üzerine temellendirilmiştir. Önerme, bir savı öne süren ya da bir durumu dile getiren bir tümce; belli bir yorumda belli bir doğruluk değeri kazanan düzgün deyim olarak tanımlanabilir (Grünberg vd., 2003). Önermeler, gözlemlere (yazılı metni okuma ya da bir konuşmayı dinleme gibi) dayanan soyutlamalardır; bellekte tutulur ve bellek sürecini yöneten kurallara uyar (Solso et al., 2005). Her önerme bir yüklemden (genellikle bir fiil ya da sıfat) ve bir ya da daha fazla sayıda ögeden (genellikle bir isim ya da ifade) oluşur (Rouet, 2006).

Kintsch, metinlerin içeriğinin anlamının, temel anlamsal birimlere, yani önermelere, ayrıştırılabileceğini belirtmiştir (Kintsch, 1974; Rouet'den, 2006). Çizelge 3.1'de, örnek bir cümle ve bu cümlelerin çözümlemesinin sonucunda elde edilen önermeler gösterilmektedir (Kintsch, 1998; Rouet'den, 2006). Her önerme, bir yüklemden (“birinci”, “sahip”, “dört” gibi) ve bir ya da daha fazla ögeden (“odacık”, “kalp” gibi) oluşmaktadır. Bazı önermeler, “Sahip (Kalp, Dört (Odacık))” önermesi gibi, başka önermelerin ögeleri olabilir.

Çizelge 3.1: Bir cümlenin önermelere ayrıştırılmış biçimi (Kintsch, 1998; Rouet'den, 2006).

Cümle	Önermeler
Kalbin dört odacığından birisi olan sağ kulakçık, oksijeni tükenmiş ve karbondioksitle dolu morumsu kanı alır.	Birinci (Odacık) Sahip (Kalp, 3) Dört (Odacık) Sağ (Kapakçık) Olan (1, 4) Morumsu (Kan) Alır (1, 6) Tükenmiş (6, Oksijen) Dolu (6, Karbondioksit)

Kintsch ve van Dijk'e (1978) göre, kavrama, bir dizi işlem döngüsü sonucunda gerçekleşir. Her döngüde, metnin küçük bir bölümü okunur ve anlamsal önermelere ayrıştırılır. Bir döngüde ayrıştırılabilecek önerme sayısı okuyucunun çalışma belleğinin kapasitesine bağlıdır. Pratikte, bir ya da iki cümlenin bir döngüde işlenebileceği kabul edilmektedir. Önermeler ayrıştırıldıktan sonra, ortak öğelere sahip olan önermeler arasındaki bağlantılar kurulur. Bir döngüde ayrıştırılan önermeler daha önceki döngülerdeki önermelerle ilişkilendirilemezse, önermeleri ilişkilendirmek için uzun süreli bellekte arama yapılır. Bu arama, okuyucunun okunan metin hakkında artalan bilgisi varsa bu bilgiyi kullanarak ve çıkarsama yaparak önermeleri ilişkilendirmek için yapılır (Rouet, 2006).

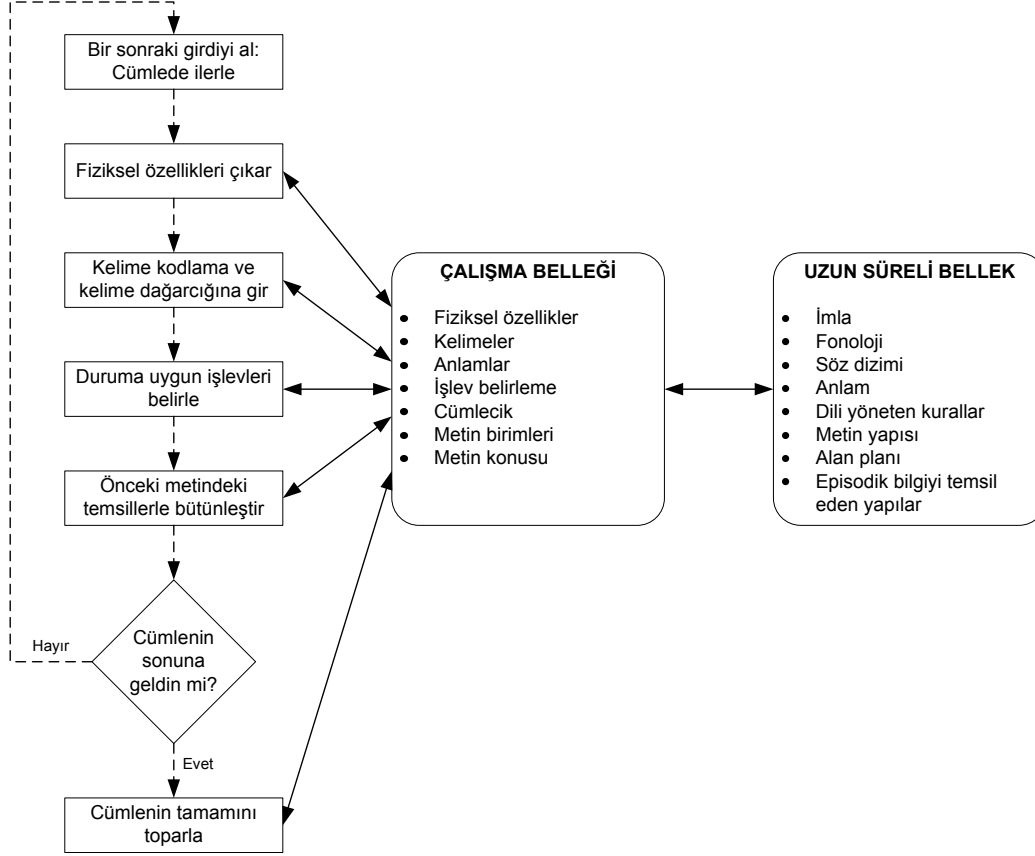
Tüm metin okunana kadar döngüler aynı şekilde işletilmeye devam eder. Sonuç olarak zihinde bir metin tabanı oluşturulur. İyi bir zihinsel temsilin oluşturulması, okuyucuya, metni tekrar edebilmesini, basit sorulara cevap verebilmesini ve yeni elde edilecek bilgilerin mevcut temsil ile tutarlılığını kontrol edebilmesini sağlayacaktır (Rouet, 2006).

Metin tabanının oluşturulması, metnin okuyucu tarafından tamamen kavrandığı anlamına gelmez. Başarılı bir kavrama için metinden elde edilen bilginin, uzun süreli bellekten getirilen artalan bilgisi ile bütünleştirilmesi ve çeşitli çıkarsamalar yapılması gerekmektedir. Bu işlemler sonucunda zihinsel ya da durum modeli oluşur (Rouet, 2006).

3.1.2 Just ve Carpenter'ın kavrama modeli

Just ve Carpenter'ın geliştirdiği kavrama modelinin temel öğeleri Şekil 3.2'de gösterilmiştir (Just and Carpenter, 1980, 1987; Solso et al.'den, 2005). Bu modelde, harflerin fiziksel özelliklerinin çıkartılması, kelimelerin kaydedilip

kelime dağarcığına ulaşılması, duruma göre işlevlerin belirlenmesi gibi işlemleri içeren farklı basamaklar bulunur (Solso et al., 2005). Bu basamaklar şeklin sol tarafında gösterilirken daha kalıcı bilişsel yapılar ve işlemler orta ve sağ tarafta gösterilmiştir.



Şekil 3.2: Okuduğunu kavramada temel yapıların ve süreçlerin şematik bir gösterimi (Just and Carpenter, 1980, 1987; Solso et al.'den, 2005).

Just ve Carpenter'ın kavrama modelinde de görüldüğü gibi okuyucu yeni bir girdi (kelime, kavram) ile karşılaştığında, bu girdiyi algıladıktan sonra, girdiyi daha önceki bilgilerle bütünleştirmeye çalışır. Bu bütünleştirme işlemini çalışma belleğini kullanarak yapar. Bu işlemler sırasında artalan bilgisi kullanılarak çıkarsama yapılacaksa uzun süreli bellek de kullanılır. Cümle sonuna gelindiğinde ise okuyucuların gözlerinin daha fazla sabitlendiği tespit edilmiş ve bu aşamada cümlede elde edilen bilgilerin genel bir bütünleştirme işlemine tabi tutulduğu sonucuna varılmıştır.

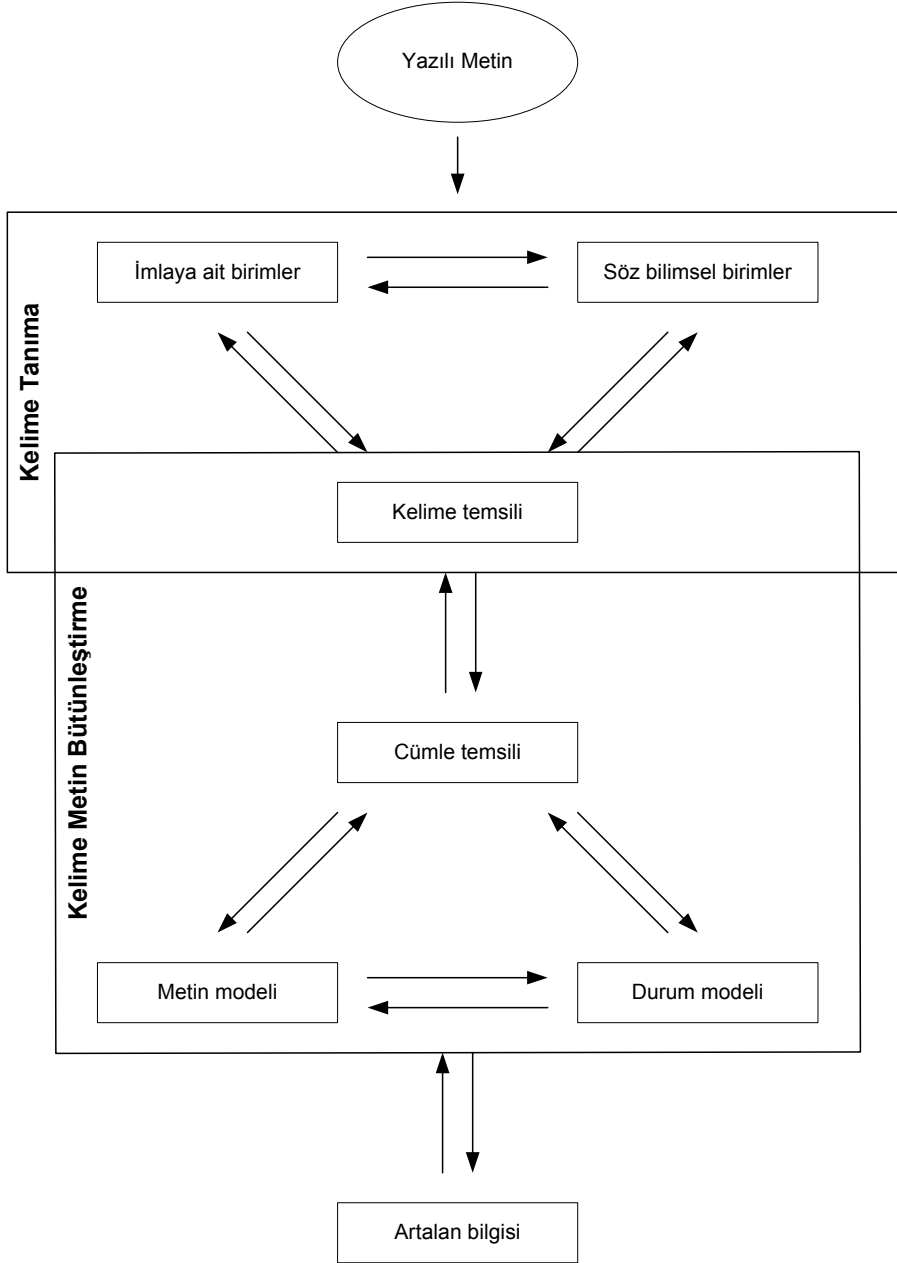
3.1.3 Perfetti'nin kavrama modeli

Perfetti'ye göre, okuma eylemindeki bilişsel süreçleri anlayabilmek için, üç ana işlem adımını incelemek gerekir (Perfetti, 1999):

1. Görsel süreçler
2. Görsel girdiyi dilsel bir temsile dönüştüren süreçler
3. Oluşturulan temsil üzerinde işlem yapan süreçler

Perfetti, bu üç ana işlem adımını kapsayan bir metin kavrama modeli önermiştir (Perfetti, 1999; Verhoeven and Perfetti, 2008). Bu model Şekil 3.3'te gösterilmiştir.

Bir metin okunurken ilk olarak kelimeler tanınır, yani görsel girdi dil bilimsel bir temsile dönüştürülür. Bir sonraki adımda, okunan metnin kavranabilmesi için kelime metin bütünleştirme işlemi gerçekleşir. Cümlelerin anlaşılabilmesi için öncelikle içindeki kelimelerin tanınması gerekir. Bir kelime tanındığında okuyucu, kelimeyi metnin daha önce oluşturulmuş temsiliyle bütünleştirmeye çalışır. Göz hareketlerini inceleyen bazı araştırmalar, cümlelerin kavranması hakkında bazı önemli sonuçlara varmıştır (Just and Carpenter, 1992). Birincisi, okunan metnin konusu hakkında artalan bilgisine sahip olan okuyucular bile, okudukları çoğu kelimeye gözlerini sabitlemişlerdir. Bu sonuç, kelime tanınmanın, okuduğunu kavrama sürecinin merkezinde olduğunu kanıtlar. İkinci önemli bulgu ise, cümle sonlarında okuyucuların gözlerinin daha uzun süre sabitlendiğidir. Bu da bütünleştirici kavrama işlemlerinin özellikle cümle sonlarında daha yoğun gerçekleştiğini kanıtlamaktadır (Verhoeven and Perfetti, 2008). Okunan metnin tamamının kavranabilmesi için her cümlelerin anlamının, metnin önceki bölümleriyle bütünleştirilmesi gerekmektedir. Bu bütünleştirme sırasında okuyucunun artalan bilgisi de kullanılır ve metin modeli ile durum modeli oluşturulur (Verhoeven and Perfetti, 2008).



Şekil 3.3: Perfetti'nin metin kavrama modeli (Verhoeven and Perfetti, 2008).

3.2 Metnin Zihinsel Temsili

Araştırmacılar, okuma sırasında, okunan metnin iki tür zihinsel temsiliinin oluştuğunu ortaya atmıştır. Bir görüşe göre metin, önermeler olarak temsil edilir ve bu temsil metin tabanlı temsil olarak adlandırılır. Diğer görüşe göre ise, metin, gerçek dünya durumlarını yansıtan ögeler tarafından temsil edilir (Stevenson, 1993). Daha sonra bu iki görüşün birleştirildiği ortak bir yaklaşım genel kabul görmüştür (Johnson-Laird, 1983; van Dijk and Kintsch, 1983; Stevenson'dan, 1993).

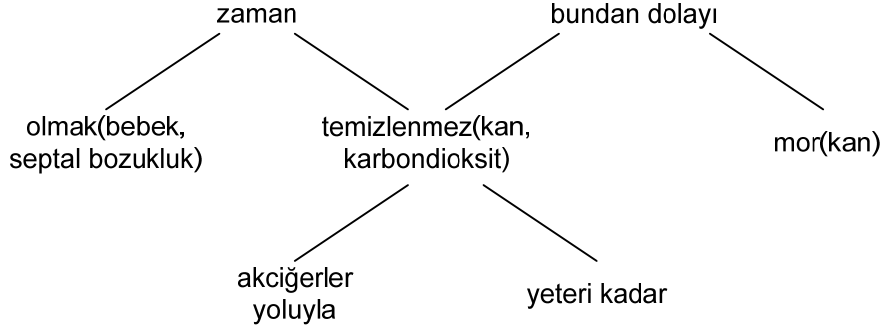
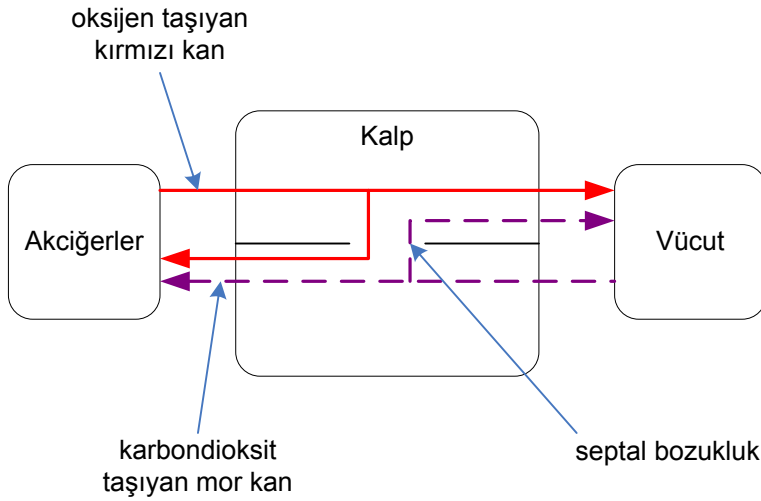
Johnson-Laird'e (1983) göre metin kavrama, hem önermesel hem de zihinsel temsili içerir (Stevenson, 1993). Önermesel temsil, öğelerinin kelime yerine kavramlar olması dışında metnin cümle yapısına çok benzerdir. Zihinsel model ise metnin betimlediği durumları temsil eder. Zihinsel model, herhangi bir dilsel yapıya değil, gerçek dünyanın bir bölümüne benzer. Zihinsel modelde, metinde bulunmayan bazı bilgiler de bulunabilir. Bu bilgiler, okuyucu tarafından yapılan çıkarsamalar ile elde edilir.

Van Dijk and Kintsch'e (1983) göre de okuma sırasında üç seviyeli temsil oluşur (Stevenson, 1993). Bu seviyeler, yüzeysel ya da sözdizimsel seviye, önermesel bir metin tabanı ve durum modeli olarak adlandırılır. Yüzeysel seviyede metin, kullanılan kelimelerle ve kelime öbekleriyle temsil edilir. Önermesel metin tabanında, metnin anlamsal içeriği temsil edilir. Bu seviye Johnson-Laird'in önermesel modeli ile uyumludur. Durum modelinde ise, metnin betimlediği durum temsil edilir ve bu modelde okuyucunun geçmiş bilgisi de bulunur. Bu seviye de Johnson-Laird'in zihinsel modeli uyumludur (Stevenson, 1993).

Şekil 3.4'te kalp hastalıkları üzerine bir metinden alınmış iki cümle ve bu cümlelerin metin tabanı ve durum modelleri gösterilmiştir (Kintsch, 1994). Metin tabanı üç önermeden ("olmak", "temizlenmez", "mor") oluşmaktadır. Her okuyucunun aynı metin tabanını oluşturması zorunlu değildir. Örneğin bir okuyucu "yeteri kadar" niteleyicisini ihmal edebilir. Durum modeli, Şekil 3.4-c'de grafik şeklinde gösterilmiştir. Grafiğin bir bölümü, metindeki bilgiden çok, okuyucunun dolaşım sistemi ile ilgili artalan bilgisi üzerine temellendirilir. Septal duvarında bir boşluk olduğu ve bundan dolayı mor kanın kırmızı kan ile karıştığı bilgisi metinden elde edilir. Metin tabanında olduğu gibi, her okuyucunun aynı durum modelini de oluşturması beklenemez. Nasıl bir durum modeli oluşacağı okuyucunun dolaşım sistemi ile ilgili artalan bilgisine de bağlıdır.

Metin:

Bir bebekte septal bozukluk olduğu zaman, kan akciğerler yoluyla karbondioksitten yeteri kadar temizlenemez. Bundan dolayı kan mor görünür.

Metin Tabanı:**Durum Modeli:**

Şekil 3.4: Metnin zihinsel temsili için bir örnek a) Metin, b) Metin tabanı, c) Durum modeli (Kintsch, 1994).

Johnson-Laird'e (1983) göre, zihinsel modellerin oluşturulması, güncellenmesi, bütünleştirilmesi ve değerlendirilmesi aşağıdaki psikolojik işlemlerle gerçekleşir (Stevenson, 1993):

1. Önerme, mevcut zihinsel modeldeki bir varlığa atıfta bulunmuyorsa bu durumda yeni bir zihinsel model oluşturulur.
2. Önerme, mevcut zihinsel modelde bulunan en az bir varlığa atıfta bulunuyorsa, önermede bulunan geri kalan bilgi zihinsel modele eklenir. Bu eklenen bilgi, varlıklardan, özelliklerden ya da ilişkilerden oluşabilir.

3. Önerme, iki farklı modeldeki varlıklara atıfta bulunuyorsa, bu iki model bütünleştirilir.
4. Önermedeki tüm öğeler modelde bulunuyorsa, modeldeki mevcut bilginin yeni önerme ile tutarlı olup olmadığı kontrol edilir.
5. Dördüncü maddedeki tutarlılık kontrolü yapılamıyorsa önermenin özelliği ya da ilişkisi modele eklenir.
6. Dördüncü maddedeki tutarlılık kontrolü sonucunda, önermenin mevcut modelle tutarlı olduğu saptanırsa, mevcut modelin önerme yanlış olacak şekilde değiştirilip değiştirilemeyeceği kontrol edilir. Bu işlem tündengelimli çıkarsamaları olanaklı kılar.
7. Dördüncü maddedeki tutarlılık kontrolü sonucunda, önermenin mevcut modelle tutarlı olmadığı saptanırsa, mevcut modelin önerme doğru olacak şekilde değiştirilip değiştirilemeyeceği kontrol edilir. Bu işlem tümevarımlı çıkarsamaları olanaklı kılar.

Kavrama sürecinde, bir ifadenin önermesel temsili, içerdiği öğelere, mevcut zihinsel modele ve artalan bilgisine bağlı olarak bu yedi işlemten birisinden geçer. Bir zihinsel modelin oluşturulması (işlem 1), güncellenmesi (işlem 2 ve 5), bütünleştirilmesi (işlem 3) ve değerlendirilmesi (işlem 4, 6 ve 7) Johnson-Laird (1983) tarafından bu şekilde açıklanır (Stevenson, 1993).

3.3 Bağdaşıklık

Yazılı metinleri, birbiriyle bağlantısız cümlelerin ya da ifadelerin oluşturduğu yığınlar olmaktan çıkarıp birleştirilmiş bir bütün olmasını sağlayan bazı özellikler vardır. Bunlar, cümlelerin ya da ifadelerin nasıl birbirlerine bağlandıklarını açıklayan bağlaşıklık (cohesion) ve bağdaşıklık (coherence)³ özellikleridir (Tanskanen, 2006).

³ Bazı Türkçe yayınlarda, “cohesion” ve “coherence” kavramlarının Türkçe karşılıkları olarak sırasıyla “bağlaşıklık” ve “bağdaşıklık” kelimeleri kullanılmaktadır (Alagözlü, 2005; Köksal, 2009). Bazı yayınlarda ise “cohesion” ve “coherence” kavramlarının Türkçe karşılıkları olarak sırasıyla “bağdaşıklık” ve “tutarlılık” kelimeleri kullanılmaktadır (Karatay, 2010; Onursal, 2003). Cambridge Learner's İngilizce - Türkçe sözlükte, “coherent” kelimesinin Türkçe karşılıklarından

Bağlaşıklık ve bağdaşıklık kavramlarının arasında bir fark olduğu konusunda araştırmacılar arasında genel bir uzlaşma vardır. Fakat aradaki farkın ne olduğu konusunda tam bir uzlaşma bulunmamaktadır. Genel olarak kabul edilen görüşe göre bağlaşıklık, metni oluşturan parçalar arasındaki bağlantıları kuran, metin yüzeyindeki dilbilgisel öğeler ile ilgilidir. Diğer taraftan bağdaşıklık, sadece metnin bir özelliği değildir. Bağdaşıklık, metin ile okuyucu arasındaki etkileşimin sonucu oluşur. İki kavram birbirinden farklı olmakla birlikte, Tanskanen'e göre bağlaşıklık, bağdaşıklığa katkıda bulunur (Tanskanen, 2006).

Dil öğelerinin açıkladıkları özellik ve durumlar arasındaki anlam bağıntılarına bağdaşıklık denir. Bağdaşıklık, metin ile okuyucu arasındaki etkileşimin sonucu olduğu için öznel bir kavramdır. Dolayısıyla bir metin, bir okuyucu için bağdaşık olurken başka bir okuyucu için bağdaşık olmayabilir. Bağdaşıklığın algılanması, okuyucunun artalan bilgisine bağlıdır.

Hobbs'dan (1985) ve Kehler'den (2002) faydalanarak Wolf and Gibson'un oluşturduğu örnek bir bağdaşıklık ilişki listesi Çizelge 3.2'de gösterilmiştir (Wolf and Gibson, 2004).

Çizelge 3.2'deki örnek cümlelerde, “a” ve “b” harfleriyle işaretlenmiş olan cümle bölümleri arasında anlam bağıntıları bulunmaktadır. Bu anlam bağıntılarının tipi, yani bağdaşıklık ilişkisi ise çizelgenin ilk sütununda gösterilmiştir.

birisi olarak “bağdaşık” kelimesi verilmektedir (Cambridge University Press, 2009). Ayrıca, “cohesion” ya da “cohesive” kelimelerinin Türkçe karşılıklarında “bağdaşıklık” ile ilgili bir Türkçe karşılık yer almamaktadır (Cambridge University Press, 2009). İnkılap Yayınevi tarafından yayımlanan Türkçe - İngilizce Bilimsel ve Teknik Terimler Sözlüğü'nde ve Fono tarafından yayımlanan Türkçe - İngilizce Teknik Terimler Sözlüğü'nde “bağdaşık” kelimesinin İngilizce karşılıklarından biri olarak “coherent” kelimesi verilmektedir (Tuğlacı, 2001; Bayram, 1993). Bu nedenlerden dolayı bu tezde, “cohesion” ve “coherence” kavramlarının Türkçe karşılıkları olarak sırasıyla “bağlaşıklık” ve “bağdaşıklık” kelimeleri kullanılmıştır.

Çizelge 3.2: Bağdaşıklık ilişkileri (Wolf and Gibson, 2004).

Bağdaşıklık İlişkisi	Bağlaç	Örnek Cümle
Sebep sonuç	Çünkü	[Uçuşumuz ertelendi] _a , [çünkü havaalanındaki hava koşulları kötüydü] _b .
Gerçekleşmemiş beklenti	Rağmen, ama	[Hava koşulları iyi olmasına rağmen] _a [uçuşumuz ertelendi] _b .
Şart	Eğer	[Eğer yeni yazılım çalışırsa] _a [herkes mutlu olacak] _b .
Benzerlik	Benzer şekilde	[Bugün İzmir’de hava güzeldi] _a , [Benzer şekilde Ankara’da da hava güzeldi] _b .
Karşıtlık	Ama	[Ahmet seçimlerde A Partisi’ni destekliyordu] _a , [ama eşi A Partisi’ne karşıydı] _b .
Detaylandırma	Ayrıca	[Mars’a gönderilecek mekik basına tanıtıldı] _a , [Ayrıca mekiğin bu yıl sonunda fırlatılacağı açıklandı] _b .
Atıf	-e göre	[Hava durumu haberlerine göre] _a [hava yarın açık olacak] _b .
Zamansal sıra	-den önce, -den sonra	[Yatmaya gitmeden önce] _a [banyo yaptım] _b .

3.4 Başarılı Metin Kavrama ve Bağdaşıklık

Okunan materyalin başarılı olarak kavranabilmesi, metnin bağdaşık bir bellek temsilinin oluşturulmasına bağlıdır. Bağdaşık bir temsil oluşturmak için okuyucu, metni oluşturan her bileşeni yorumlamalı ve metindeki diğer bileşenlerle ve zihnindeki mevcut kavramlarla arasındaki anlamlı bağlantıları yakalamalıdır. Sonuç olarak oluşan temsil, metindeki bileşenleri ya da bu bileşenlerle bağlantılı başka bileşenleri içeren düğümlerden ve bu düğümler arasındaki anlamsal ilişkileri içeren bağlantılardan meydana gelir. Bu düğümler ve bağlantılar bir ağ oluşturur. Kavrama üzerine yapılan araştırmalar sonucunda, yoğun olarak bağlantı içeren metinlerin daha az yoğun bağlantı içeren metinlere göre daha bağdaşık olduğu ve daha çok bağlantıya sahip olan bileşenlerin daha sık ve daha hızlı hatırlanabildiği rapor edilmiştir (van den Broek et al., 2002).

3.5 Bağdaşıklığın Ölçülmesi

Bir metnin bağdaşıklığının ölçülmesi için yarı otomatik ve otomatik çeşitli yöntemler önerilmiştir. Yarı otomatik yöntemlere örnek olarak, metnin önermesel temsili çözümlenerek bir metnin bağdaşıklığı ölçülebilir. Fakat bu yaklaşımdaki problem, metnin önermesel temsili oluşturmanın çok fazla çaba gerektirmesidir. Metnin önermesel temsili oluşturulması için yarı otomatik yöntemler bulunmaktadır (Turner, 1987; Foltz et al.’den, 1998). Bu yöntemler de çok ciddi çaba gerektirdiği için çözümlenebilen metinlerin büyüklüğü oldukça sınırlıdır.

Yarı otomatik yöntemlerle yapılan okuduğunu kavrama deneylerinde kullanılan metinler genellikle 50 ile 1000 arasında kelimeden oluşmaktadır (Foltz et al., 1998).

Okunabilirlik ölçüleri gibi otomatik yöntemler ise metnin bağdaşıklığını ölçmek için kullanılan bir başka yaklaşımdır. Bu otomatik yöntemler, önermesel temsil ile metnin bağdaşıklığının ölçülmesine göre çok daha az çaba gerektirmektedir. Fakat bu otomatik yöntemlerin metnin bağdaşıklığını ölçme konusunda başarılı olmadıkları rapor edilmiştir (Britton and Gulgoz, 1991; Kintsch and Vipond, 1979; Foltz et al.'den, 1998).

Metnin önermesel temsilinin çözümlenmesi ve okunabilirlik ölçülerinin kısıtlarından dolayı daha az çaba gerektiren ve daha başarılı sonuçlar üreten yöntemler araştırılmıştır. Foltz et al., metnin bağdaşıklığını ölçmek için Gizli Anlam Çözümleme yöntemini (Landauer and Dumais, 1997) kullanmıştır (Foltz et al., 1998). Foltz et al., Gizli Anlam Çözümleme yöntemini kullanarak metin bölümleri arasındaki anlamsal ilintililiği hesaplamıştır. Hesaplanan anlamsal ilintililik değerleri metnin bağdaşıklığı için bir tahminde bulunmak için kullanılmıştır (Foltz et al., 1998).

Foltz et al.'nin önerdiği yöntemde, metnin bağdaşıklığı hakkında bir tahmin yapabilmek için, metindeki bölümlerin birbiriyle anlamsal olarak ne kadar bağlantılı olduğu Gizli Anlam Çözümleme yöntemi ile tespit edilmektedir. Aralarındaki anlamsal ilintililik değerleri ölçülen metin bölümleri, kelime, cümle, paragraf ya da uzun bir metin olabilir. Metnin önermesel temsilinin çözümlenmesi yöntemiyle bağdaşıklığın tahmin edilmesi için çalışma belleğinde depolanan önermeler arasındaki anlamsal ilintililik ölçülür (Kintsch, 1998; Foltz et al.'den, 1998). Foltz et al. çalışmalarında, Gizli Anlam Çözümleme yöntemiyle bir metindeki ard arda gelen cümleler arasındaki anlamsal ilintililiği ölçmüştür. Anlamsal ilintililiği ölçmek için cümlelerin seçilmesinin nedeni, cümlelerin genellikle 3-7 arasında önerme içermesidir ve bir cümlenin içerdiği önerme sayısının çalışma belleğinde bir anda depolanabilecek bilgi miktarıyla tutarlı olmasıdır (Foltz et al., 1998). Bu yaklaşım, cümle sonlarında okuyucuların gözlerinin daha uzun süre sabitlendiği ve bütünleştirici kavrama işlemlerinin özellikle cümle sonlarında daha yoğun gerçekleştiğini kanıtlayan araştırmalar tarafından da desteklenir (Verhoeven and Perfetti, 2008). Ayrıca otomatik bir yöntem için metinlerin içindeki cümlelerin ayrıştırılması gerçekleştirim açısından oldukça kolaydır.

Foltz et al., bir metnin bağdaşık olarak nitelendirilebilmesi için komşu cümleler arasında yüksek seviyeli anlam örtüşmeleri olması gerektiğini belirtmiştir. Foltz et al.'ye göre, iki cümle arasındaki anlam örtüşmesinin seviyesi, iki cümle arasındaki anlamsal ilintililik değeri ölçülerek tespit edilebilir. İki cümle arasındaki anlamsal ilintililik ne kadar yüksekse iki cümle arasındaki anlam örtüşmesi o kadar yüksektir. Buna göre, bir metnin ard arda gelen cümleleri arasındaki anlamsal ilintililik değerleri ne kadar yüksekse, o metnin komşu cümleleri arasındaki anlam örtüşmeleri o kadar fazladır ve metin de o derecede yüksek bir bağdaşıklık değerine sahiptir. Foltz et al. cümleleri birer vektör olarak temsil etmiş ve cümle çiftleri arasındaki uzaklıkları ölçerek anlamsal ilintililiği hesaplamışlardır. Cümleleri temsil eden vektörler oluşturulurken İngilizce’de çok geçen ve cümlenin anlamına katkıda bulunmayan kelimeler (the, and, from gibi) göz ardı edilmiştir. İki cümle vektörü arasındaki anlamsal ilintililik kosinüs benzerliği⁴ yöntemiyle ölçülmüştür. Örneğin, aynı terimlere ve aynı terim frekanslarına sahip iki vektörün kosinüs benzerliği 1, anlamsal olarak birbiriyle bağlantılı hiçbir ortak kelimenin bulunmadığı iki vektörün kosinüsü 0, hem anlamsal olarak bağlantılı hem de bağlantısız kelimelerden oluşan iki vektörün kosinüsü ise 0 ile 1 arasında bir değer olacaktır (Foltz et al., 1998).

Foltz et al. yöntemini, Britton ve Gulgoz’un deneyinden (Britton and Gulgoz, 1991) alınan dört metin üzerinde sınamıştır. Elde edilen sonuçlara göre,

⁴ İki kelime vektörü arasındaki benzerliği hesaplayan yaklaşımlardan birisi “kosinüs benzerliği”dir (Salton et al., 1975). Bu yaklaşıma göre kv_i ve kv_j vektörleriyle temsil edilen iki kelime vektörü arasındaki benzerlik aşağıdaki şekilde hesaplanır:

$$\frac{kv_i \times kv_j}{\sqrt{kv_i \times kv_i} \times \sqrt{kv_j \times kv_j}}$$

İki vektörün çarpımı olan $kv_i \times kv_j$ aşağıdaki şekilde hesaplanır:

$$\sum_{k=0}^n \sum_{l=0}^n a_{ik} \times a_{jl}$$

a_{ik} ($1 < k < n$), k kelimesinin i belgesindeki ağırlığını belirten negatif olmayan bir sayıdır. Bir i belgesi ise bu ağırlık değerlerinden oluşan bir vektör olarak temsil edilir:

$$kv_i = \{ a_{i1}, a_{i2}, \dots, a_{in} \}$$

Genellikle bir kelimenin ağırlığı Terim Frekansı – Ters Belge Frekansı (Term Frequency – Inverse Document Frequency) ile belirlenir. Buna göre a_{ik} değeri aşağıdaki formülle hesaplanır:

$$tf_{ik} \times \log_2 \frac{N}{n_k}$$

Formüldeki bileşenlerin açıklamaları ise aşağıdaki gibidir:

tf_{ik} : k kelimesinin i belgesindeki frekansı

$\log_2 \frac{N}{n_k}$: ters belge frekansı, k kelimesinin tüm belgelerde ne kadar sıklıkla bulunduğunu belirtir.

N : belge sayısı

n_k : k kelimesinin en az bir kere geçtiği belge sayısı

yöntem kullanılarak elde edilen sonuçların insanların metinleri değerlendirme sonuçlarıyla uyumlu olduğu rapor edilmiştir (Foltz et al., 1998).

Lapata ve Barzilay da, Foltz et al. gibi ard arda gelen cümleler arasındaki anlamsal benzerliği ölçerek metnin bağdaşıklığını tahmin etmeye çalışmışlardır (Lapata and Barzilay, 2005). Lapata ve Barzilay'a göre her cümle, kelime ya da kelime öbeklerinden oluşan bir küme ile temsil edilebilir ve bu kümeye sadece isimler ve fiiller dahil edilebilir. Bağdaşıklığın ölçülebilmesi için komşu cümleler arasındaki anlamsal benzerliğin ölçülebilmesi gerekmektedir. Buna göre bir M metnin bağdaşıklığı (3.1) numaralı formülle ölçülebilir (Lapata and Barzilay, 2005).

$$\text{Bağdaşıklık}(M) = \frac{\sum_{i=1}^{n-1} \text{Benzerlik}(C_i, C_{i+1})}{n - 1} \quad (3.1)$$

Benzerlik(C_i, C_{i+1}) fonksiyonu C_i ve C_{i+1} cümleleri arasındaki anlamsal benzerliği temsil etmektedir. n sayısı ise metindeki cümle sayısını ifade etmektedir.

3.6 Özet ve Değerlendirme

İnsanların, okuduklarını nasıl kavradıklarını açıklamak için çeşitli modeller önerilmiştir (Kintsch, 1974; Kintsch and van Dijk, 1978; Kintsch, 1998; Just and Carpenter, 1980, 1987; Perfetti, 1999; Verhoeven and Perfetti, 2008). Bu modellerin en önemli ortak yönü, okuduğunu kavrama sürecinde, ard arda gelen bilgileri daha önceki bilgilerle bütünleştirerek bir bellek temsili oluşturulduğunun kabul edilmesidir. Johnson-Laird'in, zihinsel modellerin oluşturulması, güncellenmesi, bütünleştirilmesi ve değerlendirilmesi için gerçekleştiğini belirttiği psikolojik işlemler de, elde edilen bilgilerin önceki bilgilerle ilişkilendirildiği görüşünü desteklemektedir (Johnson-Laird, 1983). Başarılı bir kavrama süreci sonunda, bağdaşık bir bellek temsili oluşturulmuş kabul edilmektedir (van den Broek et al., 2002). İnsanın, bağdaşık bir bellek temsili oluşturabilmesi için, okunan metnin içerdiği kelimelerin, önermelerin, cümlelerin birbirleriyle anlamsal olarak bağlantılı olması gerekmektedir. Böylece okunan her birim (kelime, önerme, cümle) kısa süre önce okunan birim ile ilişkilendirilmeye çalışılır. Bu işlem çalışma belleğinde gerçekleşir. Çalışma belleğinin kapasitesi sınırlı olduğu için, ard arda gelen birimlerin birbirleriyle ilişkilendirilmesi ve bir zihinsel temsilin parçası olarak uzun süreli belleğe aktarılması gerekmektedir. Ard

arda gelen birimlerin birbiriyle bağlantısız olması durumunda bazı bilgiler zihinsel temsile eklenemeyecek, çalışma belleğinden çıkartılarak unutulacaktır. Bu durum da kavrama sürecini olumsuz etkileyecektir. Özellikle metnin birimleri arasındaki bağlantıların gevşek olması durumunda, metnin konusu hakkında artalan bilgisine sahip olmayan okuyucular bundan daha da olumsuz etkileneceklerdir. Çünkü metnin konusu hakkında bilgi sahibi olan okuyucular, metin birimleri arasındaki gevşek bağlantıları, artalan bilgilerini kullanarak yaptıkları çıkarsamalarla kuvvetlendirip başarılı bir kavrama gerçekleştirebilirler. Fakat okunan metin hakkında artalan bilgisi olmayan okuyucuların, metni başarılı olarak kavrayabilmeleri için okunan metnin bağdaşık olması gerekmektedir.

Metnin bağdaşıklığını değerlendirmek için çeşitli yöntemler önerilmiştir. Bu yöntemlerden metnin yarı otomatik olarak önermesel temsiline oluşturulması yöntemi çok fazla çaba gerektirdiği için ölçeklenebilir değildir. Bu yöntemle bağdaşıklığı değerlendirilen metinlerin uzunlukları 50 ile 1000 kelime arasında değişmektedir. Metnin bağdaşıklığı konusunda otomatik olarak değerlendirme yapabileceği düşünülen okunabilirlik ölçüleri ise yapılan deneylerde başarılı sonuçlar vermemiştir. Bundan dolayı metinlerin bağdaşıklığını değerlendirmek için hem güvenilir sonuçlar veren hem de ölçeklenebilir yöntemler araştırılmıştır.

Foltz et al. bağdaşık bir metinde cümleler arasındaki anlamsal ilintiliğin, Lapata ve Barzilay ise anlamsal benzerliğin belirli bir seviyenin üzerinde olması gerektiğini belirtmişlerdir (Foltz et al., 1998; Lapata and Barzilay, 2005). Metinde ard arda gelen cümleler arasındaki anlamsal ilintiliğin ve benzerliğin ölçülerek metnin bağdaşıklığının değerlendirildiği deneylerde başarılı sonuçlar elde edildiği görülmüştür. Bilişsel psikoloji alanında yapılan çalışmalar da bu kabulü desteklemektedir. Yapılan deneylerde okuyucuların cümle sonlarında gözlerini daha fazla sabitledikleri, bu sırada da cümlede elde ettikleri bilgiyi önceki bilgilerle bütünleştirdikleri tespit edilmiştir. Bu bulgular doğrultusunda araştırmacılar, bir metnin bağdaşıklığını otomatik olarak ölçebilmek için cümleler arasındaki bağlantıları değerlendirme yöntemini benimsemişlerdir. Foltz et al., cümleler arasındaki bağlantıları değerlendirmek için istatistiksel bir anlamsal ilintililik ölçüm yöntemini (Gizli Anlam Çözümleme), Lapata ve Barzilay ise literatürdeki çeşitli anlamsal benzerlik ölçüm yöntemlerini tercih etmiştir. Foltz et al., kelimelerin bir derlemde geçme sıklıklarına bakarak kelimeler arasındaki anlamsal ilintiliği istatistiksel olarak hesaplamaktadır. Dolayısıyla, bu hesaplama yönteminde insan bilgisini temsil eden bir kavramsal model kullanılmamaktadır. Lapata ve Barzilay'ın tercih ettiği anlamsal benzerlik, anlamsal ilintiliğin özel

bir durumdur. Anlamsal olarak birbirine benzer kavramlar, birbirleriyle ortak yönleri ile ilintilidir. Fakat iki kavram arasında farklı ilişkilerden kaynaklanan bağlantılar da bulunabilir. Aralarında parça-bütün ya da zıtlık ilişkisi bulunan iki kavram, anlamsal olarak benzer olmamasına rağmen anlamsal olarak ilintili olabilir. Bir okuyucu bir metni okurken, kavramlar arasında sadece benzerlik ilişkisini kurduğunu söylemek yanlış olacaktır. Okuyucunun kavramlar arasındaki çok çeşitli ilişkileri kurarak bir zihinsel temsil oluşturması, mevcut okuduğunu kavrama modellerine daha uygun düşmektedir. Bundan dolayı bu çalışmada, bir metnin bağdaşıklığını değerlendirirken ard arda gelen cümleler arasındaki anlamsal benzerliğin ya da anlamsal ilintililiğin istatistiksel olarak ölçülmesi yerine anlamsal ilintililiğin bir kavramsal model kullanılarak ölçülmesi tercih edilmiştir. Buna göre, bir M metninin bağdaşıklığını ölçmek için (3.2) numaralı formülün kullanılması önerilmektedir.

$$\text{Bağdaşıklık}(M) = \frac{\sum_{i=1}^{n-1} \text{Anlamsal İlntililik}(C_i, C_{i+1})}{n - 1} \quad (3.2)$$

İki cümle arasında “benzerlik” bağdaşıklık ilişkisi olması durumunda, metin bağdaşıklığını ölçmek için cümleler arasındaki benzerliğin otomatik olarak hesaplanması doğru ya da doğruya yakın bir sonuç verecektir. Fakat iki cümle arasında “karşıtlık” bağdaşıklık ilişkisi olması durumunda, iki cümle arasındaki benzerliğin ölçülmesi, bağdaşıklık tahmini için doğru olmayan bir sonuç verecektir. Bunun nedeni, karşıtlık ilişkisi içinde olan iki cümle arasındaki benzerliğin de düşük olmasından kaynaklanacaktır. Oysa metin bağdaşıklığı tahmini için, iki cümle arasındaki ilintililik ölçümü yapılması durumunda, iki cümle arasındaki karşıtlık ilişkisi de ilintililik bağlantısı oluşturduğundan, daha doğru sonuçlar elde edilecektir. Örneğin, Çizelge 3.2’de benzerlik ilişkisi için verilen örnekte, iki cümle arasındaki benzerlik yüksektir. Bu iki cümle arasındaki benzerlik otomatik olarak ölçüldüğünde yüksek bir değer elde edilecek ve iki cümle arasındaki bağdaşıklığın da bundan dolayı yüksek olduğu sonucuna varılacaktır. Fakat Çizelge 3.2’de karşıtlık ilişkisi için verilen örnekte ise, iki cümle arasında yapılacak benzerlik ölçümü, aralarında benzerlik bağdaşıklık ilişkisi olan cümlelere göre daha düşük çıkacaktır. Bu sonuçtan hareket edilerek iki cümle arasındaki bağdaşıklığın düşük olduğu sonucuna varılacaktır. Oysa iki cümle arasındaki ilintililik ilişkisinin otomatik olarak ölçülmesi durumunda ilintililik değeri, benzerlik değerine göre daha yüksek çıkacaktır. Bunun nedeni, “desteklemek” ve “karşı olmak” eylemlerinin birbirinin karşıtı olduğundan dolayı benzer olmamalarından, fakat aralarında zıtlık ilişkisi olmasından dolayı

birbirleriyle ilintili olmalarından kaynaklanacaktır. Bu durumda da iki cümle arasındaki ilintililiğin otomatik olarak ölçülmesi durumunda, iki cümle arasındaki ilintililik değeri yüksek çıkacak ve buna bağlı olarak da iki cümle arasındaki bağdaşıklık yüksek olduğu tahmin edilecektir. Bu örnekte de görüldüğü gibi, iki kavramın ilintili olma durumu, benzer olma durumunu kapsadığı ve buna ek olarak daha fazla sayıda ilişkiyi tanım gereği bünyesinde barındırdığı için, bağdaşıklık değerlendirmesi için daha doğru sonuçlar vereceği düşünülmektedir. Verilen örnek, Çizelge 3.3'te özetlenmiştir.

Çizelge 3.3: Benzerlik ve ilintililik ölçümlerinin bağdaşıklık tahmin başarısı için iki senaryo.

Senaryo	1. Cümle	2. Cümle	Benzerlik	İlintililik	Bağdaşıklık Tahmini	Sonucun Doğruluğu
1	Bugün İzmir'de hava güzeldi.	Benzer şekilde Ankara'da da hava güzeldi.	<i>Yüksek</i>	<i>Yüksek</i> (Çünkü benzer olan kavramlar aynı zamanda ilintilidir)	Benzerlik ölçümü ile: <i>Yüksek</i> İlintililik ölçümü ile: <i>Yüksek</i>	Benzerlik ölçümü ile: <i>Sonuç DOĞRU</i> İlintililik ölçümü ile: <i>Sonuç DOĞRU</i>
2	Ahmet seçimlerde A Partisi'ni destekliyordu.	Ama eşi A Partisi'ne karşıydı.	<i>Düşük</i>	<i>Yüksek</i> (Çünkü cümlelerin fiilleri arasında zıtlık ilişkisi var)	Benzerlik ölçümü ile: <i>Düşük</i> İlintililik ölçümü ile: <i>Yüksek</i>	Benzerlik ölçümü ile: <i>Sonuç YANLIŞ</i> İlintililik ölçümü ile: <i>Sonuç DOĞRU</i>

Anlamsal ilintililiğin ve benzerliğin ölçümü için çeşitli yöntemler önerilmiştir. Bir sonraki bölümde, iki kavram arasındaki ilintililiği ya da benzerliği ölçen yöntemler açıklanmıştır.

4 İLİNTİLİLİK VE BENZERLİK

İlntililik ve benzerlik, özellikle yapay zeka ve psikoloji alanlarında çok uzun zamandan beri üzerinde çalışılan kavramlardır. Psikoloji açısından benzerlik, bilgi ve davranış teorilerinde temel bir rol oynar. Benzerlik, insanların nesneleri sınıflandırması, kavramlar oluşturmaları ve genellemeler yapmaları için gerekli temeli oluşturur (Tversky, 1977). İlntililik; ilgisi, ilişkisi, bağı olma durumunu ifade etmektedir. Benzerlik ise, nitelik, görünüş ve yapı bakımından bir başkasına benzeme veya ona eş olma durumunu ifade etmektedir. Benzerliğin, ilntililiğin özel bir durumu olduğu söylenebilir (Resnik, 1985). Benzer kavramlar, birbirleriyle ortak yönleri ile ilntilidir. Örneğin, “araba” ve “bisiklet” kavramları birbirine benzerdir ve birbiriyle ilntilidir. İlntililik, kavramlar arasında daha fazla tür ilişki türünü barındırır. Örneğin, “araba” ile “tekerlek” kavramları parça-bütün ilişkisi ile birbiriyle ilntilidir (Budnitsky and Graeme, 2001). “Sıcak” ve “soğuk” kavramları zıtlık ilişkisi ile birbiriyle ilntilidir (Budnitsky and Graeme, 2001). Fakat bu kavramlar benzer değildir.

4.1 İlntililik ve Benzerlik Ölçümü

Birçok alanda, ilntililik ve benzerlik ölçümü yapılarak farklı problemlere çözümler getirilmektedir. Kullanıldığı alana göre ilntililik ve benzerlik ölçümü birçok varlık arasında yapılabilir. Bu varlıklara örnek olarak; kelime, cümle, metin, kavram, ontoloji, metin verilebilir.

Anlamsal web’in gelişimiyle birlikte, sadece içeriğe değil web’deki servislere de daha akıllı erişim mümkün olmaya başlamıştır. Web servislerinin ontoloji tabanlı olarak tanımlanmasıyla birlikte, kullanıcıların ve etmenlerin otomatik olarak web servislerini keşfetmeleri, tetiklemeleri ve birleştirmeleri mümkün olmaktadır. Web servislerinin anlamsal olarak etiketlenebilmesi için bir hizmet ontolojisi olan OWL-S, World Wide Web konsorsiyumu tarafından önerilmiştir (Martin et al., 2004). Böylece web servislerinin keşfedilmesinin, tetiklenmesinin ve birleştirilmesinin otomatikleştirilmesi, ontoloji kavramları arasındaki benzerlik ölçümüne indirgenmiştir. Web servisleri için yapılan bu çalışmalar, şebeke servisleri için de geçerlidir.

“Kelime anlamı belirginleştirme”, birden fazla anlama sahip bir kelimenin, bir cümlede hangi anlamda kullanıldığının belirlenmesi işlemine denir. Kelime anlamının sağlıklı olarak belirginleştirilmesi, otomatik metin tercümesi, bir

sorguya göre arama yapma, soru cevaplama gibi işlemlerin sonuçlarının başarılı olabilmesi için bir ön koşuldur (Leacock and Chodorow, 1998). Leacock ve Chodorow, WordNet'teki (Fellbaum, 1998) anlamsal ilişkileri kullanarak kelimeler arasındaki anlamsal benzerlikleri hesaplamışlar ve bu sonuçları kelime anlamı belirginleştirme için kullanmışlardır (Leacock and Chodorow, 1998).

Bir bilgi ihtiyacını gidermek için, genellikle elektronik ortamda saklanan geniş bir koleksiyondan, yapısal olmayan bir materyalin (genellikle metin) bulunması işlemine bilgiye erişim denilmektedir (Manning, 2008). Bilgiye erişimin gerçekleşebilmesi için, kullanıcı tarafından verilen sorgunun, arama yapılan koleksiyondaki bilgi ile kıyaslanması gerekir. Bu kıyaslama işlemi de özünde bir benzerlik hesaplamasıdır. Bilgiye erişimin başarısını değerlendirmek için en çok kullanılan iki oran “kesin isabet” ve “erişim isabeti” oranlarıdır. Mao ve Chu, kesin isabet ve erişim isabeti oranlarını yükseltmek için, bir tıbbi bilgi kaynağı olan UMLS®'yi (Unified Medical Language System®) (United States National Library of Medicine) kullanarak kelimelerin benzerliklerini hesaplamışlardır (Mao and Chu, 2002).

World Wide Web'in (WWW) kullanım alanları genişledikçe WWW'deki varlıklar arasındaki benzerliklerin ölçülmesi daha da önemli duruma gelmektedir. Örneğin, iki farklı alışveriş sitesinde bulunan ürünün aynı olup olmadığı, sosyal ağ uygulamalarında benzer ilgi alanlarına sahip kullanıcıların bulunması ya da kullanıcının kendi ilgi alanlarına yakın kullanıcı gruplarını arayıp bulabilmesi, benzerlik ölçümünün kullanılabileceği problemlere örnek olarak verilebilir (Sahami, 2006).

World Wide Web (WWW) dağıtık ve heterojen bir sistemdir. Anlamsal web, WWW'deki verinin daha yapısal ifade edilmesiyle bilgisayarlara daha akıllı işlemler yaptırmayı amaçlar. Bu vizyonun merkezinde ontoloji kavramı bulunur. WWW'nin dağıtık ve heterojen yapısıyla uyumlu olarak oluşturulacak farklı ontolojilerin birbiriyle birleştirilmesi, hizalanması gibi işlemler, anlamsal web'in hayata geçmesiyle daha da önemli duruma gelecektir. Ontolojilerin birleştirilmesi ve hizalanması için, ontolojideki kavramlar ya da kavram kümeleri arasındaki benzerliklerin bulunması önemlidir. Bunun için farklı ontolojilerin kavramları arasındaki benzerliği ölçen çeşitli yöntemler öneren çalışmalar yapılmıştır (Euzenat and Valtchev, 2004).

Zhang et al., belge kümeleme kalitesinin ontolojilerin kullanımıyla birlikte arttığını belirtmişlerdir (Zhang et al., 2008). Zhang et al., ontoloji tabanlı benzerlik ölçüm yöntemlerini medikal alandaki belgeleri kümelemek için sınıamışlardır.

Otomatik kısa cevap puanlama, metin benzerlik ölçümünün uygulandığı özel alanlardan birisidir. Amaç, sorulara verilen kısa cevapların, bir ya da daha fazla doğru cevapla karşılaştırılarak otomatik olarak puanlanmasıdır. Verilen cevaplar ile doğru cevaplar arasındaki benzerliğin hesaplanması için benzerlik ölçme yöntemleri kullanılmaktadır (Mohler and Mihalcea, 2009).

İlintililik ve benzerlik ilişkilerinin biçimlendirilmesi ve ölçülmesi problemi üzerine yapay zeka ve psikoloji alanlarında uzun zamandır çok sayıda çalışma yapılmıştır (Budanitsky and Graeme, 2001). Bu çalışmalarda çok sayıda değişik yaklaşım önerilmiştir.

Mevcut ilintililik ve benzerlik ölçüm yöntemlerinin benimsediği iki temel yaklaşım bulunmaktadır. Yöntemler bu yaklaşımlardan birisini ya da her ikisini birden bünyelerinde barındırmaktadırlar. Bu yaklaşımlardan birincisi, kavramların ortak ve farklı yönleridir (Schwering, 2008). İki kavramın ortak yönleri ne kadar fazla, farklılıkları ise ne kadar az ise iki kavram o kadar benzerdir. Lin, benzerlik kavramının biçimsel tanımını üç önermeye dayandırmıştır (Lin, 1998).

- İki kavram arasındaki benzerlik, iki kavramın ortak yönlerine bağlıdır. İki kavramın ortak yönleri arttıkça iki kavram arasındaki benzerlik de artar.
- İki kavram arasındaki benzerlik, iki kavramın farklı yönlerine bağlıdır. İki kavramın farklı yönleri arttıkça iki kavram arasındaki benzerlik de azalır.
- İki kavram arasındaki azami benzerlik, iki kavram aynı olduğunda elde edilir ve bu durum iki kavramın ne kadar ortak yönü olduğundan bağımsızdır.

Yaklaşımlardan ikincisi ise anlamsal uzaklıktır (Schwering, 2008). İki kavram arasındaki anlamsal uzaklık arttıkça iki kavramın benzerliği azalır. İki kavram arasındaki uzaklık, kavramların temsil edilme biçimlerine göre farklılık

gösterir. Örneğin, kavramların geometrik bir uzayda bir nokta olarak temsil edilmesi durumunda, bu noktalar arasındaki uzaklık iki kavram arasındaki uzaklığı temsil edebilir. Bir çizge ya da ağaç veri yapısında birer düğüm olarak temsil edilen kavramların arasındaki anlamsal uzaklık ise düğümler arasında kenarlar üzerinden elde edilen patikanın uzunluğu ile temsil edilebilir.

İlntililik ve benzerlik ölçüm yöntemleri farklı araştırmacılar tarafından farklı şekillerde sınıflandırılmıştır. Bu çalışmada benzerlik ölçüm yöntemleri, patika tabanlı, bilgi içeriği tabanlı, özellik tabanlı, geometrik model tabanlı ve karma olmak üzere beş sınıfta incelenmiştir. Patika tabanlı benzerlik ölçüm yöntemleri, iki kavram arasındaki benzerliği, bir taksonomideki ya da çizgedeki iki kavram arasındaki düğüm ve kenar sayısına göre hesaplamaktadır. Bilgi içeriği tabanlı benzerlik ölçüm yöntemleri, kavramların bir derlemde geçme sıklığına bağlı olarak benzerlik hesaplaması yapmaktadır. Özellik tabanlı benzerlik ölçüm yöntemleri, kavramları, sahip oldukları özellikler ile ifade edip kavramların özellik kümelerini karşılaştırarak bir benzerlik değeri hesaplamaktadır. Geometrik model tabanlı benzerlik ölçüm yöntemleri, kavramları n boyutlu bir uzayda birer nokta olarak temsil ederek ve bu noktalar arasındaki uzaklığı çeşitli şekillerde hesaplayarak iki kavram arasındaki benzerliği tahmin etmektedir. Karma yöntemler ise ilk dört başlıkta sınıflanan yöntemlerden bazılarını birleştirerek bir benzerlik hesaplama yöntemi önermektedir.

4.1.1 Patika tabanlı ilintililik ve benzerlik ölçüm yöntemleri

Rada et al., iki kavram arasındaki uzaklığın, iki kavramı temsil eden düğümleri ayıran kenar sayısı ile doğru orantılı olacağı gözleminden yola çıkmıştır. Buna göre Rada et al., iki kavram arasındaki uzaklığın, kavramlar arasındaki en kısa patikanın uzunluğu cinsinden ifade edilebileceğini, (4.1) numaralı formülde görüldüğü gibi, belirtmiştir (Rada et al., 1989). Rada et al., kavramlar arasındaki uzaklığı hesaplarken sadece genelleme ve özelleştirme ilişkilerini dikkate almanın yeterli olacağını belirtmiştir.

$$\text{Uzaklık}_{\text{Rada}}(K_1, K_2) = K_1 \text{ ve } K_2 \text{ arasındaki en kısa patikanın uzunluğu} \quad (4.1)$$

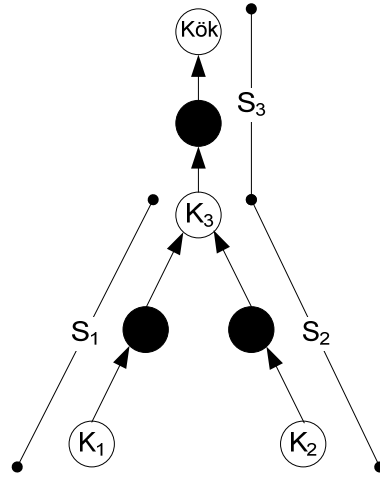
Rada et al., birden fazla kavramdan oluşan birleşik kavramların arasındaki uzaklığı da hesaplamak için bir yöntem önermiştir. Buna göre, atomik

kavramlardan (AK) oluşan K_1 ve K_2 arasındaki uzaklık (4.2) numaralı formülde gösterildiği şekilde hesaplanmaktadır.

$$\begin{aligned} & \text{Uzaklık}_{\text{Rada}}(AK1_1 \wedge \dots \wedge AK1_k, AK2_1 \wedge \dots \wedge AK2_m) \\ &= \frac{1}{k \times m} \sum_{i=1}^k \sum_{j=1}^m \text{Uzaklık}_{\text{Rada}}(AK1_i, AK2_j) \end{aligned} \quad (4.2)$$

Wu ve Palmer, otomatik çeviri yapabilmek için fiillerin anlamlarını hiyerarşiler yardımıyla temsil etmiştir (Wu and Palmer, 1994). Bir dilden diğer dile çeviri yapılırken, fiillerin karşılıkları, fiillerin hiyerarşilerdeki benzerliklerine göre seçilmektedir. Wu ve Palmer'a göre iki kavramın benzerliği, kavramların hiyerarşi içindeki yakınlıklarına bağlıdır. Buna göre Wu ve Palmer'ın önerdiği kavram benzerliği (4.3) numaralı formülde gösterilmiştir.

$$\text{Kavramsal Benzerlik}_{\text{Wu}}(K_1, K_2) = \frac{2 \times S_3}{S_1 + S_2 + 2 \times S_3} \quad (4.3)$$



Şekil 4.1: Wu ve Palmer'ın örnek kavram hiyerarşisi (Wu and Palmer, 1994).

Şekil 4.1'deki örnek kavram hiyerarşisine göre, K_3 kavramı, K_1 ve K_2 kavramlarının en yakın ortak ata kavramıdır. S_1 ve S_2 , sırasıyla K_1 ve K_2

kavramlarının K_3 kavramı ile arasındaki kavram yani düğüm sayısıdır. S_3 ise K_3 kavramı ile kök kavram arasındaki kavram sayısıdır (Wu and Palmer, 1994).

Slimani et al., Wu ve Palmer'ın formülü kullanıldığında bazı durumlarda kardeş kavramların, ata-çocuk ilişkisi olan kavramlara göre birbirlerine daha fazla benzediğini gözlemlemiştir ve bunun bilgi çıkarımı ilkeleriyle bağdaşmadığını belirtmiştir (Slimani et al., 2006). Buna göre benzerlik hesaplama yöntemini (4.4) numaralı formüldeki gibi değiştirmişlerdir.

$$\text{Benzerlik}_{\text{Slimani}} = \frac{2 \times S_3}{S_1 + S_2} + \text{CF}(K_1, K_2) \quad (4.4)$$

$\text{CF}(K_1, K_2)$ fonksiyonu, K_1 ve K_2 kardeş iki kavramın cezalandırma faktörüdür ve (4.5) numaralı formülde gösterilmiştir. Bu fonksiyon kullanılarak kardeş kavramların, ata-çocuk ilişkisi olan kavramlara göre birbirlerine daha çok benzeme olasılığı ortadan kaldırılmaktadır.

$$\text{CF}(K_1, K_2) = (1 - \lambda) \times \frac{1}{(\min(S_1, S_2) - S_3) + \lambda \times (|S_1 - S_2| + 1)} \quad (4.5)$$

(4.4) ve (4.5) numaralı formüllerdeki K_1 ve K_2 kavramları ile S_1 , S_2 , S_3 değişkenleri Şekil 4.1'deki örneğe uygundur. Cezalandırma faktöründeki λ katsayısı, iki kavram arasında ata-çocuk ilişkisi varsa 1 yoksa 0 değerini alacaktır. $\min(S_1, S_2)$, K_1 ve K_2 kavramlarının en yakın ortak kavramlarına olan uzaklıklarının en az değerini ifade eder (Slimani et al., 2006).

Leacock ve Chodorow, kelime anlamı belirginleştirme için kullanılan derlem tabanlı yöntemlere alternatif olarak WordNet kavramlar dizinindeki (Fellbaum, 1998) anlamsal ilişkileri kullanmayı önermişlerdir (Leacock and Chodorow, 1998). Bunun için Leacock ve Chodorow, WordNet kavramlar dizinindeki isim hiyerarşilerindeki en kısa patikaları kullanarak anlamsal benzerliği hesaplamıştır. Tüm sözcükler arasında bir patika bulabilmek için, her sözcüğün bulunduğu hiyerarşinin en tepesindeki düğüm, bir kök düğüme bağlanır. Böylece her sözcüğün hiyerarşi ağacı bir kök düğüm altında toplanmış olur ve her sözcük arasında bir patika bulunabilir. İki kavram arasındaki anlamsal benzerlik,

en kısa patika ve hiyerarşinin derinliği kullanılarak (4.6) numaralı formüle göre hesaplanır.

$$\text{Anlamsal Benzerlik}_{\text{Leacock}} = -\log\left(\frac{P(K_1, K_2)}{2D}\right) \quad (4.6)$$

P, K_1 ve K_2 kavramları arasındaki en kısa patikanın uzunluğunu düğüm sayısı cinsinden, D ise hiyerarşinin en fazla derinliğini temsil eder.

Mao ve Chu, bilgiye erişim başarısını yükseltmek için, bir tıbbi bilgi kaynağı olan UMLS[®]'yi (Unified Medical Language System[®]) (United States National Library of Medicine) kullanarak kelimelerin benzerliklerini hesaplamışlardır (Mao and Chu, 2002). Mao ve Chu, benzerlik hesaplaması için, kavramlar arasındaki en kısa patikanın uzunluğuna ve karşılaştırılan kavramların alt-kavramlarının sayısına bağlı olan (4.7) numaralı formülü önermiştir.

$$\text{Kavramsal Benzerlik}_{\text{Mao}}(K_1, K_2) = \frac{c}{P(K_1, K_2) \times \log_2(1 + A(K_1) + A(K_2))} \quad (4.7)$$

$P(K_1, K_2)$, K_1 ve K_2 kavramları arasındaki en kısa patikanın uzunluğunu, $A(K_1)$ ve $A(K_2)$ sırasıyla K_1 ve K_2 kavramlarının alt-kavramlarının sayısını yani kavramın genelliğini belirtir. c sayısının değeri K_1 ve K_2 kavramlarının arasındaki ata-çocuk ilişkisine göre belirlenir. Örneğin, K_1 kavramının K_2 kavramının tek atası ve K_2 kavramının da K_1 kavramının tek çocuğu olduğu bir durumda K_1 ve K_2 kavramları birbirine çok benzerdir ve bu durumda c sayısının değeri 1 ya da 1'e yakın bir değer olabilmektedir (Mao and Chu, 2002).

Nguyen ve Al-Mubaid, biyomedikal alanında kullanılmak üzere, (4.8) numaralı formülde gösterilen ontoloji tabanlı bir benzerlik formülü önermiştir (Nguyen and Al-Mubaid, 2006).

$$\begin{aligned}
\text{Benzerlik}_{\text{Nyugen}} &= \log_2([eky(K_1, K_2) - 1] \times [D - \text{derinlik}(\text{eyoa}(K_1, K_2))]) \\
&\quad + 2) \quad (4.8)
\end{aligned}$$

$eky(K_1, K_2)$, K_1 ve K_2 kavramları arasındaki en kısa yolu, $\text{derinlik}(\text{eyoa}(K_1, K_2))$, K_1 ve K_2 kavramlarının en yakın ortak atalarının (eyoa) hiyerarşideki derinliğini (kavram sayısı olarak), D ise taksonominin derinliğini ifade etmektedir.

Aydoğan et al., web’de önerilen hizmetler ile kullanıcı isteklerinin benzerliklerini ölçmek için ontolojideki kavramlar arasındaki benzerliklerden faydalanmışlardır (Aydoğan and Yolum, 2007). Aydoğan et al., taksonomideki iki kavramın benzerliğini ölçmek için RP benzerlik algoritmasını önermişlerdir. Bu algoritmayı önerirken aşağıdaki kabullerden yola çıkmışlardır:

- Bir kavramın bir seviye yukarıdaki atası, o kavrama, daha üst seviyedeki atasından daha benzerdir. Bir kavramdan daha genel kavramlara ilerledikçe o kavramdan uzaklaşılır. Kavramlar genelleştikçe birbirlerine daha az benzerler.
- Bir kavram, bir seviye yukarıdaki atasına, kendisine kardeş olan bir kavramdan daha fazla benzerdir.
- Bir kavram, kendine kardeş olan bir kavrama, bir seviyeden daha uzaktaki atalarından daha benzerdir.

Algoritma 4.1: RP benzerlik algoritması.

```

Benzerlik RP (K1, K2)
m > n > m2; m, n ∈ R[0, 1]

1: Benzerlik = 1
2: Eğer K1 ve K2 aynıysa
3: {
4:   Benzerlik değerini döndür
5: }
6: OrtakAta = OrtakAtayıBul(K1, K2)
   //Ortak ata K1 ve K2 kavramlarının en yakın ortak ata
   //kavramıdır.
7: N1 = UzaklığıHesapla(OrtakAta, K1)
8: N2 = UzaklığıHesapla(OrtakAta, K2)
   //N1 ve N2 kavramlar arasındaki bağlantı sayısıdır.
9: Eğer (OrtakAta ve K1 aynıysa) veya (OrtakAta ve K2 aynıysa)
10: {
11:   Benzerlik = Benzerlik x m(N1 + N2)
12: }
13: Değilse
14: {
15:   Benzerlik = Benzerlik x n x m(N1 + N2 - 2)
16: }
17: Benzerlik değerini döndür

```

Yukarıda gösterilen algoritmada, K_1 ve K_2 kavramlarının benzerliğini ölçmek için K_1 kavramından K_2 kavramına doğru taksonomi üzerinde ilerlenir. K_1 kavramından K_2 kavramına doğru gidildikçe her düğümde kavramlar arasındaki benzerlik azalır. Fakat yazarlara göre benzerlik her düğümde aynı oranda azalmaz. Algoritmaya göre azalma katsayısı, bir kavramdan bir üst seviyedeki atasına giderken m , bir kavramdan kardeş kavrama giderken n , bir kavramdan iki üst seviyedeki ata kavramına giderken m^2 olacaktır. Yazarların kabullerine göre m 'nin ve n 'nin değerleri 0 ile 1 arasındadır ve $m > n > m^2$ eşitsizliği geçerli olacaktır.

Ge ve Qiu, anlamsal web servisleri ve anlamsal şebeke servisleri arasındaki benzerliği ölçmek için ontoloji kavramları arasındaki benzerliği hesaplayan bir yöntem önermişlerdir (Ge and Qiu, 2008). Ge ve Qiu, kavram benzerliğini, kavramlar arasındaki kalıtım ilişkilerine göre hesaplamaktadırlar. Buna göre iki kavramı girdi olarak alan ve çıktı olarak anlamsal benzerlik değerini veren dört adımdan oluşan bir algoritma önermişlerdir.

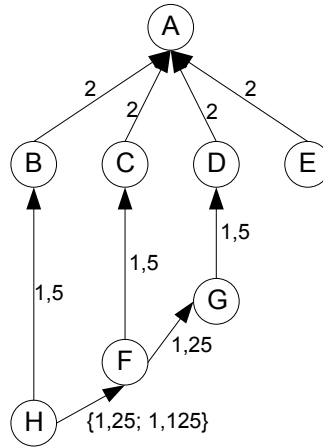
Birinci adım, ağırlık atama adıdır. Bu adımda, ontolojinin kök kavramı ile diğer kavramlar arasındaki patikaların ağırlıkları belirlenir. Ağırlık atama fonksiyonu olarak aşağıdaki (4.9) numaralı formül kullanılmıştır.

$$\text{Ağırlık}(K_1, K_2) = 1 + \frac{1}{\alpha^{\text{derinlik}(K_2)}} \quad (4.9)$$

Bu formülde, $\text{derinlik}(K)$ fonksiyonu, K kavramının ontolojinin kök kavramından uzaklığını, α , ontoloji hiyerarşisinde kök kavramdan uzaklaştıkça ağırlığın ne kadar azalacağını belirten katsayıyı belirtmektedir. Ge ve Qiu çalışmalarında α katsayısının değerini 2 olarak belirlemişlerdir.

Yazarlara göre (4.9) numaralı formülün taşıdığı iki önemli özellik vardır:

1. Üst seviye kavramlar arasındaki anlamsal fark, daha alt seviyedekilere göre daha yüksek olacaktır. Daha genel kavramlar daha özel kavramlara göre daha az benzerdir.
2. İki kardeş kavram arasındaki uzaklık, ata-çocuk arasındaki uzaklıktan daha fazladır.



Şekil 4.2: Patikaların ağırlıkları.

(4.9) numaralı formüle göre, ağırlıkların nasıl hesaplandığına ilişkin bir örnek Şekil 4.2’de verilmiştir. Şekilde de görülebileceği gibi bazı bağlantılar için birden fazla ağırlık değeri olabilir. Örneğin H ve F kavramları arasındaki ağırlık değeri için (H, F, C, A) patikası üzerinden $\text{Ağırlık}(H, F) = 1 + 1/22 = 1,25$ ve (H, F, G, D, A) patikası üzerinden $\text{Ağırlık}(H, F) = 1 + 1/23 = 1,125$ değerleri hesaplanabilir.

İkinci adımda düğüm rota çizelgesi oluşturulur. Şekil 4.2'deki örnek için düğüm rota çizelgesi, Çizelge 4.1'de gösterilmiştir.

Çizelge 4.1: Düğüm rota çizelgesi.

Düğüm	Rota
A	Kök kavram
B	$\{(B,A)(2)\}$
C	$\{(C,A)(2)\}$
D	$\{(D,A)(2)\}$
E	$\{(E,A)(2)\}$
F	$\{(F,C,A)(1,5; 2)\} \{(F,G,D,A)(1,25; 1,5; 2)\}$
G	$\{(G,D,A)(1,5; 2)\}$
H	$\{(H,B,A)(1,5; 2)\} \{(H,F,C,A)(1,25; 1,5; 2)\}$ $\{(H,F,G,D,A)(1,125; 1,25; 1,5; 2)\}$

Üçüncü adımda anlamsal uzaklık hesaplaması aşağıdaki algoritma kullanılarak yapılır.

Algoritma 4.2: Anlamsal uzaklık hesaplama algoritması.

```

Girdiler:  $K_1, K_2$  kavramları
Çıktı:  $Anlamsal\_Uzaklik(K_1, K_2)$ 
1: Eğer  $K_1$  ve  $K_2$  aynı kavramsa
2:   {
3:      $Anlamsal\_Uzaklik(K_1, K_2) = 0$ 
4:   }
5: Değilse Eğer  $K_1$  ve  $K_2$  arasında dolaysız bir patika varsa
6:   {
7:      $Anlamsal\_Uzaklik(K_1, K_2) = Ağırlık(K_1, K_2)$ 
8:   }
9: Değilse Eğer  $K_1$  ve  $K_2$  arasında dolaylı bir patika varsa
10:  {
11:     $Anlamsal\_Uzaklik(K_1, K_2) = \sum_{K \in KYol(K_1, K_2)} Ağırlık_K(K_1, K_2)$ 
12:  }
    //  $KYol$   $K_1$  ve  $K_2$  arasındaki en kısa patikayı belirtir
13: Değilse
14:   {
15:      $Anlamsal\_Uzaklik(K_1, K_2) =$ 
       $\min\{Anlamsal\_Uzaklik(K_1, K_0)\} + \min\{Anlamsal\_Uzaklik(K_2, K_0)\}$ 
16:   }

```

Dördüncü adımda anlamsal benzerlik hesaplaması yapılır. Anlamsal benzerlik ile anlamsal uzaklık arasında ters ilişki vardır. Buna göre yazarlar anlamsal benzerlik hesaplaması için (4.10) numaralı formülü kullanmaktadır.

$$\text{Anlamsal Benzerlik}_{Ge} = \frac{1}{p \times \text{Anlamsal_Uzaklık} + 1} \quad (4.10)$$

p ve Anlamsal_Uzaklık , anlamsal uzaklığın anlamsal benzerliğe olan etkisini belirler. p değeri deneyde ya da konu alanı uzmanları tarafından belirlenir.

Strube ve Ponzetto, WordNet (Fellbaum, 1998) için geliştirilmiş ilintililik ölçüm yöntemlerini, Vikipedi’de kullanıcılar tarafından kavramlara atanan kategorileri kullanarak sınımlardır (Strube and Ponzetto, 2006). Vikipedi’deki kategoriler, kendi aralarında birbirleriyle ilişkilendirildikleri için bir hiyerarşik çizge oluştururlar. Dolayısıyla Vikipedi’deki kavramlar arasında kategori çizgesi kullanılarak bir patika bulunabilir. İki kavram arasındaki ilintililik, iki kavram arasındaki patika kullanılarak hesaplanır. Strube ve Ponzetto, Resnik’in (Resnik, 1995), Wu ve Palmer’in (Wu and Palmer, 1994), Leacock ve Chodorow’un (Leacock and Chodorow, 1998) ilintililik ölçüm yöntemlerini WordSimilarity-353 test kümesi (Finkelstein et al., 2002) üzerinde test etmişler ve Vikipedi kullanımıyla birlikte ilintililik ölçümlerinin insan ölçümleriyle daha uyumlu olduğunu gözlemlemişlerdir (Strube and Ponzetto, 2006). Strube ve Ponzetto, bu sonucu Vikipedi’nin kapsamının genişliği ile açıklamaktadır.

Wang et al., gen ontolojisindeki⁵ kavramların benzerliklerini ölçmek için bir yöntem önermişlerdir (Wang et al., 2007). Bu yöntemde bir kavramın anlamı, o kavramın gen ontolojisi çizgesindeki yeri ve tüm ata kavramlarıyla olan ilişkisi tarafından belirlenmektedir. Gen ontolojisindeki bir K kavramının anlamı, K kavramının yönlü döngüsüz çizgedeki tüm ata kavramları ve kendisi ile K kavramını ata kavramlarına bağlayan anlamsal ilişkiler cinsinden temsil edilebilir. Bir K kavramının anlamsal değeri, K kavramının ata kavramlarının K kavramına olan anlamsal katkılarıyla hesaplanmaktadır. K kavramına yakın olan ata kavramlar, uzak olanlara göre, K kavramının anlamına daha çok katkıda bulunmaktadır. Buna göre bir a kavramının K kavramına olan anlamsal katkısı (AK) (4.11) numaralı formülde gösterildiği gibi ifade edilmiştir.

⁵ <http://www.geneontology.org> (Erişim tarihi: 14.03.2010)

$$AK_K(K) = 1 \text{ (Bir kavramın kendisine anlamsal katkısı 1'dir)} \quad (4.11)$$

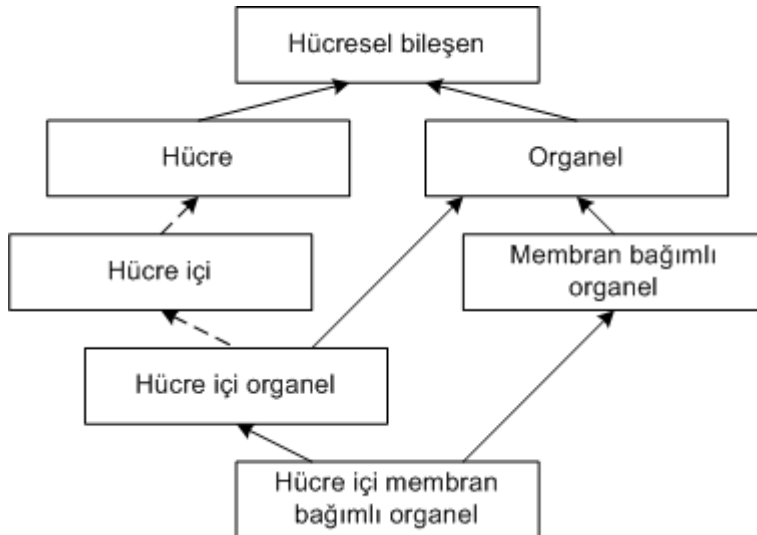
$$AK_K(a) = \max \{akf \times AK_K(a') \mid a' \in \text{çocukKavramlar}(a)\} \text{ eğer } a \neq K$$

Anlamsal katkı faktörü (akf), a kavramı ile a kavramının çocuk kavramlarından birisi olan a' arasındaki kenar içindir. akf, sıfır ile bir arasında bir değer alabilmektedir. Buna göre bir K kavramının anlamsal değeri (AD), (4.12) numaralı formülle hesaplanmaktadır.

$$AD(K) = \sum_{a \in \text{AtaK}_a} AK_K(a) \quad (4.12)$$

AtaK_a, a kavramının gen ontolojisindeki tüm ata kavramlarını ve kendisini kapsayan kümeyi ifade etmektedir. Buna göre, K kavramının anlamsal değeri, bu kavramının tüm ata kavramlarının kendisine olan anlamsal katkılarının toplamı olarak ifade edilmektedir.

Wang et al.'nin örnek olarak kullandığı gen ontolojisinden bir kesit Şekil 4.3'te gösterilmiştir.



Şekil 4.3: Gen ontolojisindeki “Hücre içi membran bağımlı organel” kavramının yönlü döngüsüz çizgesi.

Bu örneğe göre, gen ontolojisindeki özellik-genellik ilişkisinin anlamsal katkı faktörü 0,8, parça bütün ilişkisinin anlamsal katkı faktörü 0,6 olarak kabul edildiğinde, “Hücre içi membran bağımlı organel” kavramının yönlü döngüsüz çizgesindeki kavramların anlamsal katkı değerleri Çizelge 4.2’de, “Hücre içi organel” kavramının yönlü döngüsüz çizgesindeki kavramların anlamsal katkı değerleri ise Çizelge 4.3’te gösterilmiştir.

Çizelge 4.2: “Hücre içi membran bağımlı organel” kavramının yönlü döngüsüz çizgesindeki kavramların anlamsal katkı değerleri.

Kavram	Anlamsal Katkı Değeri
Hücre içi membran bağımlı organel	1,000
Hücre içi organel	0,800
Membran bağımlı organel	0,800
Hücre içi	0,480
Hücre	0,288
Organel	0,640
Hücresel bileşen	0,512

Çizelge 4.3: “Hücre içi organel” kavramının yönlü döngüsüz çizgesindeki kavramların anlamsal katkı değerleri.

Kavram	Anlamsal Katkı Değeri
Hücre içi organel	1,000
Hücre içi	0,600
Hücre	0,360
Organel	0,800
Hücresel bileşen	0,640

Anlamsal katkı ve anlamsal değer hesaplamaları kullanılarak gen ontolojisindeki iki kavramın benzerliği (4.13) numaralı formülle ölçülmektedir.

$$\begin{aligned}
 & \text{Anlamsal Benzerlik}_{\text{Wang}}(K_1, K_2) \\
 &= \frac{\sum_{a \in \text{Ata}K_{K_1} \cap \text{Ata}K_{K_2}} (AK_{K_1}(a) + AK_{K_2}(a))}{AD(K_1) + AD(K_2)} \quad (4.13)
 \end{aligned}$$

(4.13) numaralı formüldeki $AK_{K_1}(a)$, K_1 kavramına göre a kavramının anlamsal katkısını ifade etmektedir. Bu formül, gen ontolojisindeki iki kavramın anlamsal benzerliğini, bu kavramların ontolojideki yerlerini ve kavramların ata kavramlarıyla olan anlamsal ilişkilerini göz önüne alarak hesaplamaktadır. Şekil 4.3’teki örnek ontoloji kesitine göre “hücre içi membran bağımlı organel” ve

“hücre içi organel” kavramları arasındaki anlamsal benzerlik 0,7727 olarak hesaplanır.

4.1.2 Bilgi içeriği tabanlı ilintililik ve benzerlik ölçüm yöntemleri

Resnik, patika tabanlı ölçüm yöntemlerinde, kavram hiyerarşilerindeki bağlantıların, eşit uzaklıkları temsil ettikleri kabulünün yapıldığını belirtmiştir. Fakat gerçek hayatta bu uzaklıkların eşit olamayacağını düşünmüş ve “bilgi içeriği” kavramını ortaya atmıştır. Bir kavramın bilgi içeriği, o kavramın bir derlemde geçme sıklığına bağlıdır.

Resnik, kavram benzerliğinin iki kavramın sahip olduğu ortak bilgi içeriğine bağlı olduğunu belirtmiştir (Resnik, 1985). Ortak bilgi içeriği ise iki kavramın en yakın ortak ata kavramı ile ifade edilebilir. Resnik, bilgi içeriğini, bir kavramın bir metinde geçme olasılığı ile ifade etmektedir. Bir kavramın bir metinde geçme olasılığı (4.14) numaralı formülle hesaplanmaktadır.

$$\text{Olasılık}(K) = \frac{\text{Frekans}(K)}{N} \quad (4.14)$$

(4.14) numaralı formülde, N metinde geçen toplam isim sayısını ifade etmektedir. Kavramın frekansı ise o kavramın ve alt-kavramlarının metinde geçme sıklığını ifade etmektedir. Frekans hesaplaması (4.15) numaralı formülde gösterilmiştir.

$$\text{Frekans}(K) = \sum_{n \in \text{alt-kavramlar}(K)} n \text{ kelimesinin metinde geçme sayısı} \quad (4.15)$$

Olasılık(K) fonksiyonu taksonomide üst seviyelere çıktıkça artan bir fonksiyondur. Taksonominin kök kavramının fonksiyon değeri 1 olacaktır. Resnik, formülünün başarısını sınıadığı deneyde, taksonomideki kavramların olasılık değerlerini hesaplamak için Brown Derlemi’ni (Francis and Kucera, 1964) kullanmıştır. Olasılık fonksiyonu ile bilgi içeriği (Bİ) arasındaki ilişki ise (4.16) numaralı formülde gösterilmiştir.

$$B\dot{I}(K) = -\log \text{Olasılık}(K) \quad (4.16)$$

İki kavram arasındaki benzerlik ise (4.17) numaralı formülle ifade edilmiştir.

$$\text{Benzerlik}_{\text{Resnik}}(K_1, K_2) = \max_{k \in A(K_1, K_2)} [B\dot{I}(k)] \quad (4.17)$$

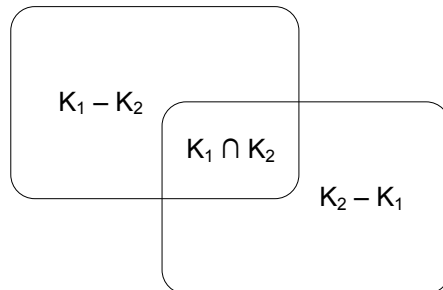
$A(K_1, K_2)$, K_1 ve K_2 kavramlarının ortak ata kavramlarının kümesidir.

Lin, iki kavramın içerdiği ortak bilgi içeriği yanında her kavramın bilgi içeriğini de anlamsal benzerlik hesaplaması için kullanmıştır (Lin, 1998). Lin'in benzerlik hesaplaması (4.18) numaralı formülde gösterilmiştir.

$$\text{Anlamsal Benzerlik}_{\text{Lin}}(K_1, K_2) = \frac{2 \times \max_{k \in A(K_1, K_2)} [B\dot{I}(k)]}{\log B\dot{I}(K_1) + \log B\dot{I}(K_2)} \quad (4.18)$$

4.1.3 Özellik tabanlı ilintililik ve benzerlik ölçüm yöntemleri

Patika tabanlı ve bilgi içeriği tabanlı ölçüm yöntemleri, kavramların özelliklerini dikkate almaz. Kavramların özellikleri, o kavram hakkında değerli bilgiler içerir. Özellik tabanlı yaklaşımlar nesneleri, özellikleri ile temsil ederler. Benzerlik ise bir özellik eşleme süreci olarak tanımlanır (Tversky, 1977).



Şekil 4.4: İki özellik kümesi arasındaki ilişki (Tversky, 1977).

Şekil 4.4'te özellik kümesi ile temsil edilen iki kavramın arasındaki ilişki grafiksel olarak gösterilmiştir. Bu şekle göre iki kavram arasındaki benzerlik üç değişkenli bir fonksiyon cinsinden (4.19) numaralı formüldeki gibi ifade edilebilir (Tversky, 1977).

$$\text{Benzerlik}_{\text{Tversky}} = F(K_1 \cap K_2, K_1 - K_2, K_2 - K_1) \quad (4.19)$$

Benzerlik fonksiyonu, 0 ve 1 arasında değer alacak şekilde normalize edilerek (4.20) numaralı formüldeki gibi yazılabilir.

$$\text{Benzerlik}_{\text{Tversky}} = \frac{f(K_1 \cap K_2)}{f(K_1 \cap K_2) + \alpha \times f(K_1 - K_2) + \beta \times f(K_2 - K_1)} \quad (4.20)$$

Andreasen et al.'nin özellik tabanlı benzerlik ölçüm formülü (4.21) numaralı formülde gösterilmiştir (Andreasen et al., 2003).

$$\begin{aligned} \text{Benzerlik}_{\text{Andreasen}}(K_1, K_2) &= p \times \frac{|A\ddot{O}(K_1) \cap A\ddot{O}(K_2)|}{|A\ddot{O}(K_1)|} + (1 - p) \\ &\times \frac{|A\ddot{O}(K_1) \cap A\ddot{O}(K_2)|}{|A\ddot{O}(K_2)|} \end{aligned} \quad (4.21)$$

$A\ddot{O}(K_1)$ ve $A\ddot{O}(K_2)$, sırasıyla K_1 ve K_2 kavramlarının ata kavramlarının özellik kümelerini temsil etmektedir.

d'Amato et al., betim mantığı ile (ALC⁶ kullanılarak) tanımlanmış bir bilgi tabanında kavramlar arasındaki benzerliği ölçmek için bir yöntem önermiştir (d'Amato et al., 2005). Bu yöntem, Tversky'nin benzerlik tanımının temel ilkelerine dayandırılmıştır:

⁶ ALC (Attributive Language with complements), betim mantığı dil ailesinin bir üyesidir.

1. İki kavramın ortak özellikleri, kavramlar arasındaki benzerliği artırır.
2. İki kavramın farklı özellikleri, kavramlar arasındaki benzerliği azaltır.
3. Kavramların ortak özelliklerinin benzerlik üzerindeki etkisi, aynı kavramların farklı özelliklerinin benzerlik üzerindeki etkisinden fazladır.

Buna göre d'Amato et al.'nin önerdiği anlamsal benzerlik hesaplama yöntemi (4.22) numaralı formülde gösterilmiştir.

$$\begin{aligned}
 \text{Benzerlik}_{d'Amato}(K_1, K_2) &= \frac{|(K_1 \cap K_2)^I|}{|K_1^I| + |K_2^I| - |(K_1 \cap K_2)^I|} \\
 &\times \max\left(\frac{|(K_1 \cap K_2)^I|}{|K_1^I|}, \frac{|(K_1 \cap K_2)^I|}{|K_2^I|}\right) \quad (4.22)
 \end{aligned}$$

(4.22) numaralı formüldeki K_1 ve K_2 , ALC mantığı ile ifade edilmiş bir içlemsel bilgideki (T-Box⁷) iki kavrama karşılık gelmektedir. K_1^I ve K_2^I ise sırasıyla K_1 ve K_2 kavramlarının kaplamasını (A-Box⁸'daki örnekleri) ifade etmektedir.

Çizelge 4.4'teki bilgi tabanı, iki kavramın benzerliğinin hesaplanması için d'Amato et al. tarafından verilen örnekten uyarlanmıştır (d'Amato et al., 2005).

⁷ Betim mantığı ile tanımlanmış bir bilgi tabanının iki bileşeninden birisidir. TBox, yani içlemsel bilgi, bir konu alanındaki kavramları (terminoloji) tanımlar.

⁸ Betim mantığı ile tanımlanmış bir bilgi tabanının iki bileşeninden birisidir. ABox, yani kaplamsal bilgi, içlemsel bilgi (TBox) ile ilgili bağlantılı örnek tanımlamalarını içerir.

Çizelge 4.4: Örnek bir bilgi tabanının içlemsel ve kaplamsal bilgi bileşenleri.

TBox: T = { Bayan $\equiv$ İnsan $\cap$ Kadın Bay $\equiv$ İnsan $\cap$ Erkek Ebeveyn $\equiv$ İnsan $\cap \exists$ ÇocuğuVar.İnsan Anne $\equiv$ Bayan $\cap$ Ebeveyn $\exists$ ÇocuğuVar.İnsan Baba $\equiv$ Bay $\cap$ Ebeveyn Çocuk $\equiv$ İnsan $\cap \exists$ EbeveyniVar.Ebeveyn BüyükEbeveyn $\equiv$ Ebeveyn $\cap \exists$ ÇocuğuVar.($\exists$ ÇocuğuVar.İnsan) Kardeş $\equiv$ Çocuk $\cap \exists$ EbeveyniVar.($\exists$ ÇocuğuVar ≥ 2) Yeğen $\equiv$ İnsan $\cap \exists$ BüyükEbeveyniVar.Ebeveyn $\cup \exists$ AmcasıVar.Amca Kuzen $\equiv$ Yeğen $\cap \exists$ AmcasıVar.($\exists$ ÇocuğuVar.İnsan) } 	
ABox: A = { Bayan(Kişi ₁), Bayan(Kişi ₂), Baba(Kişi ₃), Baba(Kişi ₄), Baba(Kişi ₅), Anne(Kişi ₆), Anne(Kişi ₇), Çocuk(Kişi ₈), Kardeş(Kişi ₉), Kardeş(Kişi ₁₀), EbeveyniVar(Kişi ₁ , Kişi ₇), EbeveyniVar(Kişi ₃ , Kişi ₅), EbeveyniVar(Kişi ₉ , Kişi ₆), EbeveyniVar(Kişi ₇ , Kişi ₄), EbeveyniVar(Kişi ₁₀ , Kişi ₅), EbeveyniVar(Kişi ₂ , Kişi ₇), EbeveyniVar(Kişi ₂ , Kişi ₃), EbeveyniVar(Kişi ₈ , Kişi ₆), EbeveyniVar(Kişi ₆ , Kişi ₄), KardeşiVar(Kişi ₃ , Kişi ₁₀), KardeşiVar(Kişi ₉ , Kişi ₈), KardeşiVar(Kişi ₇ , Kişi ₆), KardeşiVar(Kişi ₁₀ , Kişi ₃), KardeşiVar(Kişi ₂ , Kişi ₁), KardeşiVar(Kişi ₈ , Kişi ₉), ÇocuğuVar(Kişi ₃ , Kişi ₂), ÇocuğuVar(Kişi ₄ , Kişi ₇), ÇocuğuVar(Kişi ₄ , Kişi ₆), ÇocuğuVar(Kişi ₇ , Kişi ₂), ÇocuğuVar(Kişi ₇ , Kişi ₁), ÇocuğuVar(Kişi ₅ , Kişi ₁₀), ÇocuğuVar(Kişi ₅ , Kişi ₃), ÇocuğuVar(Kişi ₆ , Kişi ₈), AmcasıVar(Kişi ₉ , Kişi ₇), AmcasıVar(Kişi ₈ , Kişi ₇) } 	

d'Amato et al.'nin benzerlik formülüne göre “BüyükEbeveyn” ve “Baba” kavramları arasındaki benzerlik (4.23) numaralı formüldeki gibi hesaplanacaktır.

$$\begin{aligned}
& \text{Benzerlik}_{d'Amato}(\text{BüyükEbeveyn}, \text{Baba}) \\
&= \frac{|\text{BüyükEbeveyn} \cap \text{Baba}|}{|\text{BüyükEbeveyn}| + |\text{Baba}| - |\text{BüyükEbeveyn} \cap \text{Baba}|} \\
&\times \max\left(\frac{|\text{BüyükEbeveyn} \cap \text{Baba}|}{|\text{BüyükEbeveyn}|}, \frac{|\text{BüyükEbeveyn} \cap \text{Baba}|}{|\text{Baba}|}\right) \quad (4.23)
\end{aligned}$$

Çizelge 4.4'teki bilgi tabanına göre “BüyükEbeveyn” kavramının tanımına uyan örnekler Kişi₄ ve Kişi₅'tir. “Baba” kavramının tanımına uyan örnekler ise

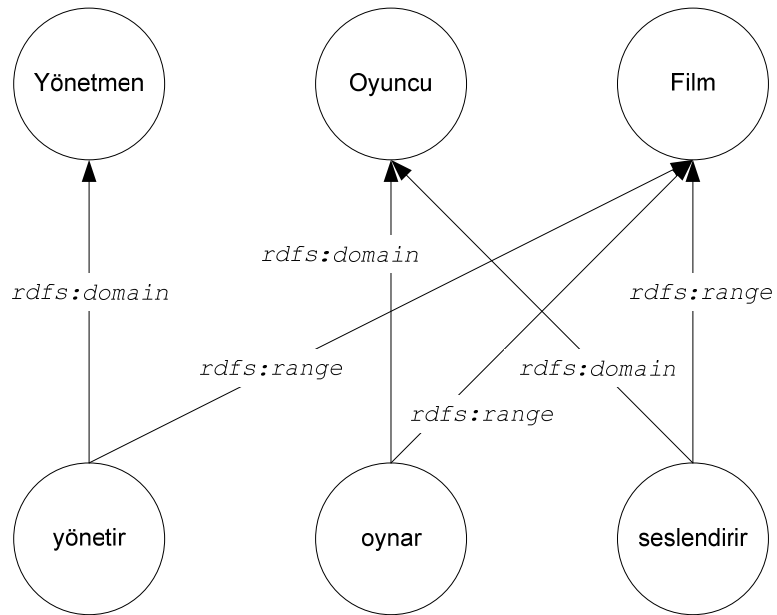
Kişi₃, Kişi₄ ve Kişi₅'tir. Buna göre $|BüyükEbeveyn \cap Baba|$ bileşenin değeri hem “BüyükEbeveyn” kavramının tanımına hem de “Baba” kavramının tanımına uyan örneklerin sayısı olacağı için 2’dir (Kişi₄ ve Kişi₅). $|BüyükEbeveyn|$ bileşenin değeri, “BüyükEbeveyn” kavramının tanımına uyan örneklerin sayısı, yani 2 (Kişi₄ ve Kişi₅) olacaktır. $|Baba|$ bileşenin değeri ise, “Baba” kavramının tanımına uyan örneklerin sayısı, yani 3 (Kişi₃, Kişi₄ ve Kişi₅) olacaktır. Bu değerlere göre “BüyükEbeveyn” ve “Baba” kavramları arasındaki benzerlik, (4.24) numaralı formülde gösterildiği gibi 0,67 olarak hesaplanacaktır.

$$\text{Benzerlik}_{d/Amato}(BüyükEbeveyn, Baba) = \frac{2}{2+3-2} \times \max(\frac{2}{2}, \frac{2}{3}) = 0,67 \quad (4.24)$$

4.1.4 Geometrik model tabanlı ilintililik ve benzerlik ölçüm yöntemleri

Geometrik modellerde, kavramlar n boyutlu bir uzayda birer nokta olarak temsil edilir. İki kavram arasındaki benzerlik, iki kavramı temsil eden noktalar arasındaki uzaklık ile ters orantılıdır.

Tous ve Delgado, anlamsal benzerliği hesaplamak için RDF ve OWL dil öğelerini anlamsal referans olarak kullanmışlardır (Tous and Delgado, 2006). Nesnelerin, diğer nesnelere bu dil öğeleri kullanılarak nasıl bağlandığı bir vektör uzayı ile temsil edilmiştir.



Şekil 4.5: Örnek bir ontoloji (Tous and Delgado, 2006).

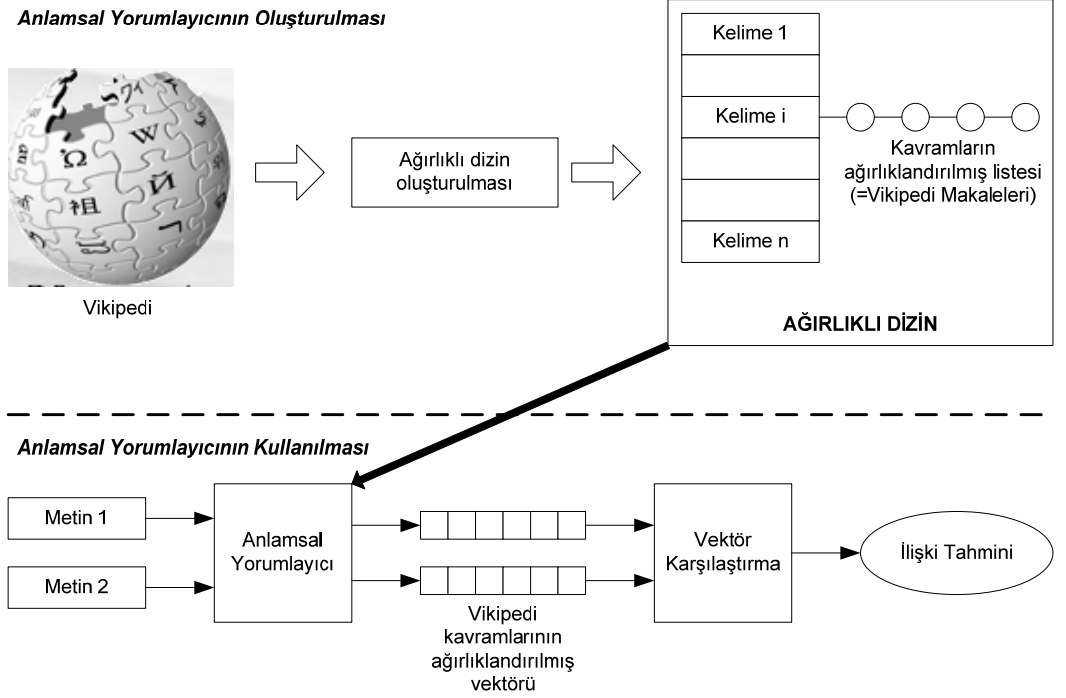
Şekil 4.5’te örnek bir ontoloji verilmiştir. Bu ontolojiye göre kullanılan ontoloji dil ögeleri `rdfs:domain` ve `rdfs:range`’dir. Nesneler arasındaki ilişkileri temsil etmek için iki boyutlu vektörler kullanılır. Bu vektörlerin ilk boyutu `rdfs:domain`, ikinci boyutu ise `rdfs:range` dil ögesini temsil eder. Buna göre “yönetir” özelliği ile “yönetmen” sınıfı arasındaki ilişki $\{1, 0\}$ vektörüyle, “oynar” özelliği ile “film” sınıfı arasındaki ilişki ise $\{0, 1\}$ vektörüyle temsil edilir. Bu ilişki vektörleri, bir başka vektöre yerleştirildiğinde aşağıdaki gibi üç boyutlu bir vektör elde edilir. Vektördeki satır sırası: “yönetmen”, “oyuncu”, “film”, “yönetir”, “oynar”, “seslendirir”.

$$A = \begin{pmatrix} (0,0) & (0,0) & (0,0) & (0,0) & (0,0) & (0,0) \\ (0,0) & (0,0) & (0,0) & (0,0) & (0,0) & (0,0) \\ (0,0) & (0,0) & (0,0) & (0,0) & (0,0) & (0,0) \\ (1,0) & (0,0) & (0,1) & (0,0) & (0,0) & (0,0) \\ (0,0) & (1,0) & (0,1) & (0,0) & (0,0) & (0,0) \\ (0,0) & (1,0) & (0,1) & (0,0) & (0,0) & (0,0) \end{pmatrix}$$

Kavramları temsil eden her bir satır yani vektör başka bir vektörle skaler olarak çarpıldığında ve çıkan değer normalize edildiğinde sonuç olarak bir benzerlik değeri elde edilir. Örneğin, “yönetir” ve “oynar” kavramlarını temsil eden vektörlerin skaler çarpımı sonucunda 1, “oynar” ve “seslendirir” kavramlarını temsil eden vektörlerin skaler çarpımı sonucunda 2 elde edilir.

Tous ve Delgado, benzerlik hesaplama yöntemini ontoloji hizalama alanında kullanmak için tasarlamışlardır. Yöntem, iki farklı ontolojideki kavramlar arasındaki benzerliği ya da yukarıdaki örnekteki gibi aynı ontolojideki farklı kavramlar arasındaki benzerliği tahmin edebilmektedir.

Gabrilovich ve Markovitch, bilgisayarların doğal dili işleyebilmeleri için alana özgü bilgiye erişimlerinin olması gerektiğini belirtmişlerdir. Buradan yola çıkarak metnin anlamını Vikipedi tabanlı kavramlar cinsinden temsil etmişlerdir (Gabrilovich and Markovitch, 2009).



Şekil 4.6: Bilgi-tabanlı anlamsal yorumlayıcı (Gabrilovich and Markovitch, 2009).

Şekil 4.6’da Gabrilovich ve Markovitch’in anlamsal yorumlayıcının oluşturulması ve bu yorumlayıcının kullanılması gösterilmiştir. Öncelikle Vikipedi kullanılarak kavramların ağırlıklandırılmış bir listesi oluşturulur. Daha sonra iki metni girdi olarak alan anlamsal yorumlayıcı, ağırlıklı dizini kullanarak metinlerin birer vektör temsilini oluşturur. Son adımda bu vektörler karşılaştırılarak iki metin arasındaki ilişki hakkında bir tahmin elde edilir.

Gabrilovich ve Markovitch, bir kelimenin anlamının iki yöntemle temsil edilebileceğini belirtmişlerdir (Gabrilovich and Markovitch, 2009). Birinci yöntem, insanların anlayabileceği sözlük anlamıdır. İkinci yöntem ise bilgisayarlar tarafından işlenebilecek, kelimenin bağlantılı olduğu diğer kavramlar ve bu bağlantıların gücü aracılığıyla anlam temsidir. Bu iki yönteme örnek vermek için Gabrilovich ve Markovitch “kedi” kelimesini kullanmışlardır. “Kedi” kelimesinin sözlük anlamı “dört ayaklı memeli bir hayvan”dır. Diğer taraftan “kedi” kelimesi, “kedigil” ve “evcil hayvan” kavramlarına güçlü olarak bağlı, “fare” ve “Tom ve Jerry” kavramlarına daha zayıf bağlıdır.

4.1.5 Karma ilintililik ve benzerlik ölçüm yöntemleri

Karma benzerlik ölçüm yöntemleri, patika, bilgi içeriği, özellik ve geometri tabanlı yöntemlerden bazılarını kullanarak elde ettikleri sonuçları değişik şekillerde birleştirilerek kavramlar arasındaki benzerliği ölçmektedir.

Jiang ve Conrath'ın anlamsal benzerlik ölçme yöntemi, patika tabanlı yaklaşımlarla bilgi içeriği tabanlı yaklaşımları birleştirmiştir (Jiang and Conrath, 1997). Jiang ve Conrath, bir kavram ile o kavramın ata kavramı arasındaki bağlantının gücünün, bir çocuk kavramının örneğinin bir metinde geçme olayının, o kavramın ata kavramının bir örneğinin metinde geçme olayına koşullu olasılığı⁹ olarak tanımlanır ve (4.25) numaralı formüldeki gibi ifade edilir.

$$P(K_c, K_a) = \frac{P(K_c \cap K_a)}{P(K_a)} = \frac{P(K_c)}{P(K_a)} \quad (4.25)$$

Yukarıdaki formülde, K_c 'nin tüm örnekleri aynı zamanda K_a 'nın örnekleri de olacağı için $P(K_c \cap K_a) = P(K_c)$ eşitliği geçerli olacaktır.

Çocuk ile ata kavramı arasındaki bağlantının gücü (BG), koşullu olasılığın negatif logaritması olarak (4.26) numaralı formüldeki gibi tanımlanmıştır.

$$\begin{aligned} BG(K_c, K_a) &= -\log(P(K_c|K_a)) \\ &= -(\log(P(K_c)) - \log(P(K_a))) = BI(K_c) - BI(K_a) \end{aligned} \quad (4.26)$$

Yukarıdaki formülde $BI(K)$, K kavramının bilgi içeriğini ifade etmektedir.

⁹ A olayı olduktan sonra B olayının olma olasılığı $P(B|A)$ olarak gösterilir. $P(B|A)$, B olayının, A olayına koşullu olasılığı olarak ifade edilir (İkiz vd., 1996).

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Jiang ve Conrath, bir çocuk kavram ile ata kavramı arasındaki bağlantının ağırlığını (BA), taksonominin yerel yoğunluğunu, düğüm derinliğini ve bağlantı tipini göz önüne alarak (4.27) numaralı formüldeki gibi ifade etmişlerdir.

$$BA(K_c, K_a) = \left(\beta + (1 - \beta) \frac{\bar{Y}}{Y(K_a)} \right) \left(\frac{d(K_a) + 1}{d(K_a)} \right)^\alpha [Bi(K_c) - Bi(K_a)] BTF(K_c, K_a) \quad (4.27)$$

Yukarıdaki formülde, $d(K_a)$, K_a kavramının taksonomideki derinliğini, $Y(K_a)$, K_a kavramının çocuklarının sayısını (yerel yoğunluğunu), $\bar{Y}$, taksonomideki ortalama yoğunluğu, $BTF(K_c, K_a)$ ise bağlantı tipi faktörünü ifade etmektedir. α ($\alpha \geq 0$) ve β ($0 \leq \beta \leq 1$) parametreleri ise sırasıyla düğüm derinliğinin ve yoğunluğun etkisini kontrol etmektedir.

Yukarıdaki tanımlara göre, Jiang ve Conrath iki kavram arasındaki anlamsal uzaklığı (4.28) numaralı formüldeki gibi ifade etmiştir.

$$\text{Anlamsal Uzaklık}_{\text{Jiang}}(K_1, K_2) = \sum_{\varphi \in \{\text{patika}(K_1, K_2) - \text{EYOA}(K_1, K_2)\}} BA(\varphi, \text{ataKavram}(\varphi)) \quad (4.28)$$

Yukarıdaki formülde, $\text{patika}(K_1, K_2)$, K_1 ve K_2 kavramları arasındaki en kısa patikadaki düğümleri, $\text{ataKavram}(\varphi)$, φ kavramının ata kavramını, $\text{EYOA}(K_1, K_2)$, K_1 ve K_2 kavramları arasındaki en kısa patikadaki en yakın ortak ata kavramı ifade etmektedir.

Karanastasi ve Christodoulakis, (4.29) numaralı formülde gösterilen OntoNL anlamsal ilintililik ölçümünü önermişlerdir (Karanastasi and Christodoulakis, 2007). Aşağıda gösterilen ilintililik formülü, üç alt formülden oluşmaktadır. İlk formül ortak özellikleri (oö), ikinci formül kavramsal uzaklık ölçümünü (kuö), üçüncü formül ise bağlantılı anlam ölçümünü (baö) göz önüne almaktadır.

$$\forall a_1 + a_2 + a_3 = 1, (a_1, a_2, a_3) > 0,$$

$$\text{İlintililik}_{o\ddot{o}}(K_1, K_2), \text{İlintililik}_{ku\ddot{o}}(K_1, K_2), \text{İlintililik}_{ba\ddot{o}}(K_1, K_2) \in [0,1]: \quad (4.29)$$

$$\begin{aligned} \text{İlintililik}_{Karanastasi} \\ = a_1 \times \text{İlintililik}_{o\ddot{o}} + a_2 \times \text{İlintililik}_{ku\ddot{o}} + a_3 \\ \times \text{İlintililik}_{ba\ddot{o}} \end{aligned}$$

Ortak özellik formülü OWL dilindeki `owl:ObjectProperty`, `owl:DatatypeProperty` ve `owl:inverseOf` dil öğelerini dikkate almaktadır. Ortak özellikler için formül (4.30) numaralı formülde gösterilmiştir. Bu formülde p_{ij} , K_1 ve K_2 kavramlarının birbirine bağlı olup olmadığını ifade eder. p_{i12} , K_1 ve K_2 kavramlarının K_i kavramına bağlı (`owl:ObjectProperty` ya da `owl:DatatypeProperty` ile) olup olmadığını ifade eder. $pters_{i12}$, K_1 ve K_2 kavramlarının K_i kavramına ters bağlı (`owl:inverseOf` ile) olup olmadığını ifade eder. f_a ve f_b ise kullanılan ontolojiye bağlı olan ağırlıklardır ve toplamaları bire eşittir.

$$\text{İlintililik}_{o\ddot{o}}(K_1, K_2) = \left(f_a \times \frac{\sum_{i=1}^n p_{i12}}{\sum_{i=1}^n p_{i1}} \right) + \left(f_b \times \frac{\sum_{i=1}^n pters_{i12}}{\sum_{i=1}^n p_{i12}} \right) \quad (4.30)$$

Kavramsal uzaklık ölçümü, kavramlar arasındaki patika uzunluğuna ve kavramların genelliğine bağlıdır. Patika uzunluğu (4.31) numaralı formülle ölçülmektedir.

$$\begin{aligned} \forall d_{K1} \geq 1, d_{K2} \geq 1, d_{K1} + d_{K2} > 2 : \text{PatikaUzunluğu}(K_1, K_2) \\ = \frac{d_{K1} + d_{K2}}{2 \times D} \in (0,1] \end{aligned} \quad (4.31)$$

d_{K1} ve d_{K2} , K_1 ve K_2 kavramlarının en yakın ortak ata kavramlarına olan uzaklığı, D ise ontolojinin en fazla derinliğini temsil etmektedir.

Ontolojideki kavramların genelliğini ölçmek için Karanastasi ve Christodoulakis, iki yöntem kullanmıştır. Birinci yöntemde, d_{K1} 'in $(d_{K1} + d_{K2}) / 2$ değerine yakın olması durumunda ilintililik azalır, çünkü K_1 kavramı en yakın ortak ata kavrama göre fazla özeldir. Bu yaklaşım (4.32) numaralı formülle temsil edilir.

$$\text{Genellik}_{K1} = \begin{cases} -\log \frac{2 \times d_{K1}}{d_{K1} + d_{K2}} \in (0,1], & \text{eğer } d_{K1} < \frac{d_{K1} + d_{K2}}{2} \\ 0 & , \text{eğer } d_{K1} \geq \frac{d_{K1} + d_{K2}}{2} \end{cases} \quad (4.32)$$

Genelliği ölçmek için kullanılan ikinci yöntemde ise K_1 kavramının ve ontolojideki K_1 kavramının bulunduğu hiyerarşinin en üst kavramın nesne özellik sayıları karşılaştırılır. Genellik hesaplamasında kullanılan ikinci formül (4.33)'te gösterilmiştir.

$$\forall N\ddot{O}_{K1} \leq 10 \times N\ddot{O}_{e\ddot{u}k} : \text{Genellik}_{K2} = 1 - \log \frac{N\ddot{O}_{K1}}{N\ddot{O}_{e\ddot{u}k}} \in [0, 1] \quad (4.33)$$

Yukarıdaki formülde, $N\ddot{O}_{K1}$, K_1 kavramının, $N\ddot{O}_{e\ddot{u}k}$, en üst kavramın nesne özellik sayılarını belirtmektedir.

Patika uzunluğu ve genellik formülleri kullanılarak kavramsal uzaklık ölçümü (4.34) numaralı formüldeki gibi ifade edilir.

$$\begin{aligned} & \text{İlintililik}_{ku\ddot{o}} \\ &= \frac{\text{Genellik}_{K1} + \text{Genellik}_{K2} + 1 - \text{PatikaUzunluğu}(K_1, K_2)}{3} \end{aligned} \quad (4.34)$$

Bağlantılı anlam ölçümü için WordNet kavramlar dizininden (Fellbaum, 1998) alınan isimlerin anlamları ve konu alanı ontolojilerinin `owl:label`, `owl:comment` dil öğelerinde bulunan açıklamalar kullanılır. Buna göre K_1 ve

K_2 kavramları arasındaki bağlantılı anlam ölçümü (4.35) numaralı formül ile yapılır.

$$\text{ilintililik}_{\text{bağ}}(K_1, K_2) = \frac{|A_1 \cap A_2|}{|A_1 \cap A_2| + |A_1 \setminus A_2|} \quad (4.35)$$

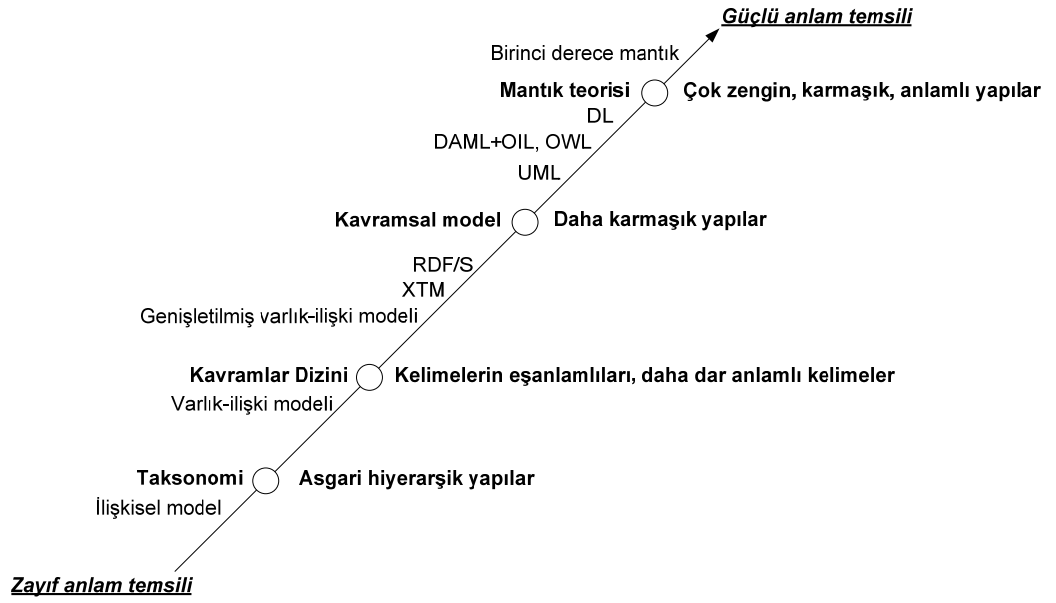
A_1 ve A_2 , K_1 ve K_2 kavramlarının WordNet kavramlar dizindeki anlamlarının kümesini temsil etmektedir.

4.2 İlintililik ve Benzerlik Ölçümünde Kullanılan Kaynaklar

Kavramlar arasındaki ilintililik ya da benzerlik ölçümü için çeşitli kaynaklar kullanılabilir. Bu kaynaklar, genel olarak yapısal ve yapısal olmayan kaynaklar olmak üzere iki grupta incelenebilir. Yapısal kaynaklar, verinin yanında bu veri hakkında veri (üst veri) ya da veriler arasında çeşitli ilişkiler barındıran kaynaklar olarak tanımlanabilir. Yapısal kaynaklara örnek olarak taksonomiler, kavram dizinleri ve ontolojiler verilebilir. Yapısal olmayan kaynaklar ise çeşitli yöntemlerle oluşturulmuş elektronik ortamda saklanan metinlerdir. Yapısal olmayan kaynaklara örnek olarak derlemeler verilebilir.

Yapısal bir kaynağın anlamsal temsil gücü, o kaynak üzerinde yapılabilecek işlemlerin niteliğini ve çeşitliliğini belirler. Örneğin, sadece genelleme ve özelleştirme ilişkilerinin bulunduğu bir taksonomiye göre eşitlik, zıtlık, bağlantı gibi ilişkileri ifade edebilen bir kavramlar dizininde daha nitelikli ve çeşitli işlemler yapılabilir.

Anlamsal temsil gücü, gerçek dünyadaki bilgiyi ne kadar gerçeğe yakın temsil ettiğinin bir ölçüsüdür. Şekil 4.7’de Daconta et al.’nin oluşturduğu, bilgi modellerini ve dilleri anlamsal zenginliklerine göre derecelendiren bir ontoloji yelpazesi gösterilmektedir (Daconta et al., 2003).



Şekil 4.7: Anlamsal temsil gücü yelpazesi (Daconta et al., 2003).

Şekil 4.7’de görüldüğü gibi ilişkisel bir veritabanı tasarımı için hazırlanmış varlık-ilişki modeli ile bir yazılım geliştirme sürecinde UML (Unified Modeling Language) ile oluşturulmuş bir sınıf diyagramının temsil güçleri birbirinden farklıdır. Yelpazede sol aşağıdan sağ yukarıya doğru gidildikçe temsil gücü artmaktadır. Böylece kavramlar hakkında daha fazla bilgi ve bu kavramlar arasında daha fazla ilişki temsil edilebilmektedir. Fakat temsil gücü arttıkça ihtiyaç duyulan çıkarsama zamanı da artmaktadır. Temsil gücü belirli bir sınırı aştığında yapılan çıkarsamaların mantıklı sınırlar içinde kalma olasılığı da azalmaktadır. Çünkü temsil gücü arttıkça, temsil edilen varlık sayısı arttıkça bilgisayarın yapması gereken hesaplama üssel olarak artar. Bu durum, eldeki güçlü temsil gücünün değerini yitirmesi anlamına gelir.

İlintililik ve benzerlik ölçümünde sıkça kullanılan taksonomiler, kavramları hiyerarşik olarak sınıflar. Bu sınıflandırma içinde özellik-genellik ilişkileri vardır.

Taksonomi için klasik örneklerden birisi, canlıların benzerlik, akrabalık, ya da köken ilişkisiyle sınıflandırıldığı biyolojideki taksonomilerdir. Şekil 4.8’de köpeğin (*Canis familiaris*) biyolojik taksonomideki yeri gösterilmiştir. Buna göre Carnivora bir çeşit Mammalia’dır. Taksonomide, Carnivora daha özel ve dar bir canlı grubunu, Mammalia ise daha genel ve geniş bir canlı grubunu temsil etmektedir. Bu ilişkiye göre her Carnivora’nın bir çeşit Mammalia olduğu sonucuna varabilirken, her Mammalia’nın bir çeşit Carnivora olduğu sonucuna varamayız.

Alem: Animalia (Hayvanlar)
Şube: Chordata
Sınıf: Mammalia
Takım: Carnivora
Aile: Canidae
Cins: Canis
Tür: Canis familiaris (Köpek)

Şekil 4.8: Köpeğin biyolojik taksonomideki yeri.

İlintililik ve benzerlik ölçümünde sıkça kullanılan bir başka kaynak da kavramlar dizinidir. Kavramlar dizini, kavramlar arasında eşitlik, zıtlık ve bağlantı gibi ilişkilerin tanımlanabildiği bir taksonomi olarak tanımlanabilir. Kavramlar dizini, bir taksonomideki genellik ve özellik hiyerarşisini ve eşitlik, zıtlık, bağlantı gibi ilişkileri ifade edebilmektedir. Çizelge 4.5’te kavramlar dizinindeki anlamsal ilişkiler, tanımları ve bu ilişkilere birer örnek gösterilmiştir (Daconta et al., 2003).

Çizelge 4.5: Kavramlar dizinindeki anlamsal ilişkiler (Daconta et al., 2003).

Anlamsal İlişki	Tanım	Örnek
Eşanlamlılık	X terimi, Y terimiyle eş ya da çok yakın anlama sahiptir.	“Daha” kelimesi “Henüz” kelimesiyle eşanlamlıdır.
Eşseslilik	X terimi, Y terimiyle aynı şekilde yazılır fakat farklı anlama sahiptir.	Doksan dokuzdan sonra gelen sayının adı olan “yüz”, Başta, alın, göz, burun, ağız, yanak ve çenenin bulunduğu ön bölüm anlamına gelen “yüz” ile eşseslidir.
Daha geniş	X teriminin anlamı Y teriminin anlamından daha geniştir.	“Canlı” kelimesi, “insan” kelimesinden daha geniş bir anlama sahiptir.
Daha dar	X teriminin anlamı Y teriminin anlamından daha dardır.	“İnsan” kelimesi, “canlı” kelimesinden daha dar bir anlama sahiptir.
Bağlantılı	X terimi, Y terimiyle bağlantılıdır.	“Çivi” kelimesi, “çekiç” kelimesiyle bağlantılıdır.

En çok bilinen kavramlar dizinlerinden birisi WordNet’tir (Fellbaum, 1998). WordNet, isimlerin, fiillerin, sıfatların ve zarfların eşanlamlı kümeler şeklinde gruplandığı ve bu gruplar arasındaki anlamsal ilişkilerin tanımlandığı İngilizce bir sözlüktür.

İnternet’in kullanım alanlarının genişlemesiyle ve kullanıcıların İnternet’in içeriğine daha fazla katkı sağlamasıyla birlikte içeriği kullanıcılar tarafından oluşturulan sitelerin sayısı ve önemi artmıştır. Vikipedi de bu oluşumlardan birisidir. Vikipedi, web tabanlı, içeriği kullanıcılar tarafından oluşturulan, çok dilli bir elektronik ansiklopedidir. Vikipedi’de birçok kavram hakkında yazılmış

makaleler bulunmaktadır. Ayrıca bu makaleler belirli sınıflara ayrılmış ve makaleler arasında bağlantılar kurulmuştur. Makalelerde geçen kavramlar üzerinde istatistiksel işlemler yapılarak ve makaleler arasındaki bağlantılar kullanılarak çeşitli benzerlik yöntemleri önerilebilir.

Yapısal olmayan bir kaynak olarak sınıflandırılabilen derlem, çok sayıda metinden ve kelimedenden oluşan, artık genellikle elektronik ortamda saklanan metin kümesi olarak tanımlanabilir. 1961 yılında Henry Kucera ve W. Nelson Francis tarafından Brown Üniversitesi'nde oluşturulan Brown Derlemi 15 farklı metin kategorisinden 1.014.312 kelime içermektedir. Bunun dışında İngiliz Ulusal Derlemi, Amerikan Ulusal Derlemi, Oxford İngilizce Derlemi gibi çok sayıda farklı dillerde farklı içeriklerde derlem bulunmaktadır. Derlemler üzerinde istatistiksel işlemler yapılarak benzerlik hesaplamaları yapılmaktadır.

4.3 Özet ve Değerlendirme

Patika tabanlı ilintililik ve benzerlik ölçüm yöntemleri, iki kavram arasındaki ilintililiği ya da benzerliği, bir taksonomideki ya da çizgedeki iki kavram arasındaki düğüm ve kenar sayısına göre hesaplamaktadır. Patika tabanlı ölçüm yöntemlerinin girdileri aşağıdaki listede özetlenmiştir:

- İki kavram arasındaki en kısa patikanın uzunluğu
- Kavramların en yakın ortak atalarına olan uzaklığı
- En yakın ortak atanın kök kavrama olan uzaklığı
- İki kavramın bulunduğu hiyerarşinin derinliği
- İki kavramın alt-kavramlarının sayısı
- İki kavramın kök kavrama olan uzaklığı
- İki kavramın ata kavramları

Patika tabanlı ilintililik ve benzerlik ölçüm yöntemleri, yukarıdaki listede de görüldüğü gibi, sadece özellik-genellik ilişkilerini dikkate almaktadır ve taksonomilerdeki kavramlar arasındaki ilintililiği ve benzerliği ölçmektedir. Wang

et al.'nin, gen ontolojisindeki kavramlar arasındaki benzerliği ölçtükleri yöntemindeki gibi bazı istisnai örneklerde ise özellik-genellik ilişkilerinin yanında parça-bütün ilişkileri de dikkate alınmaktadır (Wang et al., 2007). Bu çalışmada incelenen patika tabanlı ilintililik ve benzerlik yöntemleri Çizelge 4.6'da özetlenmiştir.

Çizelge 4.6: Patika tabanlı ilintililik ve benzerlik ölçüm yöntemleri.

Makale / Bildiri	Yöntemin Girdileri	Yöntemin Değerlendirmesi	Ölçümde Kullanılan Kaynak	Deney Sonucu
(Rada et al., 1989)	İki kavram arasındaki en kısa patikanın uzunluğu	- Kavramlar arasındaki sadece özellik-genellik ilişkileri göz önüne alınır. - Kavramlar arasındaki tüm patikalar değerlendirilir, en kısa olan patikaya göre kavramlar arasındaki uzaklık hesaplanır. - Hiyerarşideki daha özel kardeş kavramlar daha genel kardeş kavramlara göre birbirine daha çok benzemez.	Taksonomi	—
(Wu and Palmer, 1994)	- Kavramların en yakın ortak atalarına olan uzaklığı - En yakın ortak atanın kök kavrama olan uzaklığı	- Kavramlar arasındaki sadece özellik-genellik ilişkileri göz önüne alınır. - Hiyerarşideki daha özel kardeş kavramlar daha genel kardeş kavramlara göre birbirine daha çok benzerler.	Taksonomi	Miller ve Charles'ın 30 kelime çifti ile korelasyon 0,803 (Lin, 1998)
(Slimani et al., 2006)	- Kavramların en yakın ortak atalarına olan uzaklığı - En yakın ortak atanın kök kavrama olan uzaklığı - Hiyerarşideki iki kardeş kavramın cezalandırma faktörü	- Kavramlar arasındaki sadece özellik-genellik ilişkileri göz önüne alınır. - Hiyerarşide, ata-çocuk ilişkisi olan kavramlar, kardeş kavramlara göre birbirlerine daha çok benzerler.	Taksonomi	—
(Leacock and Chodorow, 1998)	- İki kavram arasındaki en kısa patikanın uzunluğu - İki kavramın bulunduğu hiyerarşinin derinliği	- Kavramlar arasındaki sadece özellik-genellik ilişkileri göz önüne alınır. - Tüm hiyerarşilerin kök düğümleri, bir seviye üstte tek bir kök düğüme bağlanır; böylece tüm kavramlar arasında bir patika olması sağlanır.	Taksonomi (WordNet'teki isim hiyerarşileri)	—

Makale / Bildirir	Yöntemin Girdileri	Yöntemin Değerlendirmesi	Ölçümde Kullanılan Kaynak	Deney Sonucu
(Mao and Chu, 2002)	- İki kavram arasındaki en kısa patikanın uzunluğu - İki kavramın alt-kavramlarının sayısı	- Kavramlar arasındaki sadece özellik-genellik ilişkileri göz önüne alınır. - İki kavram arasındaki benzerlik, kavramların arasındaki patikanın uzunluğuna ve kavramların genelliğine bağlıdır.	Taksonomi (UMLS®)	—
(Nguyen and Al-Mubaid, 2006)	- İki kavram arasındaki en kısa patikanın uzunluğu - En yakın ortak atanın kök kavrama olan uzaklığı - İki kavramın bulunduğu hiyerarşinin derinliği	- Kavramlar arasındaki sadece özellik-genellik ilişkileri göz önüne alınır. - İki kavram arasındaki benzerlik, kavramların arasındaki patikanın uzunluğuna, kavramların genelliğine ve iki kavramın en yakın ortak atalarına olan uzaklıklarına bağlıdır.	Taksonomi	Biyomedikal alanındaki 30 kavram çiftinin uzmanlar tarafından puanlanmış benzerlikleri ile korelasyon 0,836
(Aydoğan and Yolum, 2007)	- Kavramların en yakın ortak atalarına olan uzaklığı - m ve n parametreleri	- Kavramlar arasındaki sadece özellik-genellik ilişkileri göz önüne alınır. - Kavramlar genelleştikçe birbirlerine daha az benzerler. - Bir kavram, bir seviye yukarıdaki atasına, kendisine kardeş olan bir kavramdan daha fazla benzerdir. - Bir kavram, kendine kardeş olan bir kavrama, bir seviyeden daha uzaktaki atalarından daha benzerdir.	Taksonomi	—
(Ge and Qui, 2008)	- İki kavramın kök kavrama olan uzaklığı α parametresi	- Kavramlar arasındaki sadece özellik-genellik ilişkileri göz önüne alınır. - Hiyerarşideki daha özel kardeş kavramlar daha genel kardeş kavramlara göre birbirine daha çok benzerler. - Hiyerarşide, ata-çocuk ilişkisi olan kavramlar, kardeş kavramlara göre birbirlerine daha çok benzerler.	Taksonomi	—

Makale / Bildiri	Yöntemin Girdileri	Yöntemin Değerlendirmesi	Ölçümde Kullanılan Kaynak	Deney Sonucu
(Wang et al., 2007)	- İki kavramın ata kavramları - İki kavramın ata kavramlarıyla arasındaki anlamsal ilişkiler	- Kavramlar arasındaki özellik-genellik ilişkilerinin yanında parça bütün gibi başka ilişkiler de göz önüne alınır. Bu anlamsal ilişkiler için sıfırla bir arasında anlamsal katkı faktörü belirlenir. - Kavramların tüm ata kavramları göz önüne alınır (çoklu kalıtım ilişkileri).	Taksonomi	—

Bilgi içeriği tabanlı ilintililik ve benzerlik ölçüm yöntemleri, kavramların bir derlemde geçme sıklığına bağlı olarak ilintililik ya da benzerlik hesaplaması yapmaktadır. Bu tür yöntemler, patika tabanlı yöntemlerin, tüm özellik-genellik bağlantılarının eşit uzaklıkları temsil ettiği fikrine karşı çıkar. Kavramların benzerliklerinin hesaplanması için derlemlerden elde edilen istatistiksel bilgiler kullanılır. Bu çalışmada incelenen bilgi içeriği tabanlı ilintililik ve benzerlik yöntemleri Çizelge 4.7’de özetlenmiştir.

Çizelge 4.7: Bilgi içeriği tabanlı ilintililik ve benzerlik ölçüm yöntemleri.

Makale / Bildiri	Yöntemin Girdileri	Yöntemin Değerlendirmesi	Ölçümde Kullanılan Kaynak	Deney Sonucu
(Resnik, 1985)	Kavramın ve alt-kavramlarının metinde geçme sıklığı ve bilgi içeriği	- Hiyerarşilerdeki bağlantıların, eşit uzaklıkları temsil etmediğini kabul eder. - Kavram benzerliğinin iki kavramın sahip olduğu ortak bilgi içeriğine bağlı olduğunu belirtir. Bilgi içeriği, bir kavramın bir metinde geçme olasılığı ile ifade edilir. - Hesaplama için bir derlemden elde edilen istatistiksel bilgiler kullanılır.	Derlem	Miller ve Charles’ın 30 kelime çifti ile korelasyon 0,774 (Budanitsky and Graeme, 2001) Resnik’in, Miller ve Charles’ın deneyini tekrarladığı 30 kelime çifti ile korelasyon 0,9015 (Resnik, 1985)
(Lin, 1998)	Kavramın ve alt-kavramlarının metinde geçme sıklığı ve bilgi içeriği	Resnik’in tespitlerine ek olarak, kavram benzerliği, karşılaştırılan iki kavramın ortak bilgi içeriğine de bağlıdır.	Derlem	Miller ve Charles’ın 30 kelime çifti ile korelasyon 0,834

Özellik tabanlı ilintililik ve benzerlik ölçüm yöntemleri, kavramları, sahip oldukları özellikler ile ifade edip kavramların özellik kümelerini karşılaştırarak bir benzerlik değeri hesaplamaktadır. Kavramların benzer ve farklı özellikleri ile kavramların ata kavramlarının benzer özellikleri benzerlik ölçümü için kullanılır. Bu çalışmada incelenen özellik tabanlı ilintililik ve benzerlik yöntemleri Çizelge 4.8’de özetlenmiştir.

Çizelge 4.8: Özellik tabanlı ilintililik ve benzerlik ölçüm yöntemleri.

Makale / Bildiri	Yöntemin Girdileri	Yöntemin Değerlendirmesi	Ölçümde Kullanılan Kaynak	Deney Sonucu
(Tversky, 1977)	• Kavramların özellik kümelerinden elde edilen kesişim ve fark kümeleri • α ve β parametreleri	• Kavramların özellikleri, o kavram hakkında değerli bilgiler içerir. • Nesneler, özellikleri ile temsil edilir. • Benzerlik, bir özellik eşleme süreci olarak tanımlanır.	Taksonomi, Derlem	—
(Andreassen et al., 2003)	• Kavramların özellik kümelerinden elde edilen kesişim kümeleri • p parametresi	• Tversky’nin tespitlerine ek olarak, ortak ata kavramların da özellikleri dikkate alınır.	Taksonomi, Derlem	—
(d’Amato et al., 2005)	• Betim mantığı ile ifade edilmiş bir bilgi tabanındaki kavramların kaplamaları ve bu kaplamaların kesişimleri	• Tversky’nin benzerlik tanımının temel ilkelerine dayanır. • İki kavramın ortak özellikleri, kavramlar arasındaki benzerliği artırır. • İki kavramın farklı özellikleri, kavramlar arasındaki benzerliği azaltır. • Kavramların ortak özelliklerinin benzerlik üzerindeki etkisi, aynı kavramların farklı özelliklerinin benzerlik üzerindeki etkisinden fazladır.	Betim mantığı ile tanımlanmış ontoloji	—

Geometrik tabanlı benzerlik ölçüm yöntemleri, kavramları n boyutlu bir uzayda birer nokta olarak temsil ederek ve bu noktalar arasındaki uzaklığı çeşitli şekillerde hesaplayarak, iki kavram arasındaki benzerliği tahmin etmektedir. Bu yöntemlerde genellikle kavramlar bir kelime ya da kavram vektörü ile temsil edilir ve bu vektörler arasındaki kosinüs benzerliği hesaplanır. Bu çalışmada incelenen geometrik tabanlı ilintililik ve benzerlik yöntemleri Çizelge 4.9’da özetlenmiştir.

Çizelge 4.9: Geometrik tabanlı ilintililik ve benzerlik ölçüm yöntemleri.

Makale / Bildiri	Yöntemin Girdileri	Yöntemin Değerlendirmesi	Ölçümde Kullanılan Kaynak	Deney Sonucu
(Tous and Delgado, 2006)	- Kavramlar arasındaki RDF ve OWL dil öğeleri ile tanımlanmış ilişkiler - İlişkilerin yönleri	Tüm ilişkiler, anlamlarından bağımsız olarak, geometrik uzayda aynı sayı ile temsil edilir.	RDFS/OWL ile oluşturulmuş ontoloji	—
(Gabrilovich and Markovitch, 2009)	Kavramların Vikipedi’de ilişkili olduğu kavramlar	Kavramların anlamının temsil edildiği vektörler arasındaki kosinüs benzerliği hesaplanır.	Vikipedi	WordSimilarity -353 test kümesi ile korelasyon 0,72

Karma yöntemler, diğer yöntemlerden birkaçını birleştirerek ilintililik ya da benzerlik ölçümü yapmaktadır. Jiang ve Conrath, patika tabanlı yöntem ile bilgi içeriği tabanlı yöntemi birleştirmiştir (Jiang and Conrath, 1997). Karanastasi ve Christodoulakis ise patika tabanlı yöntemi, OWL öğelerinin anlamlarını dikkate alarak kullanmıştır (Karanastasi and Christodoulakis, 2007). Bu çalışmada incelenen karma ilintililik ve benzerlik yöntemleri Çizelge 4.10’da özetlenmiştir.

Elektronik ortamdaki veri hacmi arttıkça ve insanların bu verileri işleyerek daha karmaşık sonuçlara ulaşma ihtiyaçları ortaya çıktıkça, verinin daha yapısal ve daha zengin anlamsal öğelerle temsil edilmesi de bir zorunluluk haline gelmiştir. Bu çerçevede, daha zengin anlamsal temsil gücü olan bilgi temsil dilleri ortaya çıkmıştır. OWL, bu diller içinde en fazla bilinenlerden ve kullanılanlardan biridir. Bundan dolayı yeni yöntemlerden bazıları, OWL dil öğelerinin anlamlarını ilintililik ve benzerlik ölçümlerinde kullanmaktadır. Bu tez kapsamında da OWL dilinin öğeleri göz önüne alınarak, OWL ile geliştirilmiş bir ontolojideki iki kavram arasındaki anlamsal ilintililiğin ölçülmesi için bir yöntem önerilmiştir. Bu yöntem ve yöntemin geliştirilme süreci bir sonraki bölümde anlatılmıştır. Anlamsal ilintililik ölçümünde OWL öğelerinin anlamları göz önüne alındığı için Ek 2’de OWL dili tanıtılmıştır ve dil öğelerinin anlamları kısaca açıklanmıştır.

Çizelge 4.10: Karma ilintililik ve benzerlik ölçüm yöntemleri.

Makale / Bildiri	Yöntemin Girdileri	Yöntemin Değerlendirmesi	Ölçümde Kullanılan Kaynak	Deney Sonucu
(Jiang and Conrath, 1997)	• Kavramın ve alt-kavramlarının metinde geçme sıklığı ve bilgi içeriği • Kavramların taksonomideki derinlikleri • Kavramların yerel yoğunlukları (çocuk kavramlarının sayısı) • Taksonominin ortalama yoğunluğu • Bağlantı tip faktörü • Kavramlar arasındaki en kısa patika • α ve β parametreleri	• Patika tabanlı yaklaşımlarla bilgi içeriği tabanlı yaklaşımları birleştirir.	Derlem, Taksonomi	Miller ve Charles'ın 30 kelime çifti ile korelasyon 0,828
(Karanastasi and Christodoulakis, 2007)	• owl:ObjectProperty, owl:DatatypeProperty ve owl:inverseOf dil öğeleri ile tanımlanmış özellik kümeleri • Kavramların en yakın ortak atalarına olan uzaklığı • İki kavramın bulunduğu hiyerarşinin derinliği • İki kavram arasındaki en kısa patikanın uzunluğu • owl:label, owl:comment dil öğelerinde bulunan açıklamaların Wordnet'teki anlam kümeleri	• Patika tabanlı yaklaşımları, OWL öğelerinin anlamlarını da dikkate alarak kullanır.	OWL ile oluşturulmuş ontoloji	—

5 ONTOLOJİ TABANLI BİR ANLAMSAL İLİNTİLİLİK ÖLÇÜM YÖNTEMİ

Literatürde bulunan anlamsal ilintililik ve benzerlik ölçüm yöntemleri, genellikle kavramlar arasındaki çok az sayıda ilişki türünü dikkate almaktadır. Bu yöntemlerin en çok kullandığı ilişki türü ise özellik-genellik ilişkileridir. Anlamsal web vizyonu ile daha zengin anlamsal temsil gücüne sahip dillerle geliştirilen ontolojiler gittikçe yaygınlaşmaktadır. Bundan dolayı, ilintililik ve benzerlik ölçümlerini daha başarılı yapabilmek için yeni ontoloji dillerinin ögelerini dikkate almak mantıklı olacaktır.

OWL dili, ontoloji geliştirmek için W3C tarafından kullanılması önerilen bir standarttır. OWL dili, verinin anlamının bilgisayarlar tarafından işlenebilecek şekilde tanımlanabilmesini sağlar ve mantık temelli çıkarsama olanakları sağlar.

Bu bölümde OWL dilinin ögelerini kullanarak anlamsal ilintililik ölçümü yapan bir yöntemin geliştirilme süreci anlatılmaktadır.

5.1 Bilişsel Temel

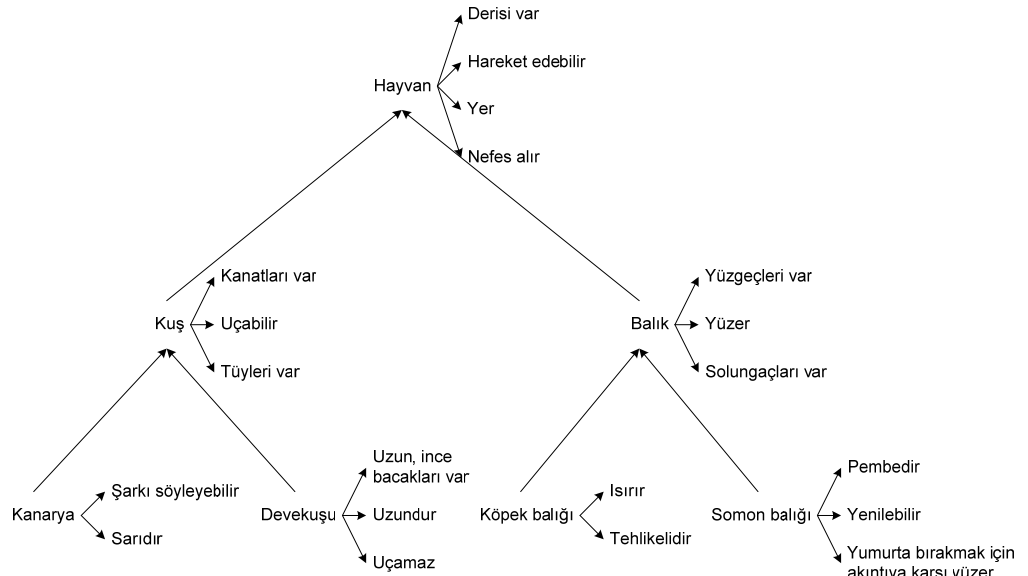
İnsan, çevresinden çeşitli şekillerde elde ettiği işlenmemiş bilgiyi düzenleyerek depolar. Bilgi (knowledge), işlenmemiş bilginin (information) düzenlenmiş, bütünleştirilmiş ve bellekte depolanmış halidir (Solso et al., 2005).

Felsefenin bir alanı olan bilgi felsefesi (epistemoloji), “bir şeyi bilmek” (bildirimsel bilgi) ile “nasıl yapılacağını bilmek” (işlemsel bilgi) arasında ayrım yapar (Solso et al., 2005). Bilişsel psikoloji alanında da, insanın uzun süreli belleğinde, bu iki tür bilginin temsil edildiği savunulmaktadır (Stevenson, 1993). Bildirimsel bilgi, kavramlar, bu kavramların örnekleri, olaylar gibi olguları kapsar. İşlemsel bilgi ise, iki sayıyı çarpmak ya da araba kullanmak gibi bir şeyin nasıl yapılacağını belirleyen bilgileri kapsar. Bu çalışma kapsamında, iki kavram arasındaki ilintililiğin ölçülmesi hedeflendiği için bildirimsel bilgi ve bu bilginin nasıl temsil edildiği konusu üzerinde durulacaktır.

Bilginin, insan belleğinde nasıl temsil edildiğini açıklayan bazı teoriler bulunmaktadır. Bu teorilere göre, bellekte depolanan bilgi, detaylı bilişsel yapılar şeklinde temsil edilir. Kümeleme modeli, küme kuramsal model, anlamsal özellik karşılaştırma modeli, ağ modeli (Solso et al., 2005) ve Nesne-Özellik-İlişki

modeli (Wang, 2007) bu modellerden bazılarıdır. Bu çalışmada, geliştirilen yöntemi etkilediği için, ağ modeli, bu modelin geliştirilmesiyle ortaya çıkan “etkinleştirmenin yayılması teorisi” (spreading activation theory) ve Nesne-Özellik-İlişki modeli üzerinde durulacaktır.

Anlamsal ağlar, bir kavrama ait bilgilerin birlikte depolandığı yönlü bir çizge olarak tanımlanabilir. Yönlü çizgedeki düğümler kavramları, düğümler arasındaki bağlantılar ise kavramlar arasındaki ilişkiyi temsil etmektedir (Stevenson, 1993). Bu modele göre her kavramın yeri, diğer kavramlarla olan ilişkisine göre belirlenir (Solso et al., 2005). Şekil 5.1’de örnek bir anlamsal ağ gösterilmiştir. Bu örneğe göre “kanarya” kavramı, “şarkı söyleyebilen sarı bir kuş” bilgisi ile birlikte depolanır. “Kanarya”, daha genel bir kavram olan “kuş” sınıfına ait bir kavramdır. “Kuş” kavramı, “kanatları var”, “uçabilir” ve “tüyleri var” bilgileriyle birlikte depolanır. Bu bilgilerin, her ayrı kuş türüyle birlikte depolanması gerekmez. “Kanarya uçabilir” önermesinin doğruluğunun denetlenmesi, “kanarya”nın “kuş” sınıfına ait bir kavram olması ve “kuş” kavramının “uçabilme” özelliğine sahip olması bilgisinin geri getirilmesi yoluyla gerçekleşir (Solso et al., 2005).



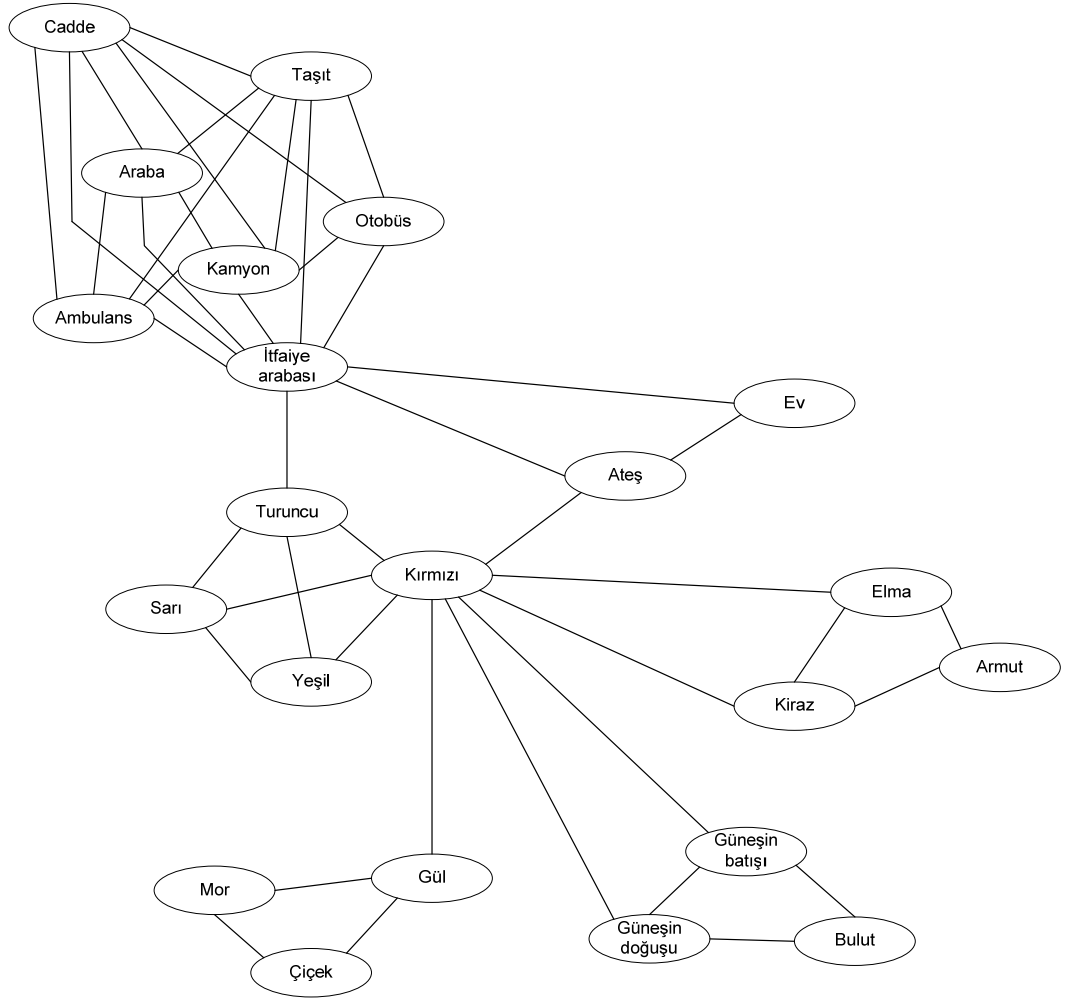
Şekil 5.1: Örnek bir anlamsal ağ (Collins and Quillian, 1969; Solso et al.’den, 2005).

Anlamsal ağlar, bilginin bellekte hiyerarşik bir yapıda ve özelliklerle birlikte depolandığını savunmuştur. Collins ve Loftus, bazı yönlerden eleştirilen bu hiyerarşik modeli geliştirerek “etkinleştirmenin yayılması teorisi”ni önermiştir (Collins and Loftus, 1975; Stevenson’dan, 1993).

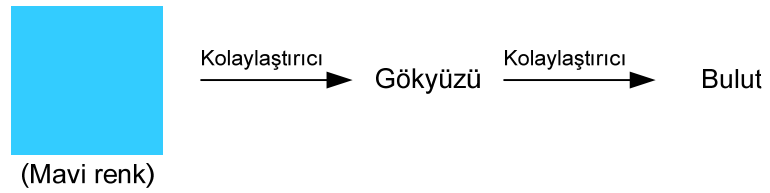
Bellekte depolanan bilginin geri getirilmesi, çevremizdeki uyaranları algıladıktan sonra onları, artalan bilgimizle bütünleştirerek anlamlandırmanın için önemlidir. Bilginin, bellekte bir ağ yapısında depolandığı ve geri getirme işleminin, ağ üzerinde “etkinleştirmenin yayılması” ile sağlandığı belirtilmiştir (Anderson, 1983). Collins ve Loftus tarafından geliştirilen “etkinleştirmenin yayılması teorisi”, karmaşık bir çağrışım ağı üzerine inşa edilmiştir (Collins and Loftus, 1975; Solso et al.’den, 2005).

Şekil 5.2’de örnek bir çağrışım ağı gösterilmiştir. Bu ağda, elipsler kavramları, çizgiler ise çağrışımları gösterir. Kavramlar arasındaki çağrışımın gücü, bağlantı çizgilerinin uzunluğuyla belirtilir. Buna göre, “kırmızı” ile “güneşin doğuşu” arasındaki gibi uzun çizgiler, bu iki kavram arasında daha zayıf bir çağrışım olduğunu gösterirken, “kırmızı” ile “ateş” arasındaki gibi daha kısa çizgiler ise bu iki kavram arasındaki çağrışımın daha güçlü olduğunu göstermektedir (Solso et al., 2005).

Collins ve Loftus’un modelinde, bir kavramın kullanımını takiben bir başka ilişkili kavramın hatırlanmasını ve geri getirilmesini açıklayan, kavramlar arasında yayılan bir etkinleştirme vardır. Örneğin, bir kişiye mavi renkli bir nesne gösterilirse, bu kişinin “mavi” kavramını hatırlaması, mavi renkli bir nesne gösterilmemesi durumundakine göre daha hızlı olacaktır. Ayrıca, mavi rengin görülmesi, “gökyüzü” gibi bu kavramla bağlantılı diğer kavramların tanınmasını kolaylaştıracaktır. Bu örnekte, “mavi” kavramı, “gökyüzü” kavramı için kolaylaştırıcı görevi yapar. “Gökyüzü” kavramı da hatırlandıktan sonra bu kavramın bağlantılı olduğu bir başka kavrama geçiş yapılabilir. Şekil 5.3’te gösterildiği gibi, “mavi” kavramı “gökyüzü” kavramının, “gökyüzü” kavramı da “bulut” kavramının hatırlanmasını kolaylaştırır (Solso et al., 2005).



Şekil 5.2: Anlamsal bilginin işlenmesinde etkinleştirmenin yayılması teorisi (Collins and Loftus, 1975; Solso et al.'den, 2005).



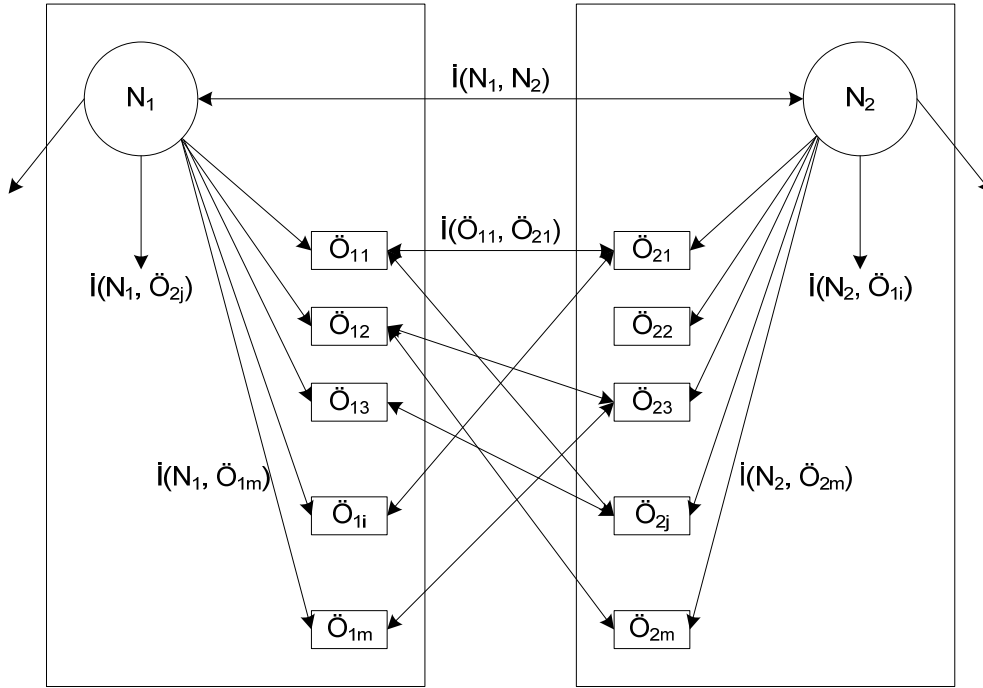
Şekil 5.3: “Mavi” kavramından yola çıkılarak “bulut” kavramının etkinleştirmenin yayılmasıyla hatırlanması (Solso et al., 2005).

Collins ve Loftus’un modeline göre, bellekteki ağlardaki her düğümün (yani kavramın) bir etkinleştirme seviyesi vardır. Bu etkinleştirme seviyesi düğümler arasında yayılır. Ağın bir bölümü belirli bir etkinleştirme seviyesini aşarsa, insan o bölgedeki kavramlar üzerinde yoğunlaşır. Etkinleştirmenin yayılması, bir düğümden bir başka düğüme geçtikçe azalır ve bir düğümün etkinleştirme seviyesi çok çabuk düşebilir. Bu iki varsayım, bir düşüncenin bir düğümü

etkinleştirilmesiyle birlikte bellekteki tüm kavram ağlarının etkinleştirilmesi gibi gerçek dışı bir durumun oluşmamasını açıklar. Ayrıca düğümlerin ve aradaki bağlantıların farklı etkinleştirme kapasitelerinin olduğu varsayılır. Buna göre, geçmişte çok fazla ve sık etkinleştirilmiş yüksek kapasiteli düğümler ve bağlantılar daha kolay etkinleştirilir (Stevenson, 1993).

Nesne-Özellik-İlişki modeli, bilginin, uzun süreli bellekte nasıl temsil edildiğini açıklayan teorilerden bir diğeridir (Wang, 2007). Nesne, bir kavramın ya da varlığın soyutlamasıdır. Özellik, nesnenin niteliklerini temsil eder. İlişki; nesne ile nesne ya da nesne ile ilişki ya da ilişki ile ilişki arasındaki bağlantıdır.

İki nesne için Nesne-Özellik-İlişki modeli Şekil 5.4'te gösterilmiştir. Şekilde, iki nesne arasındaki ilişki ($\dot{I}(N_1, N_2)$), iki özellik arasındaki ilişki ($\dot{I}(\ddot{O}_{11}, \ddot{O}_{21})$) ve nesne ile özellik arasındaki ilişki ($\dot{I}(N_1, \ddot{O}_{2j})$) için örnekler verilmiştir.



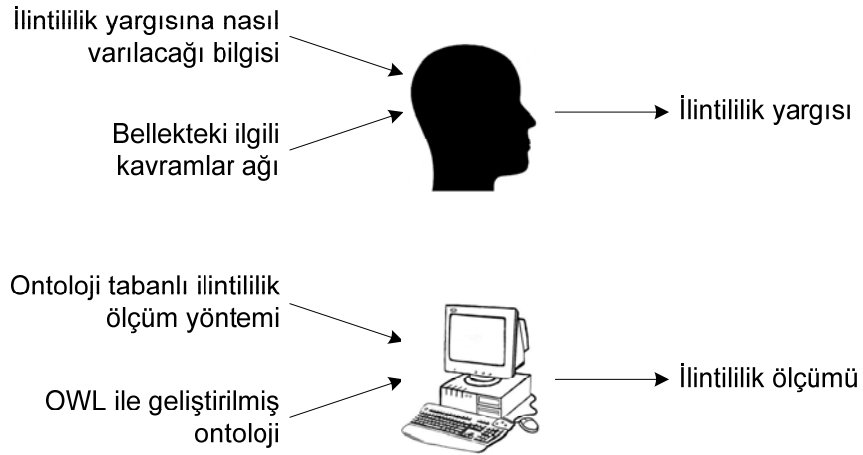
Şekil 5.4: Bellek mimarisinin Nesne-Özellik-İlişki modeli (Wang, 2007).

Özetle, bilginin temsilini açıklayan modellerdeki ana fikir, kavramların birbirleriyle bağlantılı şekilde bir ağ yapısında bellekte depolandığıdır ve kavramlar arasında çağrışım olduğudur (Solso et al., 2005).

5.2 Zihinsel Temsil ve Ontoloji

Anlamsal ilintililiğin ya da benzerliğin otomatik olarak ölçülebilmesi için çeşitli kaynaklardan faydalanılabilir (Bkz. Bölüm 4.2, Sayfa 56). Anlamsal web vizyonu çerçevesinde geliştirilen bir ontoloji temsil dili olan OWL, güçlü anlamsal temsil gücü ve kabul edilebilir zaman sınırları içindeki çıkarsama yetenekleriyle, çeşitli iş alanlarında ve araştırma çevrelerinde geniş kullanım alanları bulmaktadır. OWL ile geliştirilmiş bir ontoloji de anlamsal ilintililik ya da benzerlik ölçümü için bir kaynak olarak kullanılabilir.

Bu çalışma kapsamında, insanın iki kavram arasındaki ilintililik derecesini yargılama süreci, bilgisayar ortamında taklit edilmeye çalışılacaktır. Şekil 5.5'te de gösterildiği gibi, insan, ilintililik yargısına varmak için, bu işi nasıl yapacağı bilgisini (işlemsel bilgi) ve belleğindeki ilgili kavramlar ağını (bildirimsel bilgi) kullanır. Bu davranışı taklit etmek için, insanın ilintililik yargısına nasıl vardığını içeren bilgi yerine ontoloji tabanlı bir ilintililik ölçüm yöntemi kullanılacaktır. Bu ölçüm yöntemi, OWL dil öğelerinin ilintililik ölçümü için nasıl değerlendirileceğini ve ontolojideki iki kavram arasındaki ilintililiğin nasıl hesaplanacağını içermektedir. Diğer taraftan, insanın ilintililik yargısı için kullandığı kavramlar ağı yerine ise OWL ile geliştirilmiş ontolojiler kullanılacaktır.



Şekil 5.5: İnsanın ilintililik yargısı ve bilgisayar ortamında ilintililik ölçümü.

OWL ile geliştirilmiş ontolojilerin, insan belleğinde depolanan kavramsal ağları taklit etmek için kullanılması, OWL dil öğelerinin temsil yetenekleri ile bilginin bellekte nasıl temsil edildiğini açıklayan teoriler arasındaki örtüşmeler nedeniyle mantıklıdır. Bilginin bellek temsilini açıklayan teoriler, bilginin, ilişkili

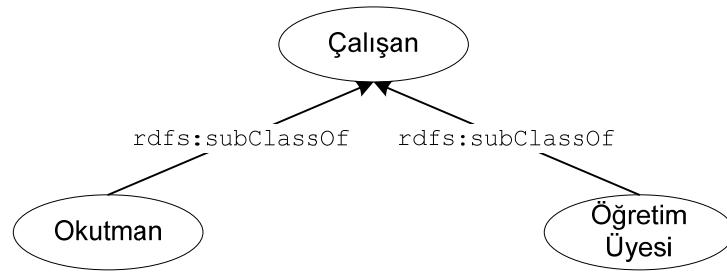
olduğu bilgilerle bağlantılı olarak bellekte depolandığını ortaya koymaktadır. Ayrıca, insan zihninin bilgiyi, nesneler, özellikler ve ilişkiler temelinde temsil ettiğini ortaya koyan teoriler de bulunmaktadır (Wang, 2007). Bu bilgiler ışığında, OWL dilinin, sınıf, nesne, özellik ve ilişki tanımlanmasına ve bilgiyi temsil etmek için bir çizge oluşturulmasına izin vermesinden dolayı, OWL ile oluşturulmuş bir ontolojinin, insan zihnindeki kavramsal ağın yerine geçmesi mantıklı görünmektedir.

5.3 Ontoloji Tabanlı Anlamsal İlintililik Ölçüm Yönteminin Geliştirilmesi

OWL ile oluşturulmuş bir ontolojideki iki kavram arasındaki ilintililiğin ölçülmesi için, insanın belleğinden bilgiyi geri getirmesini açıklayan “etkinleştirmenin yayılması teorisi”nden faydalanılmıştır. Bu teoriye göre, iki kavram arasındaki çağrışımın gücü, iki kavramın zihinde ne kadar yakın depolandığıyla doğru orantılıdır (Bkz. Bölüm 5.1, Sayfa 66 ve Şekil 5.2, Sayfa 69). Şekil 5.2’deki örneğe göre, “kırmızı” kavramını algılayan bir kişinin, “turuncu” ya da “ateş” kavramlarını belleğinden geri çağırabilmesi, “güneşin doğuşu” kavramını geri çağırabilmesine göre daha kolaydır ve daha yüksek olasılığa sahiptir. Bu yaklaşım, patika tabanlı ilintililik ve benzerlik ölçüm yöntemlerinin de temelini oluşturmaktadır. Rada et al., iki kavram arasındaki uzaklığın, iki kavramı temsil eden düğümleri ayıran kenar sayısı ile doğru orantılı olacağı gözleminden yola çıkmıştır. Buna göre Rada et al., iki kavram arasındaki uzaklığın, kavramlar arasındaki en kısa patikanın uzunluğu cinsinden ifade edilebileceğini belirtmiştir (Rada et al., 1989). Rada et al., kavramlar arasındaki uzaklığı hesaplarken sadece özellik-genellik ilişkilerini dikkate almanın yeterli olacağını belirtmiştir. Rada et al. gibi birçok araştırmacı, ilintililik ya da benzerlik tahmini yaparken, kavramların temsil edildiği kaynaktaki sadece özellik-genellik ilişkilerini göz önüne almışlardır (Wu and Palmer, 1994; Slimani et al., 2006; Leacock and Chodorow, 1998; Mao and Chu, 2002; Nguyen and Al-Mubaid, 2006; Aydoğan and Yolum, 2007; Ge and Qui, 2008). Wang et al. gibi bazı araştırmacılar ise, özellik-genellik ilişkilerinin yanında parça-bütün ilişkilerini de dikkate almışlardır (Wang et al., 2007). Bu çalışmanın çıkış noktası olarak, OWL dilinin zengin dil öğeleri göz önüne alınarak ilintililik ya da benzerlik ölçümü yapılması önerilmektedir. İlintililik ya da benzerlik ölçümü yapılırken iki kavram arasındaki hangi OWL dil öğelerinin dikkate alınması gerektiği bir sonraki bölümde anlatılmıştır.

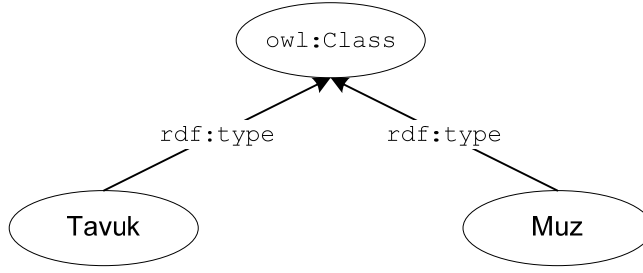
5.3.1 Ontolojik patika

Patika tabanlı ilintililik ve benzerlik ölçüm yöntemlerinin, iki kavram arasındaki ilintililiğin ve benzerliğin, iki kavram arasındaki patikanın yorumlanması ile ölçülebileceği fikri bu çalışmada geliştirilen yöntemin de temelini oluşturmaktadır. Bu yaklaşım, “etkinleştirmenin yayılması teorisi” tarafından da desteklenmektedir (Bkz. Bölüm 5.1, Sayfa 66 ve Şekil 5.2, Sayfa 69). Buna göre, OWL ile oluşturulmuş bir ontolojideki iki kavram arasındaki ilintililik, iki kavram arasındaki en kısa ontolojik patikanın yorumlanması ile hesaplanacaktır. Örneğin, Şekil 5.6’da gösterilen ontoloji kesitinde, “okutman” ile “öğretim üyesi” kavramları arasındaki patika, iki `rdfs:subClassOf` dil ögesinden oluşmaktadır. Dolayısıyla, “okutman” ile “öğretim üyesi” kavramları arasındaki ilintililik, `rdfs:subClassOf` dil ögelerinin yorumlanmasıyla hesaplanacaktır.



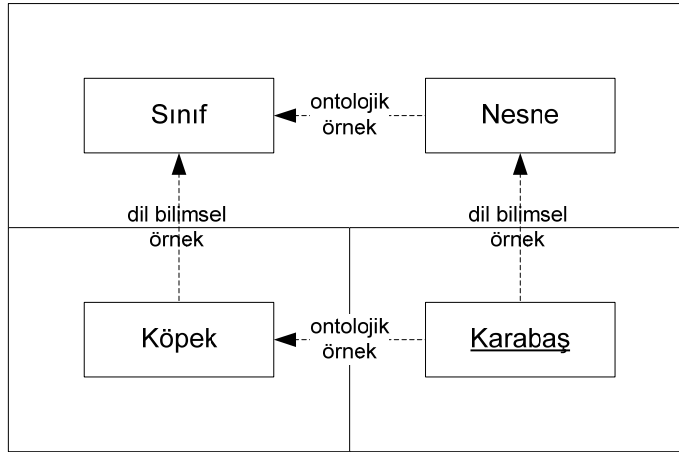
Şekil 5.6: OWL ile geliştirilmiş “üniversite” konu alanı ile ilgili bir ontolojinin kesiti.

Şekil 5.7’de, “yemek” konu alanı için OWL ile geliştirilmiş bir ontolojiden bir kesit gösterilmektedir. Bu örneğe göre, “tavuk” ve “muz” kavramları arasındaki ilintililik, iki `rdf:type` dil ögesinin yorumlanmasıyla elde edilecektir. Bu örnekte, bir önceki örnekten farklı olarak, “tavuk” ve “muz” kavramları arasındaki bağlantıyı sağlayan `owl:Class` ve `rdf:type` dil ögeleri, “yemek” konu alanında karşılığı olmayan kavramlardır. Bundan dolayı, “tavuk” ve “muz” kavramlarının birer “sınıf” olması, insanın ilintililik yargısı açısından bakıldığında bir anlam ifade etmemektedir. İki kavramın da birer “sınıf” olması, bu kavramlar arasında bir ilintililik bulunacağı anlamına gelmemektedir.



Şekil 5.7: İlintililik ölçümü için anlamsız bir patika örneği.

Atkinson ve Kühne, bilgi modellemede, kavramlar arasındaki ilişkileri, “dil bilimsel örnek” ve “ontolojik örnek” olmak üzere iki grup altında incelemiştir (Atkinson and Kühne, 2003). “Dil bilimsel örnek” ve “ontolojik örnek” ilişkilerinin gösterildiği örnek bir bilgi modeli kesiti Şekil 5.8’de gösterilmiştir.



Şekil 5.8: “Dil bilimsel örnek” ve “ontolojik örnek” ilişkilerinin gösterildiği örnek bir bilgi modeli kesiti.

Şekil 5.8’de gösterilen örnekte, “Karabaş” kavramı dil bilimsel açıdan bakıldığında bir nesne, fakat ontolojik açıdan bakıldığında ise bir “köpek”tir. Bu model temel alınarak, insan tarafından yapılacak bir ilintililik değerlendirmesinde, “Karabaş”ın bir köpek olduğu göz önüne alınacakken, “Karabaş”ın bilgi modelleme alanında bir “nesne” olduğu ise göz önüne alınmayacaktır. Aslında bilgi modelleme konusunda bilgi sahibi olmayan bir kişi, “sınıf” ve “nesne” kavramlarının bilgi modelleme bağlamındaki teknik anlamları konusunda fikir sahibi bile olmayacaktır.

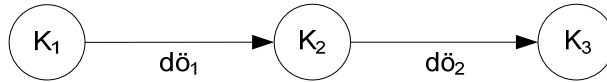
Özet olarak, insanın belleğindeki kavram ağlarını taklit etmek amacıyla bir ontoloji incelendiğinde, bu ontolojideki her tanımlamanın değerlendirmeye alınmaması gerektiği görülür. İnsanın ilintililik yargısı taklit edilmek istendiğinde,

ontoloji içindeki dil bilimsel ilişkiler anlamlı olmamaktadır. Ayrıca, ontolojideki, gerçek hayatta karşılığı olmayan ve dile özgü üst kavramlar olan “sınıf”, “nesne” gibi kavramlar da, ilintililik ölçümü için dikkate alınmamalıdır. Buna göre, bir patikanın ontolojik olması için, gerçek hayatta karşılığı olan kavramları ve ilişkileri içermesi gerekmektedir. Bu tez kapsamında, bu tip patikaları isimlendirmek için, Atkinson ve Kühne’nin çalışmasından (Atkinson and Kühne, 2003) yola çıkılarak “ontolojik patika” kullanılmıştır.

“Ontolojik patika” kavramı literatürde başka çalışmalarda da bulunmaktadır. Örneğin, OntoSearch bir ontolojideki iki kavram arasındaki en iyi ontolojik patikayı bulmaktadır. İki kavram arasındaki ilişki, aradaki patikanın uzunluğuna ve aradaki kenarların tipine bağlıdır (Onyshkevych, 1997; English and Nirenburg’dan, 2007). Bu tez çalışması kapsamında da Onyshkevych’in yaklaşımında olduğu gibi iki kavram arasındaki patika yorumlanmaktadır. Fakat bu tez çalışmasında farklı olan, bu patikalar bulunurken Atkinson ve Kühne’nin dil bilimsel ve ontolojik örnek ayırımından yola çıkılarak, sadece gerçek hayatta karşılığı olan ontolojik ilişkilerin “ontolojik patika”yı oluşturması gerektiğinin savunulmasıdır.

5.3.2 Ontolojik patika ile ilintililik ölçümü

İki kavram arasındaki anlamsal ilintililik, iki kavram arasındaki ontolojik patika üzerindeki ilişkilerin yorumlanması sonucunda elde edilecektir. Bu çalışma kapsamında, OWL ile oluşturulmuş ontolojiler kullanıldığı için, ontolojik patika üzerindeki ilişkiler OWL dil öğeleri ile ifade edilecektir. Şekil 5.9’da iki kavram arasındaki patika gösterilmiştir. K_1 ve K_3 kavramları, K_2 kavramı üzerinden $d\ddot{o}_1$ ve $d\ddot{o}_2$ “dil öğeleri” ($d\ddot{o}$) ile birbirine bağlanmıştır.

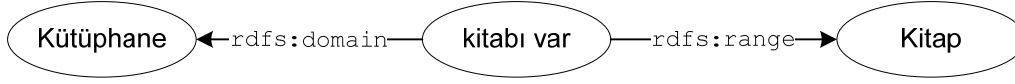


Şekil 5.9: K_1 ve K_3 kavramları arasındaki patika.

Şekil 5.9’daki örneğe göre, K_1 ve K_3 kavramları arasındaki anlamsal ilintililik, iki kavram arasındaki “dil öğeleri”nin ($d\ddot{o}$) “anlamsal ilintililik katsayıları” (aik) çarpılarak bulunacaktır. Buna göre K_1 ve K_3 kavramları arasındaki anlamsal ilintililik, (5.1) numaralı formülde gösterildiği gibi hesaplanacaktır.

$$\text{Anlamsal İlintililik } (K_1, K_3) = \text{aik}(d\ddot{o}_1) \times \text{aik}(d\ddot{o}_2) \quad (5.1)$$

Şekil 5.10’da OWL ile geliştirilmiş “kütüphane” konu alanı ile ilgili bir ontolojinin kesiti gösterilmiştir. Bu kesite göre “kitabı var” nesne özelliğinin konu alanı “kütüphane”, menzili ise “kitap” sınıfıdır.



Şekil 5.10: OWL ile geliştirilmiş “kütüphane” konu alanı ile ilgili bir ontolojinin kesiti.

Şekil 5.10’daki örneğe göre, “kütüphane” ve “kitap” kavramları arasındaki anlamsal ilintililik (5.2) numaralı formülde gösterildiği gibi hesaplanacaktır.

$$\text{Anlamsal İlintililik (kütüphane, kitap)} = \quad (5.2)$$

$$\text{aik}(\text{rdfs:domain}) \times \text{aik}(\text{rdfs:range})$$

Anlamsal ilintililik ve benzerlik ölçüm yöntemlerinde, kavramlar arasındaki ilişkilerin anlamına bağlı olarak, her ilişki türü için bir katsayı atama işlemi çeşitli çalışmalarda yapılmıştır. Örneğin Wang et al., özellik-genellik ilişkileri için anlamsal katkı faktörünü 0,8, parça-bütün ilişkileri için anlamsal katkı faktörünü 0,6 olarak belirlemiştir (Wang et al., 2007). Mazuel ve Sabouret, kavramlar arasındaki ilişkilerin gücüne göre birer katsayı belirlemiştir (Mazuel and Sabouret, 2008). Mazuel ve Sabouret, parça-bütün ilişkisi için ilişki gücü katsayısını 0,4 olarak belirlemiştir. Bu çalışmalardaki katsayılar, yapılan ilintililik ve benzerlik ölçümlerinin, çeşitli sonuçlarla kıyaslanarak, ölçüm yönteminin başarısı en fazla olacak şekilde belirlenmektedir. Örneğin, bir benzerlik yöntemindeki katsayılar, yöntem kullanılarak elde edilen sonuçlar ile insan yargısı ile belirlenmiş benzerlik değerleri arasındaki korelasyonun en fazla olacağı şekilde belirlenebilmektedir. Bu çalışmada, OWL dil öğelerinin anlamsal ilintililik katsayıları, insanların ilintililik yargılarına göre belirlenmiştir. Bu işlemde önce, OWL dil öğeleri ontolojik patika yaklaşımı açısından değerlendirilmiştir ve hangi dil öğelerinin anlamsal ilintililik katsayılarının belirleneceği tespit edilmiştir.

5.3.3 Anlamsal ilintililik ölçümü için OWL dil öğelerinin değerlendirilmesi

OWL dilinin öğeleri Çizelge 5.1’de listelenmiştir. Bu bölümde, OWL dil öğeleri, ontolojik patika yaklaşımı açısından incelenmiştir. Bu inceleme sonucunda, ontolojik patika üzerinde bulunabilecek OWL dil öğeleri belirlenmiştir. Daha sonra, bu dil öğelerinin kavramlar arasındaki ilintiliğe yaptıkları katkı sayısal olarak ifade edilmiştir.

Çizelge 5.1: OWL dil öğeleri.

Sınıf ile ilgili dil öğeleri	Birey ile ilgili dil öğeleri
•owl:Class •rdfs:subClassOf •owl:equivalentClass •owl:disjointWith •owl:unionOf •owl:complementOf •owl:intersectionOf •owl:oneOf, owl:DataRange •owl:DeprecatedClass	•owl:sameAs •owl:differentFrom •owl:AllDifferent •owl:distinctMembers
	Kısıtlama ile ilgili dil öğeleri
	•owl:Restriction •owl:onProperty •owl:allValuesFrom •owl:someValuesFrom •owl:minCardinality •owl:maxCardinality •owl:cardinality •owl:hasValue
Özellik ile ilgili dil öğeleri	Ontoloji ile ilgili dil öğeleri
•rdfs:subPropertyOf •rdfs:domain •rdfs:range •owl:ObjectProperty •owl:DatatypeProperty •owl:equivalentProperty •owl:inverseOf •owl:TransitiveProperty •owl:SymmetricProperty •owl:FunctionalProperty •owl:InverseFunctionalProperty •owl:DeprecatedProperty	•owl:Ontology •owl:imports •owl:versionInfo •owl:priorVersion •owl:backwardCompatibleWith •owl:incompatibleWith
	Etiketleme özellikleri
	•rdfs:label •rdfs:comment •rdfs:seeAlso •rdfs:isDefinedBy •owl:AnnotationProperty

owl:Class dil ögesi, ontolojideki bir kaynağın OWL dili açısından bir sınıf olduğunu belirtir. Benzer şekilde, owl:ObjectProperty ve owl:DatatypeProperty dil öğeleri ontolojideki bir kaynağın OWL dili açısından bir özellik olduğunu belirtir. Dolayısıyla bu dil öğeleri, ontolojideki kavramların OWL dilindeki tiplerini belirtmektedir ve ontolojik patika üzerinde bulunmamalıdır.

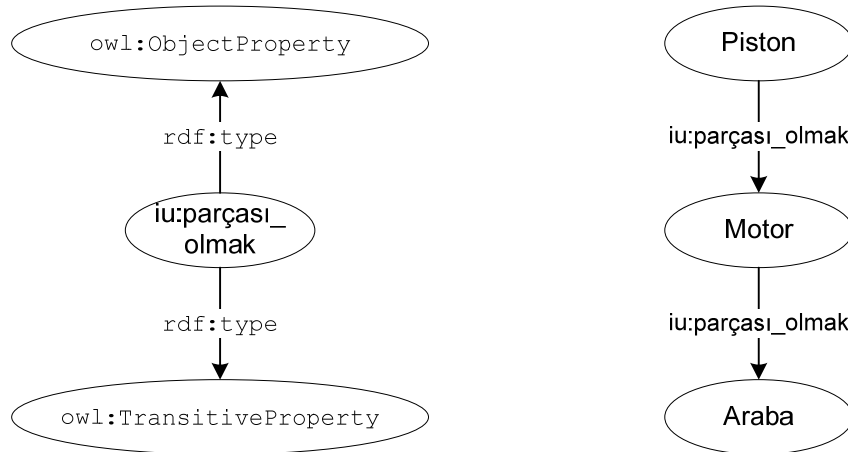
OWL dilinde taksonomi oluşturmak için `rdfs:subClassOf` dil ögesi kullanılır. Böylece kavramlar arasındaki özellik-genellik ilişkileri ifade edilir. Benzer şekilde, `rdfs:subPropertyOf` dil ögesi bir özelliğin bir başka özelliğin alt özelliği olduğunu belirtir. Dolayısıyla özellikler arasındaki özellik-genellik ilişkilerini ifade eder. Bu iki dil ögesi de, kavramlar arasında gerçek dünyada da bulunan ilişkileri temsil eder ve ontolojik patika üzerinde bulunabilir.

`owl:equivalentClass` ve `owl:equivalentProperty` dil ögeleri, sırasıyla iki sınıfın ve iki özelliğin aynı kaplama sahip olduğunu belirtir. Dolayısıyla bu dil ögeleri ile iki sınıf ya da iki özellik arasındaki gerçek dünyada da bulunan eşdeğer olma özelliği temsil edilir. `owl:sameAs` dil ögesi, iki bireyin aynı olduğunu belirtir. OWL Full dilinde `owl:sameAs` dil ögesi kullanılarak iki sınıfın eşit olduğu da ifade edilebilir. `owl:equivalentClass`, `owl:equivalentProperty` ve `owl:sameAs` dil ögeleri, temsil ettikleri gerçek dünya ilişkilerinden dolayı ontolojik patika üzerinde bulunabilir.

İki sınıfın ayrık olma durumu `owl:disjointWith` dil ögesi ile ifade edilir. Bir sınıfa ait olan bir birey, bu sınıfla ayrık olduğu belirtilen bir başka sınıfın bireyi olamaz. İki bireyin birbirinden farklı olduğu `owl:differentFrom` dil ögesi ile ifade edilir. `owl:AllDifferent` ve `owl:distinctMembers` dil ögeleri, ikiden fazla bireyin birbirinden farklı olduğunu tanımlamak için kullanılır. Dolayısıyla `owl:AllDifferent` ve `owl:distinctMembers` dil ögeleri ile yapılan bir tanımlama, anlamsal açıdan `owl:differentFrom` dil ögesi kullanılarak ifade edilir. Sonuç olarak iki kavram arasındaki farklılık ilişkisini ifade ettiği için `owl:disjointWith` ve `owl:differentFrom` dil ögelerinin, kavramlar arasındaki ilintililiğe yaptıkları katkının sayısal olarak ifade edilmesi gerekmektedir. `owl:AllDifferent` ve `owl:distinctMembers` dil ögeleri ile yapılan bir tanımlama ise `owl:differentFrom` dil ögesi cinsinden ifade edileceği için `owl:AllDifferent` ve `owl:distinctMembers` dil ögeleri kapsam dışında bırakılacaktır.

`owl:TransitiveProperty`, `owl:SymmetricProperty`, `owl:FunctionalProperty` ve `owl:InverseFunctionalProperty` dil ögeleri, bir ontolojide tanımlanan özelliklerin niteliklerini belirtir. Bir özelliğin geçişli, simetrik, işlevsel ya da ters işlevsel olması, o özelliklerin ait olduğu gerçek dünya bölümünde, kavramlar arasında bir ilintililik değerlendirmesi yapmak için gerekli değildir. Fakat bu niteliklerden yola çıkılarak yapılacak

çıkarsama sonucunda elde edilen yeni ifadeler kavram ilintililiği açısından önem taşır (Bu dil öğeleri ile yapılabilecek çıkarsamalar için Bkz. Ek 2). Dolayısıyla, `owl:TransitiveProperty`, `owl:SymmetricProperty`, `owl:FunctionalProperty` ve `owl:InverseFunctionalProperty` dil öğelerinin kavramlar arasındaki ilintililiğe yaptıkları katkının sayısal olarak ifade edilmesi gerekmemektedir. Şekil 5.11’de, OWL ile geliştirilmiş “ulaşım” konu alanı ile ilgili bir ontolojinin kesiti gösterilmektedir. Bu ontoloji kesitine göre, “iu” isim uzayındaki “parçası_olmak” nesne özelliği geçişli bir özelliktir. Bu nesne özelliğinin geçişli bir özellik olması, ilintililik ya da benzerlik yargısına katkıda bulunacak, “ulaşım” konu alanı ile ilgili herhangi bir bilgi içermemektedir. Fakat “piston” kavramı, “motor” kavramının, “motor” kavramı da “araba” kavramının parçası ise ve “parçası_olmak” özelliği geçişli bir özellik ise bu durumda “piston” kavramının da “araba” kavramının bir parçası olduğu sonucuna varılabilir. Varılan bu sonuç, ilintililik ya da benzerlik yargısına katkıda bulunabilir ve “ulaşım” konu alanı ile ilgili bir bilgi içerir.



Bilgi tabanındaki ifadeler:

iu:parçası_olmak rdf:type owl:ObjectProperty
 iu:parçası_olmak rdf:type owl:TransitiveProperty
 Piston iu:parçası_olmak Motor
 Motor iu:parçası_olmak Araba

Çıkarsanan ifadeler:

Piston iu:parçası_olmak Araba

Şekil 5.11: OWL ile geliştirilmiş “ulaşım” konu alanı ile ilgili bir ontolojinin kesiti.

OWL dilinde, özelliklerin konu alanları ve menzilleri tanımlanabilmektedir. Bir özelliğin konu alanı, o özelliğin hangi sınıflara ait olabileceğini belirtir ve `rdfs:domain` dil ögesi ile tanımlanır. Bir özelliğin menzili ise, o özelliğin hangi değerleri alabileceğini belirtir ve `rdfs:range` dil ögesi ile tanımlanır. Bir özelliğin menzili bir sınıf ya da ilkel bir veri tipi olabilir. `rdfs:domain` ve `rdfs:range` dil ögeleri, kavramlar ve özellikleri arasındaki ilişkileri tanımladığı için kavramlar arasındaki ilintililiğe yaptıkları katkının sayısal olarak ifade edilmesi gerekmektedir.

Özelliklerin, konu alanlarından menzillerine doğru bir yönü vardır. Bir özelliğin tersi `owl:inverseOf` dil ögesi ile ifade edilir ve o özelliğin yönünü değiştirir. İki özelliğin birbirinin tersi olma durumu (örneğin, “çocuk sahibi olmak” ile “ebeveyn sahibi olmak” özellikleri birbirinin tersidir), gerçek dünyada da bulunan bir ilişki olduğu için `owl:inverseOf` dil ögesi ontolojik patika üzerinde bulunabilir.

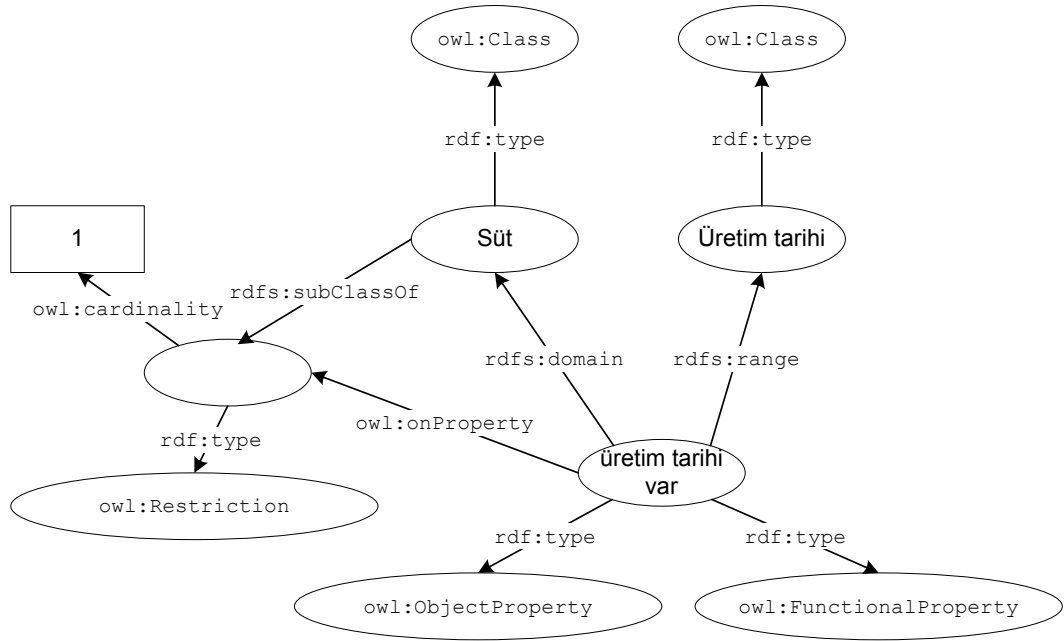
`owl:DeprecatedClass` ve `owl:DeprecatedProperty` dil ögeleri, bir ontolojinin sürümleri arasındaki uyumluluğu sağlamak için kullanılır. Bu iki dil ögesi, bir sınıfın ya da özelliğin yeni oluşturulan ontolojilerde kullanılmaması gerektiğini belirtir. Dolayısıyla bu dil ögeleri bir ontolojideki sürümler arasındaki uyumluluk açısından gereklidir ve gerçek dünya açısından anlamlı değildir.

`owl:Ontology`, `owl:imports`, `owl:versionInfo`, `owl:priorVersion`, `owl:backwardCompatibleWith` ve `owl:incompatibleWith` dil ögeleri ontoloji belgesi ve sürümü ile ilgili bilgilerin tanımlanmasını sağlar. Dolayısıyla bu dil ögeleri, kavramlar arasındaki ilintililik ölçümü için gerekli değildir.

`rdfs:label`, `rdfs:comment`, `rdfs:seeAlso`, `rdfs:isDefinedBy` ve `owl:AnnotationProperty` dil ögeleri, bir ontolojideki sınıfların, özelliklerin ve bireylerin etiketlenmesini sağlar. Bu dil ögeleri de, kavramlar arasındaki ilintililik ölçümü için gerekli değildir.

Bazı dil ögeleri ise direk olarak sınıflar, özellikler ya da bireyler arasında bulunmaz. Şekil 5.12’de gösterilen örnekteki `owl:onProperty` dil ögesi, “üretim tarihi var” özelliği ile `owl:Restriction` tipindeki isimsiz bir düğüm arasında yer almaktadır. Dolayısıyla, `owl:onProperty` dil ögesinin, anlamsal ilintililiğe katkısını, aralarında `owl:onProperty` dil ögesi bulunan kavram

çiftlerini insanların değerlendirmesi sonucunda elde etmek zordur. owl:minCardinality, owl:maxCardinality ve owl:cardinality dil öğeleri ise bir özelliğin alabileceği değerlerin sayısını kısıtlar. Dolayısıyla bu dil öğelerinin menzilleri birer sayıdır. Dolayısıyla bu dil öğelerinin ontolojik patika üzerinde yer alması mümkün değildir. Örneğin, iki özelliğin sadece birer değer alması bu iki özelliği anlamsal açıdan ilintili kılmaz. Bu nedenlerden dolayı, owl:unionOf, owl:complementOf, owl:intersectionOf, owl:oneOf, owl:DataRange, owl:Restriction, owl:onProperty, owl:allValuesFrom, owl:someValuesFrom, owl:minCardinality, owl:maxCardinality, owl:cardinality ve owl:hasValue dil öğeleri kapsam dışında bırakılacaktır.



Şekil 5.12: OWL ile geliştirilmiş “içecek” konu alanı ile ilgili bir ontolojinin kesiti.

OWL dil öğelerinin ontolojik patika üzerinde olup olmadıklarının yanında, bu öğelerin simetrik olup olmadığının da araştırılması gerekmektedir.

İki kavram arasındaki anlamsal ilintililiğin ya da benzerliğin simetrik olması durumu, ölçüm yönteminin kullanılma alanına göre değişiklik göstermektedir. Örneğin, anlamsal benzerlik yöntemleri, bilgi erişimi alanında sorgu genişletme için kullanılmaktadır. Bir kullanıcı, sorgu cümlecğine “hayvan” kelimesini eklediğinde, “hayvan” kavramı ile özellik ilişkisi bulunan kavramlar da sorgu cümlecğine otomatik olarak eklenerek sorgu genişletilir. Böylece “hayvan” kavramı ile ilgili sonuçlar bekleyen kullanıcıya aynı zamanda “kedi”, “köpek”

gibi kavramları içeren sonuçlar da sunulacaktır. Her “kedi” ve her “köpek” birer “hayvan” olduğu için sonuçlar kullanıcı için anlamlı olacaktır. Fakat sorgu cümlecğinde “kedi” kelimesi bulunan bir kullanıcıya, “hayvan” kavramını içeren sonuçlar sunulduğunda, yüksek olasılıkla kullanıcının ilgilenmeyeceği sonuçlar da kendisine sunulmuş olacaktır. Çünkü her “kedi” bir “hayvan” iken, her “hayvan” bir “kedi” değildir. Buna göre bilgi erişimi alanında kullanılacak bir anlamsal benzerlik formülünün simetrik olması yüksek olasılıkla istenmeyecektir (Knappe, 2005).

Bu çalışma kapsamında geliştirilecek anlamsal ilintililik yönteminde bazı dil öğelerinin simetrik olmama durumu bulunmaktadır. `rdfs:subClassOf` ve `rdfs:subPropertyOf` dil öğeleri özellik-genellik ilişkilerini temsil ettikleri ve bu ilişkilerin bir yönü (özelden genele ya da genelden özele) olduğu için bu dil öğelerinin anlamsal ilintililik ölçümü açısından simetrik olup olmadığının araştırılması gerekmektedir. Benzer şekilde, `rdfs:range` ve `rdfs:domain` dil öğelerinin de yönü olduğu için bu dil öğelerinin de simetrik olup olmadığının belirlenmesi gerekmektedir. `owl:equivalentClass`, `owl:equivalentProperty`, `owl:inverseOf`, `owl:disjointWith` ve `owl:differentFrom` dil öğeleri ise tanımları gereği simetrik olduğu için bu dil öğelerinin anlamsal ilintililik açısından da simetrik olduğu kabul edilecektir. Bu durum OWL dilinin çıkarsama kuralları ile de desteklenmektedir. Örneğin bir bilgi tabanında “A `owl:equivalentClass` B” ifadesi varsa yapılan çıkarsama sonucunda aynı bilgi tabanına “B `owl:equivalentClass` A” ifadesi de eklenecektir.

Sonuç olarak, `rdfs:subClassOf`, `rdfs:subPropertyOf`, `owl:equivalentClass`, `owl:equivalentProperty`, `owl:inverseOf`, `owl:disjointWith`, `owl:differentFrom`, `rdfs:range`, `rdfs:domain` dil öğelerinin, kavramlar arasındaki ilintililiğe yaptıkları katkının sayısal olarak belirlenmesine karar verilmiştir. Bu öğelerden `rdfs:subClassOf`, `rdfs:subPropertyOf`, `rdfs:range` ve `rdfs:domain` dil öğelerinin simetrik olup olmadıkları da değerlendirilecektir.

5.3.4 OWL dil öğelerinin anlamsal ilintililik katsayılarının belirlenmesi

OWL dil öğelerinin anlamsal ilintililik katsayılarının belirlenmesi için, aralarında bir önceki bölümde belirlenen OWL dil öğelerinin bulunduğu 143 kavram çifti oluşturulmuştur. Bu kavram çiftleri arasındaki ilintililik, 59 katılımcı

tarafından değerlendirilmiştir. Elde edilen sonuçlara göre OWL dil öğelerinin anlamsal ilintililik katsayıları belirlenmiştir. Bu bölümde, yapılan çalışmada kullanılan veri toplama aracının nasıl oluşturulduğu, veri toplama çalışmasının nasıl yapıldığı, verilerin nasıl çözümlendiği ve OWL dil öğelerinin anlamsal ilintililik katsayılarının nasıl belirlendiği anlatılmıştır.

5.3.4.1 Veri toplama aracı

OWL dil öğelerinin anlamsal ilintililik katsayılarının belirlenmesi için, kavram çiftleri oluşturmak amacıyla toplam 201 ontoloji çözümlenmiştir. Bu ontolojilerin 179 tanesi, “Ulusal Havacılık ve Uzay Dairesi” (NASA) tarafından “Dünya ve Çevre Terminolojisi için Anlamsal Web” (SWEET)¹⁰ projesi kapsamında geliştirilmiştir. 13 ontoloji, Protege Ontoloji Kütüphanesi¹¹’nden, 9 ontoloji de Tones Ontoloji Kütüphanesi¹²’nden alınmıştır.

Ontolojileri çözümlmek için Java programlama dili kullanılarak bir program geliştirilmiştir. Bu program ontolojilerdeki tüm ifadeleri yüklemlerine göre sınıflandırarak listelemektedir. Ontolojileri işlemek için Jena kütüphanesi¹³ (McBride, 2002) kullanılmıştır. Ontolojiler çözümlenirken, amaç sadece mevcut ifadeleri listelemek olduğu için, Jena kütüphanesi ile birlikte herhangi bir çıkarsama motoru kullanılmamıştır.

Çözümlenen ontolojilerde toplam 9730 sınıf, 1340 nesne özelliği, 669 veri tipi özelliği ve 3673 birey bulunmaktadır. Bu ontolojilerde bulunan OWL dil öğelerinin dağılımı Çizelge 5.2’de gösterilmiştir. Bu dağılıma göre, seçilen 9 OWL dil ögesi toplam öğelerin %58’ini oluşturmaktadır.

¹⁰ Semantic Web for Earth and Environmental Terminology (SWEET): <http://sweet.jpl.nasa.gov/ontology> (Erişim tarihi: 07.08.2010)

¹¹ Protege Ontology Library: http://protegewiki.stanford.edu/wiki/Protege_Ontology_Library#OWL_ontologies (Erişim tarihi: 07.08.2010)

¹² TONES Ontology Repository: <http://www.tonesproject.org> (Erişim tarihi: 07.08.2010)

¹³ Jena Web Sitesi: <http://jena.sourceforge.net/index.html> (Erişim tarihi: 15.01.2011)

Çizelge 5.2: Seçilen 201 ontolojideki OWL dil öğelerinin dağılımı.

OWL Dil Öğesi	Toplam	Toplam İçindeki Yüzdesi
<code>rdfs:subClassOf</code>	12098	%38,54
<code>owl:onProperty</code>	5770	%18,38
<code>owl:allValuesFrom</code>	2253	%7,18
<code>owl:someValuesFrom</code>	2207	%7,03
<code>rdfs:range</code>	1640	%5,22
<code>rdfs:domain</code>	1582	%5,04
<code>owl:intersectionOf</code>	1156	%3,68
<code>owl:hasValue</code>	1031	%3,28
<code>owl:equivalentClass</code>	997	%3,18
<code>owl:disjointWith</code>	939	%2,99
<code>rdfs:subPropertyOf</code>	547	%1,74
<code>owl:unionOf</code>	439	%1,40
<code>owl:inverseOf</code>	261	%0,83
<code>owl:distinctMembers</code>	111	%0,35
<code>owl:oneOf</code>	110	%0,35
<code>owl:maxCardinality</code>	55	%0,18
<code>owl:minCardinality</code>	50	%0,16
<code>owl:complementOf</code>	46	%0,15
<code>owl:cardinality</code>	42	%0,13
<code>owl:equivalentProperty</code>	28	%0,09
<code>owl:sameAs</code>	23	%0,07
<code>owl:differentFrom</code>	9	%0,03
TOPLAM	31394	%100

201 ontolojide bulunan ifadeler taranarak 143 kavram çifti oluşturulmuştur. Seçilen kavram çiftlerinin arasında, 9 OWL dil öğesinden biri ile kurulmuş bir ilişki bulunmaktadır. Kavram çiftleri oluşturulurken, mümkün olduğu kadar, aynı OWL dil öğesi ile birbirine bağlanmış kavram çiftlerinin farklı konu alanlarından seçilmesine gayret edilmiştir. Böylece OWL dil öğelerinin anlamsal ilintililik katsayılarının farklı konu alanları için değerlendirilmesinin sağlanması ve sonuçta konu alanından bağımsız anlamsal ilintililik katsayıları elde edilmesi hedeflenmiştir. Oluşturulan kavram çiftlerinin, aralarında ilişki kuran OWL dil öğelerine göre dağılımı Çizelge 5.3'te gösterilmiştir.

Çizelge 5.3: Oluşturulan kavram çiftlerinin, aralarında ilişki kuran OWL dil öğelerine göre dağılımı.

OWL Dil Öğesi	Bulunma Sayısı
owl:disjointWith	29
rdfs:subClassOf (özelden genele)	26
rdfs:subClassOf (genelden özele)	23
rdfs:domain (konu alanından özelliğe)	12
rdfs:domain (özellikten konu alanına)	11
owl:equivalentClass	10
rdfs:range (özellikten menzile)	7
rdfs:subPropertyOf (özelden genele)	6
owl:inverseOf	5
rdfs:subPropertyOf (genelden özele)	5
owl:equivalentProperty	4
rdfs:range (menzilden özelliğe)	3
owl:differentFrom	2
TOPLAM	143

Oluşturulan kavram çiftleri Ek 3'te listelenmiştir.

5.3.4.2 Veri toplama çalışması

Veri toplama çalışması, ön çalışma ve asıl çalışma olmak üzere iki aşamadan oluşmaktadır. Ön çalışma aşamasında katılımcılardan anket açıklama sayfasının (Bkz. Ek 4) ve kavram çiftlerinin (Bkz. Ek 3) anlaşılabilirliğini değerlendirmeleri istenmiştir. Asıl çalışmada ise, ön çalışmada elde edilen yorumlar doğrultusunda son haline getirilmiş anket, katılımcılara dağıtılmış ve elde edilen veriler çözümlenmiştir.

Ön çalışma, 3 psikolog ve 3 bilgisayar/elektrik-elektronik mühendisi ile yapılmıştır. Bunun nedeni, ilintililik ve benzerlik konusu üzerine psikoloji ve bilgisayar alanında araştırmaların ve uygulamaların olmasıdır. Ön çalışmada, katılımcılar anketin açıklama sayfasını ve kavram çiftlerini anlaşılabilirlik bakımından değerlendirmiştir. Katılımcılardan ilk olarak, kavram çiftleri arasındaki ilintililiği değerlendirmek için, açıklama sayfasının yeterince bilgilendirici ve anlaşılır olup olmadığını yorumlamaları istenmiştir. Daha sonra ise, kavramların, gerçekten ifade edilmek istenenleri yansıtır yansıtmadığı sınanmıştır. Ön çalışma sonucunda, anketin açıklama yazısı iyileştirilmiştir ve bazı kavram çiftlerindeki kelimeler değiştirilmiştir.

Asıl çalışmada ise, 66 katılımcıya, 143 kavram çiftinin aralarındaki ilintiliği değerlendirip 0 ile 10 arasında bir puan vermeleri istenmiştir. 0, iki kavram arasında hiç ilintilik olmaması durumunu, 10 ise iki kavram arasında en fazla ilintiliğin olma durumunu ifade etmektedir. Katılımcılara, ilintilik ve benzerlik konusunda kısa bir bilgi verildikten sonra ilintiliği daha kolay değerlendirmeleri için, ilintilik değerini belirlerken, ilk kavramı okuduklarında ikinci kavramın akıllarına gelme olasılığını ya da ikinci kavramın akıllarına ne kadar kolay geldiğini düşünmeleri önerilmiştir. Bu açıklamaların yapıldığı anketin açıklama yazısı Ek 4'te verilmiştir.

143 kavram çifti tamamen İngilizce terimlere sahip ontolojiler arasından seçilmiştir. Daha sonra bu terimler Türkçe'ye çevrilmiştir. Çeviri işlemi sırasında, elde edilecek sonuçları etkileyecek seviyede bir hata yapıp yapılmadığını değerlendirmek için kavram çiftlerinin Türkçe sürümü, ana dili Türkçe olan 59 katılımcı, İngilizce sürümü ise ana dili İngilizce olan 9 katılımcı tarafından değerlendirilmiştir.

5.3.4.3 Veri çözümleme

Veri çözümleme aşamasında öncelikle Türkçe ve İngilizce sürümler için elde edilen sonuçlar arasında anlamlı bir fark olup olmadığı değerlendirilmiştir. Daha sonra toplanan veriler içinde uç değerler belirlenmiştir ve bu değerler sonuç kümesinden çıkarılmıştır.

Kavram çiftlerinin Türkçe ve İngilizce sürümleri için elde edilen sonuçlar arasında anlamlı bir fark olup olmadığı bağımsız gruplar t-testi kullanılarak kontrol edilmiştir. Testin sonuçları Çizelge 5.4'te gösterilmiştir. Bağımsız gruplar için t-testi sonuçlarına göre ana dili Türkçe olan katılımcıların yaptığı değerlendirmeler ($\text{Ort}=6,55$) ile ana dili İngilizce olan katılımcıların değerlendirmeleri ($\text{Ort}=6,15$) arasında anlamlı bir fark bulunamamıştır ($p>0,05$). Bağımsız gruplar için t-testi sonuçlarına göre iki veri kümesi arasında anlamlı fark olduğu durumlarda p değeri 0,05'ten küçük olmalıdır. İki sonuç kümesi arasında anlamlı bir fark bulunmadığı için bu adımdan sonra 66 katılımcının sonuçları birlikte değerlendirilmiştir.

Çizelge 5.4: 143 kavram çiftinin Türkçe ve İngilizce sürümlerinin değerlendirme sonuçları için yapılan bağımsız gruplar t-testi sonuçları.

	N	Ortalamalar	Standart Sapmalar	t	P
Türkçe Sürüm	143	6,55	1,74	1,837	0,067
İngilizce Sürüm	143	6,15	1,94		

Bir sonraki adımda, 66 katılımcının sonuçları içinde uç değerlerin olup olmadığı kontrol edilmiştir. Aynı veri kümesi içinde, diğer veriler dikkate alındığında aykırı ve/veya beklenmedik olan ya da önceki bilgilerimizle karşılaştırıldığında mantıklı olmayan değerlere uç değer denir. Uç değerler, verilerin normal dağılımdan sapmasına neden olabilmektedir. Bundan dolayı, bir veri kümesi içinde sonuca önemli derecede olumsuz etkileri olan uç değerlerin belirlenmesi ve bunların ayıklanması gerekmektedir.

Uç değerlerin olup olmadığını değerlendirmek ve varsa bu değerlerin hangileri olduğunu tespit etmek için sonuçlar z dönüşümü yapılarak aritmetik ortalaması 0, standart sapması 1 olan standart normal dağılışa çevrilmiştir. Birden fazla kavram çifti için z-değeri -3,29'un altında ve +3,29'un üzerinde değerleri olan katılımcıların değerlendirmeleri sonuç kümesinden çıkarılmıştır. Uç değerlere sahip olduğu için değerlendirmeleri sonuç kümesinden çıkarılan katılımcı sayısı 7'dir. Böylece OWL dil öğelerinin anlamsal ilintililik katsayılarının belirlenmesi için 59 katılımcının sonuçları göz önüne alınmıştır.

5.3.4.4 OWL dil öğelerinin anlamsal ilintililik katsayılarının belirlenmesi

Her kavram çifti için katılımcıların verdiği 0 ile 10 arasındaki değerlerin ortalaması alınarak kavram çiftlerinin ortalama ilintililik değerleri hesaplanmıştır. Daha sonra kavram çiftleri, aralarındaki OWL dil öğesine göre gruplanarak her gruptaki kavram çiftlerinin ilintililik değerlerinin de ortalaması alınmıştır. Bulunan her değer 10'a bölünerek elde edilen değer 0 ile 1 arasında olması sağlanmıştır. Toplanan verilerin çözümlenmesi sonucunda elde edilen OWL dil öğelerinin anlamsal ilintililik katsayıları Çizelge 5.5'te gösterilmiştir.

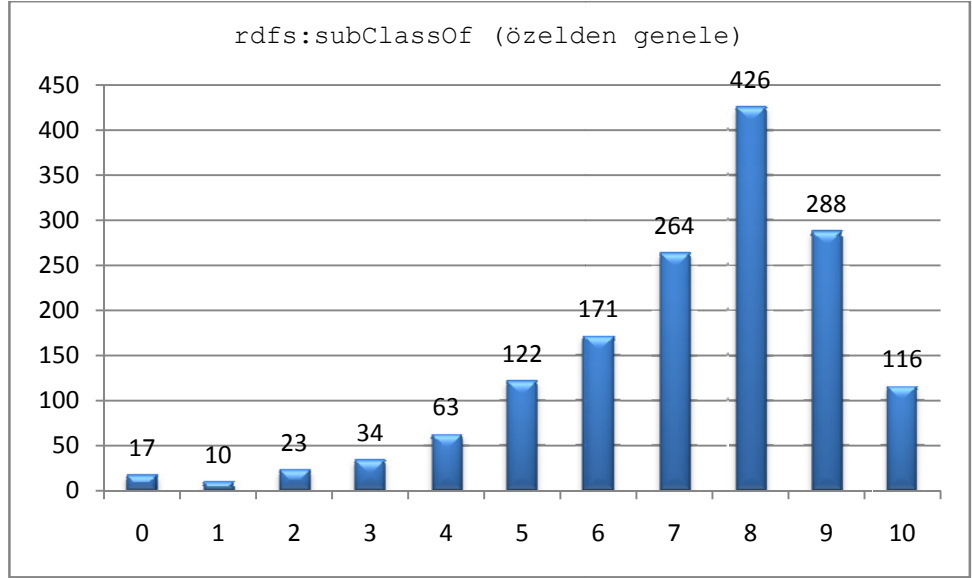
Çizelge 5.5: OWL dil öğelerinin anlamsal ilintililik katsayıları.

OWL Dil Ögesi	Anlamsal İlintililik Katsayısı (aik)
owl:inverseOf	0,85254
owl:equivalentClass	0,80228
rdfs:subClassOf (özelden genele)	0,72149
rdfs:subPropertyOf (özelden genele)	0,70409
rdfs:domain (özellikten konu alanına)	0,69768
rdfs:subClassOf (genelden özele)	0,69384
owl:equivalentProperty	0,68156
rdfs:subPropertyOf (genelden özele)	0,66135
owl:disjointWith	0,65774
rdfs:domain (konu alanından özelliğe)	0,58072
rdfs:range (özellikten menzile)	0,50690
rdfs:range (menzilden özelliğe)	0,30593
owl:differentFrom	0,19194

Çizelge 5.5'teki anlamsal ilintililik katsayısı değerlerine göre, Şekil 5.6'da gösterilen ontoloji kesitindeki “okutman” ve “öğretim üyesi” kavramları arasındaki anlamsal ilintililik, (5.3) numaralı formüldeki gibi hesaplanacaktır.

$$\begin{aligned}
 \text{Anlamsal İlintililik("Okutman", "Öğretim Üyesi")} &= \\
 \text{aik}(\text{rdfs:subClassOf (özelden genele)}) \times & \\
 \text{aik}(\text{rdfs:subClassOf (genelden özele)}) &= 0,72149 \times 0,69384 = 0,50060
 \end{aligned}
 \tag{5.3}$$

Toplanan verilerin OWL dil öğeleri bazında dağılım çözümlemesi de yapılmıştır. Şekil 5.13'te, arasında `rdfs:subClassOf` (özelden genele) dil öğesi ile ilişki kurulmuş iki kavram arasındaki ilintililik için katılımcıların verdiği değerlerin dağılımı gösterilmektedir. Şekilde, cevapların bir değer çevresinde yoğunlaştığı görülmektedir. Bu durum, katılımcıların anlamsal ilintililik yargılarıyla, kavramlar arasındaki OWL dil öğeleri arasında bir bağlantı olduğunu doğrulamaktadır.



Şekil 5.13: Arasında `rdfs:subClassOf` (özelden genele) dil ögesi ile ilişki kurulmuş iki kavram arasındaki ilintililik için katılımcıların verdiği değerlerin dağılımı.

Katılımcıların iki kavram arasındaki ilintililik için verdiği değerlerin OWL dil ögeleri bazında dağılımları Ek 4'te gösterilmiştir. `rdfs:range` dışındaki dil ögeleri için dağılımların belli değerler çevresinde yoğunlaştığını söylemek mümkündür. `rdfs:range` dil ögesi ile ilgili dağılımın, diğer dil ögelerinkine göre biraz daha bozuk olması, bu dil ögesinin anlamsal ilintililik katsayısını ölçmek için kullanılan kavram çiftlerinin özellikle Türkçe'de algılanmasının biraz daha zor olmasından kaynaklandığı düşünülebilir.

Simetrik olmayan OWL dil ögelerinin anlamsal ilintililik katsayıları Çizelge 5.6'da gösterilmiştir. Bulunan değerlere göre, `rdfs:subClassOf` ve `rdfs:subPropertyOf` dil ögelerinde ilişkinin yönünün anlamsal ilintililik katsayısına etkisi, `rdfs:domain` ve `rdfs:range` dil ögelerinkine kıyasla çok daha azdır.

Çizelge 5.6: Simetrik olmayan OWL dil öğelerinin anlamsal ilintililik katsayıları.

OWL Dil Ögesi	Yön	Anlamsal İlintililik Katsayısı (aik)
rdfs:subClassOf	özelden genele	0,72149
	genelden özele	0,69384
rdfs:subPropertyOf	özelden genele	0,70409
	genelden özele	0,66135
rdfs:domain	özellikten konu alanına	0,69768
	konu alanından özelliğe	0,58072
rdfs:range	özellikten menzile	0,50690
	menzilden özelliğe	0,30593

5.4 Ontoloji Tabanlı Anlamsal İlintililik Ölçüm Yönteminin Başarısının Değerlendirilmesi

Anlamsal ilintililik ve benzerlik yöntemlerinin başarısının değerlendirilmesi için en çok kullanılan yaklaşım, elde edilen sonuçların insanların ilintililik yargılarıyla kıyaslanmasıdır. Yöntemlerin başarısını sınamak için araştırmacılar tarafından çeşitli kelime çifti grupları oluşturulmuştur.

Rubenstein ve Goodenough, kelime benzerlikleri ile bağlam arasındaki ilişkiyi ölçmek için, benzerlik değerlendirmeleri katılımcılar tarafından yapılmış 65 kelime çiftinden bir deney kümesi oluşturmuşlardır (Rubenstein and Goodenough, 1965). 65 kelime çifti, anlamsal olarak ilişkisiz kelimelerden eşanlamlı kelimelere kadar farklı çiftlerden oluşturulmuştur. Deney kümesinde, teknik kelimelere yer verilmemiştir. Bunun nedeni hem bu kelimelerin benzerliklerini değerlendirmenin zorluğu hem de deneyin amacının İngilizce dilinin genel özelliklerini değerlendirecek olması olarak açıklanmıştır.

Deneyde, lisans öğrencilerinden oluşan iki grup kullanılmıştır. İlk gruptaki 15 denek, iki hafta arayla iki oturumda değerlendirme yapmıştır. İlk oturumda, deneklerden 48 kelime çiftinin benzerliklerini değerlendirmeleri istenmiştir. İlk oturumda değerlendirilen 48 kelime çiftinin 36 tanesi, çalışma için en son seçilen 65 kelime çiftinde bulunmaktadır. İlk grubun, kelime çiftlerinin benzerliklerini değerlendirdiği iki oturum sonucunda elde edilen sonuçlar arasındaki korelasyon 0,85 olarak bulunmuştur. Katılımcılar, kelime çiftlerinin benzerlikleri için sıfır ile dört arasında değerler vermişlerdir. Sıfır değeri iki kelime arasında hiç benzerlik olmadığını, dört ise iki kelimenin aynı olduğunu ifade etmektedir.

İkinci grup, birinci grup ile aynı özelliklere sahip 36 denekten oluşturulmuştur ve sadece ikinci oturuma katılarak 65 kelime çiftinin benzerliğini değerlendirmiştir. Birinci ve ikinci grubun değerlendirme sonuçları arasındaki korelasyon 0,99 olarak bulunmuştur. İki grubun değerlendirme sonucu arasındaki korelasyon çok yüksek olduğu için, sonuçlar birleştirilmiştir ve 65 kelime çifti için iki gruptan toplam 51 katılımcının değerlendirmelerinin ortalamaları alınmıştır.

Miller ve Charles, yaptıkları kelime çifti değerlendirme çalışmasında, Rubenstein ve Goodenough'un deneyinde kullanılan 65 kelime çiftinin içinden seçtikleri 30 çifti kullanmıştır (Miller and Charles, 1991; Allen and Wu'dan, 2005). Ana dili İngilizce olan 38 üniversitesi öğrencisi, seçilen 30 kelime çiftinin benzerlik değerlendirmesini yapmıştır. Elde edilen sonuçların, Rubenstein ve Goodenough'un sonuçlarıyla korelasyonu 0,97 olarak bulunmuştur.

Finkelstein et al.'nin deneyinde, Miller ve Charles'ın 30 kelime çiftini kapsayan 353 kelime çiftinin benzerlikleri, üniversite mezunu ve ana dili İngilizce olan ya da İngilizce'yi çok iyi konuşan 29 katılımcı tarafından değerlendirilmiştir (Finkelstein et al., 2002). Kelime çiftlerinin 153 tanesi 13 katılımcı tarafından, 200 tanesi 16 katılımcı tarafından değerlendirilmiştir. Kelime çiftini oluşturan kelimelerin birbiriyle hiçbir ilgisi yoksa değerlendirme sonucu 0, iki kelime tamamen aynıysa değerlendirme sonucu 10 olarak belirlenmiştir. Deneyin sonuçları ile Miller ve Charles'ın deneyinin sonuçları arasındaki korelasyon 0,95 olarak ölçülmüştür.

Bu çalışma kapsamında geliştirilen anlamsal ilintililik yönteminin başarısını sınamak için 24 kavram çifti oluşturulmuştur. Bu kavram çiftleri, seçilen 201 ontoloji taranarak oluşturulmuştur ve İngilizce'den Türkçe'ye çevrilmiştir. Kavram çiftlerinin, Türkçe'de ifade edildiğinde katılımcılar tarafından kolay anlaşılabilir olmalarına dikkat edilmiştir. Kavram çiftlerinin aralarında mümkün olduğu kadar çeşitli ilişki türleri (özellik-genellik, zıtlık, farklılık gibi) bulunması sağlanmıştır. Ayrıca, kavram çiftlerinin bir konu alanına özgü teknik terimler içermemesine dikkat edilmiştir.

24 kavram çifti arasındaki ilintililik 66 katılımcı tarafından değerlendirilmiştir. Bu katılımcılardan 57'sinin ana dili Türkçe'dir ve bu katılımcılar kavram çiftlerinin Türkçe sürümünü değerlendirmiştir. Katılımcıların

geriye kalan 9'unun ise ana dili İngilizce'dir ve bu katılımcılar da kavram çiftlerinin İngilizce sürümünü değerlendirmiştir.

Kavram çiftlerinin, İngilizce ve Türkçe sürümlerinin değerlendirme sonuçları arasında anlamlı bir fark olup olmadığını değerlendirmek amacıyla elde edilen verilere bağımsız gruplar t-testi yapılmıştır. Testin sonuçları Çizelge 5.7'de gösterilmiştir. Bağımsız gruplar için t-testi sonuçlarına göre ana dili Türkçe olan katılımcıların yaptığı değerlendirmeler (Ort=5,97) ile ana dili İngilizce olan katılımcıların değerlendirmeleri (Ort=5,98) arasında anlamlı bir fark bulunamamıştır ($p>0,05$). Bağımsız gruplar için t-testi sonuçlarına göre iki veri kümesi arasında anlamlı fark olduğu durumlarda p değeri 0,05'ten küçük olmalıdır.

Çizelge 5.7: 24 kavram çiftinin Türkçe ve İngilizce sürümlerinin değerlendirme sonuçları için yapılan bağımsız gruplar t-testi sonuçları.

	N	Ortalamalar	Standart Sapmalar	t	P
Türkçe Sürüm	24	5,97	2,56	-0,013	0,990
İngilizce Sürüm	24	5,98	2,57		

Kavram çiftlerinin Türkçe ve İngilizce sürümleri arasında anlamlı bir fark olmaması, İngilizce'den yapılan çeviri işleminin, Türkçe sürüm ile veri toplama için herhangi bir problem oluşturmadığını göstermektedir. Ayrıca bağımsız gruplar t-testi sonucunda iki sürüm arasında anlamlı bir fark bulunmadığı ve Türkçe ile İngilizce sürümlerin aynı kavram çiftleri arasındaki ilintililik değerlerini ölçtüğü için bu iki sürümün değerlendirmeleri sonucunda elde edilen veri kümeleri birleştirilerek sonuçlar birlikte çözümlenmiştir.

Kavram çiftlerinin Türkçe ve İngilizce sürümlerinin değerlendirme sonuçlarında uç değerlerin olup olmadığı kontrol edilmiştir. Uç değerlerin olup olmadığını değerlendirmek ve varsa bu değerlerin hangileri olduğunu tespit etmek için sonuçlar z dönüşümü yapılarak aritmetik ortalaması 0, standart sapması 1 olan standart normal dağılışa çevrilmiştir. Birden fazla kavram çifti için z-değeri -3,29'un altında ve +3,29'un üzerinde değerleri olan katılımcıların değerlendirmeleri sonuç kümesinden çıkarılmıştır. Uç değerlere sahip olduğu için değerlendirmeleri sonuç kümesinden çıkarılan katılımcı sayısı 3'tür. Böylece OWL dil öğelerinin anlamsal ilintililik katsayılarının belirlenmesi için 63 katılımcının sonuçları göz önüne alınmıştır.

Katılımcılar, kavram çifti arasındaki ilintililiği 0 ile 10 arasında bir sayı vererek değerlendirmiştir. 0, iki kavramın hiç ilintili olmama durumunu, 10 ise iki kavramın en üst düzeyde ilintili olması durumunu temsil etmektedir. Kavram çiftleri arasındaki insan yargısıyla belirlenmiş ilintililik değerleri, tüm katılımcıların verdiği değerlerin ortalaması 10'a bölünerek hesaplanmıştır.

Katılımcıların değerlendirmeleri ile ontoloji tabanlı yöntemin otomatik hesapladığı değerler arasındaki ilişkiyi anlamak için Pearson Momentler Çarpımı Korelasyon tekniği kullanılarak çözümleme yapılmıştır. Çözümleme sonucunda, katılımcı değerlendirmeleri ile hesaplanan değerler arasındaki korelasyonun ($r=0,685$; $p<0,01$) anlamlı ve pozitif yönde olduğu görülmüştür.

Çizelge 5.8'de, bazı benzerlik ölçüm yöntemlerinin ürettiği sonuçların, insanların verdiği benzerlik değerleriyle arasındaki korelasyonlar gösterilmiştir. Bu çizelgede görüldüğü gibi, en yüksek değer 0,85 olarak ölçülmüştür. Ayrıca, sonuçların sınıandığı, insanlar tarafından değerlendirilen kelime çifti sayısı arttıkça korelasyon değerlerinin düştüğü görülmektedir. Finkelstein et al.'nin 353 kelime çifti ile yapılan karşılaştırmalarda, Gabrilovich ve Markovitch'in yöntemi dışındaki diğer yöntemlerin başarı oranlarının oldukça düşük olduğu görülmektedir. Gabrilovich ve Markovitch'in yönteminin sınıandığı Finkelstein et al.'nin 353 kelime çiftinin 153 tanesi 13 denek tarafından, 200 tanesi 16 denek tarafından değerlendirilmiştir. Dolayısıyla denek sayısı arttığında yöntemin sonuçlarının ne kadar başarılı olacağı bilinmemektedir.

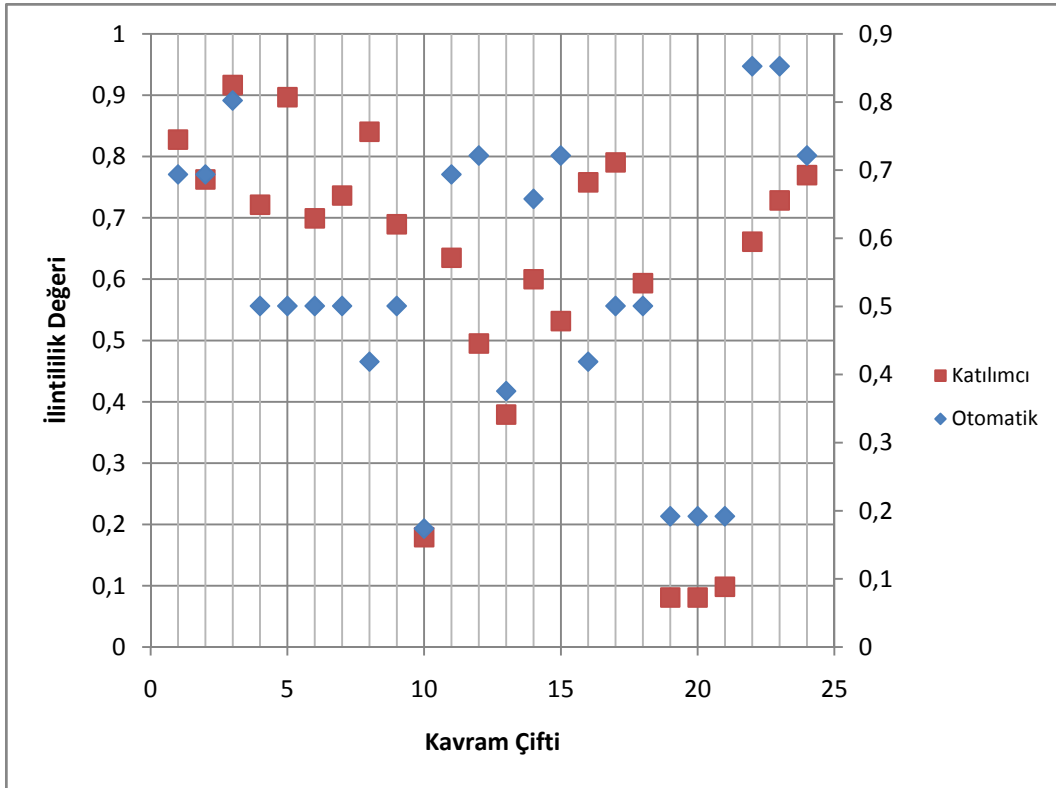
Çizelge 5.8'deki sonuçlara göre, bu tez kapsamında geliştirilen yöntem en iyi korelasyon sonucuna sahip değildir. Fakat Çizelge 5.8'deki tüm korelasyon değerlerinin benzerlik ölçüm yöntemleri için olduğu vurgulanmalıdır. Benzerliğin, ilintililiğin özel durumu olduğu düşünüldüğünde ve ilintililiğin bünyesinde çok daha fazla sayıda ve karmaşık ilişkiler barındırdığı düşünüldüğünde, ilintililik tahmininin benzerlik tahminine göre daha zor olduğu sonucuna varılmaktadır. Bundan dolayı, geliştirilen yöntemin bir dereceye kadar başarılı olduğu söylenebilir.

Katılımcıların ilintililik değerleri ile otomatik olarak hesaplanan ilintililik değerlerinin grafiği Şekil 5.14'te gösterilmiştir. Bu grafiğe göre otomatik yöntemin, insanın ilintililik yargısını belirli bir düzeyde taklit edebildiği görülmektedir. İlintililik değerlerinin mutlak değil bağıl değerler olduğu

düşünüldüğünde otomatik yöntemin genel olarak insan yargısının eğilimini yansıttığı söylenebilir.

Çizelge 5.8: Benzerlik ölçüm yöntemlerinin ürettiği sonuçların insanların verdiği benzerlik değerleriyle korelasyonları.

	Rubenstein ve Goodenough'un 65 kelime çifti (Rubenstein and Goodenough, 1965)	Miller ve Charles'ın 30 kelime çifti (Miller and Charles, 1991)	Finkelstein et al.'nin 353 kelime çifti (Finkelstein et al., 2002)
(Resnik, 1985)	0,78 (1)	0,77 (1)	0,37 (2)
(Rada et al., 1989)	-	0,64 (3)	0,25 (3)
(Wu and Palmer, 1994)	-	0,80 (4)	-
(Jiang and Conrath, 1997)	0,82 (1)	0,85 (1)	0,34 (2)
(Lin, 1998)	0,82 (1)	0,83 (4)	0,36 (2)
(Leacock and Chodorow, 1998)	0,84 (1)	0,82 (1)	0,36 (2)
(Gabrilovich and Markovitch, 2009)	-	-	0,72 (5)
Kaynaklar (1) (Budnitsky and Graeme, 2001) (2) (Bollegala et al., 2009) (3) (Mazuel and Sabouret, 2008) (4) (Lin, 1998) (5) (Gabrilovich and Markovitch, 2009)			



Şekil 5.14: Katılımcıların ilintililik değerleri ile otomatik olarak hesaplanan ilintililik değerleri.

24 kavram çifti için, katılımcıların ilintililik değerleri ile otomatik olarak hesaplanan ilintililik değerleri Ek 6'da listelenmiştir.

5.5 Uygulama

Bu tez kapsamında, ontoloji tabanlı anlamsal ilintililik yönteminin başarısını sınamak için bir uygulama geliştirilmiştir. Uygulama, Java programlama dili kullanılarak geliştirilmiştir. Uygulamada çeşitli araçlar kullanılmıştır. Uygulamada kullanılan araçların sürümleri ve web sayfaları Ek 7'de gösterilmiştir.

5.5.1 Uygulamada kullanılan araçlar

Uygulamada, ontoloji dosyalarının okunması, ontolojinin bellekte temsil edilmesi ve ontolojiler üzerinde kavramlar arasındaki patikaların bulunması gibi çeşitli işlemlerin yapılması için Jena (Carroll et al., 2004) kullanılmıştır. Çıkarılma işlemleri için Pellet (Sirin et al., 2007) OWL DL çıkarılma motoru kullanılmıştır.

5.5.1.1 Jena ve Pellet

Jena (Carroll et al., 2004), anlamsal web uygulamaları geliştirmek için kullanılan bir Java çerçevesidir. Jena, RDF, RDFS, OWL ve SPARQL için bir programlama ortamı sunar. Jena çerçevesi aşağıdaki işlevleri sunmaktadır (Jena Semantic Web Framework, 2011):

- RDF uygulama programlama arayüzü
- RDF/XML, N3 ve N-Triples ile temsil edilen RDF belgelerinin oluşturulması ve okunması
- OWL uygulama programlama arayüzü
- Bellekte ve kalıcı ortamlarda saklama
- SPARQL sorgu motoru

Bu çalışma kapsamında Jena, ontoloji dosyalarının okuması ve ontolojinin bellekte temsil edilmesi için kullanılmıştır. OWL ile oluşturulmuş ontolojiler üzerinde Jena OWL uygulama programlama arayüzü kullanılarak kavramlar arasındaki patikalar bulunmuştur.

Çıkarsama işlemleri için bir OWL DL çıkarsama motoru olan Pellet (Sirin et al., 2007) kullanılmıştır.

5.5.2 Uygulamanın ayrıntıları

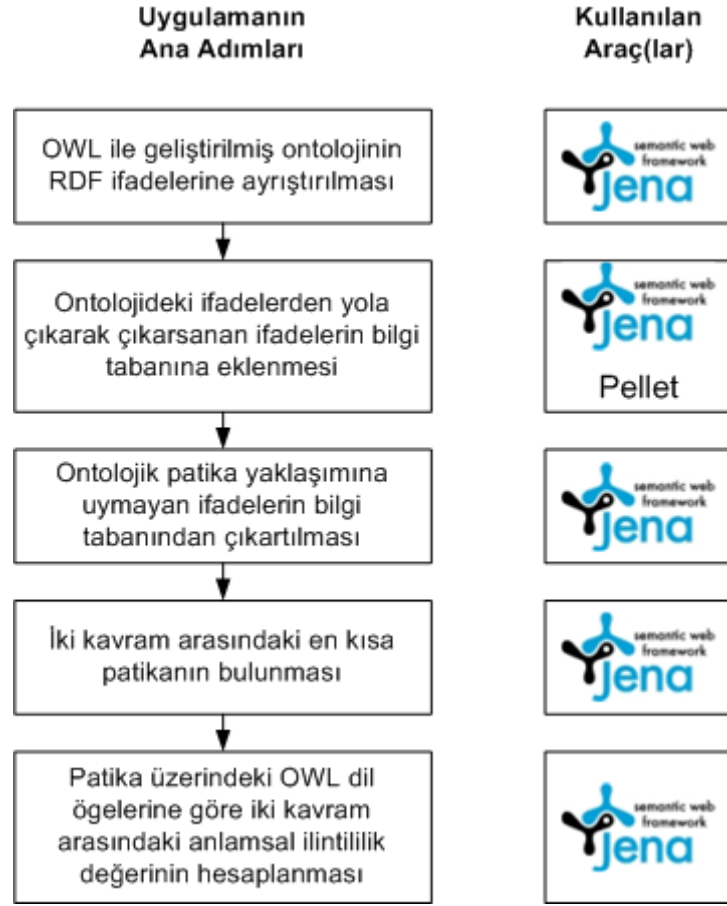
Uygulama, OWL ile geliştirilmiş bir ontolojideki iki kavram arasındaki tüm ontolojik patikaları bularak, her patika için anlamsal ilintililik değeri hesaplamaktadır. İki kavram arasındaki anlamsal ilintililik değeri olarak, tüm patikalar için hesaplanan değerlerin en büyüğü alınmaktadır. Anlamsal ilintililik değerinin hesaplanması için izlenen adımlar ve bu adımlarda kullanılan araçlar Şekil 5.15’te gösterilmiştir.

Anlamsal ilintililik değerinin hesaplanması için öncelikle OWL ile oluşturulmuş ontolojideki ifadeler ayrıştırılarak RDF ifadelerine dönüştürülür. Böylece tüm ifadeler bilgi tabanında “özne – yüklem – nesne” düzeninde bulunur. Bu adımdan sonra, ontolojideki ifadelerden yola çıkılarak çıkarsanan ifadeler bilgi tabanına eklenir. Bu iki adımda, bir OWL DL çıkarsama motoru olan Pellet (Sirin et al., 2007) kullanılmıştır.

Bir sonraki adımda, ontolojik patika yaklaşımına uygun olmayan RDF ifadeleri bilgi tabanından çıkarılmaktadır. Örneğin, tüm sınıfların `owl:Thing` sınıfının alt sınıfı olduğunu ya da `owl:Nothing` sınıfının tüm sınıfların alt sınıfı olduğunu belirten ifadeler bilgi tabanından çıkarılır. Çünkü bu ifadeler tamamen OWL diline aittir ve insanın ilintililik yargısıyla ilişkileri yoktur.

Dördüncü adımda iki kavram arasındaki tüm ontolojik patikalar bulunur. Patika arama işlemi için derinliğine arama algoritması (depth first search) kullanılmıştır.

Son adımda ise tüm ontolojik patikalar için anlamsal ilintililik değerleri hesaplanır ve en büyük değer iki kavram arasındaki anlamsal ilintililik değeri olarak alınır.



Şekil 5.15: OWL ile oluşturulmuş bir ontolojideki iki kavram arasındaki anlamsal ilintililiğin hesaplayan uygulamanın ana adımları ve bu uygulamada kullanılan araçlar.

Algoritma 5.1’de, bir ontolojideki iki kavram arasındaki patikaları bulmak ve bu patikalara göre iki kavram arasındaki en büyük anlamsal ilintililik değerinin nasıl hesaplandığı; Ek 8’de ise algoritmayı gerçekleştirmek için kullanılan sınıflar, değişkenler ve yordamlar gösterilmiştir. Anlamsal ilintililik hesaplama yordamı (AnlamsalIlintililikHesapla), ilintililik hesabı yapmak için kullanılacak ontolojinin URI’sini ve anlamsal ilintililik ölçümünün yapılacağı kelime çiftini girdi olarak alır. Patika bulma yordamı (PatikaBul), ontolojideki iki kavram arasındaki tüm ontolojik patikaları bulur. Bu yordam özyineli olarak kendini çağırır. Patikalar aranırken derinliğine arama algoritması kullanılmıştır. Ontolojik patika üzerinde bulunma olasılığı bulunmayan ifadeler IfadeleriListele yordamında ayıklanır. Bu yordam, sadece ontolojik patika üzerinde bulunma olasılığı bulunan ifadeleri sonuç olarak döndürür. Patika değeri hesaplama yordamı (PatikaDegeriHesapla), bulunan tüm patikalar için anlamsal ilintililik değerlerini hesaplar ve bu değerlerden en büyüğünü sonuç olarak döndürür.

Algoritma 5.1: Bir ontolojideki iki kavram arasındaki anlamsal ilintililik değerini hesaplayan algoritma

```

AnlamsalIlintililikHesapla(Ontoloji_URI, Kavram1, Kavram2) {
    Ontolojiyi Jena ile oku
    Kavram1'in ve Kavram2'nin URI'sini bul (Kavram1_URI, Kavram2_URI)
    PatikaBul(Kavram1_URI, Kavram2_URI, null, 0)
    return PatikaDegeriHesapla()
}

PatikaBul(baslangicDugum, bitisDugum, eklenenDugum, derinlik) {
    Eğer düğüm daha önce derinliğine arama ağacında farklı bir
    yerde ziyaret edilmişse
    {
        return
    }
    Eğer daha önce ziyaret edilen bir düğüm özyineli işlev çağırımı
    nedeniyle yeniden ziyaret edilmiyorsa
    {
        Ziyaret Edilmiş Düğümler listesine bu düğümü ve düğümün hangi
        derinlikte ziyaret edildiği bilgisini ekle
    }
    Eğer ifadeListesi boş değilse
    {
        ifadeListesi listesindeki her Kenar için
        {
            Eğer başlangıç düğümü ilk defa ziyaret ediliyorsa
            {
                Başlangıç düğümü ziyaret edilmiş düğümler listesine ekle
                Eğer başlangıç ve bitiş düğümler arasında bir ilişki varsa
                {
                    Bir patika bulunduğuna dair bilgiyi sakla (hem ifadeyi
                    içeren düğüm bilgisini hem de bu ifadenin yön bilgisini
                    - özneden nesneye ya da nesneden özneye doğru)
                }
                Eğer patika daha bulunmadıysa bir sonraki kenara geç
                {
                    PatikaBul işlevini şu andaki ifadenin öznesini ya da
                    nesnesini (ifadenin yönüne bağlı olarak), patikanın
                    biteceği kaynağı, şu andaki kenarı saklayan düğümü,
                    derinlik bilgilerini parametre olarak vererek çağır
                    (ör: PatikaBul(k.Ifade.Object, bitisDugum,
                    new Dugum(k, eklenenDugum, ++derinlik))
                }
            }
        }
    }
}

IfadeleriListele(BaslangicDugum) {
    Ontolojik patika yaklaşımına uyan tüm ifadeleri sonuç listesi
    olarak döndür
}

PatikaDegeriHesapla() {
    Tüm ontolojik patikalar için anlamsal ilintililik değerlerini
    hesapla, en yüksek anlamsal ilintililik değerini döndür
}

```

5.6 Özet ve Değerlendirme

Anlamsal web vizyonu çerçevesinde, anlamsal temsil gücü yüksek bilgi temsil dillerinin ortaya çıkmasıyla birlikte, eski bilgi temsil dilleri göz önüne alınarak tasarlanan yöntemlerin iyileştirilebilme fırsatları doğmuştur. Anlamsal ilintililik ve benzerlik ölçüm yöntemleri genellikle özellik-genellik ilişkilerini dikkate almaktadır. OWL dilinin kullanımının yaygınlaşmasıyla birlikte, anlamsal ilintililik ve benzerlik yöntemlerinde de OWL dilinin öğeleri değerlendirilmeye başlanmıştır.

Bu çalışma kapsamında geliştirilen anlamsal ilintililik ölçüm yönteminin temeli “etkinleştirmenin yayılması teorisi”ne dayanmaktadır (Bkz. Bölüm 5.1, Sayfa 66). Bu teoriye göre, bilgi bellekte bir ağ yapısında depolanmaktadır ve bilgiyi geri getirme işlemi bu ağ üzerinde “etkinleştirmenin yayılması” ile gerçekleşmektedir. Bir kavramdan yola çıkılarak diğer kavramların hatırlanma olasılığı, kavramlar arasındaki ilişkilerin kuvvet derecesine bağlıdır. Etkinleştirmenin yayılması bir kavramdan diğerine geçildikçe azalır ve bir noktada yayılma sona erer. Böylece ağdaki bir kavramın etkinleştirilmesi ile tüm kavramların geri getirilmesi gibi gerçek dışı bir durumun ortaya çıkması sağlanır.

Bu çalışmadaki ontoloji tabanlı anlamsal ilintililik ölçüm yöntemi, OWL ile geliştirilmiş ontolojileri, zihindeki çağrışım ağının yerine koymaktadır. İki kavram arasındaki anlamsal ilintililik, ontolojideki iki kavram arasındaki ilişkiye göre hesaplanmaktadır. Bu hesaplama yapılırken OWL dilinin 9 ögesi dikkate alınmaktadır. Seçilen 9 OWL dil ögesi ile ilişkilendirilmiş kavram çiftleri arasındaki ilintililiğin insanlar tarafından puanlanması ile bu 9 OWL dil ögesinin anlamsal ilintililik katsayısı belirlenmiştir. İki kavram arasındaki anlamsal ilintililik değeri, iki kavram arasındaki ontolojik patikalar üzerinde bulunan OWL dil öğelerinin katsayılarının birbiri ile çarpılması sonucunda elde edilmektedir. İki kavram arasında birden fazla ontolojik patika bulunması durumunda en yüksek anlamsal ilintililik değeri göz önüne alınmaktadır. OWL dil öğelerinin anlamsal ilintililik katsayıları 0 ile 1 arasında olduğu için iki kavram arasındaki ontolojik patika uzadıkça aradaki anlamsal ilintililik değeri düşmektedir. Aradaki patikanın üzerindeki OWL dil öğelerine bağlı olarak, bir yerden sonra aradaki anlamsal ilintililik değeri 0 olmaktadır. Bu durum da “etkinleştirmenin yayılması teorisi” ile tutarlıdır.

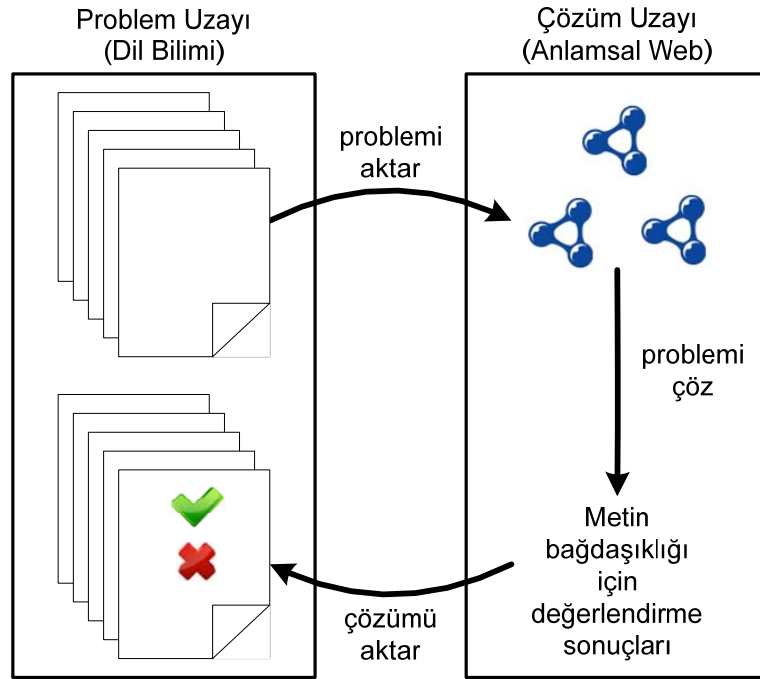
Geliştirilen ontoloji tabanlı anlamsal ilintililik yönteminin başarısı, diğer benzerlik ve ilintililik yöntemleri için de uygulandığı gibi, insanların ilintililik değerlendirmeleri ile kıyaslanarak ölçülmüştür. Bunun için 24 kavram çifti 63 katılımcı tarafından değerlendirilmiştir. Buna benzer çalışmalarda en çok kullanılan kelime çifti kümeleri, 65 (Rubenstein and Goodenough, 1965), 30 (Miller and Charles, 1991) ve 353 (Finkelstein et al., 2002) kelime çiftinden oluşmaktadır. Bu kelime çiftleri arasındaki benzerlik sırasıyla 51, 38 ve 29 katılımcı tarafından değerlendirilmiştir. Finkelstein'in 353 kelime çiftinin 153 tanesi 13, 200 tanesi 16 katılımcı tarafından değerlendirilmiştir. Bu çalışmalar ile kıyaslandığında, bu tez kapsamında oluşturulan kavram çifti sayısı en küçük gruba yakındır. Fakat kavram çiftlerini değerlendiren katılımcı sayısı tüm çalışmalardan daha yüksektir. Bu çalışmada yapılan değerlendirmede, literatürdeki diğer kelime çiftlerinin kullanılmamasının nedeni, katılımcıların bu kelime çiftleri arasındaki benzerliği değerlendirmiş olmalarıdır. Benzerlik, ilintililiğin özel bir durumu olduğu için (Bkz. Bölüm 4, Sayfa 30), bu çalışma kapsamında geliştirilen yöntemin başarısının sınanması için kavram çiftleri arasındaki ilintililik değerlerinin kullanılması daha sağlıklıdır.

Bu çalışma kapsamında geliştirilen ontoloji tabanlı anlamsal ilintililik yönteminin 24 kavram çifti için hesapladığı değerler ile katılımcıların bu kavram çiftlerinin ilintililiklerine verdikleri ortalama değerler arasındaki korelasyonun ($r = 0,685$; $p < 0,01$) anlamlı ve pozitif yönde olduğu görülmüştür. Bu değer, literatürde bulunan benzerlik ölçüm yöntemlerinin değerlendirilmesi sonucu elde edilen korelasyon değerleri (Bkz. Çizelge 5.8, Sayfa 94) arasındaki en yüksek değer değildir. İlintililiğin, benzerliğe göre bünyesinde çok daha fazla ve karmaşık ilişkiler barındırdığı düşünüldüğünde, ilintililik ölçümünün benzerlik ölçümüne göre daha karmaşık ve zor olduğu sonucuna varılabilir. Buna göre, elde edilen korelasyon değerinin diğer sonuçlarla kıyaslandığında çok düşük olmaması nedeniyle yöntemin bir dereceye kadar başarılı olduğu söylenebilir.

İlintililik ve benzerlik ölçüm yöntemlerinin başarısının sadece katılımcıların kavram çiftlerine verdiği puanlar ile değerlendirilmesi yeterli değildir. Bulunan ilintililik ve benzerlik değerlerinin bir gerçek dünya problemini çözmek için kullanılması ve yöntemin bu problemin çözümünü ne kadar iyileştirdiği gözlemlenmelidir. Bir sonraki bölümde, geliştirilen ontoloji tabanlı anlamsal ilintililik yönteminin metin bağdaşıklığını ölçmek için nasıl kullanıldığı anlatılmıştır ve yöntemin başarısının değerlendirilme sonuçları sunulmuştur.

6 ONTOLOJİ TABANLI ANLAMSAL İLİNTİLİLİK ÖLÇÜM YÖNTEMİ KULLANILARAK METİN BAĞDAŞIKLIĞININ OTOMATİK OLARAK DEĞERLENDİRİLMESİ

Metin bağdaşıklığının otomatik olarak ölçülmesini hedefleyen yöntemin çözüm yaklaşımı Şekil 6.1’de gösterilmiştir. Buna göre, öncelikle dil bilimi alanındaki problem (metnin bağdaşıklığının otomatik değerlendirilmesi problemi), anlamsal web alanına aktarılır. Bu aktarım işlemi, doğal dil ile yazılmış metinlerdeki kavramların ilgili ontolojilerdeki kavramlarla eşleştirilmesi yoluyla yapılır. Problem, anlamsal web alanında çözülür. Çözüm, kavramlar arasındaki anlamsal ilintililiği, ontolojik patikalar yardımıyla ölçerek, bu değerler yardımıyla ard arda gelen cümleler arasındaki ilintililiği hesaplayarak metnin bağdaşıklığını otomatik olarak değerlendirmektedir. Son adımda ise, anlamsal web alanında üretilen çözüm tekrar dil bilimi alanına aktarılır. Cümleler arasındaki anlamsal ilintililik değerleri, doğal dil ile yazılmış metnin bağdaşıklığı konusunda bir fikir vermektedir.



Şekil 6.1: Metin bağdaşıklığının otomatik ölçüm yöntemi için çözüm yaklaşımı.

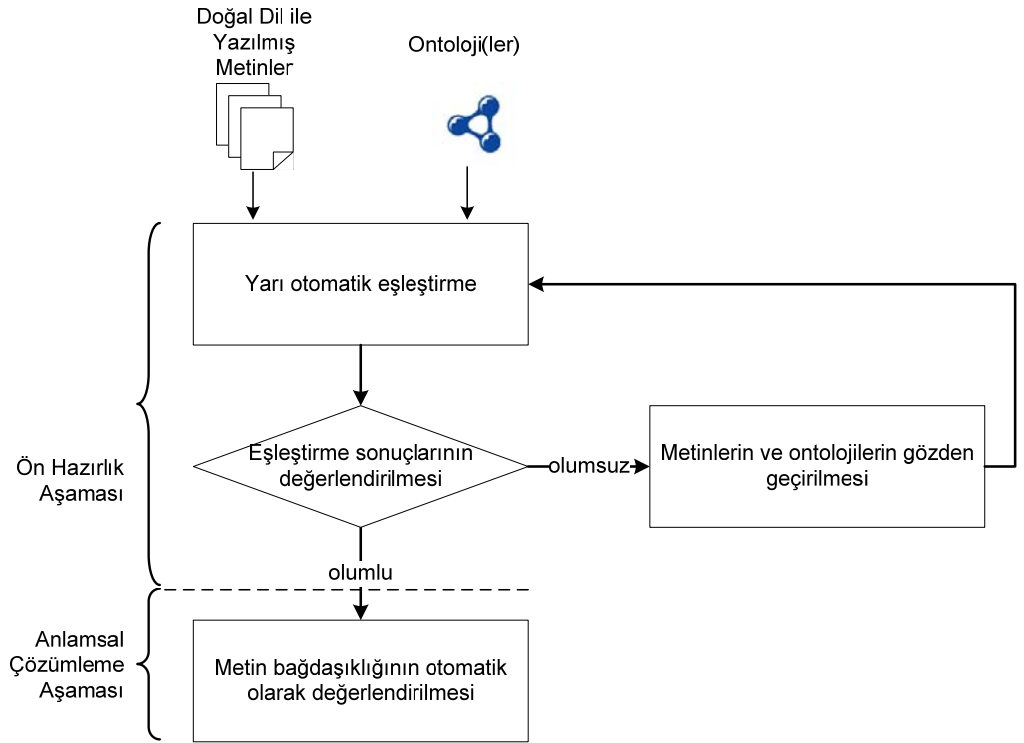
6.1 Metin Bağdaşıklığının Otomatik Olarak Değerlendirilmesi İçin Bir Yöntem

Metin bağdaşıklığının otomatik olarak değerlendirilmesi için önerilen yöntemin ana adımları Şekil 6.2’de gösterilmiştir.

- İlk adım, doğal dil cümleleriyle yazılmış metinlerdeki kelimelerin ve kelime öbeklerinin ontolojilerdeki kavramlarla, ilişkilerle ve özelliklerle eşleştirilmesi işlemini içerir. Eşleştirme işlemi, metinler yazıldıktan sonra bir yazılım yardımıyla yarı-otomatik olarak yapılabilir ya da yine bir araç yardımıyla metinler yazılırken kullanıcının denetiminde yapılabilir. Birinci seçenekte bahsedilen yazılım, otomatik eşleştirme için metin tabanlı, dil tabanlı ya da dil kaynağı tabanlı eşleştirme teknikleri üzerine yapılandırılabilir (Euzenat and Shvaiko, 2007). İkinci seçenek için ilkel bir örnek olarak “hipermetin aracılığıyla gereksinim mühendisliği” (Requirements Engineering through Hypertext – RETH) yaklaşımı gösterilebilir (Kaindl, 1993).
- İkinci adımda, eşleştirme sonuçları değerlendirilir. İçerik çözümlemesinin başarılı olabilmesi, eşleştirmenin belirli bir derecede başarılı olmasına bağlı olacaktır. Bundan dolayı eşleştirme oranının belirli bir seviyeye ulaşmış olması gerekmektedir. Bu seviye için öneriler, deneysel çalışmalar sonucunda elde edilebilir. Eşleşme oranı, basitçe, metindeki kavramların ne kadarının ontolojideki kavramlarla eşleştirildiğinin yüzdesi olarak ifade edilebilir. Eşleşme oranı tatmin edici bir seviyeye ulaşmamışsa bu durumda ontolojinin, metnin bağdaşıklığının değerlendirilmesi işleminde yeterli olamayacağı sonucuna varılabilir. Bu durumda metinler ve ontolojiler gözden geçirilir. Metinlerdeki yazım hataları düzeltilerek eşleştirme oranı yükseltilebilir. Metinlerde bulunup alan ontolojisinde bulunmayan kavramlar ontolojilere eklenebilir. Kullanılan ontolojilerin, metinlerin konu alanına uygun olmadığı durumlarda daha uygun ontolojiler seçilerek eşleştirme işlemi yeniden yapılabilir. Yarı otomatik eşleştirme ve eşleştirme sonuçlarının değerlendirilmesi adımları ön hazırlık aşamasını oluşturur.

- İçerik çözümlemesi için gerekli girdilerin hazırlandığı ön hazırlık aşamasından sonra anlamsal çözümleme aşamasına geçilir. Bu aşamanın anlamsal olarak nitelendirilmesinin nedeni, cümleler arasındaki ilintililik değerlerinin ölçülmesi için ontolojilerdeki kavramlar arasındaki anlamsal ilişkilerin kullanılmasıdır. Bu aşamada, ontolojilerdeki patikalar kullanılarak kavramlar arasındaki ilintililik değerleri ölçülür. Bu değerler kullanılarak ard arda gelen cümleler arasındaki ilintililik değerleri hesaplanır ve sonuç olarak bir metin bağdaşıklık değeri elde edilir.

Bu çalışmayı başlatan varsayım, bağdaşık bir metindeki ard arda gelen cümlelerde, ontolojilerde birbiri ile yakın ilişkisi olan kavramların bulunacağı yönündedir (Foltz et al., 1998; Lapata and Barzilay, 2005).

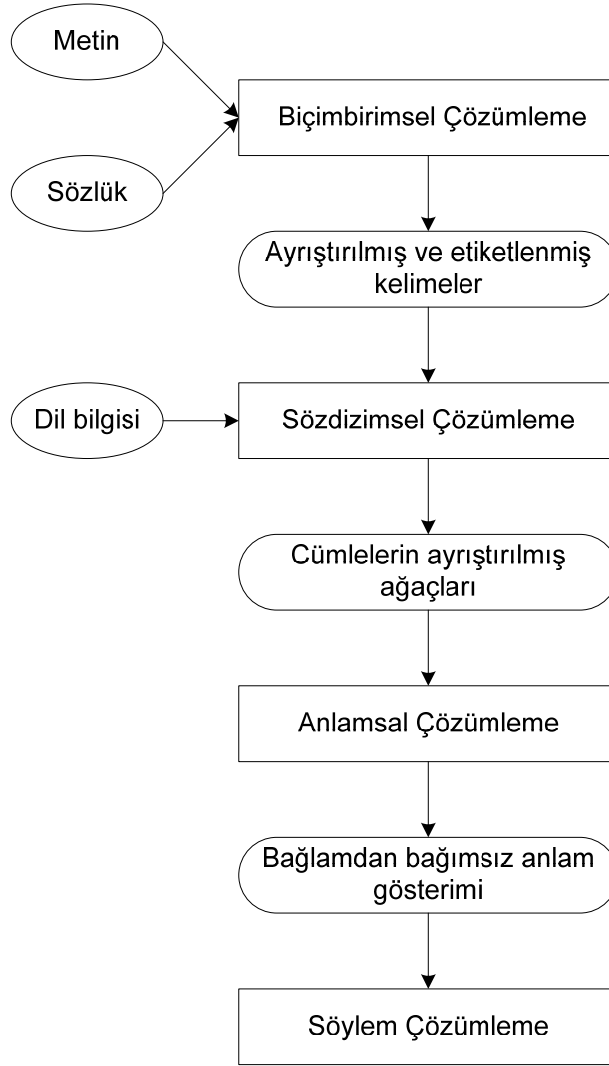


Şekil 6.2: Metin bağdaşıklığının otomatik olarak değerlendirilmesi için önerilen yöntemin ana adımları.

6.1.1 Yarı otomatik eşleştirme

Yarı otomatik eşleştirme adımı, doğal dil ile yazılmış metinlerdeki kelimeler ve kelime öbekleri, ontolojilerdeki kavramlarla, ilişkilerle ve

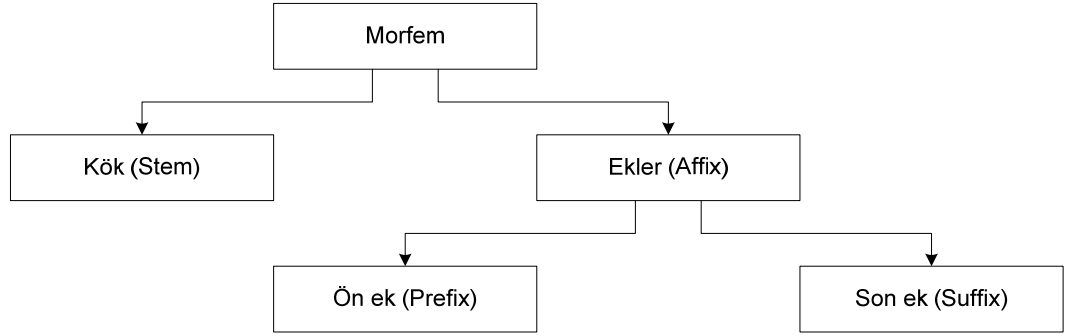
özelliklerle eşleştirilir. Bunun için öncelikle doğal dil ile yazılmış cümlelerin doğal dil anlama aşamalarının bazılarının geçmesi gerekmektedir. Şekil 6.3'te doğal dil anlamanın aşamaları gösterilmiştir. Buna göre bu çalışmada, metnin biçimbirimsel çözümlemesi ve POS (Part Of Speech) etiketlemesi yapılacaktır. Bundan sonra metindeki kavramlar, ontolojilerdeki kavramlarla eşleştirilecektir ve metin ile ilgili işlemler (ilintililik ve bağdaşıklık ölçümü) ontolojiler üzerinde yapılacaktır. Dolayısıyla metnin doğal dil anlamanın daha sonraki çözümleme aşamalarına tabi tutulması gerekmemektedir.



Şekil 6.3: Doğal dil anlamanın aşamaları (Allen, 1995).

Bu çalışma kapsamında, metinlerdeki doğal dil cümleleri biçimbirimsel çözümlemeye tabi tutulmaktadır. Biçimbirimsel çözümleme, kelimelerin morphem denilen en küçük anlamlı birimlere ayrıştırılması işlemine verilen addır. Şekil 6.4'te morfemin yapısı gösterilmektedir. Morphem, bir kökten ve eklerden

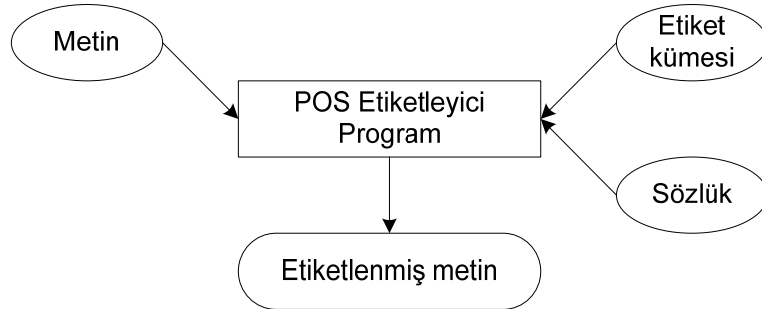
oluşmaktadır. Ekler, kullanılan doğal dilin yapısına göre, hem kökün önünde hem de arkasında olabilmektedir.



Şekil 6.4: Morfemin yapısı.

Biçimbirimsel çözümlemeden sonra, sözdizimsel çözümlemeden önce POS etiketlemesi denilen bir ara aşama bulunur. Bu aşamanın amacı, bir kelimenin ait olduğu kelime kategorisinin tespit edilmesi ve kelimelerin bu POS etiketine göre etiketlenmesidir. Örneğin, Türkçe bir metinde kelimelerin etiketlenmesi için “isim, fiil, zamir, sıfat, zarf, edat, bağlaç” gibi bir kelime kategorisi kullanılabilir.

Şekil 6.5’te, POS etiketleme işleminin girdileri ve çıktıları gösterilmiştir. Bir POS etiketleyici program, bir metin, bir etiket kümesi ve bir sözlüğü girdi olarak alır. Çıktı olarak ise etiketlenmiş bir metin verir.



Şekil 6.5: POS etiketleme.

Cümlelerin, biçimbirimsel çözümlemesi ve POS etiketlemesi yapıldıktan sonra POS etiketlerinden faydalanılarak cümlenin anlamına çok fazla katkıda bulunmayan işlev kelimeleri ve bağlaçlar ayıklanır. Bu kelimeler, etkisiz kelimeler olarak adlandırılır. Bu kelimelere örnek olarak, “acaba”, “ama”, “ancak” gibi kelimeler verilebilir. Bu kelimeler, bilgi erişimi çalışma alanında yapılan

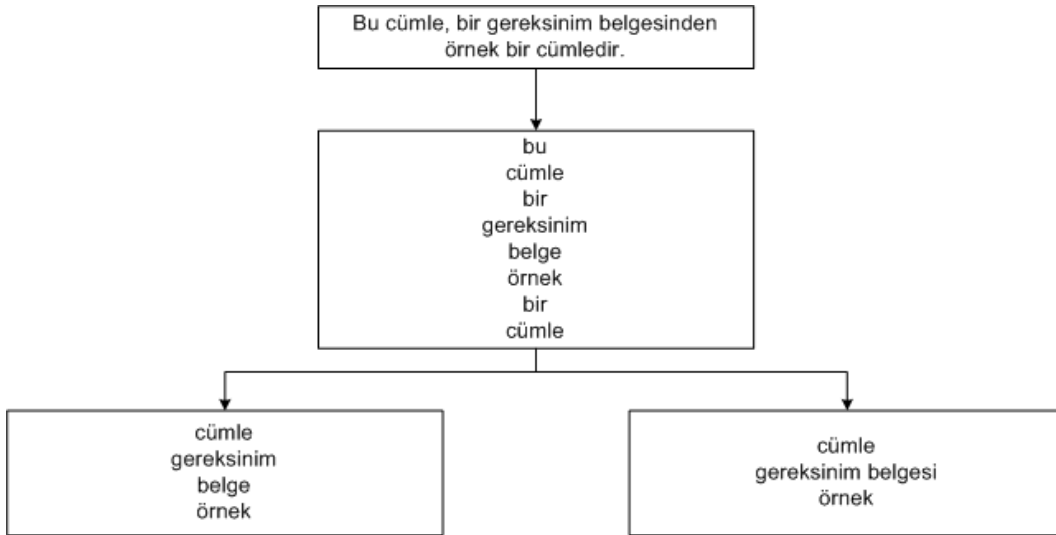
çalışmalarda da ihmal edilir. Örneğin, web sayfalarının dizinlenmesi aşamasında, dizin boyutunun çok büyümemesi için bu kelimeler dizin dışında bırakılır. Aynı şekilde kullanıcıların yazdıkları sorgulardaki bu tip kelimeler de arama işlemi yapılırken göz ardı edilir (Chakrabarti, 2003). Etkisiz kelimelerin ayıklanması sırasında çok anlamlı kelimelere dikkat edilmesi gerekmektedir. Örneğin, İngilizce’deki “can” kelimesi POS etiketine göre farklı anlamlar alabilir. Fiil POS etiketiyle işaretlenmiş “can” kelimesi bir arama sorgusu için faydasız olurken, isim POS etiketi ile işaretlenmiş “teneke kutu” anlamına gelen “can” kelimesi bir arama sorgusu için son derece önemli olabilir.

Şekil 6.6’da, cümlelerin kelime ve kavram vektörü olarak temsil edilme süreci gösterilmektedir. Cümle öncelikle eklerinden ayrılmış kelimelere ayrıştırılır ve daha sonra etkisiz kelimeler ayrıştırılarak cümle bir kelime ve kavram vektörü olarak aşağıdaki gibi temsil edilir.

$$\text{Doğal Dil Cümlesi (DDC)} = \{K_1, K_2, \dots, K_i\} ; 0 < i \leq n$$

Bir metin ise doğal dil cümlelerinden oluşur ve aşağıdaki gibi temsil edilir.

$$\text{Metin (M)} = \{DDC_1, DDC_2, \dots, DDC_i\} ; 0 < i \leq n$$



Şekil 6.6: Bir cümlelerin kelime ve kavram vektörü olarak temsili.

Metindeki cümleler kelime vektörü olarak temsil edildikten sonra, kelime vektörlerinde bulunan kelimeler ontolojilerdeki kavramlarla, ilişkilerle ve özelliklerle eşleştirilir. İlk önce sözdizimsel yani kelimelerin yazılışlarının birebir

aynı olduğu eşleştirmeler yapılır. Bu eşleştirme yapılırken kelime vektöründe ard arda gelen ve anlamlı bir kavrama karşılık gelen kelime öbeklerine de dikkat edilmelidir. Şekil 6.6’da gösterilen kelime vektöründe “gereksinim” ve “belge” kelimeleri bir araya gelerek “gereksinim belgesi” kavramını meydana getirir. Ontolojilerde “gereksinim belgesi” kavramına karşılık gelen bir sınıfın bulunma olasılığı vardır. Sözdizimsel eşleştirme yapılırken bu tip eşleştirmelere de dikkat edilir.

6.1.2 Eşleştirme sonuçlarının değerlendirilmesi

Eşleştirme sonuçlarının değerlendirilmesi aşamasında, kelime vektörlerindeki kavramların ontolojilerdeki kavramlarla, ilişkilerle ve özelliklerle yeterli derecede eşleşip eşleşmediğine karar verilir. Bunun için eşleşme oranına bakılarak karar verilir. Eşleşme oranı için (6.1) numaralı formül kullanılabilir. Bu formüldeki eşleşen kelime sayısı, ontolojilerdeki kavramlarla eşleştirilebilen kelime sayısını, toplam kelime sayısı ise etkisiz kelimeler hariç (Bkz. Bölüm 6.1.1, Sayfa 103) metindeki toplam kelime sayısını ifade etmektedir. Toplam kelime sayısına hangi kelime kategorilerine (isim, sıfat, fiil, vb.) ait kelimelerin dahil edileceği, yapılacak uygulamanın amacına göre değişebilir.

$$\text{Eşleşme Oranı (\%)} = \frac{\sum \text{Eşleşen kelime sayısı}}{\sum \text{Toplam kelime sayısı}} \times 100 \quad (6.1)$$

Eşleşme oranı ne kadar yüksek olursa, ontolojilerin, metinlerin içeriklerini o kadar iyi temsil ettiği söylenebilir (Malik et al., 2010).

6.1.3 Metinlerin ve ontolojilerin gözden geçirilmesi

Eşleştirme oranının düşük çıkması durumunda metin ve ontolojiler gözden geçirilebilir. Bu aşamanın amacı, eşleşme oranını yükseltmektir. Metindeki yazım hatalarının bulunarak düzeltilmesi gözden geçirme aşamasında yapılabilecek işlemlerden birisidir.

Ontolojilerdeki kavramların yetersiz olma durumu da bu aşamada gözden geçirilebilir. Metinlerde geçen fakat ontolojilerde bulunmayan kavramlar ontolojilere eklenerek ontolojiler genişletilebilir. Ontolojileri genişletmek için metinden ontoloji öğrenme teknikleri gibi yarı otomatik yaklaşımlar da

kullanılabilir. Ontolojinin, metin göz önüne alınarak genişletilmesi durumunda eşleşme oranı yükselecektir.

6.1.4 Metin bağdaşıklığının otomatik olarak değerlendirilmesi

Foltz et al. ve Lapata ve Barzilay bağdaşık yani anlaşılabilirliği yüksek bir metinde cümleler arasındaki anlamsal ilintililiğin ya da benzerliğin belirli bir seviyenin üzerinde olması gerektiğini belirtmişlerdir (Foltz et al., 1998; Lapata and Barzilay, 2005). Metinde ard arda gelen cümleler arasındaki anlamsal ilintililiğin ya da benzerliğin ölçülerek metnin bağdaşıklığının değerlendirildiği deneylerde başarılı sonuçlar elde edildiği görülmüştür. Bilişsel psikoloji alanında yapılan çalışmalar da bu kabulü desteklemektedir. Yapılan deneylerde okuyucuların cümle sonlarında gözlerini daha fazla sabitledikleri, bu sırada da cümlede elde ettikleri bilgiyi önceki bilgilerle bütünleştirdikleri tespit edilmiştir. Bu bulgular doğrultusunda araştırmacılar, bir metnin bağdaşıklığını otomatik olarak ölçebilmek için cümleler arasındaki bağlantıları değerlendirme yöntemini benimsemişlerdir. Foltz et al. cümleler arasındaki bağlantıları değerlendirmek için anlamsal ilintililik ölçümünü (Foltz et al., 1998), Lapata ve Barzilay ise anlamsal benzerlik ölçümünü tercih etmişlerdir (Lapata and Barzilay, 2005). Anlamsal benzerlik, anlamsal ilintililiğin özel bir durumudur. Anlamsal olarak birbirine benzer kavramlar, birbirleriyle ortak yönleri ile ilintilidir. Fakat iki kavram arasında farklı ilişkilerden kaynaklanan bağlantılar da bulunabilir. Aralarında parça-bütün ya da zıtlık ilişkisi bulunan iki kavram, anlamsal olarak benzer olmamasına rağmen anlamsal olarak ilintili olabilir. Bir okuyucu bir metni okurken, kavramlar arasında sadece benzerlik ilişkisi kurduğunu söylemek yanlış olacaktır. Okuyucunun kavramlar arasındaki çok çeşitli ilişkileri kurarak bir zihinsel temsil oluşturması, mevcut okuduğunu kavrama modellerine daha uygun düşmektedir. Bundan dolayı bu çalışmada, bir metnin bağdaşıklığını değerlendirirken ard arda gelen cümleler arasındaki anlamsal benzerliğin ölçülmesi yerine anlamsal ilintililiğin ölçülmesi tercih edilmiştir. Buna göre, bir M metninin bağdaşıklığını ölçmek için (6.2) numaralı formülün kullanılması önerilmektedir. Bu çalışmada da metin bağdaşıklığını ölçmek için (6.2) numaralı formül kullanılmıştır. Formüldeki “C” harfi “cümle”yi temsil etmektedir.

$$\text{Bağdaşıklık}(M) = \frac{\sum_{i=1}^{n-1} \text{Anlamsal İlntililik}(C_i, C_{i+1})}{n - 1} \quad (6.2)$$

Kavramlar arasındaki anlamsal ilintililik değerleri hesaplandıktan sonra, kavramların içinde geçtiği cümleler arasındaki ilintililiği hesaplamak için çeşitli yöntemler vardır. Örneğin, Okazaki et al.'nin kullandığı cümle benzerlik formülü (6.3)'te gösterilmiştir (Okazaki et al., 2003).

$$\text{Benzerlik}(C_i, C_j) = \sum_{K_i \in C_i} \sum_{K_j \in C_j} \frac{\text{Benzerlik}(K_i, K_j)}{\sqrt{|C_i||C_j|}} \quad (6.3)$$

(6.3) numaralı formüle göre, iki cümledeki tüm kavramların birbirleriyle olan benzerlik hesaplandıktan sonra elde edilen sayı, iki cümleyi oluşturan kavram sayısının geometrik ortalamasına bölünür.

Lapata ve Barzilay, iki cümle arasındaki benzerliği ölçmek için Okazaki et al.'nin kullandığı formüle benzer bir formül olan (6.4) numaralı formülü kullanmıştır (Lapata and Barzilay, 2005).

$$\text{Benzerlik}(C_i, C_j) = \frac{1}{|C_i| \cdot |C_j|} \sum_{\substack{1 \leq m \leq |C_i| \\ 1 \leq n \leq |C_j|}} \max \text{Benzerlik}(K_{i_m}, K_{j_n}) \quad (6.4)$$

$$(K_{i_m} \in C_i \text{ ve } K_{j_n} \in C_j)$$

(6.4) numaralı formülde, kavramlar arasındaki benzerlik hesaplama yöntemi, ilintililik hesaplama yöntemi ile değiştirildiğinde, bu çalışma kapsamında kullanılan, cümleler arasındaki anlamsal ilintililik hesaplama formülü (6.5) elde edilir.

$$\begin{aligned} \text{Anlamsal İlintililik}(C_i, C_j) \\ = \frac{1}{|C_i| \cdot |C_j|} \sum_{\substack{1 \leq m \leq |C_i| \\ 1 \leq n \leq |C_j|}} \max \text{Anlamsal İlintililik}(K_{i_m}, K_{j_n}) \end{aligned} \quad (6.5)$$

$$(K_{i_m} \in C_i \text{ ve } K_{j_n} \in C_j)$$

(6.5) numaralı formüle göre, üç kelimededen oluşan bir cümle ile iki kelimededen oluşan iki cümle arasındaki anlamsal ilintililik, (6.6) numaralı formülde gösterildiği şekilde hesaplanmaktadır. Bu formülde “AI”, “Anlamsal İlintililik” kavramının kısaltması olarak kullanılmıştır. Ayrıca iki kelime arasında birden fazla anlamsal ilintililik değeri hesaplanırsa bu değerlerin en büyüğü formülde kullanılmaktadır.

$$AI(C_1, C_2) = \frac{AI(K_{1_1}, K_{2_1}) + AI(K_{1_1}, K_{2_2}) + AI(K_{1_2}, K_{2_1}) + AI(K_{1_2}, K_{2_2}) + AI(K_{1_3}, K_{2_1}) + AI(K_{1_3}, K_{2_2})}{3 \times 2} \quad (6.6)$$

$$(C_1 = \{K_{1_1}, K_{1_2}, K_{1_3}\} \text{ ve } C_2 = \{K_{2_1}, K_{2_2}\})$$

(6.5) numaralı formülde, iki cümlelerin kelimeleri arasında hiçbir ilintililik olmaması durumunda formülün payı 0 olacağı için, $AI(C_1, C_2)$ 'nin alacağı en küçük değer 0'dır. İki cümledeki tüm kelime çiftleri arasındaki anlamsal ilintililik değerinin 1 olması durumunda ise $AI(C_1, C_2)$ en yüksek değerini alacaktır ve iki cümle arasındaki anlamsal ilintililik değeri 1 olacaktır. (6.6) numaralı formülde görüleceği gibi, payda bulunan tüm anlamsal ilintililik değerlerinin 1 olması durumunda payın değeri 6 olacaktır ve bu payın alabileceği en yüksek değerdir. Payda değeri de 6 olduğu için iki cümle arasındaki en yüksek ilintililik değeri 6'nın 6'ya bölünmesiyle 1 olarak elde edilecektir. (6.2) numaralı formülde, cümleler arasındaki en yüksek anlamsal ilintililik değeri 1 olacağı ve n cümleden oluşan bir metinde $(n - 1)$ kere anlamsal ilintililik hesabı yapılacağı için payın alabileceği en yüksek değer $(n - 1)$ 'dir. Paydanın değeri de $(n - 1)$ olduğu için bir metnin en yüksek bağdaşıklık değeri 1 olacaktır. Ardışık cümleler arasındaki tüm anlamsal ilintililik değerlerinin 0 olması durumunda ise pay 0 olacaktır. Bu durumda metnin bağdaşıklığı da en düşük değeri olan 0 değerini alacaktır.

6.2 Uygulama

Bu tez kapsamında, metin bağdaşıklığını otomatik olarak değerlendirmek için bir uygulama geliştirilmiştir. Uygulama, Java programlama dili kullanılarak geliştirilmiştir. Uygulamada çeşitli araçlar ve kütüphaneler kullanılmıştır.

6.2.1 Uygulamada kullanılan araçlar

Uygulamanın, doğal dil ile yazılmış metinlerin biçimbirimsel çözümlemesini ve POS etiketlemesini yapan ve metinlerdeki kavramların ontolojilerdeki kavramlarla eşleştirildiği bölümü GATE (General Architecture for Text Engineering) (Cunningham et al., 2002; Bontcheva et al., 2004) aracı kullanılarak geliştirilmiştir. Metin bağdaşıklığının otomatik olarak ölçülmesi için kullanılan ve bu çalışma kapsamında geliştirilen anlamsal ilintililik ölçüm yönteminin başarısının diğer yöntemlerle karşılaştırılabilmesi için SimPack (Bernstein et al., 2005) kütüphanesi kullanılmıştır. SimPack ve Java WordNet::Similarity kütüphaneleri kullanılarak ontolojilerdeki kavramlar arasındaki benzerlikler, Resnik'in (Resnik, 1985), Lin'in (Lin, 1998), Jiang ve Conrath'ın (Jiang and Conrath, 1997), Wu ve Palmer'ın (Wu and Palmer, 1994) ve Leacock ve Chodorow'un (Leacock and Chodorow, 1998) önerdiği yöntemlerle ölçülmüştür. Doğal dil ile yazılmış metinlerin anlamsal etiketlemesinin yapılabilmesi için gerekli ontolojiler, anlamsal web için geliştirilmiş arama motorları Swoogle (Ding et al., 2004) ve Watson (d'Aquin et al., 2007) kullanılarak bulunmuştur. Uygulamada kullanılan araçların ve kütüphanelerin sürümleri ve web sayfaları Ek 7'de gösterilmiştir.

6.2.1.1 GATE

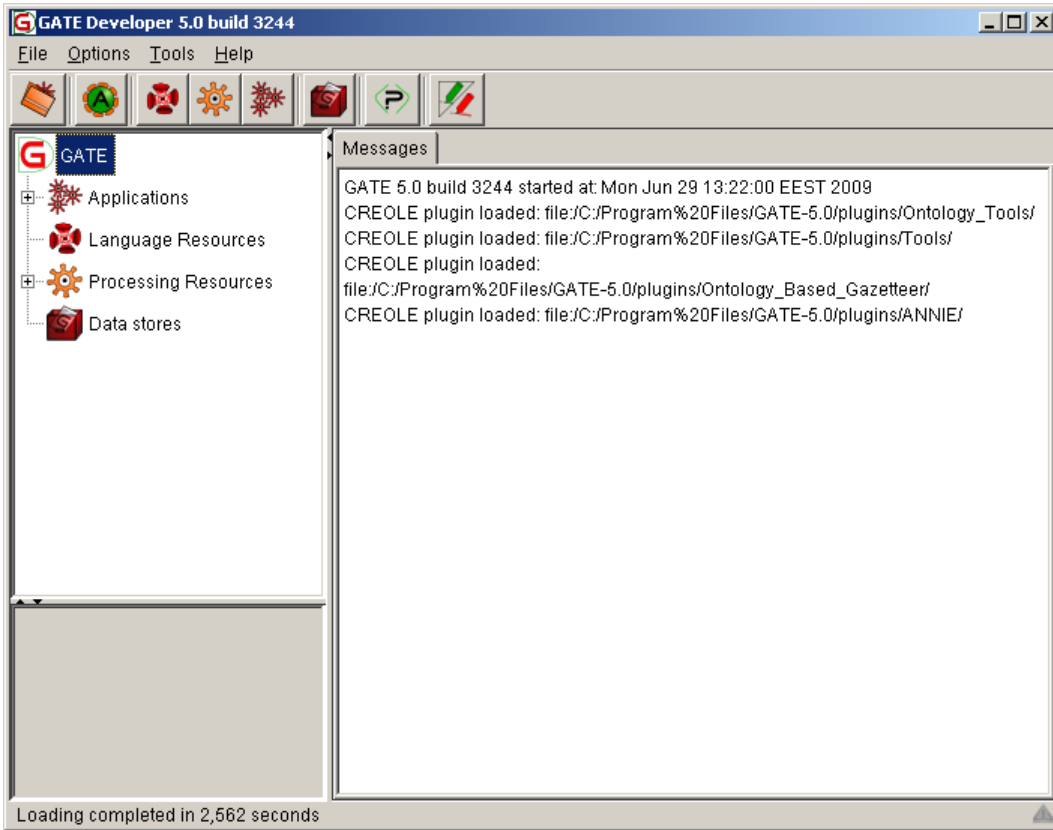
GATE (General Architecture for Text Engineering), dil mühendisliği için tasarlanmış bir mimari, çerçeve ve geliştirme ortamı olarak tanımlanabilir. Bir mimari olarak GATE, bir dil mühendisliği sisteminin organizasyonunu ve değişik sorumlulukların değişik bileşenlere atanmasını tanımlar. Bir çerçeve olarak GATE, dil mühendisliği bileşenleri için yeniden kullanılabilir gerçekleştirmeler sunar. Bir geliştirme ortamı olarak GATE, yeni bir dil mühendisliği sistemi geliştirme zamanını en aza indirmek için bazı araçlar sunar (Cunningham et al., 2002).

GATE, bileşen tabanlı bir mimariye sahiptir. GATE bileşenleri üç çeşittir (Cunningham et al., 2002):

- Dil kaynakları; sözlük, derlem, ontoloji gibi varlıkları simgeler.
- İşlem kaynakları; ayrıştırıcı, üreteç, n-gram modelleyici gibi algoritmik varlıkları simgeler.

- Görsel kaynaklar; kullanıcı arayüzündeki gösterim ve düzenleyici bileşenleri simgeler.

Şekil 6.7’de GATE’in kullanıcı arayüzü gösterilmektedir. Arayüzde de görülebileceği gibi sol taraftaki menüde dört seçenek bulunmaktadır. Dil kaynakları başlığı altında, belge (html formatında bir web sayfası, bir metin, vs.), belgelerden oluşan derlem ya da ontoloji kaynakları yer alabilir. İşlem kaynakları başlığı altında, dil kaynakları üzerinde çeşitli işlemler yapabilen bileşenler bulunur. İşlem kaynaklarına örnek olarak parçalayıcı, cümle ayırıcı, POS etiketleyici, biçimbirimsel çözümleyici gibi bileşenler verilebilir. Uygulama başlığı altında ise oluşturulan işlem kaynaklarının birleşimiyle geliştirilen uygulamalar bulunur. Uygulamalarda işlem kaynaklarının hangi sırada çalıştırılacağı kullanıcı tarafından belirtilir. Bazı işlem kaynakları çalıştırılmadan önce başka işlem kaynaklarının çalıştırılması zorunludur. Uygulamalar oluşturulurken işlem kaynaklarının ön gerekliliklerine dikkat edilmelidir. Dil kaynakları, veri depolarının içine kaydedilerek saklanabilir. Bu şekilde saklanan dil kaynakları veri depolarından tekrar geri yüklenebilir.



Şekil 6.7: GATE kullanıcı arayüzü.

6.2.1.2 SimPack

SimPack, Java dili ile geliştirilmiş bir benzerlik ölçüm kütüphanesidir (Bernstein et al., 2005). SimPack kütüphanesi kullanılarak bir ontolojideki kavramlar arasındaki benzerlik ölçülebilmektedir. Bu kütüphane içinde çeşitli çalışmalarda önerilen ölçüm yöntemleri gerçekleştirilmiştir.

Bu çalışma kapsamında, SimPack kütüphanesi kullanılarak ontolojilerdeki kavramlar arasındaki benzerlikler, Resnik'in (Resnik, 1985), Lin'in (Lin, 1998), Jiang ve Conrath'ın (Jiang and Conrath, 1997), Wu ve Palmer'ın (Wu and Palmer, 1994) ve Leacock ve Chodorow'un (Leacock and Chodorow, 1998) önerdiği yöntemlerle ölçülmüştür.

6.2.1.3 Swoogle

Swoogle (Ding et al., 2004), anlamsal web için geliştirilmiş bir arama motorudur. Swoogle, WWW'deki anlamsal web belgelerini dizinleyerek bu belgeler üzerinde arama yapılmasını sağlar.

Bu çalışma kapsamında, doğal dil ile yazılmış metinlerdeki ard arda gelen cümlelerin arasındaki anlamsal ilintililiği hesaplayabilmek için, ilk cümledeki her kelime ile bu cümleden sonra gelen cümledeki her kelime arasındaki anlamsal ilintililiğin hesaplanması gerekmektedir. Bu anlamsal ilintililik değerlerinin hesaplanması için her bir kelime çiftini barındıran ontolojilere gereksinim duyulmaktadır. Her kelime çifti için gereksinim duyulan ontolojiler Swoogle arama motoru aracılığıyla aranmıştır. Bu işlem Swoogle'ın sağladığı web servisleri kullanılarak program aracılığıyla yapılmıştır.

6.2.1.4 Watson

Watson, İnternet ortamındaki ontolojilere erişimi sağlayan, gelecek nesil anlamsal web uygulamalarının geliştirilmesini desteklemeyi amaçlayan bir uygulamadır (d'Aquin et al., 2007).

Bu çalışma kapsamında Watson, Swoogle arama motorunun da kullanıldığı gibi, doğal dil ile yazılmış metinleri anlamsal olarak etiketlemek için gerekli ontolojileri bulmak için kullanılmıştır. Bu işlem Watson'ın sağladığı web servisleri kullanılarak program aracılığıyla yapılmıştır. Çalışmada iki arama

motorunun kullanılmasının nedeni, metinleri etiketlemek için kullanılacak daha fazla sayıda ontolojiye ulaşma olasılığını ve metindeki etiketlenen kelime sayısını arttırmaktır.

6.2.1.5 Java WordNet::Similarity

WordNet::Similarity, çeşitli anlamsal ilintililik ve benzerlik ölçüm yöntemlerini gerçekleştiren bir kütüphanedir. Kütüphane ilk olarak Pedersen et al. tarafından Perl programlama dili kullanılarak oluşturulmuştur (Pedersen et al., 2004). Bu çalışmada, WordNet::Similarity kütüphanesinin Sussex Üniversitesi'nde Java programlama dili ile geliştirilen sürümü¹⁴ kullanılmıştır. Anlamsal ilintililik ve benzerlik ölçümleri yapılırken WordNet kavramlar dizini (Fellbaum, 1998) kullanılmaktadır. Ayrıca ölçüm için bir derleme gereksinim duyan yöntemler için kütüphane British National Corpus (World Sürümü), Brown Corpus, Penn Treebank (Sürüm 2) ve SemCor derlemelerini kullanabilmektedir.

Bu çalışma kapsamında, Java WordNet::Similarity kütüphanesi kullanılarak ontolojilerdeki kavramlar arasındaki benzerlikler, Resnik'in (Resnik, 1985), Lin'in (Lin, 1998), Jiang ve Conrath'ın (Jiang and Conrath, 1997), Wu ve Palmer'ın (Wu and Palmer, 1994) ve Leacock ve Chodorow'un (Leacock and Chodorow, 1998) önerdiği yöntemlerle ölçülmüştür. Resnik'in (Resnik, 1985), Lin'in (Lin, 1998) ve Jiang ve Conrath'ın (Jiang and Conrath, 1997) yöntemleri kullanılarak yapılan ölçümler British National Corpus (World Sürümü), Brown Corpus, Penn Treebank (Sürüm 2) ve SemCor derlemleri için ayrı ayrı yapılmıştır.

6.2.2 Uygulamanın ayrıntıları

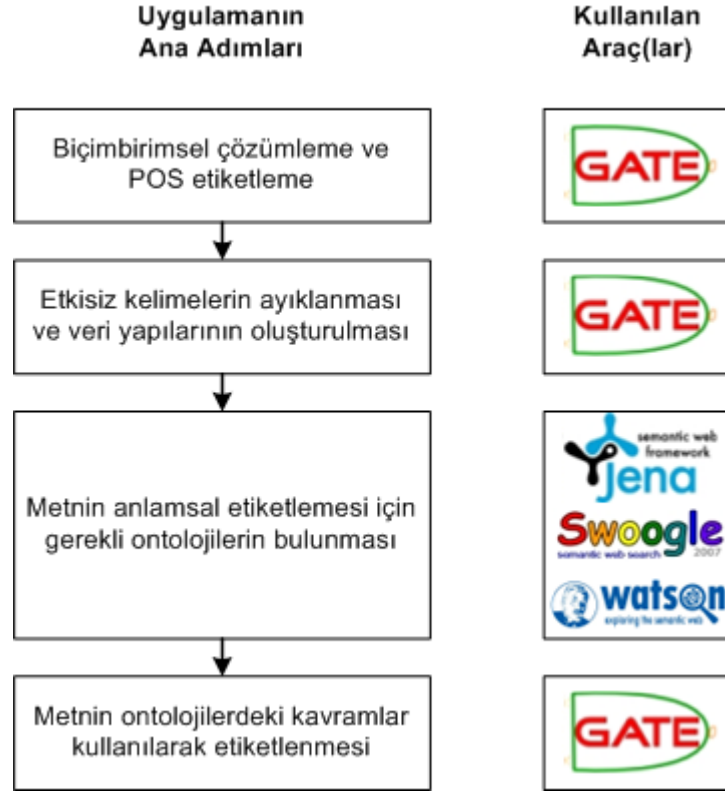
Bölüm 6.1'de anlatılan ve Şekil 6.2'de gösterilen yöntemle tutarlı olarak, ön hazırlık ve anlamsal çözümleme aşamaları için birer uygulama geliştirilmiştir.

6.2.2.1 Ön hazırlık aşaması için geliştirilen uygulama

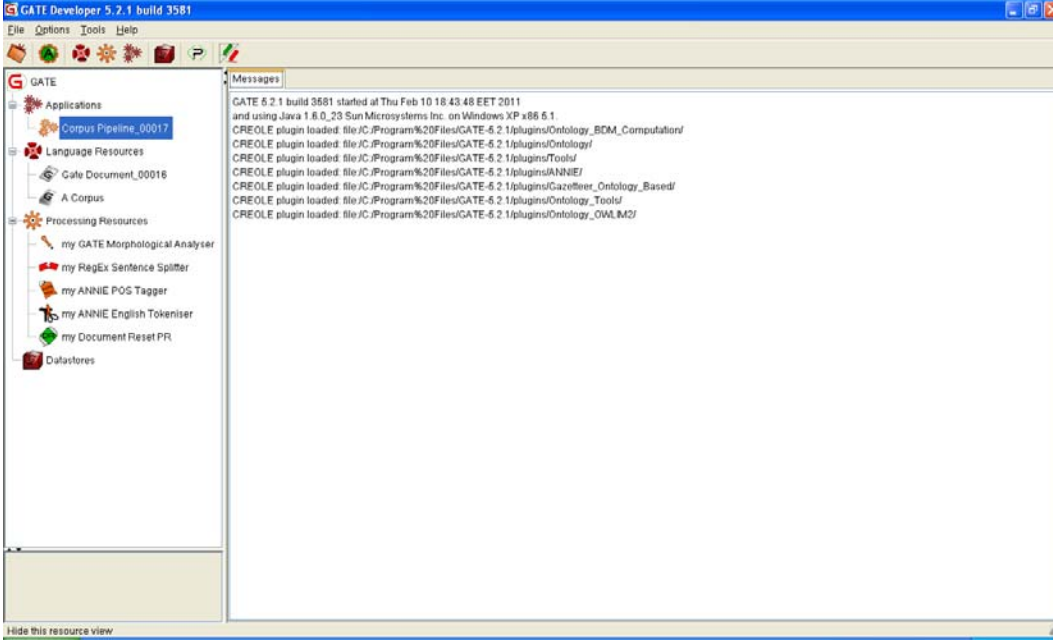
Ön hazırlık aşaması için geliştirilen uygulamanın ana adımları ve bu uygulamada kullanılan araçlar Şekil 6.8'de gösterilmiştir.

¹⁴ <http://www.cogs.susx.ac.uk/users/drh21/> (Erişim tarihi: 03.02.2011)

Doğal dil ile yazılmış metinlerin biçimbirimsel çözümlemesi ve POS etiketlemesi için GATE aracı kullanılmıştır. Şekil 6.9'da, uygulamadaki bileşenler GATE kullanıcı arayüzü aracılığıyla görüntülenmiştir. Görüntülenen bileşenler GATE kullanıcı arayüzü kullanılarak değil, tamamen GATE uygulama programlama arayüzü kullanılarak oluşturulmuştur. GATE arayüzü sadece görüntüleme ve test amaçlı olarak kullanılmaktadır.



Şekil 6.8: Ön hazırlık aşaması için geliştirilen uygulamanın ana adımları ve bu uygulamada kullanılan araçlar.



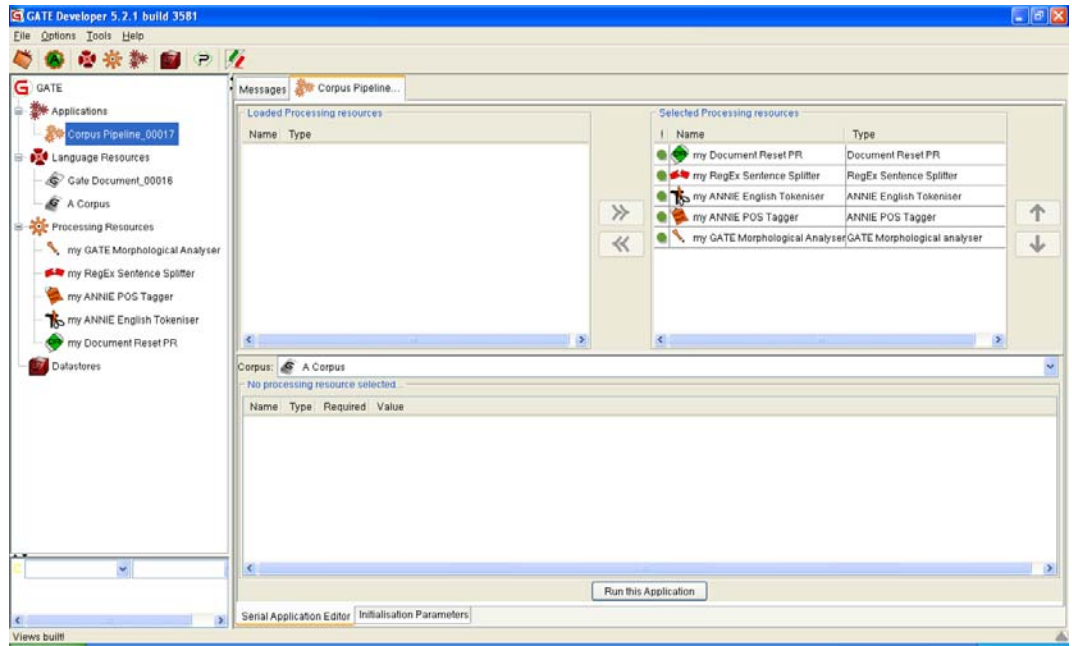
Şekil 6.9: Ön hazırlık aşaması için geliştirilen uygulamadaki bileşenlerin GATE kullanıcı arayüzündeki görünümü.

Şekil 6.9’da, dil kaynakları (language resources) başlığı altında iki kaynak görülmektedir. İkinci kaynak bir derlemdir. Bu derlem, birinci kaynak olan belgeyi içermektedir. Bir derlem birden fazla belge içerebilir. İşlem kaynakları da kendisine verilen bir derlem üzerinde işlem yapmaktadır.

Şekil 6.9’da, işlem kaynakları (processing resources) başlığı altında beş tane işlem kaynağı görülmektedir. Biçimbirimsel çözümleyici (GATE Morphological Analyser), parçalanmış bir GATE belgesini girdi olarak alır. Her parça ve o parçanın POS etiketi değerlendirilerek parçanın kökü ve ekleri ayrıştırılır. Her parçanın kökü ve ekleri o parçanın etiketleri arasına eklenir (Cunningham et al., 2009). Cümle ayırıcı (RegEx Sentence Splitter), metni cümlelere ayırmaktadır (Cunningham et al., 2002). Cümle ayırıcı bileşeni etiketleyici için gereklidir. Etiketleyici (POS Tagger), literatürde Brill etiketleyicisi olarak bilinen etiketleyicinin değiştirilmiş bir versiyonudur. Etiketleyici, her kelimeyi ya da sembolü POS etiketleriyle etiketler. Parçalayıcı (tokeniser), metni; sayı, noktalama işareti, sembol ve farklı tipte kelimeler olmak üzere parçalara bölmektedir. Belge sıfırlayıcı (Document Reset PR), bir GATE belgesinin ilk durumuna getirilmesini sağlar. Yani belgedeki her türlü etiket vs. temizlenir. Bu işlem kaynağı genellikle bir uygulamanın en başında yer alır (Cunningham et al., 2009).

Şekil 6.9’da, uygulamalar (applications) başlığı altında ise işlem kaynaklarından oluşan uygulamalar bulunur. Örnek uygulamada, bir önceki paragrafta açıklanan işlem kaynaklarının birleşimiyle bir GATE uygulaması elde edilmiştir.

Şekil 6.10’da, uygulamada kullanılan işlem kaynakları gösterilmektedir. İşlem kaynaklarının listelenme sırası aynı zamanda işlem kaynaklarının çalıştırılma sıralarıdır. Uygulamada öncelikle belge sıfırlayıcı çalıştırılır. Böylece belgede herhangi bir etiket bulunmadığından emin olunur. Daha sonra sırasıyla cümle ayırıcı ve parçalayıcı çalıştırılır. Bu iki bileşen POS etiketleyicisinin ön koşuludur. POS etiketleyicisi çalıştırdıktan sonra biçimbirimsel çözümleyici çalıştırılır. POS etiketleyicisi de biçimbirimsel çözümleyicinin ön koşuludur. Çünkü biçimbirimsel çözümleyici, parçaların POS etiketleri üzerinde işlem yapmaktadır.



Şekil 6.10: Ön hazırlık aşaması için geliştirilen uygulamadaki işlem kaynakları.

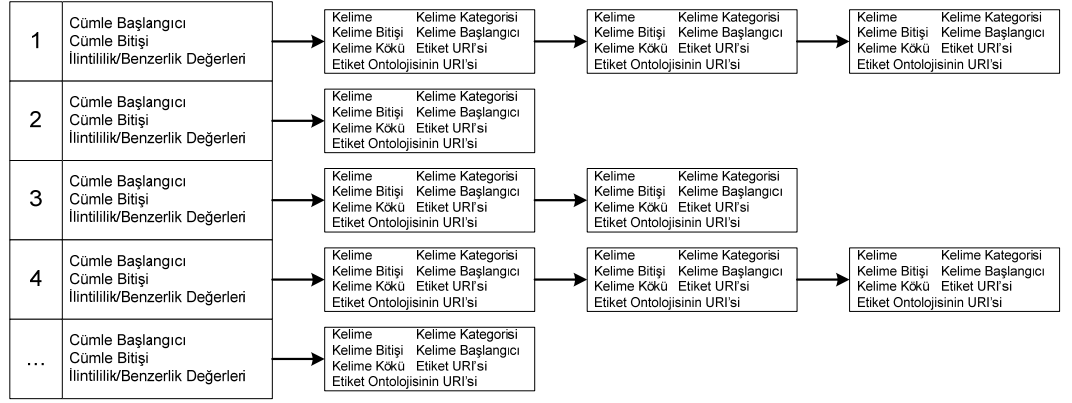
Biçimbirimsel çözümleme ve POS etiketleme işlemi tamamlandıktan sonra etkisiz kelimelerin ayıklanması ve uygulamada kullanılan veri yapılarının oluşturulması adımına geçilir. Bu adımda, cümlelerin anlamına çok fazla katkıda bulunmayan işlev kelimeleri ve bağlaçlar ayıklanır. Doğal dil ile metinlerdeki isimler, sıfatlar ve fiiller ontolojilerdeki kavramlarla eşleştirmek için dikkate alınmıştır. Çünkü bu kategorilere dahil olmayan, cümlede bağlaç (ve, veya gibi), ilgeç (ile, için gibi), belirteç (çok, pek gibi) gibi görevler yapan kelimeler,

kullanılan yöntem açısından bakıldığında cümlelerin anlamına herhangi bir katkıda bulunmamaktadır. Çünkü kullanılan yöntem, doğal dili işleyerek anlamsal bir çözümleme yapmamaktadır. Yöntem, doğal dil ile yazılmış bir metindeki bazı kelimeleri ontolojilerdeki kavramlarla eşleştirerek ve ontolojilerdeki ilişkileri kullanarak metnin bağdaşıklığı üzerinde bir tahmin yapmayı hedeflemektedir. Dolayısıyla cümlelerdeki kelimelerden, sadece ontolojilerdeki kavramlarla eşleşebilecek olan kelime kategorilerinin dikkate alınması mantıklıdır. Ontolojiler, bir konu alanındaki kavramları, kavramların özelliklerini ve bu kavramlar arasındaki ilişkileri temsil ettiğinden dolayı, cümlelerde bu ontolojik terimlere karşılık gelebilecek isimler, sıfatlar ve fiiller dikkate alınmıştır. GATE aracının sağladığı etiket listesinden (Cunningham et al., 2009) Çizelge 6.1’de gösterilenler ontolojilerdeki kavramlarla eşleştirmek için kullanılmaktadır.

Çizelge 6.1: Ontolojilerdeki kavramlarla eşleştirilen kelime kategorileri.

Kelime Kategorisi	Açıklama
NN, NNP, NNPS, NNS, NP, NPS	Tekil, çoğul ya da topluluk cins ya da özel isimleri
JJ, JJR, JJS	Sıfatlar
VBD, VBG, VBN, VBP, VB, VBZ	Her türlü kipe sahip fiiller

Cümlelerde, ontolojilerdeki kavramlarla eşleştirilebilecek kelimeler belirlendikten sonra metin Şekil 6.11’deki veri yapısıyla bellekte temsil edilir. Metindeki cümle sayısı kadar bir dizi oluşturulur ve bu dizinin her düğümünde cümlelerin başlangıç ve bitiş yerleri saklanır. Bu düğümlerde ayrıca, anlamsal çözümleme aşamasında hesaplanan bağdaşıklık değerlerinin saklanabilmesi için yer ayrılmıştır. Her cümledeki, ontolojilerdeki kavramlarla eşleşme olasılığı bulunan kelimeler de cümle düğümlerinin içinde saklanır. Her kelime düğümünde, kelimenin metinde geçen hali, kelimenin kökü (Bkz. Şekil 6.4, Sayfa 105), kelimenin kategorisi, kelimenin cümle içindeki başlangıç ve bitiş yeri saklanır. Anlamsal çözümleme aşamasında bir kelime ile ontolojideki bir kavramın eşleşmesi durumunda bu ontolojinin URI’si ve eşleşen kavramın URI’si yine kelime düğümü içinde saklanır.

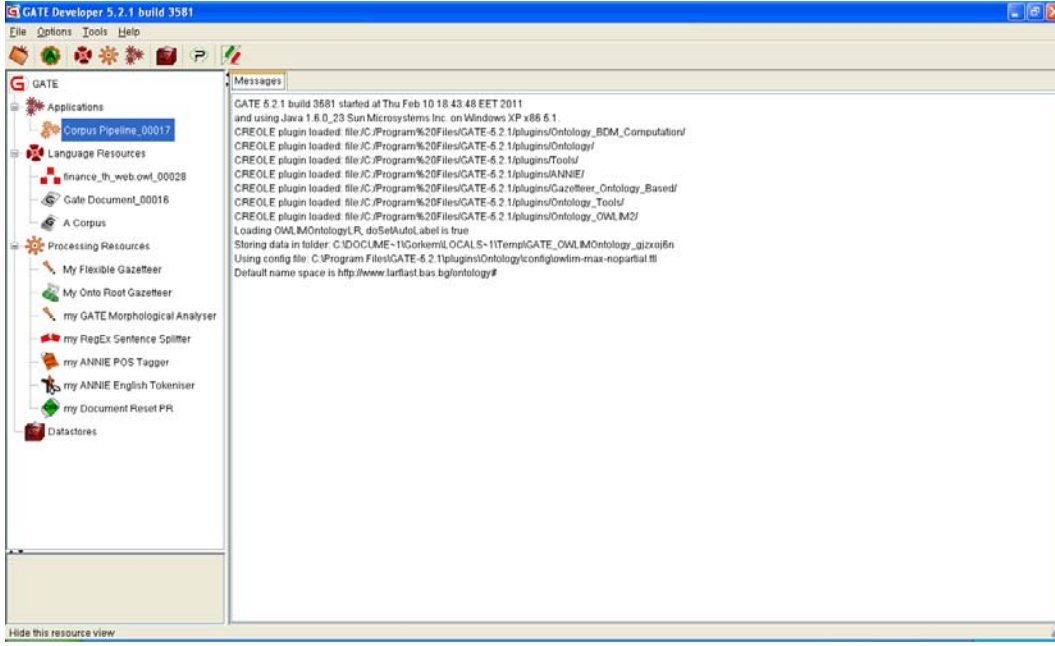


Şekil 6.11: Metnin ontolojilerle eşleştirilmeden önce bellekte temsil edildiği veri yapısı.

Ontolojilerdeki kavramlarla eşleştirilebilecek kelimeler belirlendikten sonra ardışık cümlelerin kelimeleri arasındaki ilişkiyi hesaplayabilmek için iki cümleden birer kelime alınarak bu iki kelimenin bulunduğu ontolojiler Swoogle ve Watson arama motorlarında aranır. Gelen sonuçlar içinden sadece `rdf`, `rdfs`, `daml` ve `owl` uzantılı ontolojiler dikkate alınır. Bu ontolojiler içinde, istenen kelime çiftleriyle aynı isme sahip kaynakların (özne ya da nesne) bulunup bulunmadığı kontrol edilir. Bunun yapılmasının nedeni, Swoogle ve Watson arama motorlarından gelen ontolojilerin yorum satırlarında ya da kavram isimlerinin içinde istenen kelimelerin geçme olasılığıdır. Swoogle ve Watson arama motorlarından gelen bu tip ontolojiler, metindeki kelimelerle eşleşemeyeceği için elenir.

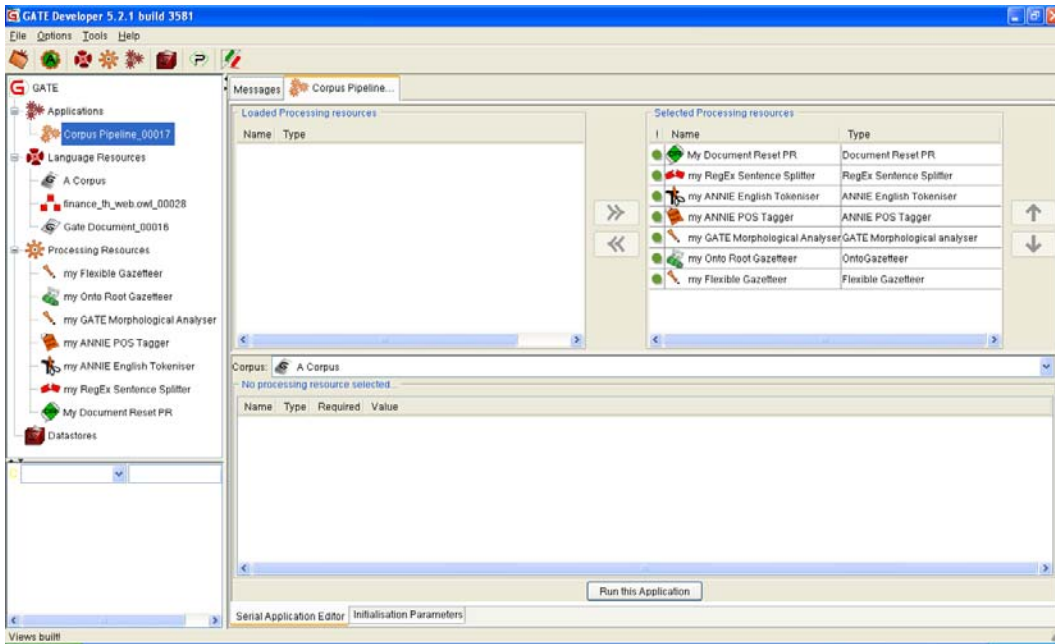
Metnin anlamsal etiketlemesi için kullanılacak ontolojiler belirlendikten sonra bu ontolojiler ve metin gözden geçirilir. Bulunan ontolojilerin GATE aracı tarafından açılıp açılmadığı kontrol edilir. Metinlerde herhangi bir yazım yanlışı varsa bunlar düzeltilir. Böylece hem eşleşme oranının yükselmesi hem de anlamsal çözümleme aşamasında kullanılan uygulamanın sağlıklı olarak çalışması sağlanır.

Doğal dil ile yazılmış metinlerin ontolojilerdeki kavramlarla eşleştirilmesi için GATE aracı kullanılmıştır. Şekil 6.12'de, uygulamadaki bileşenler GATE kullanıcı arayüzü aracılığıyla görüntülenmiştir.



Şekil 6.12: Anlamsal etiketleme için gerekli bileşenlerin GATE kullanıcı arayüzündeki görünümü.

Şekil 6.12’de görüldüğü gibi, anlamsal etiketleme için uygulamaya ontoloji dil kaynağı ve atlas işlem kaynakları (Flexible Gazetteer ve Onto Root Gazetteer) eklenmiştir. Atlas (gazetteer), şehir isimleri, haftanın günleri gibi listelerden oluşur (Cunningham et al., 2002). Listeler, sadece varlıkları değil aynı zamanda çeşitli göstergeleri de (ör: Ltd.) içerir. Bu listeler sonlu durum makinelerine dönüştürülür ve metinlerdeki parçalarla eşleştirilebilir duruma getirilir.



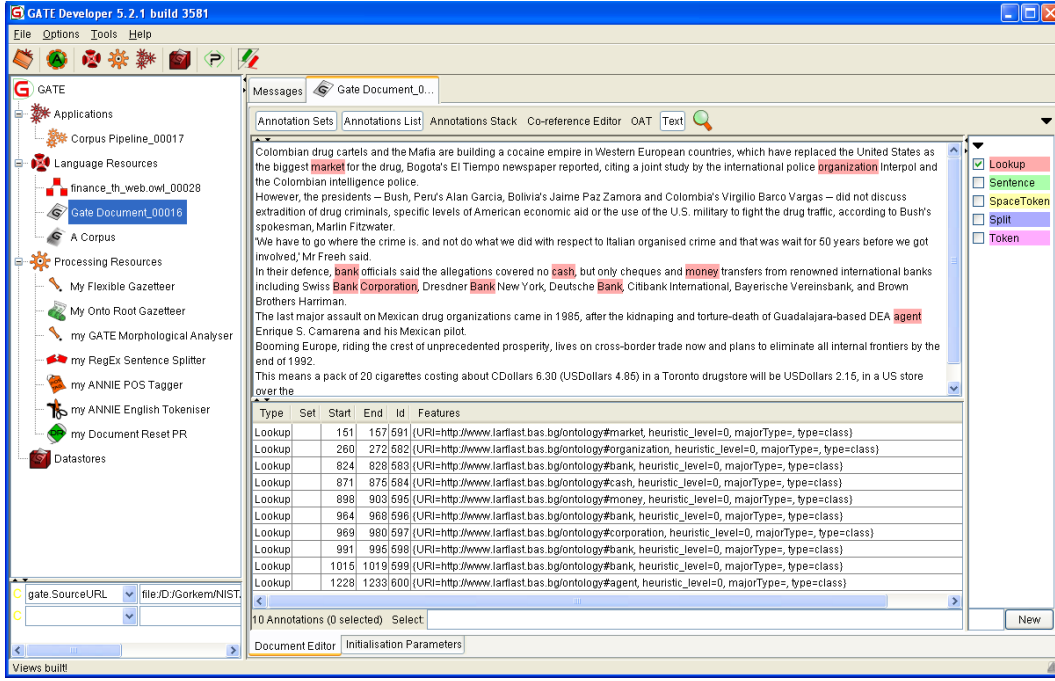
Şekil 6.13: Anlamsal etiketleme için gerekli işlem kaynakları ve bu kaynakların çalışma sırası.

Şekil 6.13’te, uygulamada kullanılan işlem kaynakları gösterilmektedir. Metnin anlamsal olarak etiketlenmesi için gerekli ontolojilerin bulunmasından sonra, metindeki kelimeleri ontolojilerdeki kavramlarla eşleştirmek için atlaslar uygulamaya eklenmiştir.

Uygulamadaki ilk beş işlem kaynağı çalıştırıldıktan sonra biçimbirimsel çözümleme ve POS etiketleme tamamlanmaktadır. Bu adımlardan sonra atlaslar çalıştırılır. Onto Root atlası, ontoloji bağlantılı etiketler yaratabilmek için kullanılan ve dinamik olarak yaratılan bir atlasır (Cunningham et al., 2009). Bunun için GATE aracına bir dil kaynağı olarak tanıtılan bir ontolojideki tüm kavramlar listeye eklenir. Daha sonra liste, üzerinde aşağıdaki işlemler yapılarak metindeki kelimelerle eşleştirilmeye hazır duruma getirilir (Cunningham et al., 2009).

- Kavram isimlerinde geçen çizgi (“-“) ve alt çizgi (“_”) karakterleri boşluk ile değiştirilir. Örneğin, “Proje-Adı” ve “Proje_Adı” isimleri listeye “Proje Adı” olarak eklenir.
- Deve stilinde (camelCase) yazılan kavram isimleri ayrıştırılır. Örneğin, “ProjeAdı” ismi listeye “Proje Adı” olarak eklenir.
- Birleşik isimlerden meydana gelen isimler işlenerek listeye birden fazla isim eklenebilir. Örneğin, “POS Tagger” kelimesi için listeye hem “POS Tagger” hem de “Tagger” kelimeleri eklenir.

Şekil 6.14’te, örnek bir metnin ontolojideki kavramlarla eşleştirilmesi işleminin sonucu, GATE kullanıcı arayüzünde görülmektedir. Şekilde kırmızı ile gösterilen kelimeler ve kelime öbekleri ontolojideki kavramlarla eşleştirilmiştir. Hangi kelimenin hangi kavramla eşleştiği bilgisi ise arayüzün orta alt kısmında bulunan “Features” başlıklı bölümde listelenmektedir. Bu listede, örneğin “market” kelimesinin “<http://www.larflast.bas.bg/ontology#market>” ontoloji kavramıyla eşleştiği görülmektedir. Ayrıca, bu ontoloji kavramının bir sınıf olduğu bilgisi de (type = class) yine GATE aracının oluşturduğu etiketten anlaşılmaktadır.



Şekil 6.14: Örnek bir metnin ontolojideki kavramlarla eşleştirilmesi.

Metin, ontolojilerdeki kavramlar ile etiketlendikten sonra Şekil 6.11'deki veri yapısı güncellenir. Kelime düğümlerindeki “Etiket URI’si” ve “Etiket Ontolojisi URI’si” isimli sahalara, ontolojilerdeki kavramlarla eşleşen kelimeler için doldurulur.

Algoritma 6.1’de, ön hazırlık aşaması için geliştirilen uygulamanın algoritması gösterilmiştir.

Algoritma 6.1: Ön hazırlık aşaması için geliştirilen uygulamanın yordamları.

BelgeyiHazirla (belge)

```

{
  GATE aracını kullanarak gerekli işlem kaynaklarını yarat
  GATE aracını kullanarak metnin biçimbirimsel çözümlemesini ve
  POS etiketlemesini yap
  Etkisiz kelimeleri ayıkla
  Cümleleri ve kelimeleri bellekte temsil etmek için veri yapısını
  oluştur ve doldur
  Her ardışık cümle çifti için
  {
    Sıradaki ilk cümledeki her kelime için
    {
      Sıradaki ikinci cümledeki her kelime için
      {
        OntAramaSonuclari = SwoogleMotorundaAra(Kelime1, Kelime2)
        Eğer OntAramaSonuclari listesi boş ise
        {
          WatsonMotorundaAra(Kelime1, Kelime2)
        }
      }
    }
  }
  Bulunan ontolojileri gözden geçirildikten sonra metni ontolojik
  terimlerle etiketle
}

```

SwoogleMotorundaAra (Kelime1, Kelime2)

```

{
  Web servisini çağırmak için URL'yi oluştur
  XML formatında gelen sonucu ayıştır
  Gelen sonuçlar içinden rdf, rdfs, daml ve owl dosyalarını seç
  Seçilen ontolojiler içinde aranan kelimelere birebir karşılık
  gelen kaynak olmayan ontolojileri sonuç listesinden çıkar
  Sonuç listesini döndür
}

```

WatsonMotorundaAra (Kelime1, Kelime2)

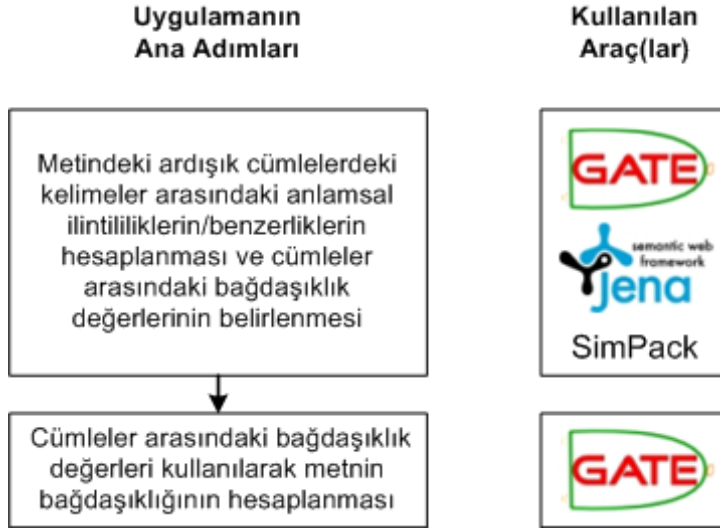
```

{
  Watson arama hizmeti için gerekli sınıfları yarat
  Arama hizmeti için gerekli parametreleri oluştur
  Gelen sonuçları içinden rdf, rdfs, daml ve owl dosyalarını seç
  Seçilen ontolojiler içinde aranan kelimelere birebir karşılık
  gelen kaynak olmayan ontolojileri sonuç listesinden çıkar
  Sonuç listesini döndür
}

```

6.2.2.2 Anlamsal çözümleme aşaması için geliştirilen uygulama

Anlamsal çözümleme aşaması için geliştirilen uygulamanın ana adımları ve bu uygulamada kullanılan araçlar Şekil 6.15'te gösterilmiştir.



Şekil 6.15: Anlamsal çözümleme aşaması için geliştirilen uygulamanın ana adımları ve bu uygulamada kullanılan araçlar.

Eşleştirme işlemi tamamlandıktan sonra ardışık cümleler arasındaki anlamsal ilintililik ve benzerlikler ölçülür. Bu ölçüm işlemi (6.5) numaralı formüle ve (6.6) numaralı örnek formüle uygun olarak yapılır. Ard arda gelen cümlelerdeki her kelime çifti için anlamsal ilintililik ve benzerlik ölçümü yapılır. İki kelimedenden biri ya da ikisi birden ontolojilerdeki bir kavramla eşleştirilmemişse iki kelime arasındaki anlamsal ilintililik ve benzerlik değeri 0 olarak belirlenir. Eğer iki kelimenin kelime kökü aynı ise bu iki kelime arasındaki anlamsal ilintililik ve benzerlik değeri 1 olarak alınır. İki kelime birden aynı ontolojideki birer kavramla eşleştirilmişse bu durumda bu çalışma kapsamında geliştirilen anlamsal ilintililik ölçüm yöntemi kullanılarak iki kavram arasındaki anlamsal ilintililik değeri hesaplanır. Aynı hesaplama SimPack kütüphanesi kullanılarak Resnik'in (Resnik, 1985), Lin'in (Lin, 1998), Jiang ve Conrath'ın (Jiang and Conrath, 1997), Wu ve Palmer'ın (Wu and Palmer, 1994) ve Leacock ve Chodorow'un (Leacock and Chodorow, 1998) önerdiği yöntemlerle de yapılır. Cümleler arasındaki anlamsal ilintililik ve benzerlik değerleri hesaplandıktan sonra Şekil 6.11'deki veri yapısındaki "İlntililik/Benzerlik Değerleri" sahaları güncellenir. Birinci cümleye ait düğümdeki ilintililik/benzerlik değerleri birinci ve ikinci cümle arasındaki anlamsal ilintililik ve benzerlik değerlerini içerir. Aynı şekilde ikinci cümleye ait düğümdeki ilintililik/benzerlik değerleri ikinci ve üçüncü cümle arasındaki anlamsal ilintililik ve benzerlik değerlerini içerir. Son cümle düğümünün "İlntililik/Benzerlik Değerleri" sahaları boş kalmaktadır. Bu değerler hesaplandıktan sonra (6.2) numaralı formüle uygun olarak metnin bağdaşıklık değerleri her bir anlamsal ilintililik ve benzerlik ölçüm yöntemi için ayrı ayrı hesaplanır.

Bir ontolojide iki kavram arasındaki ontolojik patikaları bulan ve bu patikalara göre iki kavram arasındaki en büyük anlamsal ilintililik değerini hesaplayan kod parçacıkları Bölüm 5.5.2’de açıklanmıştır ve Ek 8’de gösterilmiştir. Metin bağdaşıklık değerlerini, bu tez kapsamında geliştirilen anlamsal ilintililik ölçüm yöntemiyle ve başka benzerlik ölçüm yöntemleriyle hesaplayan yordam Ek 9’da gösterilmiştir.

Algoritma 6.2’de gösterildiği gibi, bir metnin bağdaşıklığını hesaplamak için öncelikle ardışık cümlelerdeki kelimeler arasındaki ilintililik ve benzerlik değerleri bulunur. Daha sonra bu değerlerden yola çıkılarak ardışık cümleler arasındaki bağdaşıklık değerleri bulunur. Ardışık cümleler arasındaki bağdaşıklık değerlerinin ortalaması alınarak metnin bağdaşıklığı bulunur. Metin bağdaşıklığını hesaplamak için kullanılan yordam Ek 9’da ayrıntılı olarak gösterilmiştir.

Algoritma 6.2: Metin bağdaşıklığını hesaplama algoritması.

```

BagdasiklikHesapla (Metindeki cümleleri saklayan dizi)
{
    cumleIndeks = 0
    (cumleIndeks+1 < Metindeki cümleleri saklayan dizinin uzunluğu)
    koşulu geçerliiyken
    {
        Dizideki cumleIndeks numaralı cümledeki her kelime için
        {
            Dizideki cumleIndeks+1 numaralı cümledeki her kelime için
            {
                Eğer karşılaştırılan iki kelimenin kökü aynıysa
                {
                    ilintililik/benzerlik değeri 1'dir
                }
                Eğer iki kelime de bir ontolojik terimle etiketlenmişse
                {
                    Eğer iki kelime aynı ontolojik terimle etiketlenmişse
                    {
                        ilintililik/benzerlik değeri 1'dir
                    }
                    Eğer iki kelime aynı ontolojideki iki terimle
                    etiketlenmişse
                    {
                        İlntililik/benzerlik değeri bu kelime çifti için
                        daha önce hesaplanmışsa bu değeri al; hesaplanmamışsa
                        ilintililik/benzerlik değerini hesapla
                    }
                    Eğer iki kelime aynı ontolojideki terimlerle
                    etiketlenmemişse
                    {
                        ilintililik/benzerlik değeri 0'dır
                    }
                }
                Eğer iki kelime birden ya da kelimelerden birisi
                ontolojideki terimlerle etiketlenmemişse
                {
                    ilintililik/benzerlik değeri 0'dır
                }
            }
        }
        Kelime ilintililik/benzerlik değerlerini kullanarak cümleler
        arasındaki bağdaşıklık değerlerini hesapla
        cumleIndeks = cumleIndeks + 1
    }
    Ardışık cümleler arasındaki bağdaşıklık değerlerini kullanarak
    her yöntem için metin bağdaşıklığını hesapla
}

```

6.3 Yöntemin Başarısının Değerlendirilmesi

Bu tez kapsamında önerilen yöntemin başarısını ölçmek için bir deney yapılmıştır. Bu deney, Lapata ve Barzilay'ın değişik benzerlik hesaplama

yöntemlerini kullanarak yapılan metin bağdaşıklık ölçümlerinin başarısını değerlendirdikleri deney temel alınarak hazırlanmıştır (Lapata and Barzilay, 2005). Lapata ve Barzilay'ın deneyinde 2003 yılında düzenlenen Belge Anlama Konferansı¹⁵nda oluşturulan 96 adet metin kullanılmıştır. Bu 96 metnin ne kadar bağdaşık olduğu İnternet üzerinden ulaşılan, ana dili İngilizce olan, 177 katılımcı tarafından değerlendirilmiştir. Katılımcılar, kendilerine verilen yönerge doğrultusunda, okudukları metinlerin bağdaşıklıklarını 1 ile 7 arasında bir sayı ile değerlendirmiştir. Her metin tüm katılımcılar tarafından değerlendirilmemiştir. Bir metin ortalama olarak 23 katılımcı tarafından değerlendirilmiştir. Metin sırasının etkisini ortadan kaldırmak için karşıt dengeleme yöntemi olarak Latin kare deseni (Bkz. Ek 10) kullanılmıştır. Katılımcıların bir metne verdikleri değerlerin ortalaması alınarak o metnin bağdaşıklık değeri hesaplanmıştır. Katılımcıların metinlere verdikleri bağdaşıklık değerlerinin tutarlılığını ölçmek için leave-one-out resampling yöntemi kullanılmıştır ve sonuç olarak ortalama tutarlılık değeri $r=0,768$ olarak bulunmuştur. Lapata ve Barzilay'a göre, her metnin 20'den fazla katılımcı tarafından değerlendirildiği düşünüldüğünde bu değer oldukça yüksektir (Lapata and Barzilay, 2005).

Veri toplama ve tutarlılık çözümleme işlemlerinden sonra literatürdeki çeşitli benzerlik yöntemleri kullanılarak metinlerin bağdaşıklıkları otomatik olarak ölçülmüştür. Bunun için öncelikle çeşitli benzerlik yöntemleri ve (6.4) numaralı formül kullanılarak cümleler arasındaki bağdaşıklık değerleri hesaplanmıştır. Daha sonra cümleler arasındaki bağdaşıklık değerleri ve (3.1) numaralı formül kullanılarak bir metin için her benzerlik yöntemine ait bağdaşıklık değeri hesaplanmıştır. Benzerlik yöntemlerinin, otomatik metin bağdaşıklık ölçümündeki başarısını değerlendirmek için Pearson Momentler Çarpımı Korelasyon tekniği kullanılarak çözümleme yapılmıştır. Otomatik ölçümler sonucunda elde edilen değerler ile katılımcıların verdikleri değerler

¹⁵ Belge Anlama Konferansları (Document Understanding Conferences – DUC), metin özetleme çalışma alanında bir yol haritası oluşturmayı hedeflemektedir. Bu konferanslar Advanced Research and Development Activity (ARDA) tarafından desteklenmektedir ve National Institute of Standards and Technology (NIST) tarafından düzenlenmektedir. 2008 yılından itibaren bu konferanslar Metin Çözümleme Konferansları (Text Analysis Conference – TAC) çatısı altında düzenlenmektedir. TAC, doğal dil işleme teknolojilerinin büyük ölçekli değerlendirmeleri için gerekli altyapıyı sağlamayı hedefleyen bir dizi çalışıydan oluşmaktadır.

DUC 2005 organizasyonunda, NIST katılımcıları, 50 başlık belirlemişlerdir ve bu başlıklarla ilgili Los Angeles Times ve Financial Times of London gazetelerinden, her başlık altında 25 ile 50 arasında belge oluşturacak şekilde yazılar seçmişlerdir (Dang, 2005). Daha sonra 10 NIST katılımcısı, 20 başlık için dokuzar tane, 30 başlık içinse dörder tane özet yazmışlardır.

arasındaki korelasyonlar Çizelge 6.2’de görüldüğü gibi raporlanmıştır (Lapata and Barzilay, 2005). Lapata ve Barzilay’ın bildirisinde daha fazla benzerlik yöntemi için korelasyon değeri verilmiştir. Aşağıdaki çizelgede, sadece bu çalışma kapsamında değerlendirilen benzerlik yöntemleri için verilen korelasyon değerleri listelenmiştir. Korelasyon hesaplaması için 96 metnin 58 tanesi kullanılmıştır. Diğer metinler, deneyde kullanılan benzerlik ölçüm yöntemlerinden birleştirilmiş bir model elde etmek için kullanılmıştır.

Çizelge 6.2: Lapata ve Barzilay’ın deneyinde katılımcıların bağdaşıklık değerlendirmeleri ile otomatik bağdaşıklık değerlendirmeleri arasındaki korelasyon değerleri.

İltililik/Benzerlik Ölçüm Yöntemi	Korelasyon
Resnik	0,207
Lin	0,173
Jiang ve Conrath	-0,290**
Gizli Anlam Çözümleme	0,230*
* $p < 0,05$ (2-tailed) ** $p < 0,01$ (2-tailed)	

Bu tez kapsamında yapılan deneyde, Lapata ve Barzilay’ın deneyinde kullanılan metinler ve bu metinlerin bağdaşıklık değerleri kullanılmıştır. Katılımcılardan bağdaşıklık değerlendirmeleri toplanan metinler, ön hazırlık aşaması için geliştirilen uygulama kullanılarak anlamsal olarak etiketlenmiştir. Bu uygulama kullanılarak öncelikle metindeki ardışık cümlelerde bulunan her bir kelime ya da kavram çifti için Swoogle ve Watson arama motorlarında ontoloji aranmıştır. Daha sonra bulunan ontolojiler gözden geçirilerek, metindeki kelimelerle aynı anlama sahip, fakat sözdizimsel farklılıklardan dolayı eşleşemeyecek terimler düzeltilmiştir (Örneğin, “notebook_computer” olarak yazılmış bir ontolojik terim “notebook” olarak değiştirilmiştir). Ayrıca, sözdizimsel olarak metindeki kelimelerle eşleşen fakat ontolojide farklı anlamda kullanılmış olan ontolojik terimler elenerek kelime anlamı belirginleştirme işlemi yapılmıştır. Her metin ortalama olarak 1,8 ontoloji kullanılarak etiketlenmiştir. Ön hazırlık aşamasında 96 metnin 33 tanesi etiketlenebilmiştir. Etiketlenen 33 metin Ek 11’de gösterilmiştir.

Anlamsal etiketleme için ontolojiler bulunduktan ve bu ontolojiler eşleşme işlemine hazırlandıktan sonra metindeki kelimeler ontolojik terimlerle, geliştirilen uygulama kullanılarak eşleştirilmiştir. Bu tez kapsamında yapılan deneydeki anlamsal etiketleme işleminin sonuçlarına ait bazı değerler Çizelge 6.3’te gösterilmiştir.

Çizelge 6.3: Deneyde kullanılan metinlerin anlamsal etiketleme işlemi sonuçlarına ait değerler.

	Metinlerde Bulunma Sayısı	Ontolojik Terimlerle Eşleşme Sayısı	Eşleşme Yüzdesi	Eşleşen Ontolojik Terimin Tipi		
				Sınıf	Özellik	Birey
İsim	1220	306	%25	269	1	36
Sıfat	302	24	%8	22	0	2
Fiil	531	60	%11	60	0	0
Toplam	2053	390	%19	351	1	38

33 metindeki isim, sıfat ve fiil kelime kategorisine ait tüm kelimelerin %19'u ontolojik terimlerle eşleşmiştir. Bildiğimiz kadarıyla literatürde, anlamsal etiketleme işleminin sonucunda metindeki kelimelerin ontolojik terimlerle eşleşme oranı için önerilen bir yüzde bulunmamaktadır. Ayrıca anlamsal etiketleme için hangi kelime kategorilerinin dikkate alınacağı da uygulamaya göre değişmektedir. Bu tez kapsamında yapılan deneyde, isim, sıfat ve fiil kelime kategorilerine ait kelimeler anlamsal olarak etiketlenmeye çalışılmıştır. Deney sonucunda elde edilen eşleşme oranlarının metin bazında dağılımı Çizelge 6.4'te gösterilmiştir. 33 metin içinde sadece bir metin için eşleşme oranı %4 olarak gerçekleşmiştir (Bu metinde fiil kategorisine ait kelimelerin %12'si anlamsal olarak etiketlenmiştir). Bunun dışındaki tüm metinler için eşleşme oranı %10 ve daha yüksektir. En yüksek eşleşme oranı ise %41 olarak gerçekleşmiştir. Bu tez çalışması kapsamında yapılan deneyde, Şekil 6.2'de gösterilen “eşleştirme sonuçlarının değerlendirilmesi” karar adımıındaki eşleşme oranının alt sınırı her metin için %10 olarak kabul edilmiştir. Eşleşme oranı %4 olan metnin fiil kategorisine ait eşleşme oranı %12 olduğu için bu metin de deney kümesine dahil edilmiştir. Eşleşme oranının alt sınırı başka uygulamalar için farklı değerler alabilir.

Çizelge 6.4: Metin bazında eşleşme oranlarının dağılımı.

Eşleşme Yüzdesi	Metin Sayısı
%0 - %10	3
%11 - %20	17
%21 - %30	11
%31 - %40	1
%41 - %50	1

Anlamsal etiketleme sonuçlarına göre, metinlerde en fazla isim kelime kategorisine ait kelime ontolojik bir terimle eşleştirilmiştir. Sıfat ve fiil kategorisine ait kelimeler ise isimlere göre çok daha az ontolojik terimle eşleşmiştir. Metinlerde geçen kelimelerin büyük çoğunluğu (%90) ontolojilerdeki sınıflarla eşleşmiştir. Ontolojilerde tanımlanan terimlerin ağırlıklı olarak

sınıflardan oluşması bunun bir nedeni olarak gösterilebilir. Çizelge 6.5'te, bu çalışma kapsamında çözümlenen 201 ontolojideki terimlerin tiplerinin dağılımı gösterilmiştir. Dağılım için sayılar elde edilirken anonim terimler göz önüne alınmamıştır. Çünkü bu terimlerin bir metindeki kelimelerle eşleşebilme olasılığı yoktur. 201 ontolojideki terimlerin %59'unun tipi sınıftır.

Çizelge 6.5: Seçilen 201 ontolojideki terim tiplerinin dağılımları.

	Sınıf	Özellik	Birey
Sweet Ontolojileri	5045	300	609
Protege Ontoloji Kütüphanesi	1615	738	2979
Tones Ontoloji Kütüphanesi	1476	970	53
Toplam	8136	2008	3641
Toplam İçindeki Yüzdesi	%59	%15	%26

Ön hazırlık aşamasında metinler anlamsal olarak etiketlendikten sonra metinlerin bağdaşıklık değerleri, bu tez kapsamında geliştirmiş olan anlamsal ilintililik ölçüm yöntemi, Resnik'in (Resnik, 1985), Lin'in (Lin, 1998) ve Jiang ve Conrath'ın (Jiang and Conrath, 1997) önerdiği benzerlik ölçüm yöntemleri kullanılarak hesaplanmıştır. Benzerlik hesaplamaları için kullanılan SimPack kütüphanesindeki Wu ve Palmer'ın (Wu and Palmer, 1994) ve Leacock ve Chodorow'un (Leacock and Chodorow, 1998) benzerlik yöntemlerini gerçekleştiren işlevlerin sağlıklı çalışmamasından dolayı bu yöntemler kullanılarak metin bağdaşıklıkları hesaplanamamıştır. Gizli Anlam Çözümleme yöntemi (Foltz et al., 1998) kullanılarak metin bağdaşıklığı ise Colorado Boulder Üniversitesi'nin bu yöntem için hazırladığı web sitesindeki arayüz (Latent Semantic Analysis @ CU Boulder, 2011) kullanılarak hesaplanmıştır. Örnek bir metnin Gizli Anlam Çözümleme yöntemi kullanılarak bağdaşıklık hesaplama sonucu Şekil 6.16'da gösterilmiştir. Şekilde de görüldüğü gibi, Gizli Anlam Çözümleme yöntemi kullanılarak önce cümleler arasındaki bağdaşıklık değerleri hesaplanmaktadır ve metin bağdaşıklığının bulunması için bu değerlerin ortalaması alınmaktadır.

Sentence to Sentence Coherence Comparison Results

The submitted texts' sentence to sentence coherence:

COS	SENTENCES
0.22	1: An Arctic cold wave and accompanying blizzards have killed at least 65 people throughout Europe, police and media reported on 11/23/1998.
0.28	2: Croatia reported its first cold-related death on 11/25/1998, a woman who lived alone in an unheated house.
0.17	3: The death toll from almost three weeks of sub-zero temperatures has reached 83, after a homeless man froze to death in a Warsaw park.
	4: Poland's extreme cold began November 16, with temperatures dropping to minus 26 Celsius minus 4 Fahrenheit.

Mean of the Sentence to Sentence Coherence is: 0.22

Şekil 6.16: Örnek bir metin için Gizli Anlam Çözümleme yöntemi ile yapılan bağdaşıklık ölçümü sonucu.

Elde edilen bağdaşıklık değerleri ile katılımcıların metinlerin bağdaşıklıkları için verdikleri değerler arasındaki ilişkiyi anlamak için Pearson Momentler Çarpımı Korelasyon tekniği kullanılarak çözümleme yapılmıştır. Bu deney sonucunda elde edilen korelasyon değerleri, Lapata ve Barzilay'ın deneyinde (Lapata and Barzilay, 2005) elde edilen sonuçlarla karşılaştırmalı olarak gösterilmiştir. Metin bağdaşıklıkları, Gizli Anlam Çözümleme yöntemi kullanılarak Colorado Boulder Üniversitesi'nin bu yöntem için hazırladığı web sitesindeki arayüzde bulunan tüm anlamsal uzaylar için ölçülmüştür. Aşağıdaki çizelgede Gizli Anlam Çözümleme yöntemi için verilen korelasyon değeri, elde edilen tüm korelasyonlar arasındaki en yüksek değerdir. Deneyde elde edilen metin bağdaşıklık değerleri Ek 12'deki ilk çizelgede listelenmiştir.

Çizelge 6.6: Bu tez kapsamında ontolojiler kullanılarak yapılan deneyde katılımcıların bağdaşıklık değerlendirmeleri ile otomatik bağdaşıklık değerlendirmeleri arasındaki korelasyon değerleri.

İltililik/Benzerlik Ölçüm Yöntemi	Korelasyon
Ontoloji tabanlı anlamsal iltililik ölçüm yöntemi	0,359*
Resnik	0,285
Lin	0,342
Jiang ve Conrath	0,288
Gizli Anlam Çözümleme	0,331
* p < 0,05 (2-tailed) ** p < 0,01 (2-tailed)	

Çizelge 6.6'te görüldüğü gibi, bu tez kapsamında yapılan deneyde, katılımcıların verdikleri bağdaşıklık değerleri ile en yüksek korelasyona sahip

değerler, ontoloji tabanlı anlamsal ilintililik ölçüm yöntemi ile elde edilmiştir. Ayrıca bu yöntem ile elde edilen korelasyon değeri, deney sonucunda elde edilen tüm değerler arasındaki istatistiksel olarak anlamlı tek değerdir.

Bu tez kapsamında yapılan deneyde elde edilen korelasyon değerlerinin, Lapata ve Barzilay'ın deneyinde elde edilenlere göre daha yüksek olmasının nedeni olarak, bu çalışmada ilintililik ve benzerlik ölçümü için ontolojilerin kullanılması gösterilebilir. Bu varsayımın doğruluğunu sınamak için, ontolojiler kullanılarak yapılan deney, WordNet kavramlar dizini (Fellbaum, 1998), British National Corpus (World Sürümü), Brown Corpus, Penn Treebank (Sürüm 2) ve SemCor derlemleri ile tekrarlanmıştır. Bu deneyde kelimeler arasındaki benzerlikler Java WordNet::Similarity kütüphanesi kullanılarak ölçülmüştür. Cümleler arasındaki benzerlikleri ve metin bağdaşıklığını hesaplamak için, bir önceki deneyde olduğu gibi, sırasıyla (6.5) ve (6.2) numaralı formüller kullanılmıştır. (6.5) numaralı formülde gösterilen anlamsal ilintililik değerleri olarak Resnik'in (Resnik, 1985), Lin'in (Lin, 1998), Wu ve Palmer'ın (Wu and Palmer, 1994) ve Leacock ve Chodorow'un (Leacock and Chodorow, 1998) önerdiği yöntemlerle elde edilen benzerlik değerleri kullanılmıştır. Deneyde elde edilen metin bağdaşıklık değerleri Ek 12'deki 2., 3., 4. ve 5. çizelgede, bu değerlere göre hesaplanan korelasyon değerleri ise Çizelge 6.7'de gösterilmiştir.

Çizelge 6.7: Bu tez kapsamında WordNet kavramlar dizini ve derlemler kullanılarak yapılan deneyde katılımcıların bağdaşıklık değerlendirmeleri ile otomatik bağdaşıklık değerlendirmeleri arasındaki korelasyon değerleri.

Derlem Benzerlik Ölçüm Yöntemi	British National Corpus (World Sürümü)	Brown Corpus	Penn Treebank (Sürüm 2)	SemCor
Resnik	0,049	0,045	0,062	0,049
Lin	-0,028	-0,114	-0,012	-0,067
Wu ve Palmer	0,174			
Leacock ve Chodorow	0,178			
* p < 0,05 (2-tailed) ** p < 0,01 (2-tailed)				

Resnik'in (Resnik, 1985) ve Lin'in (Lin, 1998) yöntemleri kullanılarak yapılan ölçümler hem bir kavramsal modele hem de derleme (kelimelerin bilgi içeriği değerlerini hesaplamak için) gereksinim duymaktadır. Bu yöntemlerle yapılan benzerlik ölçümleri, British National Corpus (World Sürümü), Brown Corpus, Penn Treebank (Sürüm 2) ve SemCor derlemleri için ayrı ayrı yapılmıştır. Wu ve Palmer'ın (Wu and Palmer, 1994) ve Leacock ve Chodorow'un (Leacock

and Chodorow, 1998) önerdiği yöntemler sadece bir kavramsal modele gereksinim duyduğu için, bu yöntemler kullanılarak yapılan metin bağdaşıklık değerleri, dolayısıyla da bu bağdaşıklık değerlerinin katılımcıların bağdaşıklık değerlendirmeleri ile korelasyonları tüm derlemeler için aynıdır. Java WordNet::Similarity kütüphanesindeki Jiang ve Conrath'ın (Jiang and Conrath, 1997) benzerlik yöntemini gerçekleştiren işlevlerin sağlıklı sonuçlar vermemesi nedeniyle, bu yöntem için korelasyon değeri hesaplanamamıştır.

Deneyde kullanılan dört benzerlik yöntemi ile otomatik olarak hesaplanan metin bağdaşıklık değerlerinin hiçbirisi, katılımcıların bağdaşıklık değerlendirmeleri ile istatistiksel olarak anlamlı bir korelasyona sahip olamamıştır. Ayrıca WordNet (Fellbaum, 1998) ve çeşitli derlemeler kullanılarak yapılan ölçümler sonucunda elde edilen korelasyon değerleri, ontolojiler kullanılarak yapılan deneydeki korelasyon değerlerine göre daha düşük çıkmıştır. Dolayısıyla bu tez kapsamında yapılan deneylere göre, ontolojilerin ilintililik ve benzerlik ölçümünde kullanımı elde edilen sonuçların başarısını arttırmıştır.

6.4 İlgili Çalışmalar

Serrano et al., insan bilişinin bazı yönlerinden esinlenen, metin okuma sonucunda bellekte oluşan temsili modelleyen, Bilişsel Okuma Dizinleme Modeli (Cognitive Reading Indexing Model – CRIM) adlı bir hesaplama modeli önermiştir (Serrano et. al., 2009). Bu model, anlamsal sinir ağı üzerine temellendirilmiş ve okunan metnin temsili, bellekte etkinleştirilmiş kavramlardan oluşan bir ağ olarak temsil etmiştir. Bu temsil, metin tabanlı modeline karşılık gelmektedir. Bu ağdaki kavramlar birbirleri ile ağırlıklandırılmış bağlantılar ile ilişkilendirilmiştir.

Bir metnin, Bilişsel Okuma Dizinleme Modeli'nin oluşturulma süreci Algoritma 6.3'de gösterilmiştir. Bu süreçte ilk önce metindeki etkisiz kelimeler ayıklanır ve kelimelerin kökleri bulunur. Daha sonra her kavram modele eklenir. Aynı cümlede bulunan kavramlar arasında bir bağlantı kurulur. Bir sonraki adımda, kavramlar arasındaki bağlantıların ağırlıkları belirlenir. Bu bağlantılar simetrik değildir. A kavramından B kavramına olan bağlantının ağırlığı, B kavramından A kavramına olan bağlantının ağırlığıyla aynı değildir. İki kavram arasındaki bağlantının ağırlığı, kavramların metinde geçme sıklığına ve aynı cümlede birlikte geçme sıklığına bağlıdır. Örneğin, bir metinde A kavramı 10 kez geçiyorsa ve 6 kez B kavramıyla aynı cümlede, 4 kez de C kavramıyla aynı

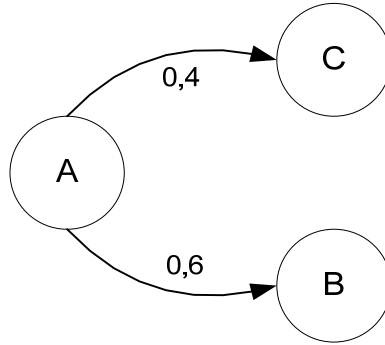
cümlede geçiyorsa, Şekil 6.17’de gösterildiği gibi, A ve B kavramlarının bağlantılarının ağırlığı 0,6, A ve C kavramlarının bağlantılarının ağırlığı 0,4 olacaktır.

Algoritma 6.3: Metnin anlamsal temsilini oluşturan algoritma.

```

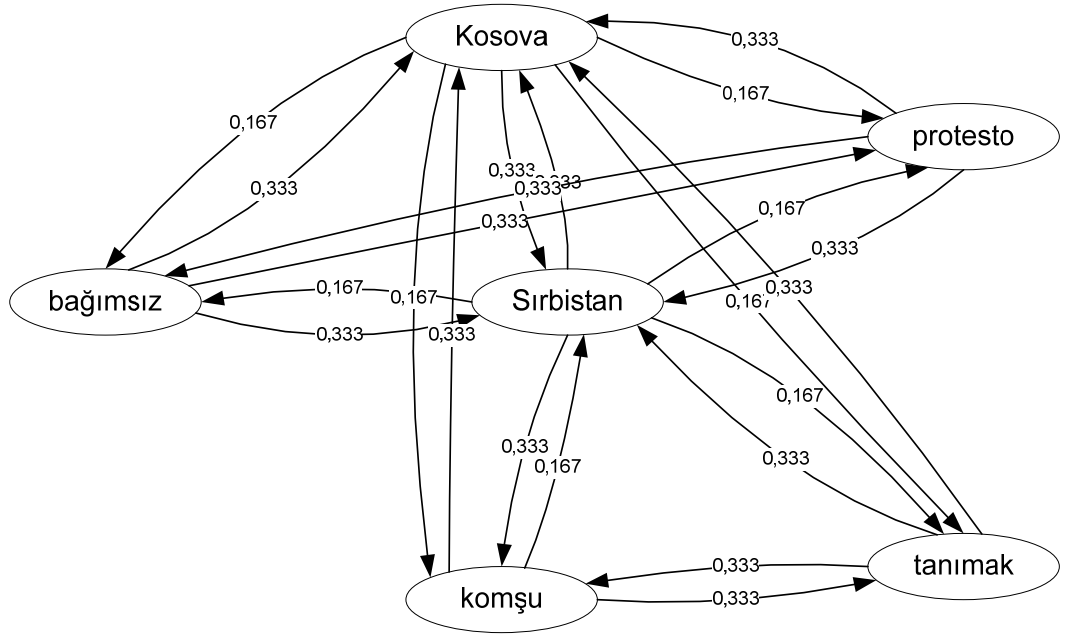
Girdi: Metin
Çıktı: Metnin anlamsal temsili
1: BilgiHavuzu = boş
2: Metindeki etkisiz kelimeleri ayıkla
3: Metindeki kelimelerin köklerini bul
4: Metindeki her cümlek için
5:   cümlek’deki her kelimej için
6:     Eğer kelimej BilgiHavuzu’nda yoksa
7:       kelimej’yi BilgiHavuzu’na ekle
8:       MetindeToplamBulunmaSayısı(kelimej) = 0
9:       MetindeToplamBulunmaSayısı(kelimej) + 1
10:    cümlek’deki kelimej’den sonra gelen her kelimeh için
11:      Eğer Bağlantı(kelimej, kelimeh) varsa
12:        Bağlantı(kelimej, kelimeh) + 1
13:        Bağlantı(kelimeh, kelimej) + 1
14:      Değilse
15:        Ekle Bağlantı(kelimej, kelimeh) + 1
16:        Ekle Bağlantı(kelimeh, kelimej) + 1
17: //Bağlantı ağırlıklarını normalize et
18: BilgiHavuzu’ndaki her kelimei için
19:   Her Bağlantı(kelimei, kelimej) için
20:     BağlantıAğırlığı(kelimei, kelimej) =
21:       Bağlantı(kelimei, kelimej) /
22:       MetindeToplamBulunmaSayısı(kelimei)

```



Şekil 6.17: Kavram ilişkilerinin ağırlıkları (Serrano et. al., 2009).

Şekil 6.18’de, örnek bir metin parçasının Bilişsel Okuma Dizinleme Modeli gösterilmiştir. Bu modele göre, iki kavram arasındaki benzerlik, iki kavram arasındaki en kısa patikanın bağlantı ağırlıklarının çarpımıyla bulunur.



Şekil 6.18: Örnek bir metin parçasının Bilişsel Okuma Dizinleme Modeli (Serrano et. al., 2009).

Serrano et al.'nin yaklaşımı ile bu tez kapsamında yapılan çalışma arasında ortak yönler bulunmaktadır. Bilişsel teoriler, iki çalışmanın da temelini oluşturmaktadır. İki çalışma da kavramlar arasındaki en kısa patikayı değerlendirmektedir ve düğümler arasındaki kenarların ağırlıklarını çarparak ilintililik ya da benzerlik hesaplamaktadır. İki çalışmadaki farklı yönlerden birincisi ise Serrano et al.'nin yönteminin benzerlik, bu tez kapsamındaki yöntemin ise ilintililik ölçmesidir. İkinci farklı yön ise düğümler arasındaki ağırlıkların hesaplanma yöntemidir. Serrano et al., kavramlar arasındaki bağlantıların ağırlıklarını istatistiksel yöntemle (kavramın metinde geçme sıklığı ve kavramların aynı cümlede geçme sıklığı) hesaplamaktadır. Bu tez kapsamında ise kavramlar arasındaki ilintililiği hesaplamak için OWL dil öğelerinin anlamsal ilintililik katsayıları kullanılmaktadır.

6.5 Özet ve Değerlendirme

Doğal dil ile yazılmış metinlerin anlamının, bu metinlerle ilişkilendirilen ontolojiler ile iyi seviyede temsil edilebilmesi için, metindeki kelimelerin ve kelime öbeklerinin mümkün olduğu kadar fazla sayıda ontolojik terimle eşleşmesi gerekmektedir. Bu eşleşme oranının hangi seviyelerde olması gerektiği ile ilgili, bildiğimiz kadarıyla, literatürde önerilen bir oran bulunmamaktadır. Bu çalışma kapsamında anlamsal olarak etiketlenen 33 metindeki eşleşme oranı, (6.1) numaralı formül ile hesaplandığında %19 olarak bulunmuştur. Eşleşme

oranlarının düşük olmasının temel nedeni, metinlerin anlamsal olarak etiketlenmesi için gerekli olan ontolojilerin, İnternet üzerinde erişilebilecek şekilde yeterli sayıda olmamasından kaynaklanmaktadır. Yeterli ontoloji olmamasının yanında, bu çalışmada kullanılabilecek ontolojilerin, bir metinde ard arda geçen iki cümledeki iki kelimeyi birden içermesi gerekmektedir. Bu durum, bu çalışma kapsamında geliştirilen yöntem, Wu ve Palmer'ın (Wu and Palmer, 1994) ve Leacock ve Chodorow'un (Leacock and Chodorow, 1998) önerdiği yöntemlerin patika tabanlı olmasından ve iki kavram arasındaki patika arama işlemi için tek bir ontolojiye gereksinim duymasından kaynaklanmaktadır. Benzer şekilde Jiang ve Conrath'ın (Jiang and Conrath, 1997) yöntemi de patika ve bilgi içeriği tabanlı yaklaşımları birleştirdiği için benzerlik hesaplayabilmek için bir ontolojiye gereksinim duymaktadır. Resnik'in (Resnik, 1985) ve Lin'in (Lin, 1998) önerdiği bilgi içeriği tabanlı yaklaşımlar ise benzerlik hesabı yaparken kavramların ata ve alt-kavramlarını dikkate aldığı için yine tek bir ontolojiye gereksinim duymaktadır. Bu nedenlerden dolayı iki kelimenin birden geçtiği ontolojileri bulmaya çalışmak eşleşme oranını daha da düşürmektedir.

Eşleşme oranını yükseltmenin yollarından birisi, aranan kelime çiftlerinin eş anlamlılarını kullanarak ontoloji arama işlemini zenginleştirmektir. Bu çalışma kapsamında, bir kelime çiftini içeren bir ontoloji bulunamadığında, kelimelerin eş anlamlılarından faydalananarak aramayı tekrarlamak için gerekli geliştirme yapılmıştır. Kelimelerin eş anlamlılarını bulmak için WordNet kavramlar dizini (Fellbaum, 1998) kullanılmıştır. Fakat Swoogle ve Watson arama motorlarında yapılan arama işlemlerinin zaman aşımı hatası ile çok fazla sonlanmasından ve eş anlamlıların arama işlemine dahil edilmesi durumunda arama motorlarına gönderilen istek sayısının çok artacak olmasından dolayı, bu işlev kullanılmadan deney yapılmıştır.

7 SONUÇ VE TARTIŞMA

Elektronik ortamda saklanan verinin hacmi büyüdükçe, veri yönetimi ve bu veri üzerinde bilgisayarların daha zekice işlemler yapma gereksinimi de artmaktadır. Anlamsal web çalışma alanı, bilgisayarlara verinin anlamını biçimsel olarak temsil eden veriyi sağlayarak, bilgisayarların veri üzerinde daha zekice işlemler yapmasını hedeflemektedir. Verinin anlamını biçimsel olarak tanımlayan ontolojiler de anlamsal web vizyonunun merkezindeki kavramdır.

Ontolojiler kullanılarak elektronik ortamdaki çeşitli kaynaklar (doğal dil ile yazılmış metinler, veritabanlarındaki alanlar, video ve ses dosyaları vs.) anlamsal olarak etiketlenebilir. Böylece bir bilgisayar, kaynağın içeriğinin anlamı hakkında işlenebilir biçimsel bir veriye erişebilir duruma gelir ve bu kaynaklar üzerinde zekice işlemler yapabilir. Örneğin, doğal dil ile yazılmış bir metnin ontolojik terimlerle etiketlenmesiyle birlikte bu metnin bağdaşıklığı hakkında bilgi sahibi olunabilir.

Bir metnin bağdaşık olup olmadığını değerlendirebilmek için insanların okuduklarını nasıl kavradıklarını anlamak gerekmektedir. İnsanların, okuduklarını nasıl kavradıklarını açıklamak için çeşitli modeller önerilmiştir (Kintsch, 1974; Kintsch and van Dijk, 1978; Kintsch, 1998; Just and Carpenter, 1980, 1987; Perfetti, 1999; Verhoeven and Perfetti, 2008). Bu modellerin en önemli ortak yönü, okuduğunu kavrama sürecinde, ard arda gelen bilgileri daha önceki bilgilerle bütünleştirerek bir bellek temsilinin oluşturulduğunu kabul etmeleridir. Johnson-Laird'in (Johnson-Laird, 1983), zihinsel modellerin oluşturulması, güncellenmesi, bütünleştirilmesi ve değerlendirilmesi için gerçekleştirdiği psikolojik işlemler de, elde edilen bilgilerin önceki bilgilerle ilişkilendirildiği görüşünü desteklemektedir. Başarılı bir kavrama süreci sonunda, bağdaşık bir bellek temsilinin oluştuğu kabul edilmektedir (van den Broek et al., 2002). İnsanın, bağdaşık bir bellek temsili oluşturabilmesi için, okunan metnin içerdiği kelimelerin, önermelerin, cümlelerin birbirleriyle anlamsal olarak bağlantılı olması gerekmektedir. Böylece okunan her birim (kelime, önerme, cümle) kısa süre önce okunan birim ile ilişkilendirilmeye çalışılır. Bu işlem çalışma belleğinde gerçekleşir. Çalışma belleğinin kapasitesi sınırlı olduğu için, ard arda gelen birimlerin birbirleriyle ilişkilendirilmesi ve bir zihinsel temsilin parçası olarak uzun süreli belleğe aktarılması gerekmektedir. Ard arda gelen birimlerin birbiriyle bağlantısız olması durumunda, bazı bilgiler zihinsel temsile eklenemeyecek, çalışma belleğinden çıkartılarak unutulacaktır. Bu durum da kavrama sürecini

olumsuz etkileyecektir. Özellikle metnin birimleri arasındaki bağlantıların gevşek olması durumunda, metnin konusu hakkında artalan bilgisine sahip olmayan okuyucular bundan daha da olumsuz etkileneceklerdir. Çünkü metnin konusu hakkında bilgi sahibi olan okuyucular, metin birimleri arasındaki gevşek bağlantıları, artalan bilgilerini kullanarak yaptıkları çıkarsamalarla kuvvetlendirip başarılı bir kavrama gerçekleştirebilirler. Fakat okunan metin hakkında artalan bilgisi olmayan okuyucuların, metni başarılı olarak kavrayabilmeleri için okunan metnin bağdaşık olması gerekmektedir (McNamara, 2001; Kendeou and van den Broek, 2007).

Bir metnin bağdaşıklığını, yani anlaşılabilirliğini, değerlendirmek için çeşitli yöntemler önerilmiştir. Bu yöntemlerden bazıları ölçeklenebilir olmadıkları için, bazıları ise ürettiği sonuçların başarısızlığı açısından eleştirilmiştir. Hem ölçeklenebilir hem de başarılı sonuçlar veren bir yöntem olarak, metnin bağdaşıklığının, metinde ard arda gelen cümleler arasındaki anlamsal ilintililiğin ya da benzerliğin ölçülerek değerlendirilmesi önerilmiştir (Foltz et al., 1998; Lapata and Barzilay, 2005).

Anlamsal benzerlik, anlamsal ilintililiğin özel bir durumudur. Anlamsal olarak birbirine benzer kavramlar, birbirleriyle ortak yönleri ile ilintilidir. Fakat iki kavram arasında farklı ilişkilerden kaynaklanan bağlantılar da bulunabilir. Aralarında parça-bütün ya da zıtlık ilişkisi bulunan iki kavram, anlamsal olarak benzer olmamasına rağmen anlamsal olarak ilintili olabilir. Bir okuyucu bir metni okurken, kavramlar arasında sadece benzerlik ilişkisi kurduğunu söylemek doğru olmayacaktır. Okuyucunun kavramlar arasındaki çok çeşitli ilişkileri kurarak bir zihinsel temsil oluşturması, mevcut okuduğunu kavrama modellerine daha uygun düşmektedir. Bundan dolayı bu çalışmada, bir metnin bağdaşıklığını değerlendirirken, ard arda gelen cümleler arasındaki anlamsal benzerliğin yerine anlamsal ilintililiğin ölçülmesi tercih edilmiştir.

Bu tez kapsamında, OWL diline özgü bir anlamsal ilintililik ölçüm yöntemi geliştirilmiştir. Literatürde çok sayıda anlamsal ilintililik ve benzerlik ölçüm yöntemi bulunmaktadır. Fakat bu yöntemlerin çoğu, kavramlar arasındaki sınırlı sayıdaki ilişki türünü dikkate almaktadır. Anlamsal web vizyonu çerçevesinde geliştirilen ontoloji temsil dilleri, zengin anlamsal temsil gücüne sahiptir. Bu tezde savunulan görüşe göre, anlamsal temsil gücü yüksek dillerle geliştirilmiş ontolojilerdeki kavramlar arasındaki ilintililiğin ve benzerliğin ölçümünde, yeni dil öğelerinin dikkate alınması gereklidir. Böylece, aralarında sadece özellik-

genellik ilişkileri olan kavramların değil, daha karmaşık ilişkilerle (farklı olma, zıtlık, eşdeğer olma gibi) birbirine bağlanmış kavramların da aralarındaki ilintililiğin ölçülmesi mümkündür.

Bu tez kapsamında geliştirilen ontoloji tabanlı anlamsal ilintililik yöntemi, bir metnin bağdaşıklığının, yani anlaşılabilirliğinin, değerlendirilmesi için kullanılmıştır. Bunun için öncelikle, doğal dil ile yazılmış metinler, ontolojiler kullanılarak anlamsal olarak etiketlenmiştir. Daha sonra, metinlerdeki ard arda gelen cümleler arasındaki anlamsal ilintililikler, ontolojilerdeki kavramlar arasındaki patikalar kullanılarak ölçülmüştür. Bu çalışmayı yönlendiren varsayım, bağdaşık bir metinde, ard arda gelen cümlelerde geçen kavramların, ontolojilerde birbiriyle yakın ilişki içinde, yani anlamsal olarak ilintili, olmaları yönündedir.

Ontoloji tabanlı anlamsal ilintililik yönteminin başarısı, genelde literatürde de uygulandığı şekilde, insanlar tarafından ilintililikleri değerlendirilmiş kavram çiftleri kullanılarak sınanmıştır. Bu değerlendirmede, insanların verdikleri ilintililik değerleri ile ontoloji tabanlı anlamsal ilintililik yöntemi ile belirlenen değerler arasındaki korelasyon hesaplamıştır. Hesaplama sonucunda, insanların değerlendirmeleri ile otomatik olarak hesaplanan değerler arasındaki korelasyonun ($r=0,685$; $p<0,01$) anlamlı ve pozitif yönde olduğu görülmüştür. Bu korelasyon değeri literatürde anlamsal benzerlik ölçüm yöntemleri için raporlanan korelasyon değerleri içindeki en yüksek değer değildir. Fakat benzerliğin, ilintililiğin özel bir durumu olduğu düşünüldüğünde, ilintililik tahmininin benzerlik tahminine göre daha zor olduğu sonucuna varılabilir. Bundan dolayı geliştirilen yöntemin, bu değerlendirmeye göre bir dereceye kadar başarılı olduğu söylenebilir. Anlamsal ilintililik ve benzerlik ölçüm yöntemlerinin başarısını sadece insanların kavramların ilintililik yargılarını puanlamaları sonucunda elde edilen veri kümesi ile değerlendirmek doğru değildir. Çünkü anlamsal ilintililik ve benzerlik ölçüm yöntemlerinin ürettiği sonuçlar bir gerçek dünya probleminin çözümünde kullanılmadığı takdirde faydasızdır. Bundan dolayı bu tez kapsamında, ontoloji tabanlı anlamsal ilintililik ölçüm yöntemi kullanılarak metin bağdaşıklığı otomatik olarak değerlendirilmiştir.

Ontoloji tabanlı anlamsal ilintililik yönteminin, metin bağdaşıklığının otomatik olarak ölçümündeki başarısını değerlendirmek için yapılan deneyde, insanların metin bağdaşıklığı için yaptıkları değerlendirmeler ile otomatik metin bağdaşıklığı ölçümü sonucunda elde edilen sonuçlar karşılaştırılmıştır. Karşılaştırma sonucunda, bu tez kapsamında geliştirilen ontoloji tabanlı anlamsal

ilintililik ölçüm yöntemi ile hesaplanan metin bağdaşıklık değerleri, insan yargıları ile istatistiksel açıdan anlamlı ilişkiye sahip tek yöntem olarak belirlenmiştir ($r=0,359$; $p<0,05$). Yapılan deneyde, Resnik'in (Resnik, 1985), Lin'in (Lin, 1998) ve Jiang ve Conrath'ın (Jiang and Conrath, 1997) yöntemi ile yapılan hesaplamalarda, bu tez kapsamında geliştirilen ontoloji tabanlı anlamsal ilintililik yöntemi ile ilintililik hesaplamak için kullanılan ontolojiler kullanılmıştır. Jiang ve Conrath'ın (Jiang and Conrath, 1997) yöntemi patika ve bilgi içeriği tabanlı yaklaşımları birleştirdiği için, Resnik'in (Resnik, 1985) ve Lin'in (Lin, 1998) önerdiği bilgi içeriği tabanlı yaklaşımlar ise benzerlik hesabı yaparken kavramların ata ve alt-kavramlarını dikkate aldığı için bu yöntemler de bir kavramsal modele gereksinim duymaktadır. Kelime çiftleri için yapılan ilintililik ve benzerlik ölçümlerinde bu dört yöntem (Resnik'in (Resnik, 1985), Lin'in (Lin, 1998), Jiang ve Conrath'ın (Jiang and Conrath, 1997) yöntemleri ve ontoloji tabanlı anlamsal ilintililik ölçüm yöntemi) için de aynı ontolojiler kullanılmıştır. Dolayısıyla bu yöntemlerin eşit koşullarda kıyaslandığı söylenebilir. Gizli Anlam Çözümleme yöntemi ise bir kavramsal modele gereksinim duymadığı için bu yöntemle yapılan ölçümler diğer yöntemlerle yapılanlardan tamamen bağımsızdır.

Anlamsal ilintililik ve benzerlik ölçümünde ontolojilerin kullanımının, metin bağdaşıklığının otomatik değerlendirilme başarısına etkisini gözlemlemek için, ontolojiler kullanılarak yapılan deney WordNet kavramlar dizini (Fellbaum, 1998) ve çeşitli derlemeler kullanılarak tekrarlanmıştır. Deney sonuçlarına göre, Resnik'in (Resnik, 1985) ve Lin'in (Lin, 1998) benzerlik ölçüm yöntemleri ile ontolojilerin kullanıldığı durumda elde edilen korelasyon değerleri ($r_{\text{Resnik}}=0,285$; $r_{\text{Lin}}=0,342$), WordNet kavramlar dizini kullanıldığı durumdakilere ($r_{\text{Resnik}}=0,062$; $r_{\text{Lin}}=-0,012$; her iki korelasyon değeri de Penn Treebank derlemi kullanılarak yapılan ölçüm sonucunda elde edilmiştir) göre daha yüksektir. Benzerlik hesaplamaları için kullanılan SimPack kütüphanesindeki Wu ve Palmer'ın (Wu and Palmer, 1994) ve Leacock ve Chodorow'un (Leacock and Chodorow, 1998) benzerlik yöntemlerini, Java WordNet::Similarity kütüphanesinde ise Jiang ve Conrath'ın (Jiang and Conrath, 1997) yöntemini gerçekleştiren işlevlerin sağlıklı çalışmamasından dolayı bu yöntemler kullanılarak bir karşılaştırma yapılamamıştır.

Metin bağdaşıklık değerlerinin otomatik hesaplanması sonucunda elde edilen değerler, bağdaşıklık değerlendirilirken kullanılan kaynağa (kavramsal model, kavramlar dizini, derlem, vs.; Bkz. Bölüm 4.2, Sayfa 56) göre

değişmektedir. Kavramlar arasındaki benzerlikleri ve metin bağdaşıklığını ölçmek için sıklıkla kullanılan yöntemlerden birisi olan Gizli Anlam Çözümleme için de bu durum geçerlidir. Çizelge 7.1’de, bu tez çalışması kapsamındaki deneyde kullanılan 33 metnin değişik Gizli Anlam Çözümleme anlamsal uzayları kullanılarak hesaplanan bağdaşıklık değerlerinin katılımcıların bağdaşıklık değerlendirmeleri ile arasındaki korelasyonlar gösterilmektedir. Çizelge 7.2’de ise 33 metnin farklı anlamsal uzaylar kullanılarak hesaplanan en düşük ve en yüksek bağdaşıklık değerleri gösterilmiştir. Bu sonuçlara göre, hesaplanan bağdaşıklık değerleri kullanılan anlamsal uzaya ve dolayısıyla bu anlamsal uzay oluşturulurken kullanılan derleme bağlıdır.

Çizelge 7.1: Katılımcıların bağdaşıklık değerlendirmeleri ile Gizli Anlam Çözümleme yöntemi ve değişik derlemler kullanılarak hesaplanan otomatik bağdaşıklık değerlendirmeleri arasındaki korelasyon değerleri.

Anlamsal Uzay	Korelasyon
HSBio	0,331
Biology_HS_betatest	0,291
Psychology_Myers_5th_ed	0,209
Cognit	0,125
Heart	0,090
General_Reading_up_to_1st_year_college	0,045
Mesoamerican	0,041
General_Reading_up_to_12th_Grade	0,022
General_Reading_up_to_03rd_Grade	0,020
General_Reading_up_to_09th_Grade	0,004
General_Reading_up_to_06th_Grade	-0,012
* p < 0,05 (2-tailed) ** p < 0,01 (2-tailed)	

Resnik’in (Resnik, 1985) ve Lin’in (Lin, 1998) benzerlik ölçüm yöntemleri ile farklı kaynak kullanılması durumunda da elde edilen sonuçlar değişmektedir (Bkz. Çizelge 6.6 ve Çizelge 6.7). Çizelge 6.7’deki değerlere göre, WordNet kavramlar dizini (Fellbaum, 1998) ile birlikte bilgi içeriği değerlerini hesaplamak için kullanılan derlemlere bağlı olarak elde edilen sonuçlar değişmektedir.

Çizelge 7.2: Metinlerin Gizli Anlam Çözümleme yöntemi ile hesaplanan en düşük ve en yüksek bağdaşıklık değerleri.

Metin	En Düşük Değer	En Yüksek Değer
D30003.M.100.T.13	0,15	0,51
D30003.M.100.T.18	0,05	0,22
D30003.M.100.T.26	0,10	0,28
D30003.M.100.T.6	0,05	0,25
D30003.M.100.T.C	0,06	0,31
D30012.M.100.T.13	0,06	0,23
D30012.M.100.T.16	0,12	0,32
D30012.M.100.T.26	0,02	0,22
D30012.M.100.T.6	0,07	0,15
D30028.M.100.T.26	0,04	0,32
D30028.M.100.T.6	0,19	0,50
D30028.M.100.T.F	0,12	0,39
D30040.M.100.T.6	0,15	0,43
D31009.M.100.T.13	0,01	0,20
D31009.M.100.T.26	0,09	0,29
D31009.M.100.T.6	0,12	0,37
D31010.M.100.T.13	0,19	0,35
D31010.M.100.T.16	0,11	0,30
D31010.M.100.T.18	0,03	0,19
D31010.M.100.T.26	0,15	0,31
D31010.M.100.T.6	0,17	0,43
D31010.M.100.T.E	0,05	0,15
D31033.M.100.T.13	0,02	0,13
D31033.M.100.T.16	0,04	0,29
D31033.M.100.T.18	0,01	0,37
D31041.M.100.T.13	0,08	0,27
D31041.M.100.T.16	0,19	0,50
D31041.M.100.T.18	0,08	0,24
D31041.M.100.T.6	0,15	0,40
D31050.M.100.T.13	0,04	0,45
D31050.M.100.T.16	0,13	0,28
D31050.M.100.T.26	0,11	0,46
D31050.M.100.T.F	0,08	0,19

Benzer şekilde, bu tez çalışması kapsamında geliştirilen yöntem kullanılarak hesaplanan bağdaşıklık değerleri de kullanılan ontolojilere göre değişmektedir. Bu durum bağdaşıklık kavramının içeriğiyle tamamen tutarlıdır. Bağdaşıklık (Bkz. Bölüm 3.3, Sayfa 21), metin ile okuyucu arasındaki etkileşimin sonucu olduğu için öznel bir kavramdır. Dolayısıyla bir metin, bir okuyucu için bağdaşık olurken başka bir okuyucu için bağdaşık olmayabilir. Bağdaşıklığın algılanması, okuyucunun artalan bilgisine de bağlıdır. Metnin bağdaşıklığı otomatik olarak hesaplanırken, kullanılan kaynak bir ölçüde okuyucunun artalan bilgisini temsil etmektedir. Şekil 5.5'te gösterildiği gibi, bu tez kapsamında geliştirilen anlamsal ilintililik ölçüm yöntemi, bellekteki kavramlar ağını, yani bir ölçüde artalan bilgisini, OWL dili ile oluşturulmuş bir ontoloji ile temsil etmektedir. Benzer şekilde, diğer ilintililik ve benzerlik ölçüm yöntemleri de değerlendirme için bir

kavramsal model, WordNet (Fellbaum, 1998) gibi bir kavramlar dizini ya da bir derlem gibi bir kaynak kullanmaktadır. Dolayısıyla artalan bilgisine göre deęiřen baędařıklığın kullanılan kaynaęa göre de deęiřmesi mantıklıdır.

7.1 Ontoloji Tabanlı Anlamsal İlintililik Ölçüm Yönteminin Kullanılabileceęi Alanlar

İlintililik ve benzerlik ölçümü yapılarak birçok alandaki farklı problemlere çözümler getirilmektedir. Bu tez kapsamında, geliştirilen ontoloji tabanlı anlamsal ilintililik yönteminin metin baędařıklığının otomatik olarak ölçülmesi için kullanımı bu çözümlere bir örnektir. Bunun dışında ilintililik ve benzerlik ölçümü, web servislerinin otomatik olarak keřfedilmesi, tetiklenmesi ve birleřtirilmesi (Martin et al., 2004; Syeda-Mahmood et al., 2005), kelime anlamı belirginleřtirme (Pedersen et al., 2005), web’deki varlıkların benzerlerinin bulunması (iki farklı alışveriř sitesinde aynı ürünün bulunması ya da sosyal aęlarda benzer ilgi alanlarına sahip kullanıcıların tespit edilmesi gibi) (Sahami, 2006), belge kümeleme işlemleri (Zhang et al., 2008) ve otomatik kısa cevap puanlama (Mohler and Mihalcea, 2009) gibi problemlerin çözümü için kullanılmaktadır. Anlamsal arama ise ilintililik ve benzerlik ölçümünün kullanıldıęı bir başka önemli alandır. Ontoloji tabanlı benzerlik ölçümlerinin kullanımıyla birlikte, bilgiye erişimin başarısını ölçen “kesin isabet” ve “eriřim isabeti” oranlarında artış gözlemlenmiřtir (Mao an Chu, 2002). Alt ontoloji çıkarma ise ilintililik ve benzerlik yöntemlerinin kullanılabileceęi bir başka alan olarak ortaya çıkmaktadır.

Pedersen et al., ilintililik ölçüm yöntemlerinin kelime anlamı belirginleřtirme işlemi için kullanımının faydalı olduęunu yaptıkları deneyler sonucunda tespit etmiřtir (Pedersen et al., 2005). Pedersen et al.’nin kelime anlamı belirginleřtirme için önerdięi yöntem, iki kavram arasındaki ilintililięi hesaplayabilen tüm ölçüm yöntemleriyle kullanılabilmektedir. Kelime anlamı belirginleřtirme işleminin başarısının kullanılan anlamsal ilintililik yönteminin kavramlar arasındaki ilintililięi ölçme başarısı ile doęru orantılı olduęu rapor edilmiřtir. Bu tez kapsamında geliştirilen anlamsal ilintililik yönteminin de kelime anlamı belirginleřtirme işlemi için kullanımının başarılı sonuç verip vermedięi arařtırmaya deęerlidir.

Web servislerinin, insan müdahalesi olmadan bilgisayarlar tarafından keřfedilmesi, tetiklenmesi ve birleřtirilmesi amacının gerçekteřtirilebilmesi için web servislerini tanımlayan özelliklerin anlamsal olarak etiketlenmesi

gerekmektedir. Böylece bilgisayarlar anlamsal etiketleri kullanarak web servislerinin içeriklerinin anlamı hakkında işlenebilir veriye sahip olacaktır ve bu veriyi kullanarak anlamsal seviyede bazı işlemler yapabilecektir. Web servislerinin keşfedilmesi işleminde anlamsal ilintililik ya da benzerlik yönteminin kullanılmasının, işlemin başarısını arttırdığı belirtilmiştir (Syeda-Mahmood et al., 2005). Syeda-Mahmood et al., kavramsal modellerdeki `equivalentClass`, `subClassOf`, `superClassOf` ve `type` ilişkilerini kullanarak kavramlar arasındaki uzaklıkları hesaplamıştır ve bu uzaklık değerlerini kullanarak gelen isteğe uygun web servislerini bulmuştur. Bu tez kapsamında geliştirilen anlamsal ilintililik ölçüm yöntemi web servislerinin keşfedilmesi için kullanılabilir. Fakat bu ölçüm yönteminin iki kavram arasındaki ilintililik değerini hesaplayabilmesi için iki kavramın aynı ontolojide ya da birbiriyle bağlantılı ontolojilerde olmasının gerekmesi kullanımı kısıtlayıcı bir koşuldur. Ayrıca Wu et al. web servislerinin anlamsal olarak etiketlenmesi için gerekli ontolojilerin oluşturulmasının çok ciddi insan emeği gerektirdiği için web servislerinin keşfedilmesinin işleminin başarısını arttırmak için Gizli Anlam Çözümleme gibi istatistiksel yöntemlerin kullanılmasını önermiştir (Wu et al., 2008). Fakat zaman içinde açık ortamda paylaşılan ontolojilerin sayılarının ve kalitelerinin artmasıyla birlikte anlamsal etiketleme için gerekli ontolojilerin oluşturulması yerine bu ontolojilerin yeniden kullanımı tercih edilebilecektir.

7.2 Metin Bağdaşıklığının Otomatik Olarak Ölçümü ve Gereksinim Mühendisliği

Yazılım geliştirme sürecinde, kaliteli gereksinim belirtilmelerinin üretilmesi, geliştirilen yazılım sisteminin başarısı açısından vazgeçilmez bir gerekliliktir. Gereksinim belirtilmelerinin kalitelerinin ölçülmesi ve iyileştirilmesi, hem üretilen yazılım sisteminin başarısız olma olasılığını azaltacaktır hem de yazılım sisteminin kullanıcı gereksinimlerini karşılama başarısını arttıracaktır. Ayrıca, yazılım geliştirme sürecinde yapılan hataların ya da eksikliklerin sürecin mümkün olduğu kadar başında fark edilmesi, yapılacak düzeltmelerin ve eklemelerin maliyetini düşürecektir. Gereksinim mühendisliği sürecinde fark edilen hataların düzeltilmesinin maliyeti, geliştirilmiş bir sistem üzerinde yapılacak düzeltmeye göre çok daha az olacaktır. Dolayısıyla gereksinim belirtilmelerinin ve modellerinin kalitelerinin ölçülmesinin önemi bu noktada ortaya çıkmaktadır. Davis et al., kaliteli gereksinim belirtilmelerinin sahip olmaları gereken bir özelliğin de “anlaşılabilirlik” olduğunu belirtmiştir (Davis et al., 1993).

Gereksinim belirtilmelerini oluřturmak iin ok eřitli yntemler ve biimsel diller bulunsa da, gereksinim belirtilmelerinin oėunluėu doėal dil ile yazılmaktadır. Luisa et al.'nin arařtırmasına gre gereksinim belirtilmelerinin %79'u doėal dil, %16'sı formlarla ve řablonlarla birlikte yapısallařtırılmıř doėal dil, %5'i ise biimsel diller kullanılarak yazılmaktadır (Luisa et al., 2002).

Gereksinim belirtilmelerinin, yazılım geliřtirme srecindeki tm paydařlar tarafından anlařılabilmesi, srecin saėlıklı yrmesi iin gerekli olan kořullardan biridir. Doėal dil ile yazılmıř gereksinim belirtilmelerinin baėdařlıklıklarının otomatik olarak deėerlendirilmesi ve baėdařık olmayan belirtilmelerin yazılım geliřtirme srecinin bařlarında dzeltilmesinin saėlanması, geliřtirilecek yazılımın kullanıcı gereksinimlerini karřılama bařarısını arttıracaktır. Bunun iin bu tez kapsamında geliřtirilen anlamsal ilintililik lm yntemi kullanılarak metin baėdařlıklığının deėerlendirilmesi onerilebilir. Gereksinim belirtilmelerinin bir konu alanını kapsaması ve bu konu alanını modelleyen ontolojilerin bulunma ya da bu ontolojilerin yazılım geliřtirme srecinde geliřtirilme olasılıėının yksek olması, doėal dil ile yazılmıř gereksinim belirtilmelerinin anlamsal olarak etiketleme iřini kolaylařtırması ve belirtilmelerdeki ok sayıdaki kelimenin ontolojik terimlerle etiketlenmesi, dolayısıyla da eřleřme oranının yksek olması beklenebilir.

7.3 Gelecek alıřmalar

Bu alıřma kapsamında geliřtirilen ontoloji tabanlı anlamsal ilintililik ynteminin bařarısı tatmin edici bulunmuř olsa da, alıřmanın geliřtirilebilecek ynleri bulunmaktadır.

Doėal dil ile yazılmıř metinlerde, baėdařlıklığa katkıda bulunan yapılardan birisi de eř gndergelerdir. Kiřilerin yerini tutan kiři adılları ya da eřitli varlıkları simgeleyen kısaltmalar eř gndergelere rnek olarak verilebilir. Eř gndergeler de ardışık cmlerler arasında anlamsal baėlantıların kurulmasına katkıda bulunur. Dolayısıyla bir metnin baėdařlıklığının deėerlendirilmesi sırasında eř gnderge zmlemesinin yapılması, elde edilecek sonuların gereėe daha yakın olmasını saėlayacaktır. Bu alıřmanın geliřtirilebilecek ynlerinden birisi, baėdařlıklığın otomatik olarak deėerlendirilmesinden nce eř gnderge zmleme iřlemi yapılarak bu sonuların deėerlendirmede gz nne alınmasıdır.

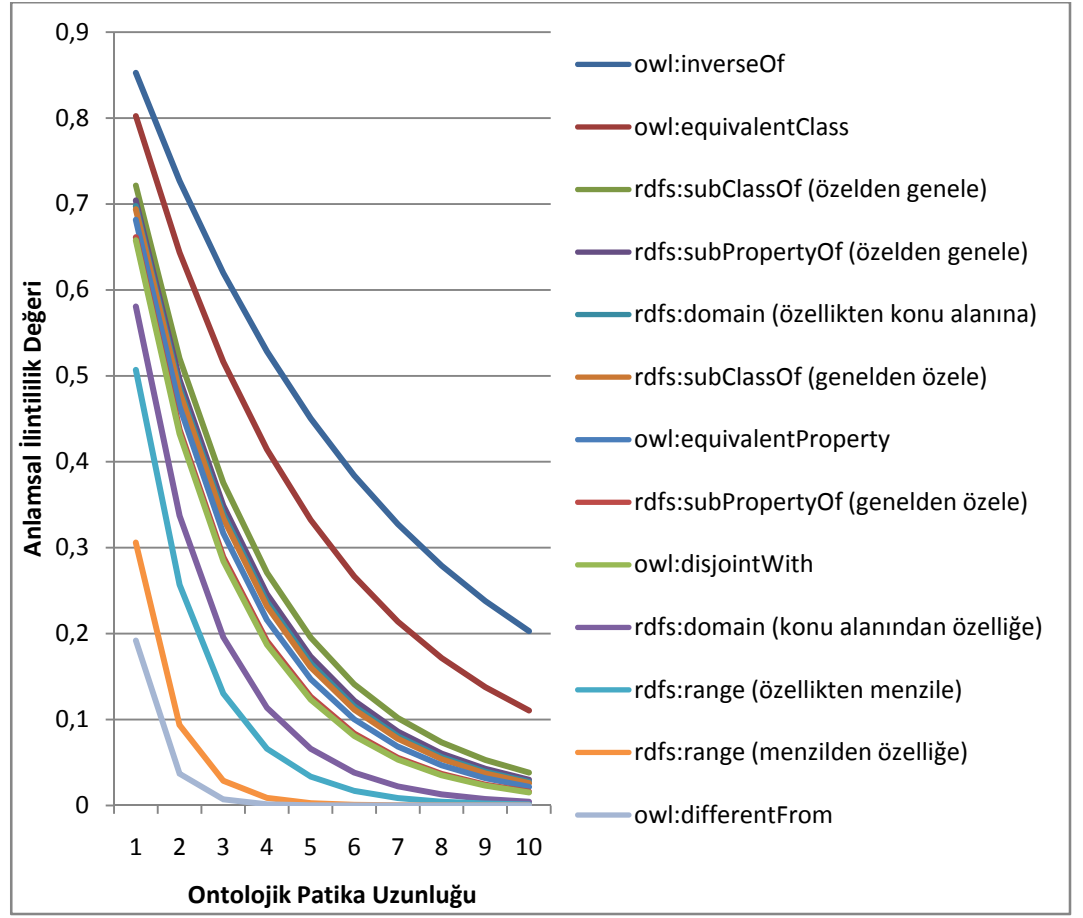
Metinlerin anlamsal olarak etiketlenmesi işlemi, metindeki kelimelerin ontolojik terimlerle sözdizimsel olarak kıyaslanması yoluyla yapılmaktadır. Bu işlem sırasında çok anlamlı kelimelerin ontolojilerdeki farklı anlamlarda kullanılan kelimelerle eşleşme olasılığı bulunmaktadır. Bunun engellenmesi için kelime anlamı belirginleştirme işlemi çeşitli yöntemler kullanılarak otomatik olarak yapılabilir. Bu çalışma kapsamında kelime anlamı belirginleştirme işlemi, yarı otomatik eşleştirme adımıyla (Bkz. Şekil 6.2, Sayfa 103) el ile yapılmıştır.

OWL dil öğelerinin anlamsal ilintililik katsayılarını belirlemek için kullanılan kavram çiftlerinin sayısı artırılarak, dil öğelerine göre dağılımın Çizelge 5.2'ye daha uyumlu duruma getirilmesi gerekmektedir. Ayrıca çözümlenen ontoloji sayısı da artırılarak elde edilen dağılımın gerçeği daha iyi yansıtmayı sağlanabilir. Benzer şekilde, anlamsal ilintililik ölçüm yönteminin başarısını sınamak için kullanılan 24 kavram çiftinin sayısının da artırılması daha güvenilir sonuçlar alabilmek için gereklidir.

Bu çalışma kapsamında, bir ontolojideki iki kavram arasındaki ontolojik patika üzerinde bulunabilecek OWL dil öğelerinden 9 tanesi için anlamsal ilintililik katsayısı belirlenmiştir. Kapsama alınmayan diğer dil öğeleri için de anlamsal ilintililik katsayılarının belirlenmesi, anlamsal ilintililik ölçüm yönteminin başarısını arttıracaktır. Örneğin, bir birey ile o bireyin ait olduğu sınıf arasındaki ilişkiyi kuran `rdf:type` dil ögesi anlamsal ilintililik katsayısı hesaplanması gereken öğeler için bir örnektir.

Bu tez kapsamında geliştirilen ontoloji tabanlı anlamsal ilintililik ölçüm yöntemi, “etkinleştirmenin yayılması teorisi” üzerine temellendirilmiştir. Bundan dolayı OWL dil öğelerinin katsayıları 0 ile 1 arasında belirlenmiştir. Böylece bir kavramdan uzaklaşıp diğer kavrama yaklaşıldıkça, aradaki ontolojik patikalarda bulunan dil öğelerine ve ilişki sayısına bağlı olarak, iki kavram arasındaki anlamsal ilintililik değerinin sıfıra yakınsaması sağlanmıştır. Fakat özellikle anlamsal ilintililik katsayısı düşük olan dil öğeleri, iki kavram arasındaki ontolojik patika üzerinde bulunduğunda, kavramlar arasındaki ilintililik değeri çok hızlı sıfıra yakınsamaktadır. Anlamsal ilintililik hesaplama formülü, yine 0 ile 1 arasında bir değer verecek, fakat sıfıra yakınsamanın yavaşlatılacağı bir şekilde değiştirilebilir. Şekil 7.1’de OWL dil ögesi bazında, ontolojik patikanın uzunluğu arttıkça anlamsal ilintililik değerinin değişimi gösterilmektedir. Bu şekle göre, özellikle anlamsal ilintililik katsayısı düşük olan öğelerin (`owl:differentFrom` gibi) bulunduğu patikalar hızla sıfıra yakınsamaktadır.

Şekil 7.1, ontolojik patika üzerinde hep aynı dil ögesi olduğu varsayılarak hazırlanmıştır.



Şekil 7.1: OWL dil ögeleri bazında ontolojik patika boyunca anlamsal ilintililik değerinin değişimi.

OWL dil ögelerinin anlamsal ilintililik katsayıları için gelebilecek bir eleştiri, bu katsayıların tüm ontolojiler ve konu alanları için aynı olmasıdır. Bu çalışma kapsamında belirlenen anlamsal ilintililik katsayıları göz önüne alınarak ve eldeki ontolojinin özelliklerine bakılarak, dil ögeleri için o ontolojiye özgü anlamsal ilintililik katsayıları belirleme fikri de araştırmaya değer görülmektedir.

Metin bağdaşıklığının katılımcılar tarafından değerlendirilmesi oldukça zor bir iştir. Bu tez kapsamında Lapata ve Barzilay'ın hazırladığı gibi bir metin kümesi ve bu metinlerin bağdaşıklık değerleri, katılımcılara anket yapılarak oluşturulmaya çalışılmıştır. Fakat elde edilen sonuçlar kendi içinde tutarlı olmadığı için bu çalışmada kullanılmamıştır. Bunun yerine, Lapata ve Barzilay'ın hazırladığı ve literatürde kabul görmüş olan metinler ve bağdaşıklık değerleri kullanılmıştır (Lapata and Barzilay, 2005; Barzilay and Lapata, 2008). Farklı bir

yöntem olarak, katılımcılara bir metin okutularak ve metnin içeriği hakkında sorular yöneltilerek, katılımcıların metni ne kadar iyi kavradıkları ölçülebilir (Kalalı, 2002). Katılımcıların verdiği cevaplar puanlanarak metnin bağdaşıklığı hakkında bir değerlendirme elde edilebilir.

Lapata ve Barzilay'ın çalışmasında (Lapata and Barzilay, 2005), katılımcıların metin bağdaşıklıkları için verdiği değerlerin ortalaması alınmaktadır. Bu ortalamalar ile değişik benzerlik yöntemleri kullanılarak otomatik olarak elde edilen metin bağdaşıklık değerleri arasındaki ilişkilerin anlaşılması için Pearson Momentler Çarpımı Korelasyon tekniği kullanılmıştır. Fakat istatistiksel açıdan, katılımcıların metinler için verdikleri bağdaşıklık değerlerinin ortalamasını almak yerine her bir katılımcının verdiği değerler ile otomatik olarak elde edilen değerleri karşılaştırmak daha doğru görünmektedir. Bunun sonucunda elde edilen korelasyon değerleri ise belirli aralıklar içinde (örneğin 0,80-1 arası çok iyi; 0,60-0,79 arası iyi gibi) raporlanabilir.

KAYNAKLAR DİZİNİ

- Alagözlü, N.**, 2005, Yılmaz Erdoğan'ın "Keşke" Öyküsü Ve Söylem Biçembilimi: Bir Yöntem ve Uygulama, *Journal of Language and Linguistic Studies*, 1(1):67-80.
- Allemang, D. and Hendler, J.**, 2008, *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*, Morgan Kaufmann Publishers Inc.
- Allen, J.**, 1995, *Natural Language Understanding (2nd Ed.)*, Benjamin-Cummings Publishing Co., Inc.
- Allen, R. B. and Wu, Y.**, 2005, Metrics for the scope of a collection: Research Articles, *Journal of the American Society for Information Science and Technology*, 56(12):1243-1249pp.
- Anderson, J. R.**, 1983, A spreading activation theory of memory, *Journal of Verbal Learning and Verbal Behavior*, 22:261-295pp.
- Andreasen, T., Bulskov, H. and Knappe, R.**, 2003, Similarity from conceptual relations, in Ellen Walker (Eds.): 22nd International Conference of the North American Fuzzy Information Processing Society, NAFIPS 2003, Chicago, Illinois USA, 179-184pp.
- Antoniou, G. and van Harmelen, F.**, 2008, *A Semantic Web Primer*, 2nd Edition, The MIT Press.
- Arlow, J. and Neustadt, I.**, 2002, *UML and the Unified Process*, Addison Wesley.
- Atkinson, C. and Kühne, T.**, 2003, Model-driven development: A metamodeling foundation, *IEEE Software* 20:36-41pp.
- Aydoğan, R. and Yolum, P.**, 2007, Learning consumer preferences using semantic similarity, 6th international Joint Conference on Autonomous Agents and Multiagent Systems (Honolulu, Hawaii, May 14 - 18, 2007), AAMAS '07, ACM, New York, 1-8pp.

KAYNAKLAR DİZİNİ (devam)

- Baclawski, K., Kokar, M.M., Kogut, P.A., Hart, L., Smith, J.E., Letkowski, J. and Emery, P.**, 2002, Extending the Unified Modeling Language for ontology development, *Software and System Modeling* 1, 142-156pp.
- Baddeley, A. D.**, 1992, Working memory, *Science*, 255:556-559pp.
- Barzilay, R. and Lapata, M.**, 2008, Modeling Local Coherence: An Entity-Based Approach, *Computational Linguistics*, 34(1), 1-34pp.
- Bayram, A.**, 1993, Türkçe - İngilizce Teknik Terimler Sözlüğü, Fono, İstanbul.
- Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F. Stein, L. A.**, 2004, OWL Web Ontology Language Reference, W3C Recommendation, 10 February 2004, <http://www.w3.org/TR/2004/REC-owl-ref-20040210/>, (Erişim tarihi: 22 Ocak 2010).
- Berners-Lee, T.**, 1998, What the semantic web can represent, <http://www.w3.org/DesignIssues/RDFnot.html>, (Erişim tarihi: 22 Ocak 2010).
- Bernstein, A., Kaufmann, E., Kiefer, C. and Bürki, C.**, 2005, Simpack: A Generic Java Library for Similiarity Measures in Ontologies (Working Paper), Department of Informatics, University of Zurich.
- Bollegala, D., Matsuo, Y. and Ishizuka, M.**, 2009., A relational model of semantic similarity between words using automatically extracted lexical pattern clusters from the web, *The 2009 Conference on Empirical Methods in Natural Language Processing*, 2, Annual Meeting of the ACL, Association for Computational Linguistics, Morristown, NJ, 803-812pp.
- Bontcheva, K., Tablan, V., Maynard, D. and Cunningham, H.**, 2004, Evolving GATE to Meet New Challenges in Language Engineering, *Natural Language Engineering*, 10 (3/4):349-373pp.

KAYNAKLAR DİZİNİ (devam)

- Bontcheva, K., Cunningham, H., Kiryakov, A. and Tablan, V.,** 2006, Semantic Annotation and Human Language Technology, Semantic Web Technology: Trends and Research, John Wiley and Sons Ltd.
- Britton, B. K. and Gulgoz, S.,** 1991, Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures, *Journal of Educational Psychology*, 83:329-345pp.
- Budanitsky, A. and Graeme H.,** 2001, Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures, In *Workshop on WordNet and Other Lexical Resources, Second Meeting of the North American Chapter of the Association for Computational Linguistics*, Pittsburgh, PA, 29-34pp.
- Cambridge University Press,** 2009, *Cambridge Learner's Dictionary English-Turkish*, 916p.
- Carroll, J. J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A. and Wilkinson, K.,** 2004, Jena: Implementing the Semantic Web Recommendations, *13th International World Wide Web Conference*, ACM, New York, USA, 74-83pp.
- Chakrabarti, S.,** 2003, *Mining the Web: Discovering Knowledge from Hypertext Data*, Morgan Kaufmann Publishers.
- Collins, A. M. and Quillian, M. R.,** 1969, Retrieval time from semantic memory, *Journal of Verbal Learning and Verbal Behavior*, 8:240-247pp.
- Collins, A. M. and Loftus, E. F.,** 1975, A spreading-activation theory of semantic processing, *Psychological Review*, 82(6):407-428pp.
- Cozby, P. C.,** 2005, *Methods in Behavioral Research*, 9th Edition, McGraw-Hill, New York, 410p.

KAYNAKLAR DİZİNİ (devam)

- Cunningham, H., Maynard, D., Bontcheva, K. and Tablan, V., 2002**, GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications, 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia.
- Cunningham, H., Maynard, D., Bontcheva, K. and Tablan, V., 2009**, Developing Language Processing Components with GATE - Version 5 (a User Guide).
- Daconta, M. C., Smith, K. T. and Obrst, L. J., 2003**, The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management, John Wiley & Sons, Inc.
- d'Amato, C., Fanizzi, N. and Esposito, F., 2005**, A semantic similarity measure for expressive description logics, CILC 2005, Convegno Italiano di Logica Computazionale 21-22 June 2005, Rome, Italy.
- d'Aquin, M., Sabou, M., Dzbor, M., Baldassarre, C., Gridinoc, L., Angeletou, S. and Motta, E., 2007**, Watson: A Gateway for the Semantic Web, Poster session of the European Semantic Web Conference, ESWC 2007.
- Dang, H. T., 2005**, Overview of DUC 2005, Document Understanding Conference Workshop 2005 (DUC 2005) at the Human Language Technology Conf./Conf. on Empirical Methods in Natural Language Processing (HLT/EMNLP).
- Davis, A., Overmyer, S., Jordan, K., Caruso, J. A. C. J., Dandashi, F. A. D. F., Dinh, A. A. D. A., Kincaid, G. A. K. G., Ledebor, G. A. L. G., Reynolds, P. A. R. P., Sitaram, P. A. S. P., Ta, A. A. T. A. And Theofanos, M. A. T. M., 1993**, Identifying and measuring quality in a software requirements specification, Proceedings of 1st International Software Metrics Symposium.
- Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R., Jhingran, A., Kanungo, T., Rajagopalan, S., Tomkins, A., Tomlin, J. A. and Zien, J. Z., 2003**,

KAYNAKLAR DİZİNİ (devam)

SemTag and Seeker: Bootstrapping the Semantic Web via Automated Semantic Annotation, 12th international conference on World Wide Web (WWW '03), ACM, New York, NY, USA, 178-186pp.

Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Peng, Y., Reddivari, P., Doshi, V. C. and Sachs, J., 2004, Swoogle: A Search and Metadata Engine for the Semantic Web, 13th ACM Conference on Information and Knowledge Management.

English, J. and Nirenburg, S., 2007, Ontology Learning from Text Using Automatic Ontological-Semantic Text Annotation and the Web as a Corpus, Proceedings of Machine Reading AAAI Spring Symposium.

Euzenat, J. and Valtchev, P., 2004, Similarity-based ontology alignment in OWL-Lite, In R. López de Mántaras and L. Saitta (Eds.), Proceedings of the 16th European conference on artificial intelligence (ECAI'2004), August 22-27, Valencia, Spain. IOS Press, 333-337pp.

Euzenat, J. and Shvaiko, P., 2007, Ontology Matching, Springer-Verlag, Heidelberg (DE).

Fellbaum, C., 1998, WordNet: An Electronic Lexical Database, Bradford Books.

Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G. and Ruppin, E., 2002, Placing search in context: The concept revisited, ACM Transactions on Information Systems, 20(1):116-131pp.

Foltz, P. W., Kintsch, W. and Landauer, T. K., 1998, The measurement of textual coherence with Latent Semantic Analysis, Discourse Processes, 25(2&3):285-307pp.

Francis, W. N. and Kucera, H., 1964, Brown corpus manual: Manual of information to accompany a standard corpus of present day edited American English, Brown University, Providence, Rhode Island.

KAYNAKLAR DİZİNİ (devam)

- Gabrilovich, E. and Markovitch, S.**, 2009, Wikipedia-based semantic interpretation for natural language processing, Journal of Artificial Intelligence Research (JAIR), 34(1):443-498pp.
- Ge, J. and Qiu, Y.**, 2008, Concept similarity matching based on semantic distance, 4th International Conference on Semantics, Knowledge and Grid, SKG, IEEE Computer Society, Washington, DC, 380-383pp.
- Gruber, T.R.**, 1993, A translation approach to portable ontology specifications, Knowl. Acquis. 5:199-220pp.
- Grünberg, T., Onart A., Grünberg, D., ve Turan, H.**, 2003, Mantık Terimleri Sözlüğü (Genişletilmiş 3. Basım), METU Press, Ankara.
- Handschuh, S., Staab, S. and Maedche, A.**, 2001, CREAM: Creating Relational Metadata with a Component-based, Ontology-driven Annotation Framework, 1st international conference on Knowledge capture (K-CAP '01), ACM, New York, NY, USA, 76-83pp.
- Hobbs, J. R.**, 1985, On the coherence and structure of discourse, CSLI Technical Report 85-37, Stanford, CA, USA.
- Horridge, M.**, 2009, A practical guide to building OWL ontologies using Protégé 4 and CO-ODE tools, Edition 1.2.
- İkiz, F., Püskülcü, H. ve Eren, Ş.**, İstatistiğe Giriş, 4. Baskı, Fakülteler Kitabevi Barış Yayınları, 1996.
- Jena Semantic Web Framework**, <http://openjena.org/index.html> (Erişim tarihi: 6 Şubat 2011).
- Jiang, J. J. and Conrath, D. W.**, 1997, Semantic similarity based on corpus statistics and lexical taxonomy, In Proceedings of the International Conference on Research in Computational Linguistic, Taiwan.

KAYNAKLAR DİZİNİ (devam)

- Johnson-Laird, P.N.**, 1983, Mental Models, Harvard University Press, Cambridge, Mass.
- Just, M. A. and Carpenter, P. A.**, 1980, A theory of reading: From eye fixations to comprehension, Psychological Review, 87(4):329-354pp.
- Just, M. A. and Carpenter, P. A.**, 1987, The Psychology of Reading and Language Comprehension, Boston: Allyn & Bacon.
- Just, M. A. and Carpenter, P. A.**, 1992, A capacity theory of comprehension: Individual differences in working memory, Psychological Review, 99:122-149pp.
- Kahan, J. and Koivunen, M. R.**, 2001, Annotea: An Open RDF Infrastructure for Shared Web Annotations, International World Wide Web Conference, Hong Kong, 623-631pp.
- Kaindl, H.**, 1993, The missing link in requirements engineering, SIGSOFT Software Engineering Notes, 18:30-39pp.
- Kalalı, F.**, 2002, The Effect of Causal Relationship on Cohesion and Recall: A Research on the Text Processing of Primary School Students, Yüksek Lisans Tezi, Mersin Üniversitesi, Sosyal Bilimler Enstitüsü, 87s (yayımlanmamış).
- Karanastasi, A. and Christodoulakis, S.**, 2007, Ontology-driven semantic ranking for natural language disambiguation in the OntoNL framework, 4th European Conference on the Semantic Web: Research and Applications (Innsbruck, Austria, June 03 - 07, 2007), E. Franconi, M. Kifer, and W. May, Eds. Lecture Notes In Computer Science, 4519. Springer-Verlag, Berlin, Heidelberg, 443-457pp.
- Karatay, H.**, 2010, Bağdaşıklık Araçlarını Kullanma Düzeyi İle Tutarlı Metin Yazma Arasındaki İlişki, Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 7(13):373-385.

KAYNAKLAR DİZİNİ (devam)

- Kehler, A.**, 2002, Coherence, Reference, and the Theory of Grammar, Stanford, CA: CSLI Publications.
- Kendeou, P. and van den Broek, P.**, 2007, The Effects of Prior Knowledge and Text Structure on Comprehension Processes During Reading of Scientific Texts, *Memory and Cognition*, 35(7), 1567–1577pp.
- Kintsch, W.**, 1974, The representation of meaning in memory, Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kintsch, W. and Vipond, D.**, 1979, Reading comprehension and readability in educational practice and psychological theory, in L. G. Nilsson (Eds.), *Perspectives on Memory Research*, Hillsdale, NJ: Erlbaum.
- Kintsch, W. and van Dijk, T. A.**, 1978, Toward a model of text comprehension and production, *Psychological Review*, 85:363-394pp.
- Kintsch, W.**, 1998, *Comprehension: A Paradigm for Cognition*, New York: Cambridge University Press.
- Kintsch, W.**, 1994, Text comprehension, memory and learning, *American Psychologist*, 49(4):294-303pp.
- Kintsch, W.**, 1998, *Comprehension: A Paradigm for Cognition*, Cambridge, MA: Cambridge University Press.
- Kiryakov, A., Popov, B., Terziev, I., Manov, D. and Ognyanoff, D.**, 2004, Semantic Annotation, Indexing, and Retrieval, *Web Semantics: Science, Services and Agents on the World Wide Web*, 2(1):49-79pp.
- Knappe, R.**, 2005, Measures of Semantic Similarity and Relatedness for Use in Ontology-based Information Retrieval, PhD Thesis, Roskilde University, Denmark, 143p (unpublished).

KAYNAKLAR DİZİNİ (devam)

- Köksal, H.**, 2009, Yabancı Dil Olarak Almanca Öğretiminde Sınav Değerlendirmesi, Bağdaşık Yapıya İlişkin Yanlış Çözümlemesi, Uluslararası 5. Balkan Eğitim ve Bilim Kongresi, Trakya Üniversitesi, Edirne, 2:75-78.
- Laclavik, M., Seleng, M., Gatial, E., Balogh, Z. and Hluchy, L.**, 2007, Ontology based Text Annotation - OnTeA, The 2007 conference on Information Modelling and Knowledge Bases XVIII, Duzi, M., Jaakkola, H., Kiyoki, Y. and Kangassalo, H. (Eds.), IOS Press, Amsterdam, The Netherlands, 311-315pp.
- Landauer, T. K. and Dumais, S. T.**, 1997, A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge, Psychological Review, 104:211-240pp.
- Lapata, M. and Barzilay, R.**, 2005, Automatic evaluation of text coherence: models and representations, In proceedings of the 19th International Joint Conference on Artificial Intelligence, 19:1085-1090pp.
- Leacock, C. and Chodorow, M.**, 1998, Combining local context and WordNet similarity for word sense identification, In: Christiane Fellbaum (Eds.) WordNet: An Electronic Lexical Database. Cambridge, MA:MIT Press.
- Lin, D.**, 1998, An information-theoretic definition of similarity, In Proceedings of the Fifteenth international Conference on Machine Learning (July 24 - 27, 1998), J. W. Shavlik, Ed. Morgan Kaufmann Publishers, San Francisco, CA, 296-304pp.
- Latent Semantic Analysis @ CU Boulder**, <http://lsa.colorado.edu/cgi-bin/LSA-sentence.html>, (Erişim tarihi: 6 Şubat 2011).
- Ludewig, J.**, 2003, Models in software engineering – an introduction, Software System Modeling 2:5-14pp.

KAYNAKLAR DİZİNİ (devam)

- Luisa, M., Mariangela, F. and Pierluigi, N. I.**, 2002, Market research for requirements analysis using linguistic tools, Technical Report 66, Informatica e Studi Aziendali, University of Trento.
- Malik, S. K., Prakash, N. and Rizvi, S.**, 2010, Semantic Annotation Framework for Intelligent Information Retrieval Using KIM Architecture, International Journal of Web & Semantic Technology, 1(4):12-25pp.
- Manning, C., Raghavan, P., and Schtze, H.**, 2008, Introduction to Information Retrieval, Cambridge University Press.
- Mao, W. and Chu, W. W.**, 2002, Free text medical document retrieval via phrased-based vector space model, in Proc. of AMIA'02, San Antonio, TX.
- Martin, D., Burstein, M., Hobbs, J., Lassila, O., McDermott, D., McIlraith, S., Narayanan, S., Paolucci, M., Parsia, B., Payne, T., Sirin, E., Srinivasan, N. and Sycara, K.**, 2004, OWL-S: semantic markup for web services, W3C Member Submission, <http://www.w3.org/Submission/OWL-S/>, (Erişim tarihi: 22 Ocak 2010).
- Mazuel, L. and Sabouret, N.**, 2008, Semantic relatedness measure using object properties in an ontology, In Proceedings of the 7th international Conference on the Semantic Web (Karlsruhe, Germany, October 26 - 30, 2008), A. Sheth, S. Staab, M. Dean, M. Paolucci, D. Maynard, T. Finin, and K. Thirunarayan, Eds. Lecture Notes In Computer Science, 5318. Springer-Verlag, Berlin, Heidelberg, 681-694pp.
- McBride, B.**, 2002, Jena: A Semantic Web Toolkit, IEEE Internet Computing, 6(6):55-59pp.
- McNamara, D. S.**, 2001, Reading Both High-Coherence and Low-Coherence Texts: Effects of Text Sequence and Prior Knowledge, Canadian Journal of Experimental Psychology, 55, 51-62pp.

KAYNAKLAR DİZİNİ (devam)

- Miller, G.A. and Charles, W.G.**, 1991, Contextual correlates of semantic similarity, *Language and Cognitive Processes*, 6(1):1-28pp.
- Mohler, M. and Mihalcea, R.**, 2009, Text-to-text semantic similarity for automatic short answer grading, In *Proceedings of the 12th Conference of the European Chapter of the Association For Computational Linguistics* (Athens, Greece, March 30 - April 03, 2009), European Chapter Meeting of the ACL, Association for Computational Linguistics, Morristown, NJ, 567-575pp.
- Nguyen, H. A. and Al-Mubaid, H.**, 2006, New ontology-based semantic similarity measure for the biomedical domain, *IEEE GrC 2006*, 623-628pp.
- Okazaki, N., Matsuo, Y., Matsumura, N. and Ishizuka, M.**, 2003, Sentence extraction by spreading activation through sentence similarity, *IEICE Trans. Information and Systems*, E86D, 9:1686-1694pp.
- Onursal, İ.**, 2003, Türkçe Metinlerde Bağdaşıklık ve Tutarlılık, 121-132, *Günümüz Dilbilim Çalışmaları*, A. Kıran, E. Korkut, S. Ağıldere (Derl.), Multilingual Yayınları, Dilbilim Dizisi, İstanbul.
- Onyshkevych, B.**, 1997, Ontosearch: Using an ontology as a search space for knowledge based text processing, PhD Dissertation, Carnegie Mellon University (unpublished).
- OWL Web Ontology Language Guide**, 2004, Smith, M. K., Welty, C. and McGuinness, D. L., Editors, W3C Recommendation, <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>, (Erişim tarihi: 22 Ocak 2010).
- OWL Web Ontology Language Use Cases and Requirements**, 2004, Heflin J. Editor, <http://www.w3.org/TR/webont-req/>, (Erişim tarihi: 22 Ocak 2010).
- Pedersen, T., Banerjee, S. and Patwardhan, S.**, 2005, Maximizing semantic relatedness to perform word sense disambiguation, University of Minnesota

KAYNAKLAR DİZİNİ (devam)

Supercomputing Institute Research Report UMSI 2005/25, Minnesota, March.

Pedersen, T., Patwardhan, S. and Michelizzi, J., 2004, WordNet::Similarity - Measuring the Relatedness of Concepts, Proceedings of Fifth Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL-04), Boston, MA, 38-41pp.

Perfetti, C. A., 1999, Comprehending written language: A blueprint of the reader, In C. Brown & P. Hagoort (Eds.), The neurocognition of language, Oxford University Press, 167-208pp.

Popov, B., Kiryakov, A., Kirilov, A., Manov, D., Ognyanoff, D. and Goranov, M., 2003, KIM - Semantic Annotation Platform, 2nd International Semantic Web Conference (ISWC2003), Florida, USA, 834-849pp.

Rada, R., Mili, H., Bicknell, E., Blettner, M., 1989, Development and application of a metric on semantic nets, IEEE Transactions on Systems, Man, and Cybernetics, 19.

Resnik, P., 1985, Using information content to evaluate semantic similarity in a taxonomy, in Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI '95), Montreal, August, 448-453pp.

Rouet, J., 2006, The Skills of Document Use: From Text Comprehension to Web-based Learning, Erlbaum, Mahwah, NJ.

Rubenstein, H. and Goodenough, J. B., 1965, Contextual correlates of synonymy, Communications of the ACM, 8(10):627-633.

Sahami, M., 2006, Mining the web to determine similarity between words, objects, and communities, Proceedings of the 19th International FLAIRS Conference (FLAIRS).

KAYNAKLAR DİZİNİ (devam)

- Salton, G., Wong, A. and Yang, C. S.**, 1975, A vector space model for automatic indexing, *Communications of the ACM* 18(11):613-620pp.
- Schwering, A.**, 2008, Approaches to semantic similarity measurement for geo-spatial data: A survey, *Transactions in GIS* 12(1):5-29pp.
- Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A. and Katz, Y.**, 2007, Pellet: A practical OWL-DL reasoner, *Web Semantics: Science, Services and Agents on the World Wide Web* 5(2):51-53.
- Slimani, T., Yaghlane, B.B. and Mellouli, K.**, 2006, A new similarity measure based on edge counting, *International Journal of Applied Science, Engineering and Technology*, 17.
- Solso, R. L., MacLin, M. K. and MacLin, O. H.**, 2005, *Cognitive Psychology* (7th ed.), Boston: Allyn & Bacon.
- Sowa, J. F.**, 1999, *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks Cole Publishing Co., Pacific Grove, CA.
- Stevenson, R. J.**, 1993, *Language, Thought, and Representation*, Wiley & Sons, Chichester, New York.
- Strube, M. and Ponzetto, S. P.**, 2006, WikiRelate! computing semantic relatedness using wikipedia, 21st National Conference on Artificial intelligence, A. Cohn, Ed. AAAI Conference On Artificial Intelligence, AAAI Press, 2:1419-1424pp.
- Syeda-Mahmood, T., Shah, G., Akkiraju, R., Ivan, A. A. and Goodwin, R.**, 2005, Searching Service Repositories by Combining Semantic and Ontological Matching, *IEEE International Conference on Web Services (ICWS'05)*, 13-20pp.

KAYNAKLAR DİZİNİ (devam)

United States National Library of Medicine, “Unified Medical Language System[®] (UMLS[®])”, <http://www.nlm.nih.gov/research/umls/> (Erişim tarihi: 18 Haziran 2010).

Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E. and Ciravegna, F., 2006, Semantic Annotation for Knowledge Management: Requirements and a Survey of the State of the Art, Journal of Web Semantics: Science, Services and Agents on the World Wide Web, 4:14–28pp.

Uschold, M. and Gruninger, M., 2004, Ontologies and semantics for seamless connectivity, SIGMOD Rec. 33:58-64pp.

Tanskanen, S., 2006, Collaborating Towards Coherence: Lexical Cohesion in English Discourse, Amsterdam: John Benjamins.

Tous, R. and Delgado, J., 2006, A vector space model for semantic similarity calculation and OWL ontology alignment, Database and Expert Systems Applications, 17th International Conference, DEXA 2006, Kraków, Poland.

Tuğlacı, P., 2001, Türkçe - İngilizce Bilimsel ve Teknik Terimler Sözlüğü, İnkılâp Kitabevi, İstanbul.

Turner, A. A., 1987, The propositional analysis system (Tech. Rep. No. 87-2), Boulder: University of Colorado, Institute of Cognitive Science.

Tversky, A., 1977, Features of similarity, in Psychological Review, 84(2):327-352pp.

van den Broek, P., Virtue, S., Gaddy, M., Tzeng, Y. and Sung, Y.-C., 2002, Comprehension of memory of science texts: inferential processes and the construction of a mental representation, In: Otero, J., Leon, J., Graesser, A.C. (Eds.), The Psychology of Science Text Comprehension, Lawrence Erlbaum Associates, Mahwah, NJ, 131-154pp.

KAYNAKLAR DİZİNİ (devam)

- van Dijk, T.A. and Kintsch, W.**, 1983, Strategies of Discourse Comprehension, Academic Press, New York.
- Verhoeven, L., and Perfetti, C. A.**, 2008, Introduction, Advances in text comprehension: Model, process and development, Applied Cognitive Psychology, 22(3):293-301pp.
- Wang, J. Z., Du, Z., Payattakool, R., Yu, P. S. and Chen, C.**, 2007, A new method to measure the semantic similarity of GO terms, Bioinformatics 23:1274-1281pp.
- Wang, Y.**, 2007, The OAR Model of Neural Informatics for Internal Knowledge Representation in the Brain, The International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), IPG Publishing, USA, 1(3):64-75pp.
- Waugh, N. C. and Norman, D. A.**, 1965, Primary Memory, Psychological Review 72(2):89-104pp.
- Wolf, F. and Gibson, E.**, 2004, Representing discourse coherence: A corpus-based analysis, Proceedings of the 20th International Conference on Computational Linguistics (COLING), Geneva, Switzerland.
- Word Add-in For Ontology Recognition**, <http://ucsdbiolit.codeplex.com/> (Erişim tarihi: 5 Şubat 2011).
- Wu, C., Potdar, V. and Chang, E.**, 2008, Latent Semantic Analysis – The Dynamics of Semantics Web Services Discovery, 346-373pp., Advances in Web Semantics I, Dillon, T. S., Chang, E., Meersman, R. and Sycara K. (Eds.), Lecture Notes In Computer Science, 4891, Springer-Verlag, Berlin, Heidelberg, 396p.
- Wu, Z. and Palmer, M.**, 1994, Verb semantics and lexical selection, 32nd Annual Meeting of the Association for Computational Linguistics, 133-138pp.

KAYNAKLAR DİZİNİ (devam)

Zhang, X., Jing, L., Hu, X., Ng, M., Jiangxi, J. X. and Zhou, X., 2008, Medical document clustering using ontology-based term similarity measures, International Journal of Data Warehousing and Mining, 4(1):62-73pp.

ÖZGEÇMİŞ

Adı Soyadı : Görkem GİRAY

Doğum Tarihi : 13.08.1976

Doğum Yeri : İzmir

Uyruğu : T.C.

Lise : Bornova Anadolu Lisesi, 1994

Lisans : Ege Üniversitesi, Bilgisayar Mühendisliği Bölümü, 1999

Yüksek Lisans : Koç Üniversitesi, İşletme (MBA), 2001

Dil : İngilizce, Almanca

EKLER

- Ek 1 Türkçe – İngilizce Kavramlar Sözlüğü
- Ek 2 Web Ontology Language (OWL)
- Ek 3 OWL Öğelerinin Anlamsal Ağırlıklarını Belirlemek için Kullanılan Kavram Çiftleri
- Ek 4 Anketin Açıklama Yazısı
- Ek 5 Katılımcıların İki Kavram Arasındaki İlintililik için Verdiği Değerlerin OWL Dil Öğeleri Bazında Dağılımı
- Ek 6 Ontoloji Tabanlı Anlamsal İlintililik Yönteminin Başarısını Ölçmek için Kullanılan Kavram Çiftleri
- Ek 7 Uygulamalarda Kullanılan Araçların ve Kütüphanelerin Sürümleri ve Web Sayfaları
- Ek 8 Bir Ontolojideki İki Kavram Arasındaki Anlamsal İlintililik Değerini Hesaplamak için Kullanılan Sınıflar, Değişkenler ve Yordamlar
- Ek 9 Metin Bağdaşıklığını Hesaplamak için Kullanılan Sınıf ve Yordam
- Ek 10 Latin Kare Deseni
- Ek 11 Deneyde Kullanılan Metinler
- Ek 12 Deneyde Elde Edilen Metin Bağdaşıklık Değerleri

Ek 1. Türkçe – İngilizce Kavramlar Sözlüğü

A

Alt-kavram	Subconcept
Anlamsal çözümleme	Semantic analysis
Anlamsal uzay	Semantic space
Anlamsal web	Semantic web
Arayüz	Interface
Artalan bilgisi	Background knowledge
Atıf	Reference
Atlas	Gazetteer

B

Bağdaşıklık	Coherence
Bağlam	Context
Bağlaşıklık	Cohesion
Belge	Document
Belge sıfırlayıcı	Document reset
Belirtim	Specification
Benzerlik	Similarity
Betim mantığı	Description logic
Betimleyici	Descriptive
Bişimbirimsel çözümleme	Morphological analysis

Bıçımsel	Formal
Bildirimsel bilgi	Declarative knowledge
Bilgi erişimi	Information retrieval
Bilgi tabanı	Knowledge base
Birey	Individual
Biricik	Unique
Biricik kimlik	Unique identifier
Biricik-isim kabulü	Unique-name assumption
Birincil bellek	Primary memory
Birleşim	Union

C

Cümle ayırıcı	Sentence splitter
---------------	-------------------

Ç

Çalışma belleği	Working memory
Çalıştay	Workshop
Çıkarsama	Inference
Çizge	Graph
Çözümleme	Analysis

D

Değişmez	Literal
Derinliğine arama	Depth first search

Derlem	Corpus
Dil bilgisi	Grammar
Dil kaynağı	Language resource
Dil ögesi	Language construct
Dizi	Array
Doğrulama	Verification
Durum modeli	Situation model
Düğüm	Node

E

Eklenti	Add-in
Erişim isabeti	Recall
Eş gönderge	Coreference
Eşdeğer	Equivalent
Etiketleme	Annotation
Etkinleştirmenin yayılması teorisi	Spreading activation theory
Etkisiz kelime	Stopword
Etmen	Agent

G

Gizli Anlam Çözümleme	Latent Semantic Analysis
-----------------------	--------------------------

H

Hesaplama	Computation
-----------	-------------

Hizalama	Alignment
i	
İçlemsel bilgi	TBox
İfade	Expression, Statement
İkincil bellek	Secondary memory
İlintililik	Relatedness
İlksav	Axiom
İnsan bilişi	Human cognition
İsim öbeği	Noun phrase
İşlemsel bilgi	Procedural knowledge
İşlem kaynağı	Processing resource
K	
Kaplam	Extension
Kaplamsal bilgi	ABox
Kardeş	Sibling
Kardinalite	Cardinality
Karşıt anlamlılık	Antonymy
Kavrama	Comprehension
Kelime anlamı belirginleştirme	Word sense disambiguation
Kelime Öbeği	Phrase
Kesin isabet	Precision

Kesişim	Intersection
Kısıtlama	Restriction
Kimlik	Identifier
Konu alanı	Domain
Korelasyon	Correlation
Kökleme	Lemmatization
Kuralcı	Prescriptive

M

Menzil	Range
Metin	Text
Metin tabanı	Textbase

N

Nesne-Özellik-İlişki modeli	Object-Attribute-Relation (OAR) model
-----------------------------	---------------------------------------

O

Okunabilirlik ölçüsü	Readability measure
Ontoloji	Ontology

Ö

Önerme	Preposition
Örnek	Instance
Özellik	Property
Özellik-genellik ilişkisi	Is-a relation

Özyineli

Recursive

P

Parça-bütün ilişkisi

Meronymy

Parçalayıcı

Tokeniser

Patika

Path

POS etiketleyici

POS Tagger

S

Sınıf

Class

Sonlu durum makinesi

Finite state machine

Sorgu genişletme

Query expansion

Söylem çözümleme

Discourse analysis

Sözdizimsel çözümleme

Syntactic analysis

Sözlük

Lexicon

Ş

Şebeke

Grid

T

Tamamlayıcı

Complement

Tekrar

Rehearsal

Terim frekansı

Term frequency

Ters Belge Frekansı

Inverse Document Frequency

Tümdengelimli

Deductive

Tümevarımlı

Inductive

U

Uç değır

Outlier

Uyaran

Stimulus

Uygulama

Applicaton

Uzun süreli bellek

Long term memory

Ü

Üst model

Metamodel

Üst veri

Metadata

V

Varlık

Entity

Y

Yeniden kullanım

Reuse

Yordam

Procedure

Yönlü çizge

Directed graph

Yönlü döngüsüz çizge

Directed acyclic graph

Yüklem

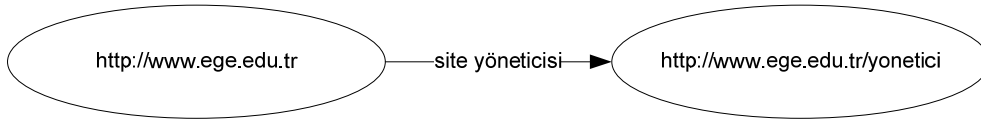
Predicate

Ek 2. Web Ontology Language (OWL)

Anlamsal web çalışma alanı, İnternet'in hayatımızdaki öneminin ve elektronik ortamdaki bilgi hacminin önlenemez artışıyla birlikte oldukça ilgi çekmeye başlamıştır. Bu çalışma alanının merkezindeki kavramlardan birisi de ontolojilerdir. Ontolojiler, bilginin anlamının biçimsel temsilini sağlayarak, şu anda tamamen insanlar tarafından yapılan, elektronik ortamdaki verinin anlamı üzerinde çıkarsama yapma işini, bir ölçüde bilgisayarlara devretmeyi hedeflemektedir.

W3C, anlamsal web vizyonu çerçevesinde, ontoloji temsil dilleri geliştirmektedir ve standartlaştırmaktadır. Resource Description Framework (RDF), RDF Schema (RDFS) ve Web Ontology Language (OWL), W3C'nin şu ana kadar ontoloji temsili için geliştirdiği ve standartlaştırdığı dillerdir.

Resource Description Framework, web'deki kaynaklar hakkında bilgi temsil etmek için kullanılan bir dildir. RDF, basit bir veri modelinden oluşur. RDF'in temel yapı ögesi "ifade"lerdir. Bir ifade, "özne", "yüklem" ve "nesne"den oluşur. Örneğin, www.ege.edu.tr web sitesinin yöneticisini belirten RDF ifadesi Şekil - 1'deki gibi olacaktır.



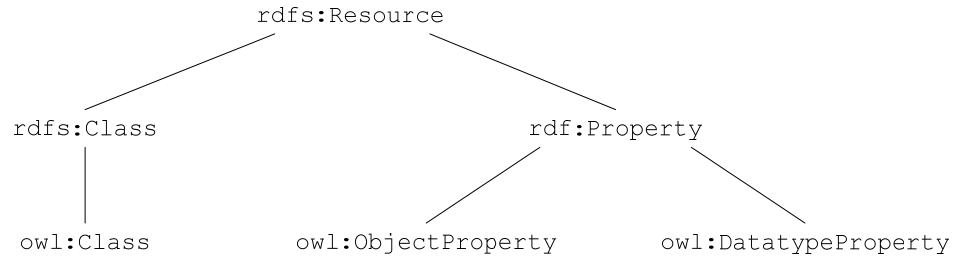
Şekil - 1: Örnek bir RDF ifadesi.

RDF'de kaynaklar Uniform Resource Identifier (URI) ile tanımlanmaktadır. Yukarıdaki ifadenin bilgisayar tarafından işlenebilmesi için XML gibi yapısal bir dille kodlanması gerekmektedir.

RDF, herhangi bir konu alanı için geliştirilmemiştir. Bir konu alanına özgü terminoloji tanımlamak için RDF Schema (RDFS) dili kullanılır. RDFS, RDF veri modelinde kullanılabilecek kelimeleri, hangi özelliklerin hangi tür nesnelere uygulanabileceğini ve nesneler arasındaki ilişkileri tanımlar.

RDF'in temsil yeteneđi, sadece özne, yüklem ve nesneden oluşan ifadelerle sınırlıdır. RDFS ise alt sınıf ve özellik hiyerarşileriyle ve özelliklerin alabileceđi değerlerin tanımıyla sınırlıdır. W3C, anlamsal web vizyonu çerçevesinde kullanılacak dillerin temsil güçlerinin RDF ve RDFS'den daha fazla olması gerektiđi sonucuna varmıştır (OWL Web Ontology Language Use Cases and Requirements, 2004). Bunun sonucunda OWL geliştirilmiştir.

OWL dili, RDF ve RDFS dilleri temel alınarak geliştirilmiştir. OWL, RDF ve RDFS dillerindeki bazı modelleme yapıları arasındaki ilişkiler Şekil - 2'de gösterilmiştir (Antoniou and van Harmelen, 2008).



Şekil - 2: OWL ve RDF/RDFS arasındaki alt sınıf ilişkileri.

OWL, W3C'nin anlamsal web vizyonu kapsamında kullanılmasını önerdiği ve teşvik ettiği ontoloji temsil dillidir.

Bilgi temsil dillerinin temsil güçleri ile bu dillerle temsil edilen ontolojilerde yapılacak çıkarsamalar için gerekli kaynak ve zaman ihtiyacı arasında ters orantı vardır. Buna göre çok güçlü temsil gücüne sahip bir dille temsil edilmiş bir ontolojide, yapılabilecek tüm çıkarsamalar kabul edilebilir bir zaman dilimi içinde yapılamayabilir. Bundan dolayı geliştirilen ontoloji temsil dillerinin temsil güçleri sınırlandırarak çıkarsama zamanları mantıklı sürelerle indirgenmektedir. Bu durumdan yola çıkılarak OWL dilinin farklı ihtiyaçlara hizmet edebilecek, farklı temsil güçlerine sahip üç sürümü kullanıma sunulmuştur.

OWL Full, tüm OWL dil öğelerini içerir. Fakat OWL Full, anlamsal temsil gücü çok yüksek olduğu için, kabul edilebilir zaman sınırları içinde tüm çıkarsamaların yapılabileceđini garanti edemez. OWL DL, OWL Full sürümündeki bazı yeteneklerin kısıtlanmasıyla elde edilmiştir ve verimli bir çıkarsama desteđi sunmaktadır. OWL Lite ise, OWL dilinin bazı öğeleri kapsam dışında bırakılarak oluşturulmuştur. OWL Lite'nın avantajları, kullanıcılar tarafından kolay öğrenilmesidir ve ontoloji araç geliştiricileri için OWL'ın diğer

sürümlerine göre daha kolay desteklenebilir olmasıdır. OWL Lite da, OWL DL gibi verimli bir çıkarsama desteği sunmaktadır. OWL Lite’ın dezavantajı ise temsil gücünün sınırlı olmasıdır (Antoniou and van Harmelen, 2008).

OWL Dil Öğeleri

OWL ontolojilerinin temel yapı taşları, “sınıf”lar, “birey”ler ve “özellik”lerdir (Horridge, 2009).

Sınıflar

OWL dilinde sınıf, bireylerden oluşan bir kümeyi temsil eder (Horridge, 2009). Bir sınıf tanımı, bir OWL sınıfını, bir sınıf ismiyle ya da isimlendirilmemiş anonim bir sınıfın bireyler kümesiyle ifade eder (Bechhofer et al., 2004).

OWL, altı farklı tür sınıf tanımına izin verir (Bechhofer et al., 2004):

1. Bir sınıf kimliği (Bir URI referansı)
2. Bir sınıfı oluşturan bireylerin listesi
3. Bir özellik kısıtlaması
4. Bir ya da daha fazla sınıf tanımının kesişimi
5. Bir ya da daha fazla sınıf tanımının birleşimi
6. Bir sınıf tanımının tamamlayıcısı

Birinci tür tanımda sınıf, bir sınıf ismi ile tanımlanır. Diğer beş tür tanımda, bireyler kümesi üzerine kısıtlar konularak anonim sınıf tanımı yapılır (Bechhofer et al., 2004).

Birinci tür sınıf tanımı için aşağıdaki örnek verilebilir:

```
<owl:Class rdf:ID="İnsan"/>
```

“İnsan” sınıfına aynı OWL belgesi içinde `rdf:resource="#İnsan"` ile, farklı OWL belgelerinden ise URI aracılığıyla (<http://www.ornekontoloji.org#İnsan>) erişilebilir.

İkinci tür sınıf tanımında, sınıfın bireyler kümesindeki bireyler listelenir. Bu liste `owl:oneOf` dil ögesi ile yapılır. Bu tür tanımda, sınıfın kapsamı tamamen belirtildiği için, bu sınıfa ait bir başka birey tanımlanamaz (OWL Web Ontology Language Guide, 2004). Örneğin, aşağıda birey kümesi “beyaz”, “gül” ve “kırmızı”dan oluşan “ŞarapRengi” isimli sınıf tanımlanmıştır.

```
<owl:Class rdf:ID="ŞarapRengi">
  <owl:oneOf rdf:parseType="Collection">
    <owl:Thing rdf:about="#Beyaz"/>
    <owl:Thing rdf:about="#Gül"/>
    <owl:Thing rdf:about="#Kırmızı"/>
  </owl:oneOf>
</owl:Class>
```

Bu tür sınıf tanımında yapılabilecek çıkarsamalardan birisi, sınıfa ait bireylerin o sınıfın birer örneği olduğudur (Allemang and Hendler, 2008). Buna göre yukarıdaki örnekten yola çıkılarak aşağıdaki üçlülerin doğru olduğu sonucuna varılabilir.

```
Beyaz rdf:type ŞarapRengi
Gül rdf:type ŞarapRengi
Kırmızı rdf:type ŞarapRengi
```

Liste türü sınıf tanımı sadece OWL DL ve OWL Full dillerinde yapılabilir (Bechhofer et al., 2004).

Üçüncü tür sınıf tanımında, bir özellik kısıtı konularak sınıf tanımı yapılır. Bu anonim sınıf, bu özellik kısıtına uyan tüm bireylerden oluşur. OWL dilinde iki tür özellik kısıtı vardır: değer kısıtı ve kardinalite kısıtı (Bechhofer et al., 2004).

Değer kısıtı, bir özelliğin alacağı değerlere kısıt uygular. Örneğin, “komşu” özelliği “bölge” tipinde bir örnek olan tüm bireyleri kapsayan bir sınıf, değer kısıtı kullanılarak tanımlanabilir. Değer kısıtı, özellik değerini her zaman kısıtlayan `rdfs:range` dil ögesinden farklıdır. Değer kısıtı, üç OWL dil ögesi kullanılarak tanımlanabilir: `owl:allValuesFrom`, `owl:someValuesFrom`, `owl:hasValue`.

Aşağıdaki örnek, “ebeveyn” özelliği, “insan” tipinde örneklerden oluşan bireylerin oluşturduğu anonim bir sınıfı tanımlamaktadır. Bu tanıma göre “ebeveyn” özelliği her zaman “insan” tipinde örneklerden oluşan bir değer almak zorunda değildir. Fakat aşağıda tanımlanan anonim sınıf, “ebeveyn” özelliği “insan” tipinde örneklerden oluşan bireylerden oluşmaktadır.

```
<owl:Restriction>
  <owl:onProperty rdf:resource="#ebeveyn" />
  <owl:allValuesFrom rdf:resource="#İnsan" />
</owl:Restriction>
```

owl:someValuesFrom dil ögesi kullanılarak, bir özelliğin, en az bir değerinin tanımlanan bir sınıf tipinde olan bireylerin oluşturduğu bir anonim sınıf tanımlanabilir (Allemang and Hendler, 2008). Buna göre aşağıdaki tanımda, “ebeveyn” özelliğinin değerinin en az bir “doktor” sınıfı tipinde bireyin olduğu, bireylerin oluşturduğu anonim sınıfı tanımlamaktadır.

```
<owl:Restriction>
  <owl:onProperty rdf:resource="#ebeveyn" />
  <owl:someValuesFrom rdf:resource="#Doktor" />
</owl:Restriction>
```

owl:allValuesFrom dil ögesi ile yapılan sınıf tanımında, bu sınıfa ait bireylerin “ebeveyn” özelliğinin bir değeri olmak zorunda değildir. Bu sınıftaki bireylerin “ebeveyn” özelliği olması durumunda ise bu özelliğin değeri mutlaka “insan” tipinde bir birey olmak zorundadır. owl:someValuesFrom dil ögesi ile yapılan sınıf tanımında ise, bu sınıftaki bireylerin, değeri “doktor” tipinde bir birey olan en az bir “ebeveyn” özelliği bulunmak zorundadır. Fakat bu bireylerin, farklı tipte birey değerleri olan “ebeveyn” özellikleri de olabilir (OWL Web Ontology Language Guide, 2004).

owl:hasValue dil ögesi, kısıtladığı özelliğin değerini bir birey ya da veri değeri olarak ifade eder (Bechhofer et al., 2004). Örneğin aşağıdaki özellik tanımı, “Türkiye”de bulunan herhangi bir nesne grubunu (futbol takımı gibi) oluşturmak için kullanılabilir.

```
<owl:Restriction>
  <owl:onProperty rdf:resource="#ülke" />
  <owl:hasValue rdf:resource="#Türkiye" />
</owl:Restriction>
```

Kardinalite kısıtları, bir özelliğin kaç farklı değer alabileceğini belirlemektedir (Allemang and Hendler, 2008). OWL dilinde kardinalite kısıtları

owl:maxCardinality, owl:minCardinality ve owl:cardinality dil ögeleri kullanılarak tanımlanır. Örneğin, iki “ebeveyn”i olan bireylerin ait olduğu sınıf aşağıdaki kısıt ile tanımlanabilir.

```
<owl:Restriction>
  <owl:onProperty rdf:resource="#ebeveyn" />
  <owl:cardinality
    rdf:datatype="&xsd;nonNegativeInteger">2</owl:cardinality>
</owl:Restriction>
```

4., 5. ve 6. tür sınıf tanımlarında, temel küme operatörleri kullanılır. Bu operatörlerin OWL dilindeki karşılığı, kesişim için owl:intersectionOf, birleşim için owl:unionOf, tamamlayıcı için owl:complementOf dil ögesidir.

Örneğin, “Şarap” tipinde olan bireylerin ve “renk” özellikleri “Beyaz” olan bireylerin kesişiminin “BeyazŞarap” sınıfını oluşturması aşağıdaki gibi ifade edilebilir.

```
<owl:Class rdf:ID="BeyazŞarap">
  <owl:intersectionOf rdf:parseType="Collection">
    <owl:Class rdf:about="#Şarap" />
    <owl:Restriction>
      <owl:onProperty rdf:resource="#renk" />
      <owl:hasValue rdf:resource="#Beyaz" />
    </owl:Restriction>
  </owl:intersectionOf>
</owl:Class>
```

“Meyve” sınıfının, “TatlıMeyve” ve “TatlıOlmayanMeyve” sınıflarının kaplamalarının birleşiminden oluştuğu aşağıdaki gibi belirtilebilir.

```
<owl:Class rdf:ID="Meyve">
  <owl:unionOf rdf:parseType="Collection">
    <owl:Class rdf:about="#TatlıMeyve" />
    <owl:Class rdf:about="#TatlıOlmayanMeyve" />
  </owl:unionOf>
</owl:Class>
```

Bir sınıfın tamamlayıcısı, o sınıfa ait olmayan tüm bireylerden oluşur (Allemang and Hendler, 2008). Örneğin, “BeyazOlmayan” sınıfı, “renk” özelliği “beyaz” olan bireyler kullanılarak aşağıdaki gibi tanımlanabilir.

```
<owl:Class rdf:ID="BeyazOlmayan">
  <owl:complementOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#renk" />
      <owl:hasValue rdf:resource="#Beyaz" />
    </owl:Restriction>
  </owl:complementOf>
</owl:Class>
```

Alt sınıflar

OWL dilinde taksonomi oluşturmak için `rdfs:subClassOf` temel dil ögesi kullanılır. Bu dil ögesi, daha özel bir sınıfı, daha genel bir sınıfa bağlar. S_2 sınıfı S_1 sınıfının alt sınıfı ise, S_2 sınıfına ait tüm bireyler aynı zamanda S_1 sınıfına da aittir (OWL Web Ontology Language Guide, 2004). Her sınıf kendisinin alt sınıfıdır (Bechhofer et al., 2004). `rdfs:subClassOf` ilişkisi geçişlidir; S_3 sınıfı S_2 sınıfının alt sınıfı ve S_2 sınıfı S_1 sınıfının alt sınıfı ise S_3 sınıfı da S_1 sınıfının alt sınıfıdır (OWL Web Ontology Language Guide, 2004). Aşağıdaki örnek, “Otobüs” sınıfının “TaşımaAracı” sınıfının alt sınıfı olduğunu göstermektedir.

```
<owl:Class rdf:ID="Otobüs">
  <rdfs:subClassOf rdf:resource="#TaşımaAracı" />
</owl:Class>
```

Bir sınıf için istenilen sayıda `rdfs:subClassOf` ilksavı yazılabilir. Örneğin, aşağıdaki ilksav, “Opera” sınıfının, bireyler kümesindeki bireylerin “metinYazarı” özellikleri en az bir değer alacak olan anonim bir OWL sınıfının alt kümesi olduğunu belirtir (Bechhofer et al., 2004).

```
<owl:Class rdf:about="#Opera">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#metinYazarı" />
      <owl:minCardinality
        rdf:datatype="&xsd;nonNegativeInteger">1
      </owl:minCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

Eşdeğer sınıflar

`owl:equivalentClass`, bir sınıf tanımını başka bir sınıf tanımına bağlar. Bu ifade, iki sınıfın aynı kaplama sahip olduğunu belirtir. Bir sınıf ilksavı birden fazla `owl:equivalentClass` ifadesi içerebilir (Bechhofer et al., 2004).

`owl:equivalentClass` ifadesi en basit olarak iki isimlendirilmiş sınıf arasındaki bireyler kümesi eşitliğini belirtir (Bechhofer et al., 2004).

```
<owl:Class rdf:about="#Otomobil">
<equivalentClass rdf:resource="#MotorluDörtTekerlekliKaraTaşıtı"/>
</owl:Class>
```

`owl:equivalentClass` dil ögesi, birinci tür sınıf tanımı ile ikinci tür sınıf tanımı ile de kullanılabilir. Örneğin, ana renkleri temsil eden “AnaRenkler” sınıfı ile “Sarı”, “Kırmızı” ve “Mavi” bireylerinden oluşan kümenin aynı olduğu aşağıdaki şekilde ifade edilebilir. `owl:equivalentClass` dil ögesi kullanılarak sınıf üyeliği için gerekli ve yeterli koşullar ifade edilir. Aşağıdaki örnekte görüldüğü gibi “AnaRenkler” sınıfı sadece listelenen üç bireyden oluşmaktadır.

```
<owl:Class rdf:ID="AnaRenkler">
  <owl:equivalentClass>
    <owl:Class>
      <owl:oneOf rdf:parseType="Collection">
        <Renk rdf:about="#Sarı"/>
        <Renk rdf:about="#Kırmızı"/>
        <Renk rdf:about="#Mavi"/>
      </owl:oneOf>
    </owl:Class>
  </owl:equivalentClass>
</owl:Class>
```

`owl:equivalentClass` dil ögesi kullanılarak, iki sınıfın kesişimi ile bir başka sınıfın birey kümelerinin aynı olduğu aşağıdaki örnekteki gibi ifade edilebilir (Bechhofer et al., 2004).

```

<owl:Class rdf:about="#İstiklalMarşı">
  <owl:equivalentClass>
    <owl:Class>
      <owl:intersectionOf rdf:parseType="Collection">
        <owl:Restriction>
          <owl:onProperty rdf:resource="#metinYazarı"/>
          <owl:hasValue rdf:resource="#MehmetAkifErsoy"/>
        </owl:Restriction>
        <owl:Restriction>
          <owl:onProperty rdf:resource="#besteci"/>
          <owl:hasValue rdf:resource="#OsmanZekiÜngör"/>
        </owl:Restriction>
      </owl:intersectionOf>
    </owl:Class>
  </owl:equivalentClass>
</owl:Class>

```

owl:equivalentClass dil ögesi kullanılarak, bir özellik kısıtlaması ile bir başka sınıfın birey kümelerinin aynı olduğu, aşağıdaki örnekteki gibi ifade edilebilir (OWL Web Ontology Language Guide, 2004).

```

<owl:Class rdf:ID="TürkYapımı">
  <owl:equivalentClass>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#üretildiğiÜlke" />
      <owl:someValuesFrom rdf:resource="#Türkiye" />
    </owl:Restriction>
  </owl:equivalentClass>
</owl:Class>

```

Yukarıdaki örnekteki tanımda, Türkiye’de üretilen şeylerin Türk yapımı olduğu belirtilmektedir.

owl:equivalentClass ve rdfs:subClassOf dil ögeleri arasındaki anlam farkı, gerekli koşul ve gerekli ve yeterli koşul arasındaki farktan ileri gelmektedir. “TürkYapımı” örneğinde owl:equivalentClass yerine rdfs:subClassOf kullanılması durumunda “TürkYapımı” sınıfına ait bireylerin Türkiye’de üretildiği sonucuna varılabilecektir. Fakat bunun tersi, Türkiye’de üretilen her şey “TürkYapımı” olmayacaktır. owl:equivalentClass dil ögesi kullanıldığında ise hem “TürkYapımı” sınıfına ait bireylerin Türkiye’de üretildiği sonucuna hem de Türkiye’de üretilen her şeyin “TürkYapımı” olduğu sonucuna varılabilmektedir. Bu bilgiler ışığında, “ S_1 subClassOf S_2 ” ve “ S_2 subClassOf S_1 ” ilksavlarının geçerli olması durumunda “ S_1 equivalentClass S_2 ” sonucuna varılabilir (OWL Web Ontology Language Guide, 2004).

`owl:equivalentClass`, iki sınıf arasındaki eşitliği ifade etmez. Sınıf eşitliği, iki sınıfın aynı kavramı karşıldığını ifade eder. Sınıf eşitliği, `owl:SameAs` dil ögesi ile ifade edilir. `owl:SameAs` dil ögesi kullanılarak sınıf eşitliği OWL Full dilinde ifade edilebilir (Bechhofer et al., 2004).

Ayrık sınıflar

Sınıfların ayrık olma durumu `owl:disjointWith` dil ögesi ile ifade edilir. Bir sınıfa ait olan bir birey, bu sınıfla ayrık olduğu belirtilen bir başka sınıfın bireyi olamaz (OWL Web Ontology Language Guide, 2004). Aşağıdaki örnek, “Erkek” ve “Kadın” sınıflarının ayrık olduğunu göstermektedir (Bechhofer et al., 2004).

```
<owl:Class rdf:about="#Erkek">
  <owl:disjointWith rdf:resource="#Kadın"/>
</owl:Class>
```

“Erkek” ve “Kadın” sınıflarına ait iki birey tanımı aşağıda gösterilmiştir.

```
<Erkek rdf:ID="Ali" />
<Kadın rdf:ID="Ayşe" />
```

“Erkek” ve “Kadın” sınıflarının ayrık olduğu bilinerek, yukarıdaki iki ilksava göre “Ali” ve “Ayşe” bireylerinin birbirinden farklı olduğu sonuna varılabilir ve aşağıdaki ifade çıkarsanabilir (Allemang and Hendler, 2008).

```
<Erkek rdf:ID="Ali">
  <owl:differentFrom rdf:resource="#Ayşe"/>
</Erkek>
```

Bireyler

Sınıfların örnekleri olan bireylerin tanımı aşağıdaki gibi yapılabilir. Bu tanıma göre “Ahmet” bireyi, “Çalışan” sınıfının bir üyesidir.

```
<Çalışan rdf:ID="Ahmet" />
```

Aynı bireyler

OWL dilinde, veritabanı sistemlerinde olduğu gibi bir bireyin her zaman aynı isme sahip olacağını ifade eden biricik-isim kabulü yoktur. Bundan dolayı

farklı kaynaklardan gelen iki bireyin aynı olduğu owl:sameAs dil ögesi ile belirtilir.

```
<Kişi rdf:ID="İstiklalMarşınınMetinYazarı">
  <owl:sameAs rdf:resource="#MehmetAkifErsay" />
</Kişi>
```

Farklı bireyler

İki bireyin birbirinden farklı olması durumu owl:differentFrom dil ögesi ile aşağıdaki gibi tanımlanır.

```
<ŞarapRengi rdf:ID="Beyaz" />

<ŞarapRengi rdf:ID="Kırmızı">
  <owl:differentFrom rdf:resource="#Beyaz"/>
</ŞarapRengi>

<ŞarapRengi rdf:ID="Gül">
  <owl:differentFrom rdf:resource="#Beyaz"/>
  <owl:differentFrom rdf:resource="#Kırmızı"/>
</ŞarapRengi>
```

İkiden fazla bireyin birbirinden farklı olduğunu tanımlamak için owl:AllDifferent dil ögesi kullanılır. Yukarıda, üç cümlecik ile yapılan tanımın tek cümlecik ile ifadesi aşağıdaki gibidir.

```
<owl:AllDifferent>
  <owl:distinctMembers rdf:parseType="Collection">
    <ŞarapRengi rdf:about="#Beyaz" />
    <ŞarapRengi rdf:about="#Kırmızı" />
    <ŞarapRengi rdf:about="#Gül" />
  </owl:distinctMembers>
</owl:AllDifferent>
```

Özellikler

Özellik, ikili bir ilişki olarak tanımlanabilir (OWL Web Ontology Language Guide, 2004). Özelliklerin konu alanı ve menzili bulunmaktadır. Özelliğin konu alanı, o özelliğe hangi nesnelerin sahip olabileceğini belirtir. Özelliğin menzili ise o özelliğin alabileceği değerleri belirtir (Antoniou and van Harmelen, 2008). OWL dilinde iki tür özellik bulunmaktadır (OWL Web Ontology Language Guide, 2004):

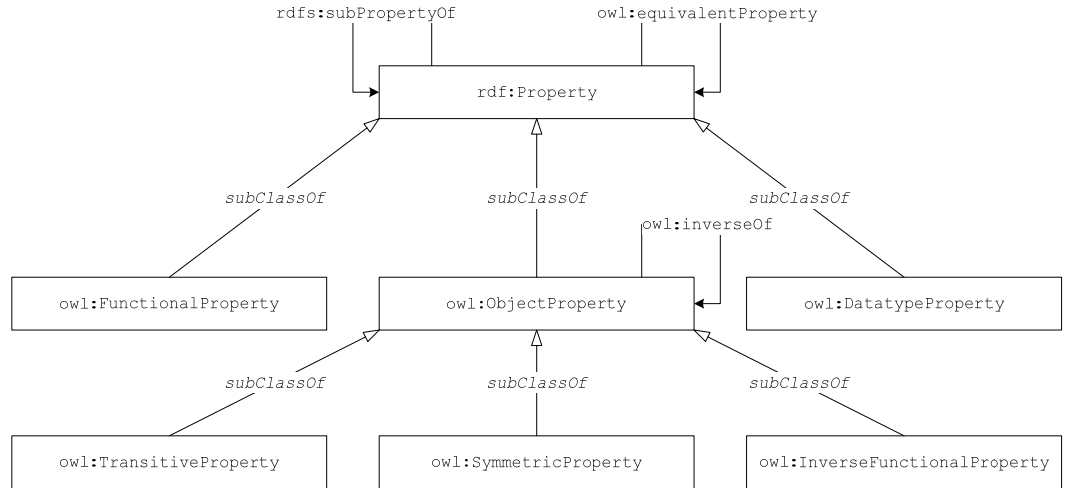
- Veri Tipi Özelliği: Sınıf örnekleri ile RDF değişmezleri ve XML Schema veri tipleri arasındaki ilişkileri ifade eder.

- Nesne Özelliği: İki sınıf örneği arasındaki ilişkileri ifade eder.

Özellik ilksavları, özelliklerin ek niteliklerini tanımlar. OWL dili özellik ilksavları için aşağıdaki dil öğelerini içerir (Bechhofer et al., 2004):

- RDF Schema dil öğeleri: `rdfs:subPropertyOf`, `rdfs:domain` ve `rdfs:range`
- Özellikler arasındaki ilişkiler: `owl:equivalentProperty` ve `owl:inverseOf`
- Global kardinalite kısıtları: `owl:FunctionalProperty` ve `owl:InverseFunctionalProperty`
- Mantıksal özellik nitelikleri: `owl:SymmetricProperty` ve `owl:TransitiveProperty`

OWL dilindeki, özellikler ile ilgili dil öğeleri arasındaki ilişkiler Şekil - 3'te gösterilmiştir.



Şekil - 3: Özellikler ile ilgili dil öğeleri.

Nesne özellikleri

Nesne özellikleri, sınıf örnekleri (nesneler) arasındaki ilişkileri tanımlar. Örneğin “öğretir” özelliği aşağıdaki gibi tanımlanabilir. Bu özelliğin konu alanı “akademikÜye” tipinde bir nesne, menzili ise “ders” tipinde bir nesne olabilir. Bir

nesne özelliği için birden fazla konu alanı ve menzil tanımlanabilir. Bu durumda geçerli konu alanı ve menzil, yapılan tanımlamaların kesişimi olacaktır (Antoniou and van Harmelen, 2008).

```
<owl:ObjectProperty rdf:ID="öğretir">
  <rdfs:domain rdf:resource="#akademikÜye"/>
  <rdfs:range rdf:resource="#ders "/>
</owl:ObjectProperty>
```

Veri tipi özellikleri

Veri tipi özellikleri, nesneleri veri tiplerine bağlar. OWL dili XML Schema veri tiplerini kullanır (Antoniou and van Harmelen, 2008). Örneğin, değeri, pozitif bir tam sayı olabilecek “yaş” veri tipi özelliği aşağıdaki gibi tanımlanabilir.

```
<owl:DatatypeProperty rdf:ID="yaş">
  <rdfs:range                                rdf:resource=
    "http://www.w3.org/2001/XMLSchema#nonNegativeInteger"/>
</owl:DatatypeProperty>
```

Alt özellikler

`rdfs:subPropertyOf` dil ögesi, bir özelliğin bir başka özelliğin alt özelliği olduğunu belirtir. $\ddot{O}_1$ özelliği, $\ddot{O}_2$ özelliğinin alt özelliği ise $\ddot{O}_1$ özelliğinin kapsamı $\ddot{O}_2$ özelliğinin kapsamının alt kümesidir (Bechhofer et al., 2004). Aşağıdaki örneğe göre, “Anne” özelliğinin kapsamındaki örnekler aynı zamanda “Ebeveyn” özelliğinin kapsamındadır.

```
<owl:ObjectProperty rdf:ID="Anne">
  <rdfs:subPropertyOf rdf:resource="#Ebeveyn"/>
</owl:ObjectProperty>
```

Örneğin, $\ddot{O}_1$ özelliği, $\ddot{O}_2$ özelliğinin alt özelliği ise ve B_1 bireyinin $\ddot{O}_1$ özelliğinin değeri B_2 bireyi ise B_1 bireyinin $\ddot{O}_2$ özelliğinin değerinin B_2 bireyi olduğu sonucuna varılabilir. Bu durum aşağıda gösterilmiştir (Allemang and Hendler, 2008).

```
 $\ddot{O}_1$  rdfs:subPropertyOf  $\ddot{O}_2$ 
 $B_1$   $\ddot{O}_1$   $B_2$ 
 $B_1$   $\ddot{O}_2$   $B_2$ 
```

`rdfs:subPropertyOf` dil ögesi hem veri tipi hem de nesne tipi özellikleri ile kullanılabilir. OWL DL dilinde, `rdfs:subPropertyOf` dil

ögesinin hem öznesi hem de nesnesi aynı tip özellik olmak zorundadır (Bechhofer et al., 2004).

Eşdeğer özellikler

İki özelliğin aynı kaplama sahip olması, bu özelliklerin eşdeğer olduğu anlamına gelmektedir (Bechhofer et al., 2004). Bu durum owl:equivalentProperty dil ögesi ile tanımlanır.

owl:equivalentProperty, iki özellik arasındaki eşitliği ifade etmez. Özellik eşitliği, iki özelliğin aynı kavramı karşıladığını ifade eder. Özellik eşitliği, owl:SameAs dil ögesi ile ifade edilir. owl:SameAs dil ögesi kullanılarak özellik eşitliği OWL Full dilinde ifade edilebilir (Bechhofer et al., 2004).

İşlevsel özellikler

İşlevsel bir özellik, her bir B bireyi için, bir biricik D değeri alır. Örneğin, bir B₁ bireyi için işlevsel bir özellik, birbirinden farklı hem D₁ hem de D₂ değerlerini alamaz (Bechhofer et al., 2004). Aşağıdaki örnek, “koca” özelliğinin işlevsel olduğunu belirtmektedir. Buna göre bir “kadın” en fazla bir “koca”ya sahip olabilecektir.

```
<owl:ObjectProperty rdf:ID="koca">
  <rdf:type      rdf:resource="&owl;FunctionalProperty" />
  <rdfs:domain  rdf:resource="#Kadın" />
  <rdfs:range   rdf:resource="#Erkek" />
</owl:ObjectProperty>
```

İşlevsel özellik tanımına göre, İÖ özelliği işlevsel ise ve İÖ özelliğinin değeri B₁ bireyi için hem B₂ hem de B₃ olabiliyorsa bu durumda B₂ ve B₃ bireylerinin aynı olduğu sonucuna varılabilir (Allemang and Hendler, 2008). Bu durum aşağıdaki ifadede gösterilmiştir.

$$\begin{aligned} \text{İÖ } \text{rdf:type owl:FunctionalProperty} \wedge B_1 \text{ İÖ } B_2 \wedge B_1 \text{ İÖ } B_3 \\ \rightarrow B_2 \text{ owl:sameAs } B_3 \end{aligned}$$

Ters işlevsel özellikler

Bir özellik, ters işlevsel olarak tanımlandığında, bu özelliğin ifadesindeki nesne, özneyi biricik olarak belirlemektedir (Bechhofer et al., 2004). Aşağıdaki

örnek, “biyolojikAnne” özelliğinin ters işlevsel olduğunu belirtmektedir. Buna göre, “biyolojikAnne” özelliğinin nesnesi, yani “insan”, özneyi, yani “kadın”, biricik olarak ifade edecektir (Bechhofer et al., 2004).

```
<owl:InverseFunctionalProperty rdf:ID="biyolojikAnne">
  <rdfs:domain rdf:resource="#Kadın"/>
  <rdfs:range rdf:resource="#İnsan"/>
</owl:InverseFunctionalProperty>
```

Ters işlevsel özellik (TİÖ) tanımına göre, bir özellik ters işlevsel ise ve bu özelliğin öznesi, B_1 bireyinin nesne olduğu bir ifade için hem B_2 hem de B_3 olabiliyorsa bu durumda B_2 ve B_3 bireylerinin aynı olduğu sonucuna varılabilir (Allemang and Hendler, 2008). Bu durum aşağıdaki ifadede gösterilmiştir.

$$\begin{aligned} \text{TİÖ } \text{rdf:type owl:InverseFunctionalProperty} \wedge B_2 \text{ TİÖ } B_1 \wedge B_3 \text{ TİÖ } B_1 \\ \rightarrow B_2 \text{ owl:sameAs } B_3 \end{aligned}$$

Ters işlevsel özellik, veritabanlarındaki anahtar sahaya benzemektedir (Bechhofer et al., 2004). Bir ters işlevsel özelliğin menzilineki değerler, bu özelliğin konu alanındaki her bir değer için birer biricik kimliktir (OWL Web Ontology Language Guide, 2004). OWL Full dilinde, veri tipi özellikleri de ters işlevsel özellik olarak tanımlanabilir. Böylece, örneğin bir metin sahası bir anahtar saha olarak kullanılabilir. OWL DL dilinde ise veri tipi özellikleri ters işlevsel özellik olarak tanımlanamaz; sadece nesne özellikleri ters işlevsel özellik olarak tanımlanabilir (OWL Web Ontology Language Guide, 2004).

Bir özellik aynı anda hem işlevsel hem de ters işlevsel olabilir. Bir özellik hem işlevsel hem de ters işlevsel ise bu özellik bire-bir özelliktir. Bu özelliğin konu alanında bulunan her bir birey için bir değer vardır ve aynı zamanda bunun tersi de doğrudur (Allemang and Hendler, 2008). Örneğin, bir vatandaşın bir kimlik numarası olduğu ve bir kimlik numarasının sadece bir vatandaşa ait olduğu durumu aşağıdaki ifadelerle gösterilebilir.

```
kimlikNo rdfs:domain Vatandaş
kimlikNo rdfs:range xsd:String
kimlikNo rdf:type owl:FunctionalProperty
kimlikNo rdf:type owl:InverseFunctionalProperty
```

Geçişli özellikler

Matematikte, $F(a, b)$ ve $F(b, c)$, $F(a, c)$ 'yi gerektiriyorsa bu fonksiyon geçişlidir (Allemang and Hendler, 2008). Buna göre, GÖ geçişli bir özellikse, (B_1 ,

B₂) ve (B₂, B₃) çiftleri bu özelliğin örnekleri ise (B₁, B₃) çiftinin de bu özelliğin bir örneği olduğu sonucuna varılabilir (Bechhofer et al., 2004). Geçişli özelliklerin tipik örneklerinden birisi parça-bütün ilişkileridir (Bechhofer et al., 2004). Örneğin, “altBölge” özelliği aşağıdaki gibi geçişli bir özellik olarak tanımlanabilir.

```
<owl:TransitiveProperty rdf:ID="altBölge">
  <rdfs:domain rdf:resource="#Bölge"/>
  <rdfs:range rdf:resource="#Bölge"/>
</owl:TransitiveProperty>
```

Yukarıdaki özellik tanımına göre “Avrasya”, “Türkiye” ve “Ege” birer bölge ise ve “Ege”, “Türkiye”nin, “Türkiye” de “Avrasya”nın alt bölgesi ise bu durumda “Ege”nin de “Avrasya”nın alt bölgesi olduğu sonucuna varılabilir.

```
<Bölge rdf:ID="Ege">
  <altBölge rdf:resource="#Türkiye" />
</Bölge>

<Bölge rdf:ID="Türkiye">
  <altBölge rdf:resource="#Avrasya" />
</Bölge>
```

Yukarıdaki iki ifadeden aşağıdaki ifade çıkarılır.

```
<Bölge rdf:ID="Ege">
  <altBölge rdf:resource="#Avrasya" />
</Bölge>
```

Simetrik özellikler

SÖ simetrik bir özellikse, (B₁, B₂) çifti bu özelliğin örneği ise (B₂, B₁) çiftinin de bu özelliğin bir örneği olduğu sonucuna varılabilir (Bechhofer et al., 2004). Örneğin, “arkadaş” özelliğinin simetrik bir özellik olduğu aşağıdaki gibi tanımlanabilir.

```
<owl:SymmetricProperty rdf:ID="arkadaş">
  <rdfs:domain rdf:resource="#İnsan"/>
  <rdfs:range rdf:resource="#İnsan"/>
</owl:SymmetricProperty>
```

Yukarıdaki tanımda da görüldüğü gibi, simetrik özelliklerin konu alanları ve menzilleri aynıdır (Bechhofer et al., 2004).

Bir özellik simetrik ise, bu özelliğin tersi de yine kendisidir sonucuna varılabilir (Allemang and Hendler, 2008). Bu durum aşağıdaki gibi gösterilebilir.

SÖ rdf: type owl: SymmetricProperty → SÖ owl: inverseOf SÖ

Ters özellikler

Özelliklerin, konu alanlarından menzillerine doğru bir yönü vardır (Bechhofer et al., 2004). Bir özelliğin tersi, o özelliğin yönünü değiştirir (Allemang and Hendler, 2008). Örneğin “Çocuk” özelliği, “Ebeveyn” özelliğinin tersi olarak aşağıdaki gibi tanımlanır.

```
<owl:ObjectProperty rdf:ID="Çocuk">  
  <owl:inverseOf rdf:resource="#Ebeveyn"/>  
</owl:ObjectProperty>
```

Ö₂ özelliği Ö₁ özelliğinin tersi ise ve B₁ bireyinin Ö₁ özelliğinin değeri B₂ ise, owl:inverseOf dil ögesinin anlamı gereği B₂ bireyinin Ö₂ özelliğinin değeri B₁ olacaktır.

$$\text{Ö}_1 \text{ owl:inverseOf } \text{Ö}_2 \wedge B_1 \text{ Ö}_1 B_2 \rightarrow B_2 \text{ Ö}_2 B_1$$

Ek 3. OWL Öğelerinin Anlamsal Ağırlıklarını Belirlemek için Kullanılan Kavram Çiftleri

Türkçe

#	Kavram 1	Kavram 2
1	Eksiklik	Fazlalık
2	İçsel	Dışsal
3	Homojen	Heterojen
4	Negatif	Pozitif
5	Yapay	Doğal
6	Eğri	Paralel
7	Yoğun	Akışkan
8	Deniz suyu	Tuzlu su
9	Asimetri	Simetri
10	Isıtma	Soğutma
11	Aşınma	Erozyon
12	Karaya bağlı buz	Yüzer buz
13	Baz	Asit
14	Askeri	Sivil
15	İniş	Kalkış
16	Azalış	Düşüş
17	Çıkarma	Fark
18	Kategori	Sınıflandırma
19	Zaman aralığı	Süre
20	H ₂ O	Su
21	Asimetri	Asimetrik
22	Buz parçacığı	Buz kristali
23	Tsunami	Deprem sonrası oluşan su dalgası
24	Dozaj	Doz
25	Sürdürülebilir	Sürdürülebilirlik
26	Sonra	Önce
27	Kuzey	Güney
28	Doğu	Batı
29	Aşağıda	Yukarıda
30	Daha büyük (>)	Küçük veya eşit (≤)
31	Tatlı	Kuru
32	Yönetici	Asistan
33	Erkek çocuk	Kız çocuk
34	Belge	Proje
35	Proje	Kişi
36	Kişi	Kurum
37	Alıcı	Satıcı
38	Vize	Pasaport
39	Bilet	Pasaport
40	Çocuk	Emekli
41	Spor	Dinlenme
42	Et	Makarna
43	Meyve	Tatlı
44	Meyve	Deniz ürünü
45	Deniz ürünü	Tavuk eti
46	Yapan	Yaratan
47	Yayınlar	Belge
48	Yer almak (located in)	Bölge
49	Meyveden yapılmış (made from fruit)	Meyve

#	Kavram 1	Kavram 2
50	Anneye sahip olmak (has mother)	Ebeveyne sahip olmak (has parent)
51	Erkek kardeşe sahip olmak (has brother)	Kardeşe sahip olmak (has sibling)
52	Çocuğa sahip olmak (has child)	Kız çocuğa sahip olmak (has daughter)
53	Etnik grup	İnsan grubu
54	Meyveden yapılmış (made from fruit)	Üzümnden yapılmış (made from grape)
55	Kapasite	Pil
56	Posta kodu	Adres
57	Para birimi	Fiyat
58	Gün	Tarih
59	Tarih	Yıl
60	Şehir	Adres
61	Saat	Zaman
62	Koltuk numarası	Koltuk
63	Dakika	Zaman
64	Kişi	Babaya sahip olmak (has father)
65	Grup	Üye
66	Belge	Başlık
67	Kişi	Soyad
68	Kişi	Tanımak
69	Hayvan	Habitata sahip olmak (has habitat)
70	Bilet	Variş tarihi
71	Turist	Geçici olarak kalmak
72	Şarap	Üzümnden yapılmış (made from grape)
73	Kütle	Ağırlık
74	Etkiye sahip olmak	Sebep olmak
75	Proton sayısı	Atom numarası
76	Başarılı olmak	Araştırma
77	Yön	-e karşı (opposite to)
78	Organizma	Öldürmek
79	Kimya	Bilim
80	Kinetik enerji	Enerji
81	Sosyoloji	Davranış bilimi
82	Oran	Bölme işlemi
83	Fizik	Bilim
84	Düzlem	Yüzey
85	Yerel saat	Zaman dilimi
86	Tornado	Fırtına
87	Güneşli	Gökyüzü durumu
88	Bulutlu	Gökyüzü durumu
89	İnek	Hayvan
90	Hayvan	Organizma
91	Balık	Deniz hayvanı
92	İnsan	Memeli
93	Yangın	Afet
94	Yetiştiricilik	Tarım
95	Sinyal	İletişim
96	Kablosuz	İletişim yöntemi
97	Mobil telefon	Telefon
98	Pil	Enerji deposu
99	Enerji	Elektrik
100	Taşıma aracı	Gemi
101	Hava taşıtı	Jet
102	Kişi	Müşteri
103	Personel	Yönetici
104	Şehir	Başkent
105	Macera	Safari
106	Tatlı meyve	Üzüm

#	Kavram 1	Kavram 2
107	Et	Kırmızı et
108	Politikacı	Milletvekili
109	Öğretim Kurumu	Üniversite
110	Taşıma aracı	Otomobil
111	Sporcu	Basketbol oyuncusu
112	Kişi	Direktör
113	Depolama aygıtına sahip olmak (has storage)	Sabit sürücüye sahip olmak (has hard drive)
114	Modeme sahip olmak (has modem)	Bağlantı aygıtına sahip olmak (has connectivity device)
115	Maddenin halleri	Gaz
116	Kimyasal durum	Uçucu
117	Katı	Donmuş
118	Maddenin halleri	Sıvı
119	Hava taşıtı	Helikopter
120	Gevşeme	Yoga
121	Konaklama	Otel
122	Pizza	Muz
123	Pil	Kapasite
124	Zaman	Dakika
125	Babaya sahip olmak (has father)	Kişi
126	Soyad	Kişi
127	Üzümden yapılmış (made from grape)	Şarap
128	-e karşı (opposite to)	Yön
129	Öldürmek	Organizma
130	Direktör	Kişi
131	Bilim	Fizik
132	Hayvan	İnek
133	Enerji deposu	Pil
134	Elektrik	Enerji
135	Müşteri	Kişi
136	Safari	Macera
137	Üniversite	Öğretim Kurumu
138	Helikopter	Hava taşıtı
139	Yoga	Gevşeme
140	Ebeveyne sahip olmak (has parent)	Anneye sahip olmak (has mother)
141	Kız çocuğa sahip olmak (has daughter)	Çocuğa sahip olmak (has child)
142	İnsan grubu	Etnik grup
143	Sabit sürücüye sahip olmak (has hard drive)	Depolama aygıtına sahip olmak (has storage)

İngilizce

#	Concept 1	Concept 2
1	Deficit	Excess
2	Internal	External
3	Homogeneous	Heterogeneous
4	Negative	Positive
5	Artificial	Natural
6	Oblique	Parallel
7	Condensed	Fluid
8	Brine	Saline Water
9	Asymmetry	Symmetry
10	Cooling	Heating
11	Weathering	Erosion
12	Fast ice	Drift ice
13	Base	Acid
14	Military	Civil
15	Landing	Takeoff
16	Decrease	Fall
17	Subtraction	Difference
18	Category	Classification
19	Time interval	Duration
20	H ₂ O	Water
21	Asymmetry	Asymmetric
22	Ice particle	Ice crystal
23	Tsunami	Tidal wave
24	Dosage	Dose
25	Sustainable	Sustainability
26	After	Before
27	North	South
28	East	West
29	Below	Above
30	greater than	less than or equal
31	Sweet	Dry
32	Manager	Assistant
33	Son	Daughter
34	Document	Project
35	Project	Person
36	Person	Organization
37	Buyer	Seller
38	Visa	Passport
39	Ticket	Passport
40	Child	Retiree
41	Sport	Relaxation
42	Meat	Pasta
43	Fruit	Dessert
44	Fruit	Seafood
45	Seafood	Fowl
46	Maker	Creator
47	Publications	Document
48	located in	Region
49	made from fruit	Fruit
50	has mother	has parent
51	has brother	has sibling
52	has child	has daughter
53	Ethnic group	Group of people
54	made from fruit	made from grape

#	Concept 1	Concept 2
55	Capacity	Battery
56	Postcode	Address
57	Currency	Price
58	Day	Date
59	Date	Year
60	City	Address
61	Hour	Time
62	Seat number	Seat
63	Minute	Time
64	Person	has father
65	Group	Member
66	Document	Topic
67	Person	Last name
68	Person	know
69	Animal	has habitat
70	Ticket	arrival date
71	Tourist	temporarily living in
72	Wine	made from grape
73	Mass	Weight
74	has effect	cause
75	has protons	Atomic number
76	success in	Research
77	Direction	opposite to
78	Organism	kills
79	Chemistry	Science
80	Kinetic energy	Energy
81	Sociology	Behavioral science
82	Ratio	Division
83	Physics	Science
84	Plain	Surface
85	Local time	Time zone
86	Tornado	Storm
87	Sunny	Sky condition
88	Cloudy	Sky condition
89	Cow	Animal
90	Animal	Organism
91	Fish	Marine animal
92	Human	Mammal
93	Fire	Disaster
94	Cultivation	Agriculture
95	Signal	Communication
96	Wireless	Communication method
97	Mobile phone	Phone
98	Battery	Energy storage
99	Energy	Electricity
100	Vehicle	Ship
101	Aircraft	Jet
102	Person	Customer
103	Personnel	Manager
104	City	Capital
105	Adventure	Safari
106	Sweet fruit	Grape
107	Meat	Red meat
108	Politician	Senator
109	Educational institution	University
110	Mean of transportation	Automobile
111	Athlete	Basketball player

#	Concept 1	Concept 2
112	Person	Director
113	has storage	has hard drive
114	has modem	has connectivity device
115	State of matter	Gas
116	Chemical state	Volatile
117	Solid	Frozen
118	State of matter	Liquid
119	Aircraft	Helicopter
120	Relaxation	Yoga
121	Accommodation	Hotel
122	Pizza	Banana
123	Battery	Capacity
124	Time	Minute
125	has father	Person
126	Last name	Person
127	made from grape	Wine
128	opposite to	Direction
129	kills	Organism
130	Director	Person
131	Science	Physics
132	Animal	Cow
133	Energy storage	Battery
134	Electricity	Energy
135	Customer	Person
136	Safari	Adventure
137	University	Educational institution
138	Helicopter	Aircraft
139	Yoga	Relaxation
140	has parent	has mother
141	has daughter	has child
142	Group of people	Ethnic group
143	has hard drive	has storage

Ek 4. Anketin Açıklama Yazısı

Türkçe

Sayın Katılımcı,

Bu anket, Ege Üniversitesi Bilgisayar Mühendisliği Bölümü'nde yürütülen bir doktora tez çalışması için yapılmaktadır. Araştırmanın amacı, iki kavram arasındaki ilintililiği otomatik olarak hesaplayan bir yöntemin hem oluşturulmasına katkıda bulunmak hem de bu yöntemin güvenilirliğini sınamak için gerekli veriyi sağlamaktır. Araştırmaya katılmayı kabul ederseniz, sizden ankette yer alan kavramlar arasındaki ilintililik değerlerini belirlemeniz istenecektir.

İlintililik; ilgisi, ilişkisi, bağı olma durumunu ifade etmektedir. İlintili kavramların, birbirine benzer olanlara göre aralarında daha fazla ilişki türünü barındırdığı gözlenir. Örneğin, “araba” ile “tekerlek” kavramları parça-bütün ilişkisi ile birbiriyle ilintilidir. “Sıcak” ve “soğuk” kavramları zıtlık ilişkisi ile birbiriyle ilintilidir. Fakat bu kavramlar benzer değildir. Benzerlik, ilintililiğin özel bir durumudur. Benzerlik; nitelik, görünüş ve yapı bakımından bir başkasına benzeme veya ona eş olma durumunu ifade etmektedir. Benzer kavramlar, birbirleriyle ortak yönleri ile ilintilidir. Örneğin, “araba” ve “bisiklet” kavramları birbirine benzer ve birbiriyle ilintilidir. Bu açıklamalar doğrultusunda bu ankette kavramlar arasındaki ilintililiğin değerlendirilmesi beklenmektedir.

Bir sonraki sayfada kavram çiftleri bulunmaktadır. Sizden, her bir kavram çifti için bir ilintililik değeri belirtmenizi rica ediyoruz. **Lütfen her değeri belirlerken, ilk kavramı okuduğunuzda ikinci kavramın aklınıza gelme olasılığını ya da ikinci kavramın aklınıza ne kadar kolay geldiğini düşününüz.** Her kavram çiftini okuduktan sonra 0'dan (= iki kavram arasında hiç ilintililik bulunmaması durumu) 10'a (= iki kavramın aynı ya da özdeş olması durumu) kadar bir değerden, sizin için en uygun olanı belirtiniz. İlintililik değeri olarak 7,5 gibi ondalıklı bir değer de verebilirsiniz. Ankette, aynı kavram çiftleri birden fazla kere fakat farklı sıralarda bulunabilir. Bunun nedeni, sonuçların tutarlılığını ölçmek değil; ilk kavramın ikinciye ne kadar çağrıştırdığını ölçmek içindir.

Anketi yaparken, anlamını bilmediğiniz herhangi bir kavram ile karşılaştığınızda sözlük kullanabilirsiniz. Sözlük kullanmanıza rağmen ilintililik

değeri için bir karar veremiyorsanız o kelime çiftinin ilintilik değerini boş bırakınız.

Anketin tamamını cevapladıktan sonra, formu **gorkemgiray@yahoo.com** adresine iletmenizi rica ediyoruz. Anket formundaki değerlendirmeleriniz ve kişisel bilgileriniz üçüncü kişilerle paylaşılmayacak, tarafımızca teslim alınacaktır. Kişisel bilgileriniz, anket katılımcılarının genel profilini belirlemek ve herhangi bir sorumuz olduğunda size ulaşabilmek amacıyla istenmektedir. Araştırmayla ilgili sorularınızı yukarıdaki e-posta adresini kullanarak bize yöneltebilirsiniz. Katılımınız için tekrar teşekkür ederiz.

Adınız Soyadınız:	
E-Posta adresiniz:	
Mesleğiniz (Öğrenci iseniz hem öğrenci olduğunuzu hem de mezun olduktan sonra sahip olacağınız mesleği yazınız):	
Meslek tecrübeniz (yıl olarak):	

Görkem Giray

Ege Üniversitesi - Bilgisayar Mühendisliği Bölümü

İngilizce

Dear Participant,

This survey is done for a research carried out at Computer Engineering Department of Ege University in Turkey. The aim of this study is to provide the necessary data for the formation of an automatic method that measures the relatedness of two concepts and the test of its reliability. If you accept to participate in the survey, you will be asked to specify values of relatedness between pairs of concepts contained in the survey.

Relatedness means the state of being related or associated. Related concepts preserve more types of relationships than similar concepts. For instance, the concepts "car" and "wheel" are related through part-whole relationship. The concepts "hot" and "cold" are related through antonym relationship. However, these concepts are not similar. Similarity represents a special case of relatedness. Similar concepts are related to each other through their common characteristics. In line with these explanations, you are expected to rate the relatedness for each concept pair in this survey.

There are pairs of concepts on the next page. You will be prompted to specify a value for relatedness of each pair of concepts. **While you are specifying a relatedness value, please think of the possibility of how easy the second concept comes to your mind if you read or hear the first concept.** After reading each pair of concept, please specify your relatedness value from 0 (zero means that two concepts are not related at all) to 10 (ten means that two concepts are the same or identical). You can specify a decimal value such as 7.5. The same concept pairs may be found more than once in the survey. The reason for this is not to check consistency of answers; the reason is to measure how much the first concept connotes the second one.

When you encounter a concept whose meaning you do not know please use a dictionary. Despite you use a dictionary, you cannot decide on a relatedness value, you can left that pair's value blank.

Once you have answered all of the survey, we ask you to forward the form to **gorkemgiray@yahoo.com** e-mail address. The survey form and your personal information will be received by us and will not be shared with any third party.

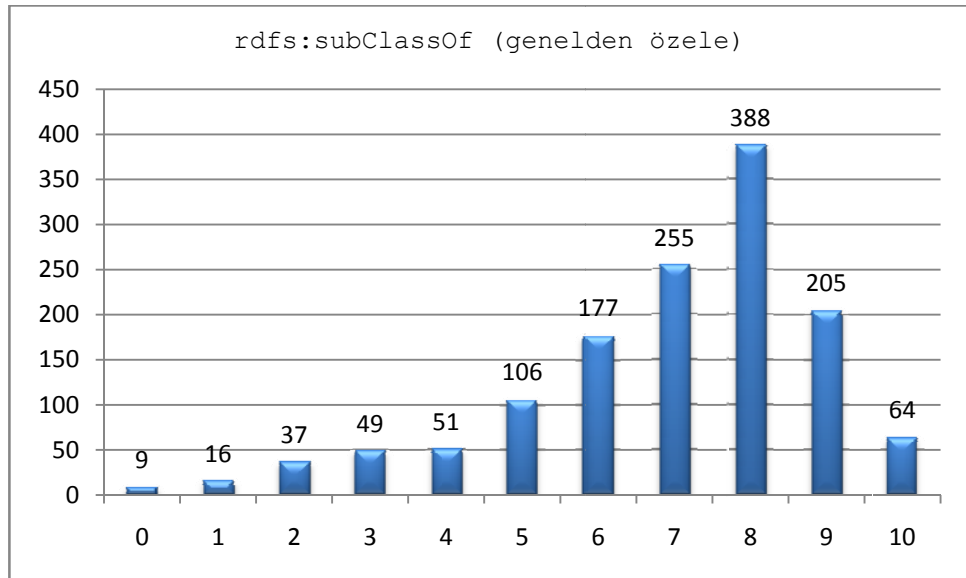
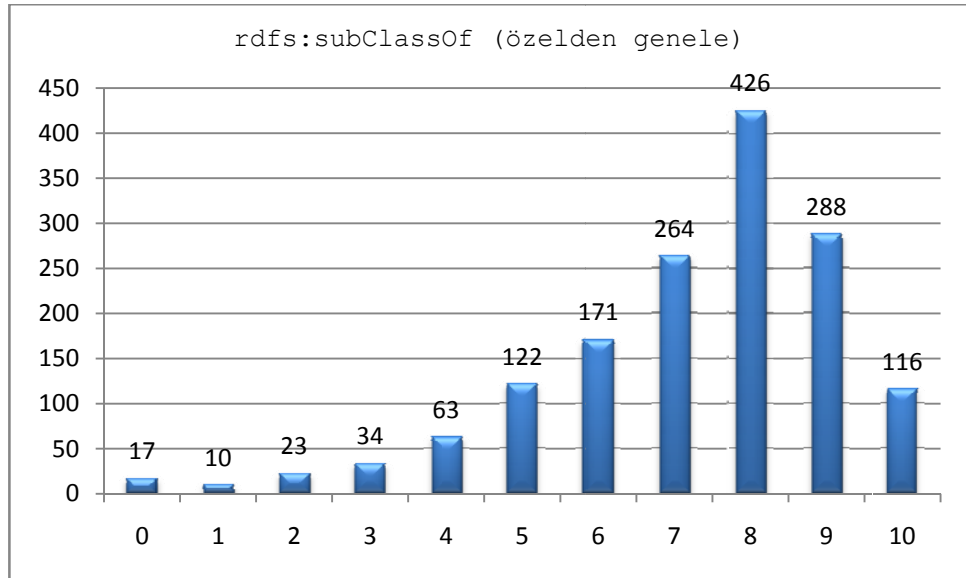
Your personal information will only be used to contact you if we have any question regarding to your answers and your profile information will be used to analyze general profile of the respondents. You can ask your questions regarding this research to us by using the e-mail address above. Thank you for your participation.

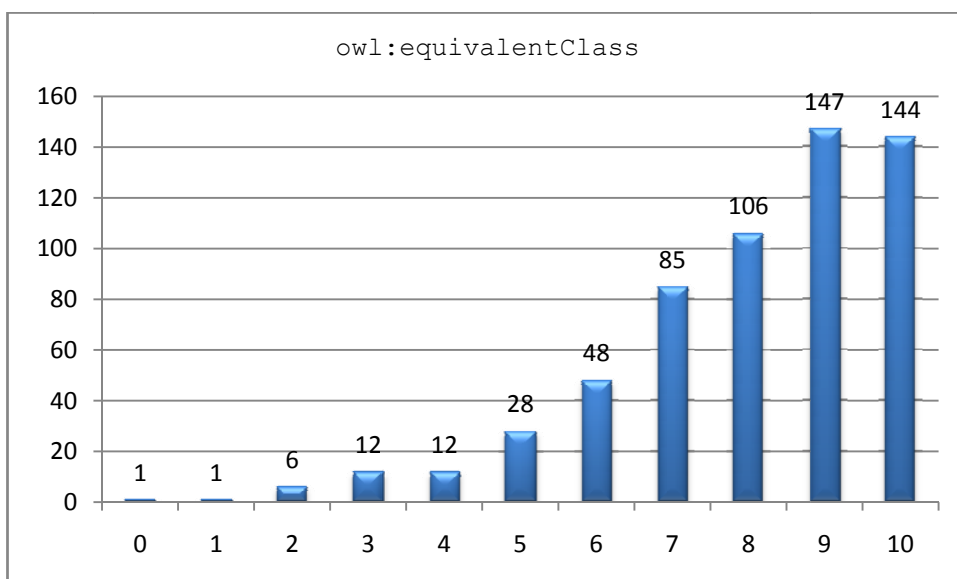
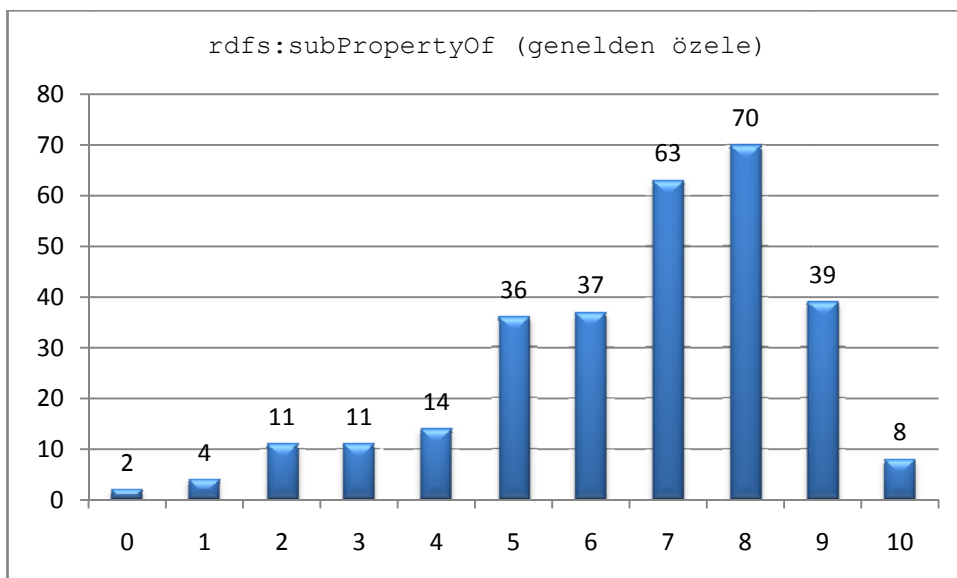
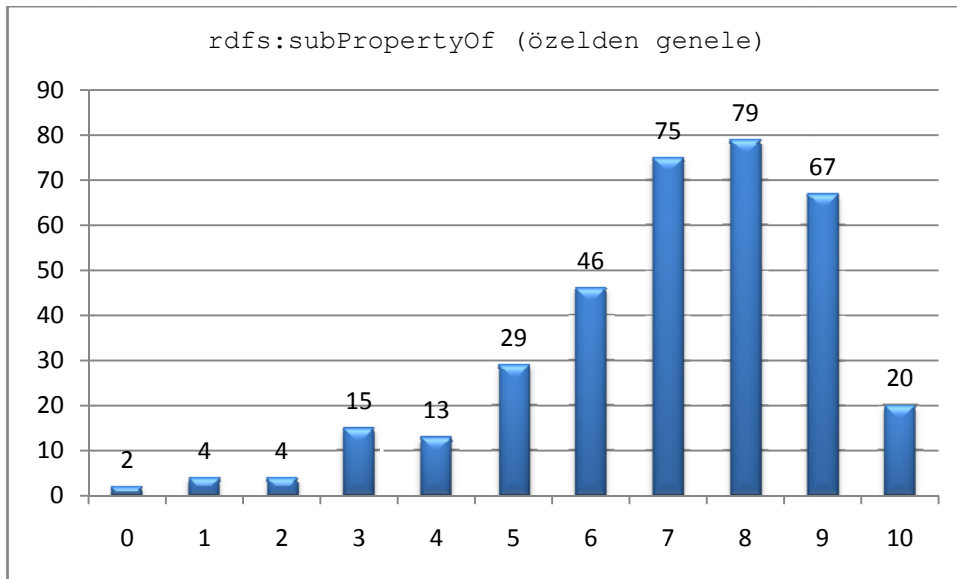
Name Surname:	
E-Mail Address:	
Profession (If you are a student please state that you are a student and state your future profession after graduation):	
Professional Experience (in years):	

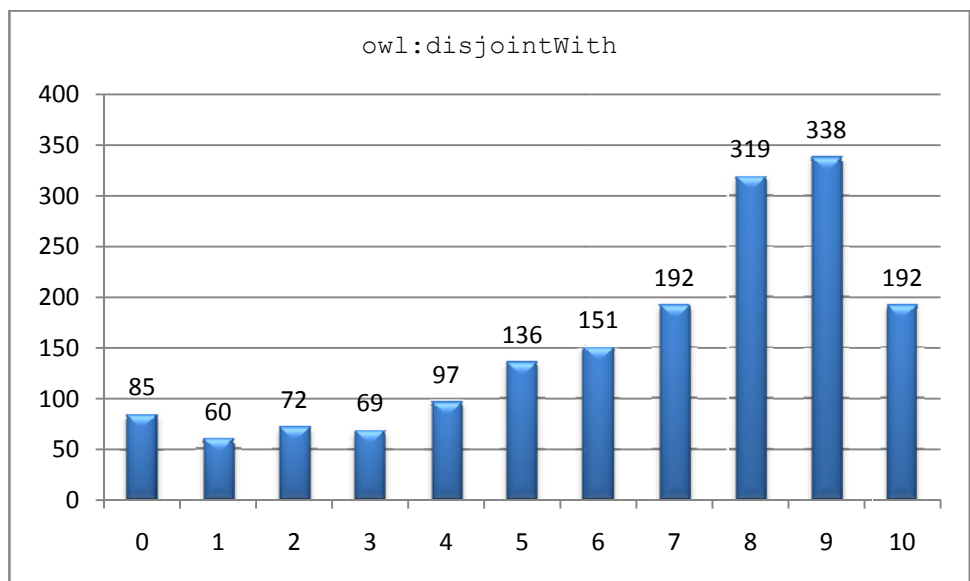
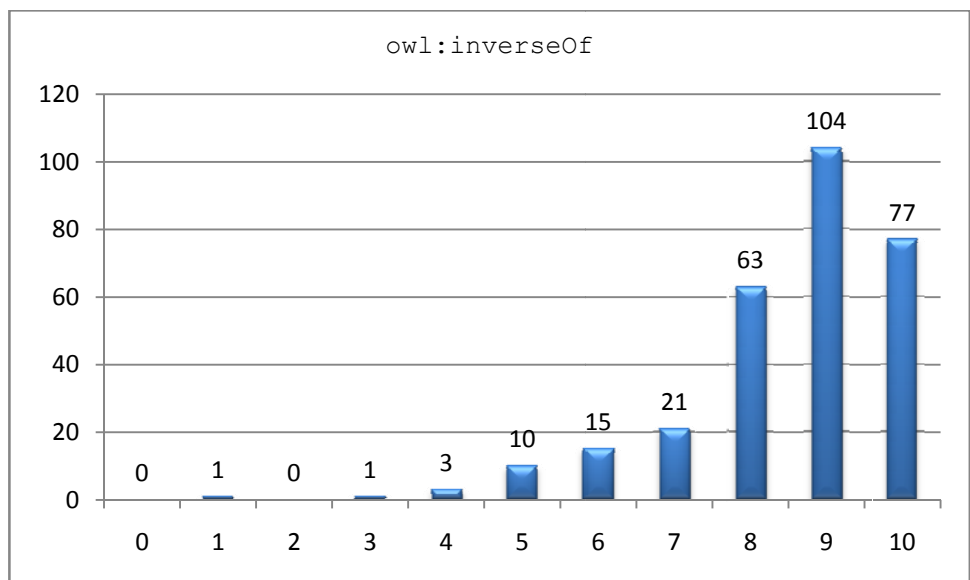
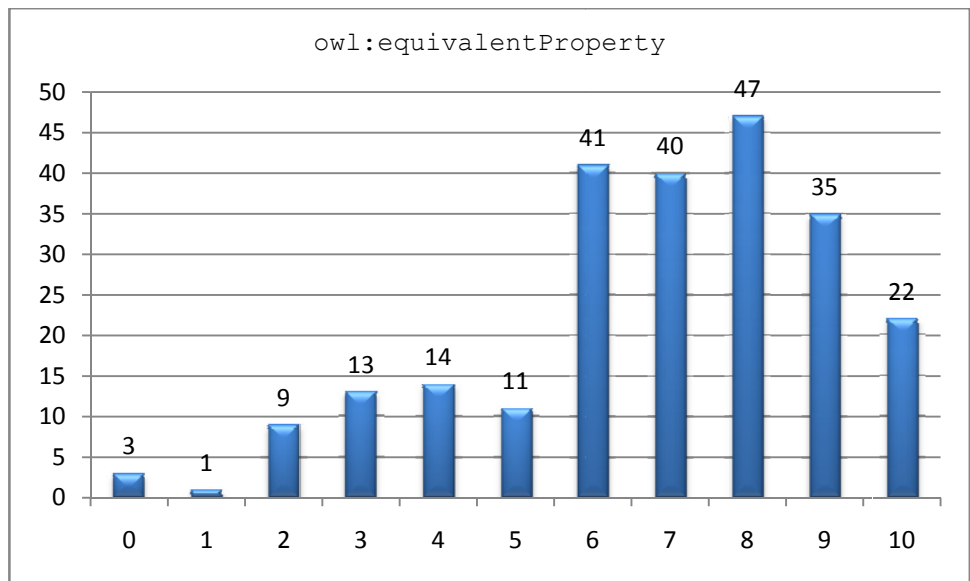
Görkem Giray

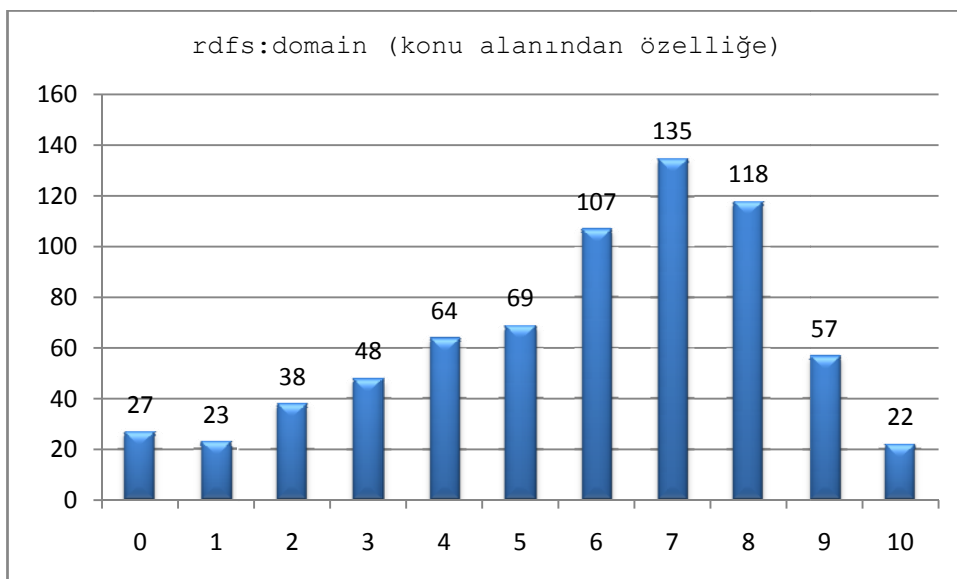
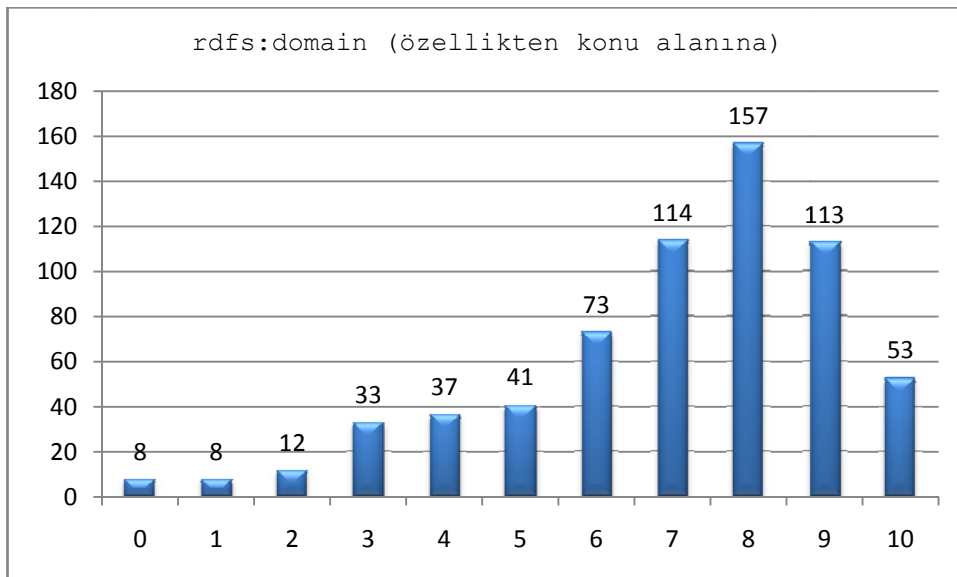
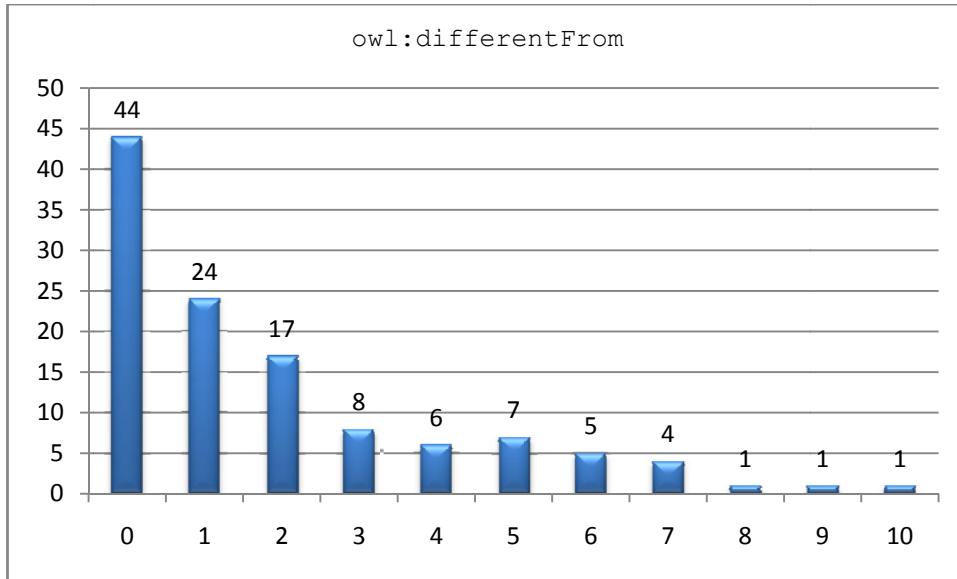
Ege University - Computer Engineering Department

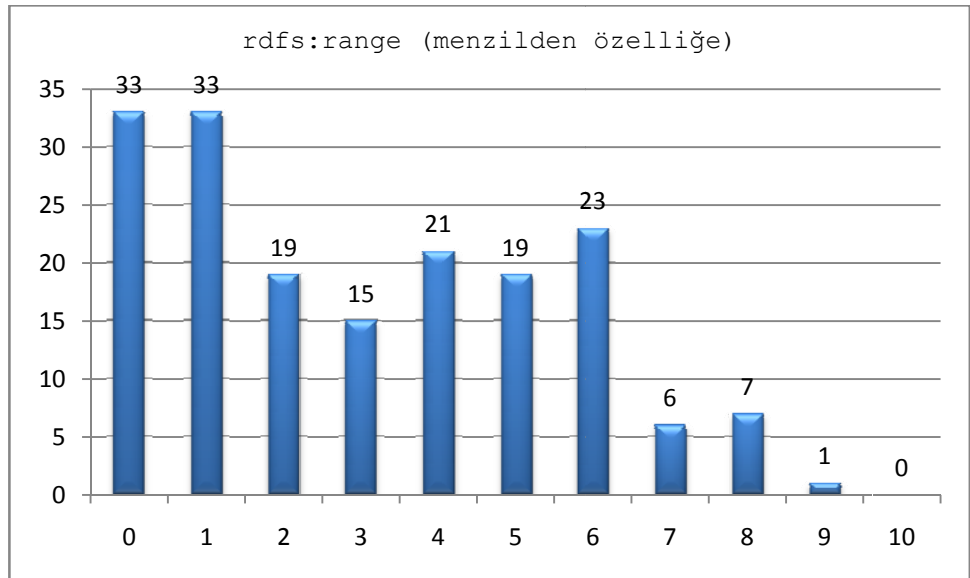
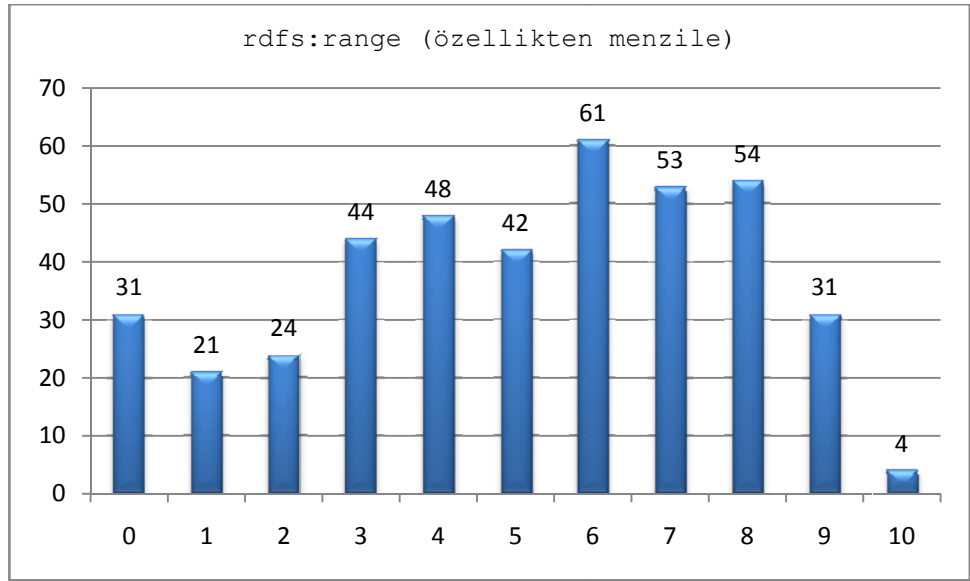
Ek 5. Katılımcıların İki Kavram Arasındaki İlintililik için Verdiği Değerlerin OWL Dil Öğeleri Bazında Dağılımı











Ek 6. Ontoloji Tabanlı Anlamsal İlintililik Yönteminin Başarısını Ölçmek için Kullanılan Kavram Çiftleri

Kavram 1	Kavram 2	Katılımcı	Otomatik
Para	Nakit	0,82777	0,69384
Para	Para birimi	0,76269	0,69384
Araba	Otomobil	0,91666	0,80228
Cadde	Bulvar	0,72142	0,50060
Erkek	Kadın	0,89682	0,50060
Gezegen	Yıldız	0,69920	0,50060
Odun	Orman	0,73650	0,50060
Kitap	Yazar	0,84047	0,41898
Nehir	Göl	0,68968	0,50060
Şehir	Nehir	0,17936	0,17388
Erkek	Baba	0,63492	0,69384
Ebeveyn	Kişi	0,49523	0,72149
Büyükanne	Kişi	0,37936	0,37557
Kırsal alan	Şehir	0,60000	0,65774
Yürüyüş yapma	Etkinlik	0,53174	0,72149
Yayın	Yazar	0,75793	0,41898
Fizik	Kimya	0,79047	0,50060
Öğrenci	Profesör	0,59365	0,50060
Tavuk	Muz	0,08095	0,19194
Kek	Hindi	0,08095	0,19194
Pizza	Kılıç balığı	0,09841	0,19194
Üst küme	Alt küme	0,66111	0,85254
Daha küçük (<)	Büyük veya eşit (≥)	0,72857	0,85254
Kaplan	Hayvan	0,76984	0,72149

Ek 7. Uygulamalarda Kullanılan Araçların ve Kütüphanelerin Sürümleri ve Web Sayfaları

Araç/Kütüphane	Kullanılan Sürüm	Web Sayfası
Jena	2.6.2	http://jena.sourceforge.net/ (Erişim tarihi: 07.10.2010)
Pellet	2.0.1	http://clarkparsia.com/pellet/ (Erişim tarihi: 07.10.2010)
Gate	5.2.1 build 3581	http://gate.ac.uk/ (Erişim tarihi: 07.10.2010)
SimPack	0.91	http://www.ifi.uzh.ch/ddis/simpack.html (Erişim tarihi: 20.10.2010)
Swoogle	3.1	http://swoogle.umbc.edu/index.php?option=com_swoogle_manual&manual=search_overview (Erişim tarihi: 15.10.2010)
Watson	-	http://watson.kmi.open.ac.uk/WS_and_API.html (Erişim tarihi: 15.10.2010)
Java WordNet::Similarity	beta.11.01	http://www.cogs.susx.ac.uk/users/drh21/ (Erişim tarihi: 03.02.2011) Kelimelerin farklı derlemeler için bilgi içeriği değerleri WordNet-InfoContent-3.0'dan (http://www.d.umn.edu/~tpederse/Data/WordNet-InfoContent-3.0.tar.gz (Erişim tarihi: 03.02.2011)) elde edilmiştir.

Ek 8. Bir Ontolojideki İki Kavram Arasındaki Anlamsal İlintililik Değerini Hesaplamak için Kullanılan Sınıflar, Değişkenler ve Yordamlar

```
//Sınıf tanımlamaları
class KaynakDugum
{
    String Dugum
    int Derinlik
}

class Kenar
{
    Statement Ifade
    boolean Yon
}

class Dugum
{
    Kenar Kenar
    Dugum AtaDugum
}

//Global değişkenler
List<KaynakDugum> ziyaretEdilmisDugumler
List<Kenar> ziyaretEdilmisKenarlar
List<Dugum> hedefDugumler

double AnlamsalIlintililikHesapla(String Ontoloji_URI, String
Kavram1, String Kavram2)
{
    Ontolojiyi Jena ile oku
    Kavram1'in URI'sini bul (Kavram1_URI)
    Kavram2'nin URI'sini bul (Kavram2_URI)

    PatikaBul(Kavram1_URI, Kavram2_URI, null, 0)
    return PatikaDegeriHesapla()
}

void PatikaBul(String baslangicDugum, String bitisDugum, Dugum
eklenenDugum, int derinlik)
{
    KaynakDugum geciciDugum //değişken tanımlama
    boolean ekle = true //değişken tanımlama

    ziyaretEdilmisDugumler listesindeki her KaynakDugum için
    {
        geciciDugum = listedeki siradaki KaynakDugum
        if (geciciDugum.Dugum = baslangicDugum)
        {
            if (derinlik != geciciDugum.Derinlik)
            {
                return
            }
            else
            {
                ekle = false
                break
            }
        }
    }
}
```

```
    }
}

if (ekle)
{
    ziyaretEdilmisDugumler.add(new KaynakDugum(baslangicDugum,
        derinlik))
}

Iterator<Kenar> ifadeListesi //değişken tanımlama
ifadeListesi = IfadeleriListele(baslangicDugum)

Kenar k //değişken tanımlama
boolean bayrak = false //değişken tanımlama
boolean ziyaretEdilmismi //değişken tanımlama

if (ifadeListesi != null)
{
    ifadeListesi listesindeki her Kenar için
    {
        bayrak = false
        k = ifade listesindeki bir sonraki Kenar
        ziyaretEdilmismi = false
    }
}

if (not ziyaretEdilmismi)
{
    //Başlangıç düğümü ilk defa ziyaret ediliyorsa bu düğümü
    //ziyaret edilmiş düğümler listesine ekle
    ziyaretEdilmisKenarlar.add(k);

    if (k.Yon)
    {
        if ((k.Ifade.Subject = baslangicDugum) AND
            (k.Ifade.Object = bitisDugum))
            //Başlangıç ve bitiş düğümler arasında bir ilişki var ve
            //ilişkinin yönü düz (Özneden nesneye doğru)
            {
                Dugum d = new Dugum(k, eklenenDugum)
                hedefDugumler listesine d düğümünü ekle
                bayrak = true
            }
    }
    else
    {
        if ((k.Ifade.Object = baslangicDugum) AND
            (k.Ifade.Subject = bitisDugum))
            //Başlangıç ve bitiş düğümler arasında bir ilişki var ve
            //ilişkinin yönü ters (Nesneden özneye doğru)
            {
                Dugum d = new Dugum(k, eklenenDugum)
                hedefDugumler listesine d düğümünü ekle
                bayrak = true
            }
    }
}

if (not bayrak)
    //Patika daha bulunmadıysa bir sonraki kenara geç
    {
        if (k.Yon)
```

```
    {
        PatikaBul(k.Ifade.Object, bitisDugum,
            new Dugum(k, eklenenDugum, ++derinlik)
        )
    }
    else
    {
        PatikaBul(k.Ifade.Subject, bitisDugum,
            new Dugum(k, eklenenDugum, ++derinlik)
        )
    }
}
}
```

List<Kenar> IfadeleriListele(String BaslangicDugum)

```
{
    List<Kenar> ifadeler //değişken tanımlama
    Statement ifade //değişken tanımlama
    Resource kaynak //değişken tanımlama

    gecici isimli listeye ontoloji modelindeki tüm ifadeleri al
    kaynak = baslangicDugum URI'sine sahip kaynağı ontoloji
        modelinden al

    if (kaynak = null)
    {
        return null;
    }

    gecici listesindeki her ifade için
    {
        ifade = listedeki siradaki Statement

        Ontolojik patika yaklaşımına uyan tüm ifadeleri
        ifadeler listesine ekle
        //Yorum ifadeleri listeye eklenmez
        //OWL dil öğeleri arasındaki ilişkiler listeye eklenmez
        //(topDataProperty, bottomDataProperty, topObjectProperty,
        //bottomObjectProperty gibi)
        //Konu alanındaki kavramların dil açısından tiplerini belirten
        //ifadeler listeye eklenmez ("Özne rdf:type owl:Class" gibi)
        //Her sınıfın owl:Thing dil ögesinin alt kümesi olduğunu
        //belirten ifadeler listeye eklenmez
        //owl:Nothing dil ögesinin her sınıfın alt kümesi olduğunu
        //belirten ifadeler listeye eklenmez
        //Hiyerarşik ilişkilerde her sınıfın sadece direk üst
        //sınıfları listeye eklenir. Çıkarıma ile elde edilen diğer
        //ilişkiler hiyerarşideki genelliğin hesaplanmasını
        //engellediği için listeye eklenmez
    }

    return ifadeler
}
```

double PatikaDegeriHesapla()

```
{
    Dugum geciciDugum
    double patikaDegeri = 1
    double minPatikaDeger = 0

    hedefDugumlar listesindeki her ifade için
    {
```

```
Dugum = listedeki siradaki düğüm
patikaDegeri = 1

do
{
    patikaDegeri = patikaDegeri *
        ifadenin_yükleminin_anlamsal_ilintililik_katsayısı
    geciciDugum = geciciDugum.AtaDugum()
} while (geciciDugum != null)
}

if ((minPatikaDeger < patikaDegeri) AND (patikaDegeri != 1))
{
    minPatikaDeger = patikaDegeri
}

return minPatikaDeger
}
```

Ek 9. Metin Bağdaşıklığını Hesaplamak için Kullanılan Sınıf ve Yordam

```
//Sınıf tanımlamaları
class EslestirmeSonucu
{
    String String1
    String String2
    double GG_Ilintililik
    double Resnik_Benzerlik
    double Lin_Benzerlik
    double JC_Benzerlik
    double LC_Benzerlik
    double WP_Benzerlik
}

void BagdasiklikHesapla (ArrayList<CumleDugum> arrCumleler)
{
    int cumleIndeks = 0
    int n_max = 0
    ArrayList<EslestirmeSonucu> es =
        new ArrayList<EslestirmeSonucu>()
    int esIndex = 0

    while (cumleIndeks + 1 < arrCumleler.size)
    {
        //Bu tez kapsamında geliştirilen yöntemle hesaplanan
        //ilintililik değerleri için
        double[] kelimeIlintililikGG =
            new double[arrCumleler[cumleIndeks].Kelimeler.size *
                arrCumleler[cumleIndeks + 1].Kelimeler.size]
        //Resnik'in yöntemiyle hesaplanan benzerlik değerleri için
        double[] kelimeBenzerlikResnik =
            new double[arrCumleler[cumleIndeks].Kelimeler.size *
                arrCumleler[cumleIndeks + 1].Kelimeler.size]
        //Lin'in yöntemiyle hesaplanan benzerlik değerleri için
        double[] kelimeBenzerlikLin =
            new double[arrCumleler[cumleIndeks].Kelimeler.size *
                arrCumleler[cumleIndeks + 1].Kelimeler.size]
        //Jiang ve Conrath'ın yöntemiyle hesaplanan benzerlik
        //değerleri için
        double[] kelimeBenzerlikJC =
            new double[arrCumleler[cumleIndeks].Kelimeler.size *
                arrCumleler[cumleIndeks + 1].Kelimeler.size]
        //Wu ve Palmer'ın yöntemiyle hesaplanan benzerlik değerleri
        //için
        double[] kelimeBenzerlikWP =
            new double[arrCumleler[cumleIndeks].Kelimeler.size *
                arrCumleler[cumleIndeks + 1].Kelimeler.size]
        //Leacock ve Chodorow'un yöntemiyle hesaplanan benzerlik
        //değerleri için
        double[] kelimeBenzerlikLC =
            new double[arrCumleler[cumleIndeks].Kelimeler.size *
                arrCumleler[cumleIndeks + 1].Kelimeler.size]

        n_max = arrCumleler[cumleIndeks + 1].Kelimeler.size

        for (int m=0; m<arrCumleler[cumleIndeks].Kelimeler.size; m++)
        {
```

```

for (int n=0; n<arrCumleler[cumleIndeks + 1].Kelimeler.size;
    n++)
{
    if (arrCumleler[cumleIndeks].Kelimeler[m].KelimeKok =
        arrCumleler[cumleIndeks + 1].Kelimeler[n].KelimeKok)
        //Karşılaştırılan iki kelimenin kökü aynıysa aradaki
        //ilintililik/benzerlik değeri 1'dir
        {
            kelimeIlintililikGG[m * n_max + n] = 1
            kelimeBenzerlikResnik[m * n_max + n] = 1
            kelimeBenzerlikLin[m * n_max + n] = 1
            kelimeBenzerlikJC[m * n_max + n] = 1
            kelimeBenzerlikWP[m * n_max + n] = 1
            kelimeBenzerlikLC[m * n_max + n] = 1
        }
    else if
        ((arrCumleler[cumleIndeks].Kelimeler[m].EtiketURI != "")
        AND
        (arrCumleler[cumleIndeks+1].Kelimeler[n].EtiketURI !=""))
        //İki kelime de bir ontolojik terimle etiketlenmişse
        {
            if (arrCumleler[cumleIndeks].Kelimeler[m].EtiketURI =
                arrCumleler[cumleIndeks + 1].Kelimeler[n].EtiketURI)
                //İki kelime aynı ontolojik terimle etiketlenmişse
                //aradaki ilintililik/benzerlik değeri 1'dir
                {
                    kelimeIlintililikGG[m * n_max + n] = 1
                    kelimeBenzerlikResnik[m * n_max + n] = 1
                    kelimeBenzerlikLin[m * n_max + n] = 1
                    kelimeBenzerlikJC[m * n_max + n] = 1
                    kelimeBenzerlikWP[m * n_max + n] = 1
                    kelimeBenzerlikLC[m * n_max + n] = 1
                }
            else if (arrCumleler[cumleIndeks].Kelimeler[m].
                EtiketOntolojiURI = arrCumleler[cumleIndeks + 1].
                Kelimeler[n].EtiketOntolojiURI)
                //İki kelime aynı ontolojideki iki terimle
                //etiketlenmişse
                {
                    esIndex = -1
                    for (int e = 0; e < es.size; e++)
                    {
                        if (((es[e].String1 =
                            arrCumleler[cumleIndeks].Kelimeler[m].KelimeKok)
                            AND (es[e].String2 = arrCumleler[cumleIndeks + 1].
                                Kelimeler[n].KelimeKok)) OR
                            ((es[e]. String2 =
                                arrCumleler[cumleIndeks].Kelimeler[m].KelimeKok)
                                AND (es[e]. String1 = arrCumleler[cumleIndeks + 1].
                                    Kelimeler[n].KelimeKok)))
                            //İlntililik/benzerlik değerinin bu kelime çifti
                            //için daha önce hesaplanıp hesaplanmadığını
                            //kontrol et
                            {
                                esIndex = e
                                break
                            }
                    }
                }
            if (esIndex = -1)
                //İlntililik/benzerlik değeri daha önce
                //hesaplanmamışsa

```

```

{
    kelimeIlintililikGG[m * n_max + n] =
        AnlamsalIlintililikHesapla(
            arrCumleler[cumleIndeks].
            Kelimeler[m].EtiketOntolojiURI,
            arrCumleler[cumleIndeks].Kelimeler[m].EtiketURI,
            arrCumleler[cumleIndeks + 1].Kelimeler[n].EtiketURI
        )
    kelimeBenzerlikResnik[m * n_max + n] =
        BenzerlikHesaplaResnik(
            arrCumleler[cumleIndeks].
            Kelimeler[m].EtiketOntolojiURI,
            arrCumleler[cumleIndeks].Kelimeler[m].EtiketURI,
            arrCumleler[cumleIndeks + 1].Kelimeler[n].EtiketURI
        )
    kelimeBenzerlikLin[m * n_max + n] =
        BenzerlikHesaplaLin(
            arrCumleler[cumleIndeks].
            Kelimeler[m].EtiketOntolojiURI,
            arrCumleler[cumleIndeks].Kelimeler[m].EtiketURI,
            arrCumleler[cumleIndeks + 1].Kelimeler[n].EtiketURI
        )
    kelimeBenzerlikJC[m * n_max + n] =
        BenzerlikHesaplaJC(
            arrCumleler[cumleIndeks].
            Kelimeler[m].EtiketOntolojiURI,
            arrCumleler[cumleIndeks].Kelimeler[m].EtiketURI,
            arrCumleler[cumleIndeks + 1].Kelimeler[n].EtiketURI
        )
    kelimeBenzerlikWP[m * n_max + n] =
        BenzerlikHesaplaWP(
            arrCumleler[cumleIndeks].
            Kelimeler[m].EtiketOntolojiURI,
            arrCumleler[cumleIndeks].Kelimeler[m].EtiketURI,
            arrCumleler[cumleIndeks + 1].Kelimeler[n].EtiketURI
        )
    kelimeBenzerlikLC[m * n_max + n] =
        BenzerlikHesaplaLC(
            arrCumleler[cumleIndeks].
            Kelimeler[m].EtiketOntolojiURI,
            arrCumleler[cumleIndeks].Kelimeler[m].EtiketURI,
            arrCumleler[cumleIndeks + 1].Kelimeler[n].EtiketURI
        )

    //Bulunan ilintililik/benzerlik degerlerini sakla
    es.add(new EslestirmeSonucu(
        arrCumleler[cumleIndeks].Kelimeler[m].KelimeKok,
        arrCumleler[cumleIndeks+1].Kelimeler[n].KelimeKok,
        kelimeIlintililikGG[m * n_max + n],
        kelimeBenzerlikResnik[m * n_max + n],
        kelimeBenzerlikLin[m * n_max + n],
        kelimeBenzerlikJC[m * n_max + n],
        kelimeBenzerlikLC[m * n_max + n],
        kelimeBenzerlikWP[m * n_max + n]))
}
else
//Daha önce bulunmuş ilintililik/benzerlik degerlerini
//al
{
    kelimeIlintililikGG[m * n_max + n] =
        es[esIndex].GG_Ilintililik
    kelimeBenzerlikResnik[m * n_max + n] =
        es[esIndex].Resnik_Benzerlik
    kelimeBenzerlikLin[m * n_max + n] =
        es[esIndex].Lin_Benzerlik
    kelimeBenzerlikJC[m * n_max + n] =
        es[esIndex].JC_Benzerlik
}

```

```

        kelimeBenzerlikLC[m * n_max + n] =
            es[esIndex].LC_Benzerlik
        kelimeBenzerlikWP[m * n_max + n] =
            es[esIndex].WP_Benzerlik
    }
else
{
    //İki kelime aynı ontolojideki terimlerle
    //etiketlenmemişse
    kelimeIlintililikGG[m * n_max + n] = 0
    kelimeBenzerlikResnik[m * n_max + n] = 0
    kelimeBenzerlikLin[m * n_max + n] = 0
    kelimeBenzerlikJC[m * n_max + n] = 0
    kelimeBenzerlikWP[m * n_max + n] = 0
    kelimeBenzerlikLC[m * n_max + n] = 0
}
}
else
{
    //İki kelime birden ya da kelimelerden birisi
    //ontolojideki terimlerle etiketlenmemişse
    kelimeIlintililikGG[m * n_max + n] = 0
    kelimeBenzerlikResnik[m * n_max + n] = 0
    kelimeBenzerlikLin[m * n_max + n] = 0
    kelimeBenzerlikJC[m * n_max + n] = 0
    kelimeBenzerlikWP[m * n_max + n] = 0
    kelimeBenzerlikLC[m * n_max + n] = 0
}
}
}

double tmpToplamIlintililikGG = 0
double tmpToplamBenzerlikResnik = 0
double tmpToplamBenzerlikLin = 0
double tmpToplamBenzerlikJC = 0
double tmpToplamBenzerlikWP = 0
double tmpToplamBenzerlikLC = 0

//Kelime ilintililik/benzerlik değerlerini kullanarak cümleler
//arasındaki bağdaşıklık değerlerini hesapla
for (int k = 0; k < kelimeIlintililikGG.length; k++)
{
    tmpToplamIlintililikGG = tmpToplamIlintililikGG +
        kelimeIlintililikGG[k];
}
arrCumleler[cumleIndeks].GGBagdasiklik(tmpToplamIlintililikGG/
    kelimeIlintililikGG.length)
for (int k = 0; k < kelimeBenzerlikResnik.length; k++)
{
    tmpToplamBenzerlikResnik = tmpToplamBenzerlikResnik +
        kelimeBenzerlikResnik[k]
}
arrCumleler[cumleIndeks].
    ResnikBagdasiklik(tmpToplamBenzerlikResnik /
        kelimeBenzerlikResnik.length)
for (int k = 0; k < kelimeBenzerlikLin.length; k++)
{
    tmpToplamBenzerlikLin = tmpToplamBenzerlikLin +
        kelimeBenzerlikLin[k]
}
arrCumleler[cumleIndeks].LinBagdasiklik(tmpToplamBenzerlikLin/

```

```
        kelimeBenzerlikLin.length)
    for (int k = 0; k < kelimeBenzerlikJC.length; k++)
    {
        tmpToplamBenzerlikJC = tmpToplamBenzerlikJC +
            kelimeBenzerlikJC[k]
    }
    arrCumleler[cumleIndeks].JCBagdasiklik(tmpToplamBenzerlikJC /
        kelimeBenzerlikJC.length)
    for (int k = 0; k < kelimeBenzerlikWP.length; k++)
    {
        tmpToplamBenzerlikWP = tmpToplamBenzerlikWP +
            kelimeBenzerlikWP[k]
    }
    arrCumleler[cumleIndeks].WPBagdasiklik(tmpToplamBenzerlikWP /
        kelimeBenzerlikWP.length)
    for (int k = 0; k < kelimeBenzerlikLC.length; k++)
    {
        tmpToplamBenzerlikLC = tmpToplamBenzerlikLC +
            kelimeBenzerlikLC[k]
    }
    arrCumleler[cumleIndeks].LCBagdasiklik(tmpToplamBenzerlikLC /
        kelimeBenzerlikLC.length)
    cumleIndeks++
}

//Ardışık cümleler arasındaki bağdaşıklık değerlerini kullanarak
//her yöntem için metin bağdaşıklığını hesapla
double tmpToplamGGBagdasiklik = 0
for (int l = 0; l < arrCumleler.size - 1; l++)
//son satırda bağdaşıklık değeri yok
{
    tmpToplamGGBagdasiklik = tmpToplamGGBagdasiklik +
        arrCumleler[l].GGBagdasiklik
}
double tmpToplamResnikBagdasiklik = 0
for (int l = 0; l < arrCumleler.size - 1; l++)
{
    tmpToplamResnikBagdasiklik = tmpToplamResnikBagdasiklik +
        arrCumleler[l].ResnikBagdasiklik
}
double tmpToplamLinBagdasiklik = 0
for (int l = 0; l < arrCumleler.size - 1; l++)
{
    tmpToplamLinBagdasiklik = tmpToplamLinBagdasiklik +
        arrCumleler[l].LinBagdasiklik
}
double tmpToplamJCBagdasiklik = 0
for (int l = 0; l < arrCumleler.size - 1; l++)
{
    tmpToplamJCBagdasiklik = tmpToplamJCBagdasiklik +
        arrCumleler[l].JCBagdasiklik
}
double tmpToplamWPBagdasiklik = 0
for (int l = 0; l < arrCumleler.size - 1; l++)
{
    tmpToplamWPBagdasiklik = tmpToplamWPBagdasiklik +
        arrCumleler[l].WPBagdasiklik
}
double tmpToplamLCBagdasiklik = 0
for (int l = 0; l < arrCumleler.size - 1; l++)
{
```

```
    tmpToplamLCBagdasiklik = tmpToplamLCBagdasiklik +  
        arrCumleler[1].LCBagdasiklik  
}
```

```
//GG bařdasıklık deęeri  
(tmpToplamGGBagdasiklik / arrCumleler.size - 1)  
//Resnik bařdasıklık deęeri  
(tmpToplamResnikBagdasiklik / arrCumleler.size - 1)  
//Lin bařdasıklık deęeri  
(tmpToplamLinBagdasiklik / arrCumleler.size - 1)  
//Jiang ve Conrath bařdasıklık deęeri  
(tmpToplamJCBagdasiklik / arrCumleler.size - 1)  
//Wu ve Palmer bařdasıklık deęeri  
(tmpToplamWPBagdasiklik / arrCumleler.size - 1)  
// Leacock ve Chodorow bařdasıklık deęeri  
(tmpToplamLCBagdasiklik / arrCumleler.size - 1)  
}
```

Ek 10. Latin Kare Deseni

Latin kare deseni yöntemiyle N adet koşulun sıralamalarının belirlenmesi sonucunda N farklı sıralama elde edilir. Yani 4 koşulun Latin kare deseni yöntemiyle sıralanması sonucunda yine 4 adet sıralama elde edilir. Latin kare yöntemiyle oluşturulan sıralamalarda, her koşul bir başka koşulun öncesinde ve sonrasında sadece bir kere bulunur. N sayısının çift olması durumunda Latin kare deseni yöntemiyle sıralamalar oluşturmak için aşağıdaki adımlar izlenmelidir (Cozby, 2005):

1. Koşul sayısı belirlenir. Bu örnekte kullanılmak üzere 6 koşulu temsil etmek için A, B, C, D, E, F harfleri kullanılmıştır.
2. İlk sıralama aşağıdaki gibi belirlenir:

A, B, S, C, S – 1, D, S – 2, E

Bu örnekteki S harfi son koşulu temsil etmektedir. 6 koşul için ilk sıralama aşağıdaki gibi olacaktır:

A, B, F, C, E, D

F harfi son koşulu (S), E harfi son koşuldan bir önceki koşulu (S – 1) temsil etmektedir.

3. İkinci sıralama için değerler, birinci sıralamadaki değerler bir arttırılarak bulunur. Son koşulun (bu örnek için F) bir arttırılması sonucunda tekrar ilk koşul (bu örnek için A) elde edilir.

B, C, A, D, F, E

4. Üçüncü adımdaki işlem diğer sıralamalar için de tekrarlanır. İkinci sıralamanın koşulları bir arttırılarak üçüncü sıralama elde edilir ve bu şekilde son sıralama elde edilene kadar devam edilir.

6 x 6 Latin kare deseni aşağıdaki gibi olacaktır:

A	B	F	C	E	D
B	C	A	D	F	E
C	D	B	E	A	F
D	E	C	F	B	A
E	F	D	A	C	B
F	A	E	B	D	C

5. Her koşul bir harfe atanarak uygulamada kullanılacak sıralamalar elde edilir.

N sayısının tek olması durumunda ise iki Latin kare deseni oluşturulur. Birinci Latin kare deseni için yukarıda anlatılan adımlar aynı şekilde izlenir. İkinci Latin kare deseni ise elde edilen birinci kare deseni ters çevrilerek elde edilir. Her satırdaki ilk koşul sonuncu, ikinci koşul sondan bir önceki olacak şekilde sıralamalar değiştirilerek ikinci Latin kare deseni elde edilir. Daha sonra elde edilen ikinci kare deseni birincinin altına eklenir. 5 koşul için 10 satırlık bir Latin kare deseni elde edilir.

Ek 11. Deneyde Kullanılan Metinler

D30003.M.100.T.13

Clearly, the Spanish authorities have every right to try and investigate these matters further. General Augusto Pinochet, who ruled Chile as a despot for 17 years, has been arrested in London after Spain asked that he be extradited for the presumed murders of hundreds of Chilean and Spanish citizens, the British authorities announced on 10/17/1998. General Augusto Pinochet, the former Chilean dictator arrested here at the request of a Spanish judge, remained sequestered under police guard on 10/18/1998, awaiting a potentially devastating court hearing to weigh his extradition on charges of genocide, terrorism and murder.

D30003.M.100.T.18

A stubbornly independent judge who more than once has challenged the Spanish government, Baltasar Garzon was ridiculed when he first set out to pursue the most notorious right-wing generals in South America for human rights crimes committed in the 1970s and '80s, offenses for which they had already been granted amnesty at home. He said the case seemed to be unprecedented and unusual. An influential lawmaker from the governing Labor Party on Saturday backed Spanish requests to question former Chilean dictator General Augusto Pinochet, in London for back surgery, on allegations of genocide and terrorism.

D30003.M.100.T.26

The Spanish authorities contend that Pinochet may have committed crimes against Spanish citizens in Chile. Former Chilean dictator General Augusto Pinochet has been arrested by British police on a Spanish extradition warrant, despite protests from Chile that he is entitled to diplomatic immunity. Baltasar Garzon, one of two Spanish magistrates handling probes into human rights violations in Chile and Argentina, filed a request to question Pinochet on Wednesday. But Britain said Pinochet did not have diplomatic immunity. Chile said, graf 3 pvs President Fidel Castro said Sunday he disagreed with the arrest in London of former Chilean dictator Augusto Pinochet, calling it a case of international meddling.

D30003.M.100.T.6

Former Chilean dictator General Augusto Pinochet has been arrested by British police on a Spanish extradition warrant, despite protests from Chile that he is entitled to diplomatic immunity. Augusto Pinochet is being called to account by Spanish authorities for the deaths, detention and torture of political opponents. Pinochet, who ousted elected President Salvador Allende in a bloody 1973 coup, remained commander-in-chief of the Chilean army until March. Since the case against Pinochet began at Madrid's National Court, the United States has agreed to provide the Spanish judges with documents related to human rights violations under the Pinochet regime.

D30003.M.100.T.C

Augusto Pinochet, former Chilean dictator, was arrested in London on 14 October 1998. Pinochet, 82, was recovering from surgery. The arrest was in response to an extradition warrant served by Baltasar Garzon, a maverick Spanish judge. Pinochet was charged with murdering thousands, including many Spaniards. Pinochet is awaiting a hearing, his fate in the balance. Chile protested, insisting that Pinochet had diplomatic immunity. Britain disagreed. The reaction of the Chilean people was mixed. American scholars applauded the arrest, saying that it set a precedent for other terrorist dictators. Castro criticized the arrest, and called it unprecedented international meddling.

D30012.M.100.T.13

Satellite operators reported no immediate problems on 11/17/1998 as the Earth passed through the tail of a comet that peppered the atmosphere with tiny bits of space junk zipping along at 155,000 mph (250,000 kph). While people around the world watched the beginning of the best meteor storm light show in three decades, the USA National Aeronautics and Space Administration and private satellite companies rotated their high-tech machinery to face away from the approaching debris. The shower, which peaks every 33 years, is caused by the Earth's passage through the long tail of the Comet Tempel-Tuttle.

D30012.M.100.T.16

Peppered the atmosphere with tiny bits of space junk zipping along at 155,000 mph (250,000 kph). Clouds interfered with the view from many sites, but many stargazers were rewarded with a spectacular view of meteors streaking through the skies in blazes of red and white as the Leonid meteor shower reached its climax early Wednesday in Asia. The shower is caused by the Earth's passage through the long tail of the Comet Tempel-Tuttle. By the time the earth collides with a dense swarm of meteoroids Tuesday night to unleash what may be the heaviest meteor bombardment since 1966, many orbiting satellites will have maneuvered into positions facing away from the storm to avoid damage.

D30012.M.100.T.26

All of these satellites came through Tuesday's meteor shower unscathed. Showers of Leonid meteors may produce hundreds or thousands of blazing meteors each hour. Some satellites in low-earth orbits can actually hide from meteoroid storms, Ozkul said. The scientists who track Temple-Tuttle do not even call it a shower, they call it a meteor storm. Satellite experts said that some damage might take days to detect, but that satellites generally seemed to have escaped disabling harm. This storm of meteors, called Leonid meteors because they come from the direction of constellation Leo, will be the first to hit the Earth since 1966 when the world's space programs were in their infancy, and its effects on satellite systems are uncertain.

D30012.M.100.T.6

People in East Asia, however, will have ringside seats, and if skies are clear, they are likely to witness a great meteor display under ideal conditions: a dark, moonless night. Some people think the Leonid storm this year will be as good as the one in 1966, which was the best one this century, but we think the 1998 shower will be more modest perhaps something between 200 and 5,000 meteors per hour. Astrophysicists, atmospheric scientists and others have been readying their experiments for a year or more in preparation for Tuesday night's expected Leonid Meteor storm.

D30028.M.100.T.26

Turkey has amassed troops on the border with Syria and warned of a military confrontation following allegations that Syria is supporting Turkish Kurdish rebels. For its part, Turkey is complaining ever more loudly about Syria's support for Kurdish insurgents in Turkey. Turkey amassed troops, pick up 4th graf pvs Egyptian President Hosni Mubarak flew to Damascus on Sunday to try to defuse growing tension between Syria and Turkey. On Saturday, Defense Minister Ismet Sezgin said Turkey would soon carry out an exercise near the Syrian border, private NTV television reported. Turkish generals began issuing public warnings to Syria last month.

D30028.M.100.T.6

President Hosni Mubarak of Egypt will go to Syria and Turkey to help cool tensions between the two countries, which appear to be edging toward a military conflict. Signaling it does not want to be involved in any potential military confrontation between Syria and Turkey, Israel is limiting routine exercises along its own border with Syria. For its part, Turkey is complaining ever more loudly about Syria's support for Kurdish insurgents in Turkey. Turkey's military-dominated National Security Council discussed problems with Syria on Wednesday, and announced afterward that Turkey was considering imposing economic, political and military sanctions against Syria.

D30028.M.100.T.F

Tensions between Turkey and Syria have reached a point causing the top Turkish military commander to say they have reached a state of undeclared war. Turkey accuses Syria of supporting Kurdish rebels in Turkey, a charge denied by Syria. Syria is angry over Turkey's closer ties to Israel and alleges that Turkey, by building dams on the Euphrates River, threatens its water supply. Egypt's President Mubarak has visited Saudi Arabia and Damascus and plans to continue to Ankara hoping to defuse the crisis. Israel claims no part in the tension. Turkey has attacked Kurdish bases in Iraq and threatens those in Syria.

D30040.M.100.T.6

Taking a major step toward statehood, the Palestinians on Tuesday inaugurated Gaza International Airport, their first gateway to the world, with cheers, tears and an outpouring of patriotism. The first Palestinian commercial flight landed at Amman's Marka Airport on Saturday, inaugurating an air route between Jordan and the autonomous Gaza Strip. Established in 1996, the tiny Palestinian airline was an outgrowth of the accords with Israel on Palestinian self-rule. Israel has threatened to close down the Palestinian-run Gaza airport over a security violation, an Israeli official said Tuesday, a move that could further undermine the already fragile peace process.

D31009.M.100.T.13

Turkey's president began trying to construct a new government on 11/25/1998 but the pieces do not neatly fit: pro-Islamic forces pressing for power and their military-backed, Western-style opponents struggling to unite. President Suleyman Demirel opened talks with party leaders just hours after the corruption-battered coalition of Prime Minister Mesut Yilmaz lost a confidence vote in parliament. One of Turkey's leading politicians warned on 11/29/1998 that a political vacuum could cripple the domestic battle to keep Islamists at bay and in the fight abroad for the extradition of a Kurdish rebel leader.

D31009.M.100.T.26

The Islamic party, Virtue, is the largest party in parliament. Last year, the military ousted Turkey's first Islamic-led government. Acting Deputy Premier Bulent Ecevit on Sunday called on Turkey's two center-right parties, led by arch-rivals acting Prime Minister Mesut Yilmaz and former Premier Tansu Ciller, to join him in a governing coalition. The no-confidence vote has dragged Turkey into grave political chaos, Yilmaz declared. Within hours of the government's collapse, Virtue leader Recai Kutan met with President Suleyman Demirel, who is charged with picking a political leader who can command a governing majority in Parliament.

D31009.M.100.T.6

Turkey's government collapsed Wednesday after losing a confidence vote in parliament over a corruption scandal that tainted the prime minister with allegations of mob ties. Turkey's president began trying to construct a new government Wednesday but the pieces do not neatly fit: pro-Islamic forces pressing for power and their military-backed, Western-style opponents struggling to unite. One of Turkey's leading politicians warned Sunday that a political vacuum could cripple the domestic battle to keep Islamists at bay and in the fight abroad for the extradition of a Kurdish rebel leader. Turkey's secular military leadership warned a pro-Islamic party Monday against dragging the army into politics.

D31010.M.100.T.13

An Arctic cold wave and accompanying blizzards have killed at least 65 people throughout Europe, police and media reported on 11/23/1998. Croatia reported its first cold-related death on 11/25/1998, a woman who lived alone in an unheated house. The death toll from almost three weeks of sub-zero temperatures has reached 83, after a homeless man froze to death in a Warsaw park. Poland's extreme cold began November 16, with temperatures dropping to minus 26 Celsius (minus 4 Fahrenheit).

D31010.M.100.T.16

The early winter weather played a role in at least 24 deaths in Romania and Bulgaria over the past three days. In Poland, 32 people died, most of them homeless or others who passed out in the cold after drinking alcohol. An Arctic cold wave and accompanying blizzards have killed at least 62 people throughout Europe. As much as 16 inches of snow capped Mt. Vesuvius as an unusually cold air mass invaded western Europe on chilly winds from the east. Closer to the center of the frigid air mass, temperatures plunged to minus 2 degrees Fahrenheit in Moscow on November 17, prompting Russian meteorologists to declare the mid-November outbreak the coldest in 20 years.

D31010.M.100.T.18

It already has claimed 20 more lives than the number of cold-related deaths in all of last winter, police say two more victims of Poland's prolonged cold front were found Thursday, raising the death toll from more than two weeks of freezing temperatures to at least 74. One man was injured by a falling icicle, but he was not hospitalized. The weather played a role in at least 24 deaths in Romania and Bulgaria over the past three days. An Arctic cold wave that hit Poland a week ago has killed 32 people, police said Monday.

D31010.M.100.T.26

Poland's extreme cold began November 16, with temperatures dropping to minus 26 Celsius (minus 4 Fahrenheit). In Poland, 32 people died, most of them homeless or others who passed out in the cold after drinking alcohol, police said. An Arctic cold wave and accompanying blizzards have killed at least 62 people throughout Europe, police and media reported Monday. Temperatures in central and eastern Poland over the weekend were as low as minus 20 degrees Celsius (minus 4 Fahrenheit). Two more victims of Poland's prolonged cold front were found Thursday, raising the death toll from more than two weeks of freezing temperatures to at least 74. Police said a 40-year-old man was found frozen to death in central Poland on Thursday morning.

D31010.M.100.T.6

An Arctic cold wave and accompanying blizzards have killed at least 65 people throughout Europe, police and media reported Monday. Temperatures in central and eastern Poland over the weekend were as low as minus 20 degrees Celsius (minus 4 Fahrenheit). Temperatures have dropped well below freezing, with minus 26 Celsius (minus 4 Fahrenheit) recorded in Poland on Sunday. The death toll from almost three weeks of sub-zero temperatures reached 85 on Saturday, after two men froze to death in their unheated homes in northeastern Poland.

D31010.M.100.T.E

An arctic cold wave, the worst in 10 years, hit parts of Europe, bringing sub-zero temperatures and killing scores of people. Hardest hit were Poland, Bulgaria, and Romania as well as parts of central and eastern France. In Poland, three weeks of sub-zero temperatures killed at least 85 people in November, 29 more than in all of the previous winter. Most of the victims were homeless whose deaths by exposure were alcohol related. Blizzards and cold temperatures also hit Bulgaria and Romania, stranding hundreds in their cars. Elsewhere, snow blanketed the Italian island of Capri for the first time in 10 years.

D31033.M.100.T.13

In a recent move that indicated he has no intention of backing away from his company's aggressive business tactics, Microsoft Corporation's chairman, Bill Gates, has proposed acquiring the software business behind 3Com Corporation's popular Palm Pilot hand-held computer, according to people who have been briefed on the discussions. The antitrust suit, initiated last October, originally focused on Microsoft's linking of its Internet Explorer World Wide Web browser with the industry-dominant Windows 95 operating system. Microsoft replies that the prosecutors are misinterpreting a routine business meeting. Demonstrating yet again the power of its franchise, the Microsoft Corporation.

D31033.M.100.T.16

After months of noisy prelude, the antitrust trial against the Microsoft Corporation opened in federal court Monday morning with a pointed personal attack on Bill Gates, the company's chairman. Microsoft, the world's leading maker of computer software, said demand for the Windows 98 operating-system exceeded its projections, particularly in Japan and Europe. A lawyer for Microsoft Corporation on Thursday portrayed the company's competitor in the Internet software business, Netscape Communications Corporation, as a willing and eager participant in deal-making that culminated in a June 1995 meeting at which the government and Netscape now say that Microsoft illegally offered to divide the market.

D31033.M.100.T.18

The New Economy is on trial here along with Microsoft Corporation Sure, the company is the one in the dock. Every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations, is hereby declared to be illegal. In a recent move that indicated he has no intention of backing away from his company's aggressive business tactics, Microsoft Corporation's chairman, Bill Gates, has proposed acquiring the software business behind 3Com Corporation's popular Palm Pilot hand-held computer, according to people who have been briefed on the discussions.

D31041.M.100.T.13

Prospective investor Cathay Pacific Airways wants the number of employees at troubled Philippine Airlines to be slashed from more than 8,000 to 5,000 but PAL has not yet agreed, Finance Secretary Edgardo Espiritu said on 11/23/1998. PAL closed for nearly two weeks on September 23 after failing to persuade its largest union to accept a management-proposed recovery plan under which its collective bargaining agreement would be suspended for 10 years in exchange for a 20 percent share of the company's stock. The union later decided to accept the plan, allowing PAL to resume domestic and international flights.

D31041.M.100.T.16

President Joseph Estrada on Saturday urged militant unionists at Philippine Airlines to accept a vote by workers approving a 10-year no-strike deal to revive the debt-laden airline. President Joseph Estrada said Saturday the financially troubled Philippine Airlines will resume its international flights on Sunday by flying him to Singapore where he will address the World Economic Forum. Philippine Airlines said Thursday it will attempt to rebuild alone after Hong Kong's Cathay Pacific Airlines pulled out of talks on acquiring a stake in the ailing Philippine flag carrier. A strike by employees precipitated the airline's near-death experience in September, when Tan shut down the carrier after its unions refused to accept a drastic cost-cutting plan.

D31041.M.100.T.18

Philippine Airlines is not just another shipwreck of a company. PAL had hoped for an infusion of capital from another airline to keep it afloat and help it resume payments on dollars 2.1 billion in debt. PAL earlier accepted a preliminary investment offer from Cathay Pacific, and the companies announced November 10 that they would draw up a comprehensive plan to rehabilitate the Philippine flag carrier. The number of labor strikes and lockouts in the Philippines has declined this year despite business difficulties caused by Asia's financial crisis, the labor department said Sunday.

D31041.M.100.T.6

PAL, Asia's oldest airline, has been unable to make payments on dollars 2.1 billion in debt after being devastated by a pilots' strike and by Asia's currency crisis. PAL earlier accepted a preliminary investment offer from Cathay Pacific, Ailing Philippine Airlines and prospective investor Cathay Pacific Airways have clashed over the laying off of PAL workers, prompting PAL to revive talks with another foreign airline, an official said Tuesday. Cathay Pacific Airways said Wednesday it had pulled out of talks to buy a stake in ailing Philippine Airlines making the uncertain future at PAL even cloudier.

D31050.M.100.T.13

The charges that Wang faces, in a trial that will almost certainly be secret and quick, could bring a sentence of five years or more. A Chinese dissident fleeing a new round of arrests of democracy activists in Shanghai arrived here on 12/09/1998 and announced that he and other opponents of the Chinese government plan a demonstration on 12/09/1998 at the United Nations to protest the crackdown. Last week the Chinese government arrested 10 members and sympathizers of the China Democracy Party, one of whom, Wang Youcai, is to go on trial December 17.

D31050.M.100.T.16

With attorneys locked up, harassed or plain scared, two prominent dissidents will defend themselves against charges of subversion Thursday in China's highest-profile dissident trials in two years. Releasing Liu Nianchun appeared to be an attempt by the government to blunt international criticism over Monday's upcoming trial of prominent dissident Xu Wenli. Xu is the third leading member of a would-be opposition political party put on trial for subversion in a three-week crackdown that has seen 30 dissidents arrested or interrogated. The dissident, Yao Zhenxian is a leader of the China Democracy Party, which was formed in June during President Clinton's visit to China.

D31050.M.100.T.26

Qin, 44, and Wang, 31, are accused of inciting the subversion of state power, apparently for helping to organize the China Democracy Party. Wang argued that his activities were legal under China's constitution and under an international covenant on political rights that China has signed, Ms Hu said. As the founder of the suppressed new political party, the China Democracy Party, he publicly announced its charter in June during President Clinton's visit to China. He said he had no information from Chinese authorities about Qin or Wang. Three other Democratic Party organizers were also detained on Monday, although two of them were released early Wednesday.

D31050.M.100.T.F

Chinese officials have finally spoken out on their dealings with dissidents trying to establish the China Democracy Party. They claimed that Xu Wenli, was suspected of activities damaging to national security. The USA has denounced China's detaining those peacefully exercising fundamental freedoms. Those arrested are expected to be tried under China's vague State Security Law. Obtaining lawyers for their trials became virtually impossible forcing some of the dissidents to defend themselves. President Zemin said that economic reforms would continue but would not be a prelude to multiparty democracy. China did release ailing Liu Nianchun from prison work camp and exiled him to the USA.

Ek 12. Deneyde Elde Edilen Metin Bağdaşıklık Değerleri

Çizelge 1

Metin	İnsan yargısı	Ontoloji tabanlı anlamsal ilintililik ölçüm yöntemi	Resnik	Lin	Jiang ve Conrath	Gizli Anlam Çözümleme
D30003.M.100.T.13	4,20000	0,02389	0,13164	0,03830	0,51245	0,15000
D30003.M.100.T.18	3,94444	0,00574	0,00574	0,00574	0,00574	0,08000
D30003.M.100.T.26	3,50000	0,01987	0,06779	0,02589	0,23681	0,10000
D30003.M.100.T.6	6,11765	0,02598	0,16826	0,04653	0,67235	0,05000
D30003.M.100.T.C	5,93750	0,02425	0,11176	0,03830	0,41813	0,21000
D30012.M.100.T.13	5,80769	0,02342	0,21492	0,04427	0,91513	0,06000
D30012.M.100.T.16	3,28571	0,00592	0,01352	0,00726	0,03507	0,12000
D30012.M.100.T.26	2,62500	0,01245	0,02497	0,01417	0,06215	0,02000
D30012.M.100.T.6	4,00000	0,02772	0,05734	0,03165	0,16515	0,07000
D30028.M.100.T.26	3,80952	0,02097	0,02351	0,02123	0,03776	0,13000
D30028.M.100.T.6	4,57143	0,03345	0,03293	0,03293	0,03293	0,20000
D30028.M.100.T.F	5,09524	0,04409	0,16233	0,04046	0,83135	0,12000
D30040.M.100.T.6	4,92857	0,01441	0,03026	0,01469	0,10715	0,15000
D31009.M.100.T.13	3,33333	0,00649	0,05063	0,01193	0,21009	0,01000
D31009.M.100.T.26	3,24000	0,00641	0,02622	0,00908	0,09516	0,09000
D31009.M.100.T.6	4,38095	0,01105	0,05448	0,01626	0,21865	0,12000
D31010.M.100.T.13	5,40000	0,01107	0,01107	0,01107	0,01107	0,21000
D31010.M.100.T.16	5,16000	0,01029	0,03766	0,01497	0,15677	0,11000
D31010.M.100.T.18	3,60000	0,01017	0,01269	0,00661	0,08837	0,03000
D31010.M.100.T.26	4,80952	0,01691	0,04052	0,02124	0,12177	0,15000
D31010.M.100.T.6	5,28571	0,02686	0,09255	0,03641	0,36347	0,18000
D31010.M.100.T.E	5,75758	0,00871	0,01010	0,00678	0,06328	0,05000
D31033.M.100.T.13	4,03846	0,01497	0,09883	0,02839	0,38744	0,02000
D31033.M.100.T.16	4,80952	0,01975	0,23201	0,05311	0,95228	0,09000
D31033.M.100.T.18	3,61111	0,01646	0,06677	0,02522	0,23779	0,01000
D31041.M.100.T.13	4,33333	0,01808	0,11091	0,03015	0,46973	0,08000
D31041.M.100.T.16	4,13636	0,01407	0,05587	0,02081	0,20256	0,23000
D31041.M.100.T.18	3,88462	0,00493	0,05978	0,01321	0,24933	0,08000
D31041.M.100.T.6	3,87500	0,02036	0,05795	0,02579	0,18964	0,18000
D31050.M.100.T.13	3,34615	0,02091	0,19430	0,04390	0,84103	0,04000
D31050.M.100.T.16	3,27273	0,01909	0,11419	0,03351	0,44712	0,13000
D31050.M.100.T.26	4,31250	0,01484	0,05046	0,01933	0,20317	0,11000
D31050.M.100.T.F	4,20000	0,01451	0,18388	0,03808	0,78275	0,08000

Çizelge 2

British National Corpus (World Sürümü)					
Metin	İnsan yargısı	Resnik	Lin	Wu ve Palmer	Leacock ve Chodorow
D30003.M.100.T.13	4,20000	0,89220	0,14097	0,24430	0,78273
D30003.M.100.T.18	3,94444	0,97432	0,13711	0,24184	0,83051
D30003.M.100.T.26	3,50000	0,89323	0,11262	0,24874	0,84742
D30003.M.100.T.6	6,11765	0,83266	0,10081	0,24194	0,82530
D30003.M.100.T.C	5,93750	0,81141	0,10431	0,24952	0,87367
D30012.M.100.T.13	5,80769	1,75084	0,19260	0,40811	1,27725
D30012.M.100.T.16	3,28571	1,47600	0,15810	0,32648	1,05576
D30012.M.100.T.26	2,62500	1,65578	0,17709	0,35378	1,16965
D30012.M.100.T.6	4,00000	1,24544	0,16417	0,33920	1,16590
D30028.M.100.T.26	3,80952	1,52093	0,16684	0,35942	1,14586
D30028.M.100.T.6	4,57143	1,42992	0,19025	0,33228	1,04827
D30028.M.100.T.F	5,09524	1,35946	0,16266	0,31435	0,93836
D30040.M.100.T.6	4,92857	1,01855	0,11407	0,26728	0,87319
D31009.M.100.T.13	3,33333	0,97307	0,12533	0,24372	0,73738
D31009.M.100.T.26	3,24000	1,13206	0,13537	0,27075	0,85737
D31009.M.100.T.6	4,38095	1,22783	0,15622	0,32068	0,97870
D31010.M.100.T.13	5,40000	1,40595	0,16132	0,33234	1,05141
D31010.M.100.T.16	5,16000	1,70973	0,18288	0,39831	1,28628
D31010.M.100.T.18	3,60000	1,47706	0,17743	0,39008	1,31465
D31010.M.100.T.26	4,80952	1,31010	0,15105	0,35015	1,17381
D31010.M.100.T.6	5,28571	1,32084	0,14383	0,32030	1,05185
D31010.M.100.T.E	5,75758	1,15338	0,13468	0,32292	1,06958
D31033.M.100.T.13	4,03846	1,29130	0,14977	0,28075	0,88714
D31033.M.100.T.16	4,80952	1,32903	0,14472	0,32761	1,03737
D31033.M.100.T.18	3,61111	1,34212	0,17375	0,29376	0,94493
D31041.M.100.T.13	4,33333	1,30831	0,15503	0,33070	1,13165
D31041.M.100.T.16	4,13636	1,00940	0,11025	0,24976	0,82463
D31041.M.100.T.18	3,88462	1,34927	0,19304	0,36285	1,10159
D31041.M.100.T.6	3,87500	1,05782	0,12720	0,27528	0,93196
D31050.M.100.T.13	3,34615	0,86907	0,09612	0,25284	0,84187
D31050.M.100.T.16	3,27273	0,73865	0,10799	0,21043	0,67449
D31050.M.100.T.26	4,31250	0,86633	0,13764	0,24724	0,79616
D31050.M.100.T.F	4,20000	1,10167	0,14418	0,30565	1,03951

Çizelge 3

Brown Corpus					
Metin	İnsan yargısı	Resnik	Lin	Wu ve Palmer	Leacock ve Chodorow
D30003.M.100.T.13	4,20000	0,89707	0,13213	0,24430	0,78273
D30003.M.100.T.18	3,94444	0,98284	0,12945	0,24184	0,83051
D30003.M.100.T.26	3,50000	0,90737	0,10617	0,24874	0,84742
D30003.M.100.T.6	6,11765	0,83886	0,09296	0,24194	0,82530
D30003.M.100.T.C	5,93750	0,81214	0,09563	0,24952	0,87367
D30012.M.100.T.13	5,80769	1,73511	0,19202	0,40811	1,27725
D30012.M.100.T.16	3,28571	1,45478	0,15713	0,32648	1,05576
D30012.M.100.T.26	2,62500	1,61393	0,17845	0,35378	1,16965
D30012.M.100.T.6	4,00000	1,24633	0,16552	0,33920	1,16590
D30028.M.100.T.26	3,80952	1,54081	0,15538	0,35942	1,14586
D30028.M.100.T.6	4,57143	1,47757	0,18266	0,33228	1,04827
D30028.M.100.T.F	5,09524	1,37362	0,15328	0,31435	0,93836
D30040.M.100.T.6	4,92857	1,00955	0,08965	0,26728	0,87319
D31009.M.100.T.13	3,33333	0,97771	0,11941	0,24372	0,73738
D31009.M.100.T.26	3,24000	1,14579	0,13498	0,27075	0,85737
D31009.M.100.T.6	4,38095	1,23423	0,14915	0,32068	0,97870
D31010.M.100.T.13	5,40000	1,39077	0,14670	0,33234	1,05141
D31010.M.100.T.16	5,16000	1,71066	0,17599	0,39831	1,28628
D31010.M.100.T.18	3,60000	1,47353	0,17612	0,39008	1,31465
D31010.M.100.T.26	4,80952	1,30049	0,14124	0,35015	1,17381
D31010.M.100.T.6	5,28571	1,28402	0,12520	0,32030	1,05185
D31010.M.100.T.E	5,75758	1,14994	0,12803	0,32292	1,06958
D31033.M.100.T.13	4,03846	1,30999	0,14902	0,28075	0,88714
D31033.M.100.T.16	4,80952	1,34412	0,14462	0,32761	1,03737
D31033.M.100.T.18	3,61111	1,36149	0,17602	0,29376	0,94493
D31041.M.100.T.13	4,33333	1,32145	0,15595	0,33070	1,13165
D31041.M.100.T.16	4,13636	1,01261	0,10714	0,24976	0,82463
D31041.M.100.T.18	3,88462	1,36166	0,19079	0,36285	1,10159
D31041.M.100.T.6	3,87500	1,06130	0,12620	0,27528	0,93196
D31050.M.100.T.13	3,34615	0,87313	0,08808	0,25284	0,84187
D31050.M.100.T.16	3,27273	0,74620	0,10494	0,21043	0,67449
D31050.M.100.T.26	4,31250	0,87596	0,13039	0,24724	0,79616
D31050.M.100.T.F	4,20000	1,10420	0,14325	0,30565	1,03951

Çizelge 4

Penn Treebank (Sürüm 2)					
Metin	İnsan yargısı	Resnik	Lin	Wu ve Palmer	Leacock ve Chodorow
D30003.M.100.T.13	4,20000	0,89576	0,13915	0,24430	0,78273
D30003.M.100.T.18	3,94444	0,96694	0,13575	0,24184	0,83051
D30003.M.100.T.26	3,50000	0,88068	0,11014	0,24874	0,84742
D30003.M.100.T.6	6,11765	0,82808	0,09986	0,24194	0,82530
D30003.M.100.T.C	5,93750	0,81209	0,10325	0,24952	0,87367
D30012.M.100.T.13	5,80769	1,77348	0,18620	0,40811	1,27725
D30012.M.100.T.16	3,28571	1,52442	0,14341	0,32648	1,05576
D30012.M.100.T.26	2,62500	1,68537	0,15270	0,35378	1,16965
D30012.M.100.T.6	4,00000	1,26656	0,15494	0,33920	1,16590
D30028.M.100.T.26	3,80952	1,52705	0,15617	0,35942	1,14586
D30028.M.100.T.6	4,57143	1,46080	0,18076	0,33228	1,04827
D30028.M.100.T.F	5,09524	1,38817	0,15430	0,31435	0,93836
D30040.M.100.T.6	4,92857	1,03357	0,11350	0,26728	0,87319
D31009.M.100.T.13	3,33333	0,96105	0,12217	0,24372	0,73738
D31009.M.100.T.26	3,24000	1,11829	0,13176	0,27075	0,85737
D31009.M.100.T.6	4,38095	1,20277	0,15122	0,32068	0,97870
D31010.M.100.T.13	5,40000	1,43373	0,15156	0,33234	1,05141
D31010.M.100.T.16	5,16000	1,73871	0,17415	0,39831	1,28628
D31010.M.100.T.18	3,60000	1,47690	0,16759	0,39008	1,31465
D31010.M.100.T.26	4,80952	1,32085	0,14418	0,35015	1,17381
D31010.M.100.T.6	5,28571	1,29572	0,12464	0,32030	1,05185
D31010.M.100.T.E	5,75758	1,16634	0,13246	0,32292	1,06958
D31033.M.100.T.13	4,03846	1,25933	0,14899	0,28075	0,88714
D31033.M.100.T.16	4,80952	1,31737	0,15125	0,32761	1,03737
D31033.M.100.T.18	3,61111	1,31715	0,17981	0,29376	0,94493
D31041.M.100.T.13	4,33333	1,28703	0,16175	0,33070	1,13165
D31041.M.100.T.16	4,13636	0,98543	0,11315	0,24976	0,82463
D31041.M.100.T.18	3,88462	1,32922	0,19418	0,36285	1,10159
D31041.M.100.T.6	3,87500	1,03782	0,13228	0,27528	0,93196
D31050.M.100.T.13	3,34615	0,85371	0,09540	0,25284	0,84187
D31050.M.100.T.16	3,27273	0,72420	0,10783	0,21043	0,67449
D31050.M.100.T.26	4,31250	0,84494	0,13217	0,24724	0,79616
D31050.M.100.T.F	4,20000	1,10083	0,14638	0,30565	1,03951

Çizelge 5

Semcor					
Metin	İnsan yargısı	Resnik	Lin	Wu ve Palmer	Leacock ve Chodorow
D30003.M.100.T.13	4,20000	0,92960	0,10467	0,24430	0,78273
D30003.M.100.T.18	3,94444	1,06298	0,09083	0,24184	0,83051
D30003.M.100.T.26	3,50000	0,94912	0,06443	0,24874	0,84742
D30003.M.100.T.6	6,11765	0,90220	0,06126	0,24194	0,82530
D30003.M.100.T.C	5,93750	0,88883	0,06563	0,24952	0,87367
D30012.M.100.T.13	5,80769	1,73919	0,15908	0,40811	1,27725
D30012.M.100.T.16	3,28571	1,46793	0,12420	0,32648	1,05576
D30012.M.100.T.26	2,62500	1,59355	0,12987	0,35378	1,16965
D30012.M.100.T.6	4,00000	1,38258	0,11036	0,33920	1,16590
D30028.M.100.T.26	3,80952	1,57243	0,10922	0,35942	1,14586
D30028.M.100.T.6	4,57143	1,56866	0,14986	0,33228	1,04827
D30028.M.100.T.F	5,09524	1,37376	0,12243	0,31435	0,93836
D30040.M.100.T.6	4,92857	1,02157	0,06364	0,26728	0,87319
D31009.M.100.T.13	3,33333	0,99888	0,08817	0,24372	0,73738
D31009.M.100.T.26	3,24000	1,18145	0,09099	0,27075	0,85737
D31009.M.100.T.6	4,38095	1,24336	0,11335	0,32068	0,97870
D31010.M.100.T.13	5,40000	1,42601	0,11212	0,33234	1,05141
D31010.M.100.T.16	5,16000	1,77596	0,11719	0,39831	1,28628
D31010.M.100.T.18	3,60000	1,52196	0,11411	0,39008	1,31465
D31010.M.100.T.26	4,80952	1,34195	0,09808	0,35015	1,17381
D31010.M.100.T.6	5,28571	1,29584	0,07714	0,32030	1,05185
D31010.M.100.T.E	5,75758	1,18329	0,08978	0,32292	1,06958
D31033.M.100.T.13	4,03846	1,37357	1,37357	0,28075	0,88714
D31033.M.100.T.16	4,80952	1,41228	0,09259	0,32761	1,03737
D31033.M.100.T.18	3,61111	1,52029	0,14842	0,29376	0,94493
D31041.M.100.T.13	4,33333	1,36743	0,11270	0,33070	1,13165
D31041.M.100.T.16	4,13636	1,04009	0,06829	0,24976	0,82463
D31041.M.100.T.18	3,88462	1,38769	0,13602	0,36285	1,10159
D31041.M.100.T.6	3,87500	1,11413	0,08411	0,27528	0,93196
D31050.M.100.T.13	3,34615	0,88643	0,06784	0,25284	0,84187
D31050.M.100.T.16	3,27273	0,80644	0,07948	0,21043	0,67449
D31050.M.100.T.26	4,31250	0,90544	0,09259	0,24724	0,79616
D31050.M.100.T.F	4,20000	1,12791	0,10391	0,30565	1,03951