

EGE ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

(DOKTORA TEZİ)

**ONTOLOJİLER VE ANLAMSAÇ ÇIKARSAMA
KULLANILARAK BİR ÖNERİ SİSTEMİ
TASARLANMASI**

Özlem ÖZGÖBEK

Tez Danışmanı: Doç. Dr. R. Cenk ERDUR

Bilgisayar Mühendisliği Anabilim Dalı

Bilim Dalı Kodu: 619.01.00

Sunuş Tarihi: 3.Nisan.2015

Bornova-İZMİR

2015

Özlem ÖZGÖBEK tarafından **DOKTORA TEZİ** olarak sunulan "**ONTOLOJİLER VE ANLAMSAL ÇIKARSAMA KULLANILARAK BİR ÖNERİ SİSTEMİ TASARLANMASI**" 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 **03.04.2015** tarihinde yapılan tez savunma sınavında aday oybirliği/oyçokluğu ile başarılı bulunmuştur.

Jüri Üyeleri:

İmza

Jüri Başkanı: Doç. Dr. R. Cenk ERDUR

.....

Raportör Üye: Prof. Dr. Oğuz DİKENELLİ

.....

Üye: Yard. Doç. Dr. Geylani KARDAŞ

.....

Üye: Yard. Doç. Dr. Korhan KARABULUT

.....

Üye: Yard. Doç. Dr. Tahir Emre KALAYCI

.....

EGE ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

ETİK KURALLARA UYGUNLUK BEYANI

E.Ü. Lisansüstü Eğitim ve Öğretim Yönetmeliğinin ilgili hükümleri uyarınca Doktora Tezi olarak sunduğum **"ONTOLOJİLER VE ANLAMSAÇ ÇIKARSAMA KULLANILARAK BİR ÖNERİ SİSTEMİ TASARLANMASI"** başlıklı bu tezin kendi çalışmam olduğunu, sunduğum tüm sonuç, doküman, bilgi ve belgeleri bizzat ve bu tez çalışması kapsamında elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara atıf yaptığımı ve bunları kaynaklar listesinde usulüne uygun olarak verdiğimi, tez çalışması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını, bu tezin herhangi bir bölümünü bu üniversite veya diğer bir üniversitede başka bir tez çalışması içinde sunmadığımı, bu tezin planlanmasından yazımına kadar bütün safhalarda bilimsel etik kurallarına uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul edeceğimi beyan ederim.

03.04.2015

.....

Özlem ÖZGÖBEK

ÖZET

ONTOLOJİLER VE ANLAMSAL ÇIKARSAMA KULLANILARAK BİR ÖNERİ SİSTEMİ TASARLANMASI

ÖZGÖBEK, Özlem

Doktora Tezi, Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı: Doç. Dr. R. Cenk ERDUR

Nisan 2015, 77 sayfa

Öneri sistemleri son yıllarda giderek artan ilgiye sahip bir araştırma alanıdır. Öneri sistemleri kullanıcıların herhangi bir kişisel çabası olmadan onlara en uygun öğeleri tavsiye eden sistemlerdir. Öneri sistemlerinin kullanım alanı oldukça geniştir. Alışveriş, müzik, film, haber, kitap gibi akla gelebilecek pek çok alanda uygulamaları mevcuttur. Ancak farklı alanlardaki farklı zorlukları nedeni ile alan bağımsız bir öneri sistemi tasarlanması oldukça zordur. Gerek kullanıcılar arasındaki benzerlikler, gerek içerikler arasındaki benzerlikler, gerekse de her ikisi birden kullanılarak geliştirilen farklı öneri sistemleri bulunmaktadır.

Bu tez kapsamında ontolojiler ve anlamsal çıkarsama kullanılarak tasarlanan bir öneri sistemi sunulmuştur. Sunulan bu yöntem diğer alanlara göre daha değişken olması sebebiyle farklı zorluklar içermekte olan haber öneri sistemleri üzerinde tasarlanmıştır. Literatürdeki haber öneri sistemleri haber kaynaklarını dikkate almamakta, her haber makalesine eşit şekilde davranmaktadır. Ancak medya kuruluşları tarafından yapılan araştırmalar göstermektedir ki haber kaynaklarının çeşitliliği oldukça önemlidir. Kişilerin haber kaynağı tercih ve beğenilerinin göz önüne alınması gerekmektedir. Yapılan literatür taramasında haber kaynaklarını göz önüne alan bir haber öneri sistemine rastlanmamıştır. Bu tez kapsamında bir haber kaynağı öneri sistemi sunulmuştur. Ontolojilerin yanı sıra kullanıcıların okudukları haber kaynaklarına ilişkin veriler kullanılarak birliktelik kuralları çıkarılmış ve bu iki yöntem karşılaştırılmıştır.

Anahtar Sözcükler: Öneri sistemleri, Haber öneri sistemi, Haber kaynağı öneri sistemi, Anlamsal çıkarsama, Ontolojiler.

ABSTRACT

DESIGNING A RECOMMENDER SYSTEM BY USING ONTOLOGIES AND SEMANTIC REASONING

ÖZGÖBEK, Özlem

PhD. in Computer Engineering

Supervisor: Assoc. Prof. R. Cenk ERDUR

April 2015, 77 pages

Recommender systems research is becoming one of the popular research fields. Unlike search engines recommender systems are the systems which recommends the most suitable items to the users without any manual effort of the user. Recommender systems have really wide application areas like shopping, movies, music, book and news. Since every application area has its own challenges, it is really hard to develop a domain independent recommender system. There are many techniques for developing the recommender systems. Using the similarities between users or content or both is a common way to make recommendations.

In this thesis, a recommender system using ontologies and semantic reasoning is proposed. The proposed system is developed in the news domain which has many additional challenges compared to other domains because of its dynamic nature. The existing news recommender systems do not take into account the news sources where the news articles come from. According to the research done by the media compaines, for a wider view of the news it is important to get the news articles from different sources. So in the news recommendation procedure it is important that the news sources are taken into account. In this thesis, a news source recommender system is proposed. Also the association rules are extracted by using the users' reading patterns. As a result, the relations between news sources gathered from this method are compared with the results from ontologies.

Keywords: Recommender systems, News recommender system, News resource recommender system, Semantic reasoning, Ontologies.

TEŞEKKÜR

Her konuda hiçbir zaman desteklerini esirgemeyen annem Nermin Özgöbek ve babam Hikmet Özgöbek başta olmak üzere, bu tezin ortaya çıkmasında çok değerli katkıları olan danışmanım Doç. Dr. R. Cenk Erdur, Prof. Dr. Jon Atle Gulla, Prof. Dr. Oğuz Dikenelli, Arne Dag Fidjestøl ve Dr. Jon Espen Ingvaldsen'e, yurtdışında geçirdiğim süre boyunca verdikleri desteklerden ötürü Ege Üniversitesi Bilgisayar Mühendisliği Bölümü bölüm başkanı Prof. Dr. Serdar Korukoğlu ve bölümdeki diğer hocalarıma, tez çalışmalarım süresince fikir alışverişinde bulunduğum ve çeşitli resmi işlemler konusunda yardımlarını esirgemeyen Birol Çiloğlugil ve Oylum Alatlı'ya, tez yazım süreci boyunca desteğini hep hissettiğim Sindre Anderssen'e ve bu süreçte bana her konuda destek olan Türkiye ve Norveç'teki tüm arkadaşlarıma çok teşekkür ederim.

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	vii
ABSTRACT	ix
TEŞEKKÜR	xi
ŞEKİLLER DİZİNİ	xviii
ÇİZELGELER DİZİNİ	xix
SİMGELER VE KISALTMALAR DİZİNİ	xxi
1 GİRİŞ	1
1.1 Tezin Literatüre Katkısı	2
1.2 Teze Genel Bakış	3
2 ÖNERİ SİSTEMLERİ	5
2.1 Öneri Yöntemleri	5
2.1.1 İçerik tabanlı filtreleme	6
2.1.2 İşbirlikçi filtreleme	7
2.1.3 Hibrid yaklaşım	8
2.1.4 Anlamsal yaklaşım	8
2.1.5 Birliktelik kuralları	9
2.2 Kullanıcı Geribildirimleri	12
2.2.1 Açık geribildirim	12

İÇİNDEKİLER (Devam)

	<u>Sayfa</u>
2.2.2 Örtük geribildirim.....	13
2.3 Öneri Sistemlerinin Zorlukları	13
2.3.1 Soğuk başlangıç sorunu	13
2.3.2 Seyrek veri sorunu	14
2.3.3 Anındalık sorunu.....	14
2.3.4 Örtük kullanıcı geribildirimi sorunu.....	14
2.3.5 Kullanıcı beğeni ve ilgi alanlarının değişmesi	15
2.3.6 Ölçeklenebilirlik	15
2.3.7 Düzensiz içerik	16
2.3.8 Kullanıcı modelleme ve profilleme.....	16
2.3.9 Gri koyun sorunu.....	16
2.3.10 Rastlantısallık sorunu.....	16
2.3.11 Kullanıcı gizliliği sorunu.....	17
2.3.12 Komşu geçişkenliği sorunu	17
2.3.13 Eş anlamlılık sorunu.....	17
2.4 Veri Setleri	18
2.4.1 Haber öneri sistemlerinde veri setleri	18
2.4.2 Öneri sistemi veri setlerinin karşılaştırılması	19

İÇİNDEKİLER (Devam)

	<u>Sayfa</u>
3 HABER ÖNERİ SİSTEMLERİ.....	23
3.1 Haber Öneri Sistemi Zorlukları ve Literatürdeki Çözüm Önerileri	24
3.1.1 Soğuk başlangıç sorunu	24
3.1.2 Anındalık	26
3.1.3 Örtük geribildirim.....	27
3.1.4 Kullanıcı tercihlerinin değişmesi	27
3.1.5 Ölçeklenebilirlik	28
3.1.6 Veri seyrekliği	29
3.2 Anlamsal Haber Öneri Sistemleri.....	30
4 ANLAMSAL HABER KAYNAĞI ÖNERİ SİSTEMİ.....	33
4.1 Kullanılan Araçlar	35
4.1.1 Protege.....	35
4.1.2 Termine.....	35
4.1.3 Orange.....	36
4.1.4 Veri seti	36
4.2 Anlamsal Haber Kaynağı Öneri Sistemi Detayları	37
4.2.1 Birliktelik kuralları.....	40
4.2.2 Haber kaynağı ontolojisi	42

İÇİNDEKİLER (Devam)

	<u>Sayfa</u>
4.2.3 Ontoloji ve birliktelik kuralları sonuçlarının karşılaştırılması	53
4.3 Durum Çalışması.....	56
4.4 SmartMedia Projesi.....	58
5 SONUÇLAR.....	61
5.1 Kısıtlamalar	62
5.2 Tartışmalar	63
5.2.1 Örtüşmeyen kurallar	63
5.2.2 Çoklu birliktelik kuralları	64
5.2.3 Duyarlılık - anma hesaplamaları	65
5.3 Edinimler.....	66
5.4 Katkıların Özeti	68
5.5 Gelecek Çalışmalar	68
KAYNAKLAR DİZİNİ	71
ÖZGEÇMİŞ	77
EKLER.....	

ŞEKİLLER DİZİNİ

<u>Şekil</u>	<u>Sayfa</u>
2.1 Öğeler arasındaki benzerlik oranı (1-10 arasında).....	6
2.2 Kullanıcının öğeleri beğenme oranı.....	6
2.3 İşbirlikçi fitreleme.....	7
2.4 Farklı öneri sistemi veri setlerinin karşılaştırılması ve özellikleri.....	22
4.1 Bir kullanıcı ve farklı haber kaynakları arasındaki ilişki.....	34
4.2 YOW veri setinin ilk bölümü detaylı kullanıcı bilgilerini içeriyor.	38
4.3 YOW veri setinin ikinci bölümü haber makalelerinin bilgilerini içeriyor. ..	39
4.4 YOW veri seti kullanılarak elde edilen Orange girdi dosyası örneği.	41
4.5 Orange arayüzü kullanılarak birliktelik kurallarının çıkarılması.	42
4.6 Orange arayüzünde birliktelik kurallarının görüntülenmesi.	42
4.7 Orange arayüzünde birliktelik kurallarının filtrelenmesi.....	43
4.8 Termine web arayüzü kullanılarak elde edilmiş analiz sonucu örneği.	44
4.9 Haber kaynağı ontolojisi sınıf hiyerarşisi örneği.....	45
4.10 Ontolojideki “Actor” (Oyuncu) sınıfının içerdiği haber kaynakları.....	46
4.11 Ontolojideki “Software” (Yazılım) sınıfındaki bir öğeyi içeren haber kaynakları.....	49
4.12 X ve Y haber kaynaklarındaki öğelerin küme diyagramı.	49
4.13 Protege’de “Software” (Yazılım) sınıfı için bir DL sorgu örneği ve sonucu.	50

ŞEKİLLER DİZİNİ (Devam)

Şekil	Sayfa
4.14 Protege’de “Technology” (Teknoloji) sınıfı için bir DL sorgu örneği ve sonucu.	51
4.15 Ontolojideki “Software” (Yazılım) sınıfının içerdiği haber kaynakları.	51
4.16 Ontolojideki sınıflardan bazılarının sahip oldukları RSS_ID sayıları.	51
4.17 Ontolojideki kategorilerden bazılarının sahip oldukları RSS_ID sayıları grafiği.	52
4.18 8172 ve 8155 RSS_ID’lerine sahip haber kaynakları arasındaki birliktelik kuralı.	54
4.19 8156 ve 8154 RSS_ID’lerine sahip haber kaynakları arasındaki birliktelik kuralı.	54
4.20 8153 ve 8195 RSS_ID’lerine sahip haber kaynakları arasındaki birliktelik kuralı.	54
4.21 8156 ve 8195 RSS_ID’lerine sahip haber kaynakları arasındaki birliktelik kuralı.	55
4.22 8172 ve 8152 RSS_ID’lerine sahip haber kaynakları arasındaki birliktelik kuralı.	55
4.23 Durum çalışması için seçilen kullanıcılar ve çeşitli kaynaklardan okuma bilgileri.	57
4.24 SmartMedia Projesi kapsamında geliştirilen haber öneri sistemi mimarisi.	59
5.1 Çoklu birliktelik kurallarına örnek - 1.	64
5.2 Çoklu birliktelik kurallarına bir örnek - 2.	64

ÇİZELGELER DİZİNİ

<u>Çizelge</u>	<u>Sayfa</u>
2.1 Birliktelik kuralları analizinde kullanılan örnek öge listesi.	9
2.2 İkili ifade sistemi ile öge çizelgesi.	10
4.1 RSS_ID'lerin ontolojide yer alma sıklığı.....	52
4.2 Ontolojiden elde edilen haber kaynakları arasındaki ilişkiler.....	53
4.3 Birliktelik kurallarından elde edilen tüm tekli kurallar.	55
4.4 Ontoloji ile tutarlı olan birliktelik kuralları.	55
4.5 Birliktelik kuralları ve ontoloji kullanılarak kullanıcılara tavsiye edilen haber kaynakları.	58
5.1 Tez çalışması kapsamındaki edinimlerin özeti.....	67

SİMGELER VE KISALTMALAR DİZİNİ

Kısaltma	Açılım
API	Application Programming Interface
CF-IDF	Concept Frequency - Inverse Document Frequency
OWL	Web Ontology Language
PLSI	Probabilistic Latent Semantic Indexing
RDF	Resource Description Framework
RSS	Really Simple Syndication
SF-IDF	Synset Frequency - Inverse Document Frequency
SOAP	Simple Object Access Protocol
SS	Semantic Similarity
TF-IDF	Term Frequency - Inverse Document Frequency
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
XML	eXtensible Markup Language

1 GİRİŞ

İnternette gün geçtikçe artan veri miktarı kullanıcıların gerçekten aradıkları bilgiye ulaşmalarını zorlaştırmaktadır. Her ne kadar arama motorları ve RSS okuyucuları internette aranan ya da ilgi çeken bilgiye ulaşılmasını kolaylaştırırsa da gerçekten aranan bilgiyi bulmak hala zor olabilmektedir. Kullanıcının aklındaki kavramı kelimelere dökmesi, doğru anahtar kelimeleri bulması her zaman kolay değildir. Kimi zaman ise aslında kullanıcılar bile ne aradıklarından emin değildirler. İlgi çeken bir şeyler bulmaya çalışır, ama nerede nasıl arayacaklarını bilemezler. Bildikleri, o an akıllarına gelen şeyleri arar ve yeni bir bilgiye, farklı sitelere bağlantılar bulmaya çalışırlar. Ancak bu yöntem çoğunlukla tesadüflere, kullanıcının o anki dikkatine ve bazı şeyleri gözden kaçırmamasına dayalıdır. Üstelik çok fazla zaman tüketmektedir. Sonuçta ise kullanıcının kendisine en uygun sonucu bulamama ihtimali yüksektir.

Öneri sistemleri, internette kullanıcılara en uygun olan bilgiyi bulmalarında kolaylık sağlaması amacı ile geliştirilmiştir. Arama motorlarının aksine öneri sistemleri, herhangi bir kullanıcı çabasına gerek duymadan, kullanıcıya en uygun bilgiyi otomatik olarak getirmeyi hedefler. Öneri sistemlerinin kişilerin tercih ve özelliklerine uygun olarak en doğru önerilerde bulunması beklenmektedir. Ancak bu görüldüğü kadar kolay değildir. Hem kişisel özelliklerin doğru bir şekilde öğrenilmesi, hem anlık durum ve duygu değişikliklerinin göz önüne alınması, hem de kullanıcı gizliliğinin korunması gerekmektedir. Diğer yandan, bir öneri sisteminin ne derece başarılı olduğu konusundaki ölçütler de tartışmalıdır. Anlık kullanıcı memnuniyetine dayanan bir ölçüt sistemi yeterince somut değildir. Diğer ölçütlerin ise pratikte ne kadar geçerli olduğu tartışmalıdır. Bir öneri sisteminden beklenen bu özelliklerin her birisi kendi içinde oldukça geniş araştırma alanına sahiptir. Öneri sistemleri, içerdikleri çok farklı problemlerle üzerinde durulması gereken bir alandır.

Öneri sistemleri uygulandıkları alana göre farklılıklar göstermektedir. Her alanda çözülmesi gereken farklı sorunlar bulunmakta ve uygulanan çözüm yöntemleri farklı alanlarda farklı sonuçlar verebilmektedir. Bu sebeple tek bir öneri sisteminin tüm alanlar için çalışması oldukça zordur. Farklı alanlar için farklı öneri sistemleri geliştirilmekte ve her alanın kendine özgü zorlukları çözülmeye çalışılmaktadır.

1.1 Tezin Literatüre Katkısı

Literatürde üzerinde çokça çalışılan film, müzik ve alışveriş öneri sistemlerinden farklı olarak haber öneri sistemleri daha farklı zorluklar içermektedir. Özellikle haber alanının oldukça değişken olması bu alanda karşılaşılan en büyük zorluklardandır. Kullanıcının beğeneceği tahmin edilen haberlerin seçilmesi ve bunun oldukça hızlı bir şekilde yapılması gerekmektedir. Haber alanı metin ağırlıklı bir alan olduğundan metin analizi, ontolojiler ve anlamsal çıkarsama yöntemleri haber öneri sistemlerinde kullanılmaktadır. Aynı zamanda kullanıcılar arasındaki benzerlikleri dikkate alan ve bu benzerliklere dayanarak önerilerde bulunan sistemler de bulunmaktadır. Ancak literatürdeki çalışmaların tamamında farklı kaynaklardan gelen haber metinleri ayrı birer öge olarak ele alınmakta, haber metninin kaynağı önemslenmemektedir. Bu durumun birkaç olumsuz etkisi vardır. İlk olarak aynı haberin birden fazla kere aynı kullanıcıya önerilmesi sorununa yol açmaktadır. Bir haber farklı kaynaklarda farklı anlatımlarla yer alabileceğinden öneri sisteminin metinler arasındaki gerçek farkı anlayabilmesi gerekmektedir. İki haber metninin aynı haberin farklı anlatımları mı, çok benzer bir konuda mı yoksa aynı konunun devamı mı olduğunu ayırt etmek gereklidir. Bu sorun literatürde rastlantısallık sorunu olarak geçmektedir ve tam olarak çözümü oldukça güçtür (Bölüm 2.3.10). Bir diğer olumsuz etki ise var olan haber öneri sistemlerinin kullanıcının takip etmek istediği ya da isteyebileceği haber kaynaklarını dikkate almamalarıdır. Bu hem kullanıcının beğenebileceği diğer kaynakları keşfetmesini engellemekte, hem de aynı haberin kullanıcının daha çok beğeneceği bir anlatımı varken daha az beğenilen metni önererek öneri kalitesini düşürmektedir. Bu tez kapsamında literatürde bu konuda herhangi bir çalışmaya rastlanmayan haber kaynağı öneri kavramı sunulmuştur. Bu sayede hem öneri kalitesinin yükseltilmesi, hem de bazı öneri sistemi zorluklarına daha iyi çözümler getirilmesi amaçlanmıştır.

Bu tez kapsamında sunulan haber kaynağı öneri sistemi için anlamsal yaklaşım içeren bir mimari önerilmektedir. Bunun için örnek bir haber kaynağı ontolojisi oluşturulmuştur. Aynı zamanda veri analizi yöntemi ile haber kaynakları arasındaki ilişkiler bulunmuş ve bu yöntem anlamsal yaklaşım ile karşılaştırılmıştır. Ayrıca öneri sistemlerinin zorlukları derlenmiştir. Sunulan yöntemlerin uygulanması ve karşılaştırılması için gerekli olan veri seti için, öneri sistemleri alanındaki veri setleri karşılaştırılmış ve haber öneri sistemi alanındaki bir veri seti seçilmiştir.

1.2 Teze Genel Bakış

Bu tez řu bölümlerden oluşmaktadır: 2. Bölüm’de genel olarak öneri sistemi kavramı ele alınmış ve öneri sistemleri her yönüyle alt başlıklarda incelenmiştir. Bölüm 2.1’de öneri yöntemleri detaylı olarak incelenmiştir. Farklı filtreleme yöntemleri ve anlamsal öneri yöntemi hakkında bilgi verilmiştir. Bölüm 2.2’de kullanıcı geribildirimlerinin önemi tartışılmış ve geribildirim çeşitleri alt başlıklarda açıklanmıştır. Bölüm 2.3’te öneri sistemlerinin zorlukları açıklanmıştır. Bu zorluklar literatürde karşılaşılan tüm öneri sistemi alanlarından derlenmiştir ve bazıları alan bağımlı bazıları ise alan bağımsızdır. Bölüm 2.4’te veri setlerinin önem ve gerekliliğı tartışılmış, öneri sistemlerinin test edilmesi ve karşılaştırılmasında kullanılan veri setleri incelenmiştir. Bölüm 3’te haber öneri sistemleri ile ele alınmıştır. Bölüm 3.1’de haber öneri sistemlerine yönelik zorluklar ve literatürdeki çözüm önerileri açıklanmış, Bölüm 3.2’de ise anlamsal haber öneri sistemlerinin literatürdeki örneklerine yer verilmiştir. Bölüm 4’te bu tez kapsamında önerilen anlamsal haber kaynağı öneri sistemi tartışılmış, sunulan yöntemin detayları, kullanılan araçlar, yöntemlerin karşılaştırılması ve durum çalışması alt başlıklarda incelenmiştir. Bölüm 5’te ise sonuçlar özetlenmiş, sonuçlarla ilgili tartışmalar yapılmış ve gelecek çalışmalar sunulmuştur.

2 ÖNERİ SİSTEMLERİ

Öneri sistemlerinin kökenleri bilgi elde etme çalışmalarına dayanmaktadır. Ancak 1990'ların ortasından itibaren öneri sistemleri ayrı bir araştırma dalı haline gelmiştir (Adomavicius and Tuzhilin, 2005). Öneri sistemleri kullanıcının herhangi bir çabası olmadan, kişinin tercih ve özelliklerine göre uygun öğeler tavsiye eden sistemlerdir. Arama motorlarının aksine kullanıcı aradığını bulmaya çalışmamakta, sistem kullanıcının istediği öğeleri bulmaya çalışmaktadır. Diğer yandan kullanıcılar her zaman ne aradıklarından emin değildirler. Örneğin; canı değişik bir film izlemek isteyen bir kullanıcı bir film kiralama servisindeki binlerce filmden birisini seçmeye çalıştığında aslında bilmediği bir şeyi aramaktadır. Bu durumda yapılacak şey seçenekleri film türlerine göre, yönetmenlere ya da oyunculara göre elemektir. Ancak bu durum kullanıcının daha önce duymadığı ama aslında hoşuna gidebilecek filmleri bulma şansını düşürecektir. Öneri sistemleri kullanıcılara daha önce bilmedikleri öğeleri de sunarak, kullanıcıların yeni şeyler keşfetme şansını arttırmaktadır.

Öneri sistemleri pek çok alana uygulanabilir sistemlerdir. Alışveriş ürünleri, film ve müzik önerileri üzerinde en çok çalışılan konular olmuştur. Pek çok araştırmaya göre alışveriş ürünlerinde yapılan kişiselleştirilmiş öneriler satışları arttırmaktadır. Günümüzde özellikle alışveriş ve haber sitelerinde sıkça rastlanan, kişiselleştirilmiş öneriler olmasa bile en yeni ya da en popüler öğelerin kullanıcılara sunulması bile etkili bir yöntemdir. Ancak öneri sistemlerinde öneriler çoğunlukla kişiye özel olmaktadır. Öneri sistemlerinin bir diğer araştırma dalı ise grup öneri sistemleridir. Grup önerilerinde bireysel farklılıklar yerine bir grup kullanıcının ortak özellikleri dikkate alınmaktadır. Bu araştırma alanı kullanıcıların nasıl gruplanacağı, grupların kendine özgü özellikleri, gruptaki birey sayısı ve bireylerin kişisel farklılıkları gibi oldukça değişik zorluklar barındırmaktadır.

2.1 Öneri Yöntemleri

Öneri sistemlerinin geliştirilmesi için kullanılan farklı yöntemler bulunmaktadır. Kimi öneri sistemleri önerilerde bulunmak için kullanıcılar arasındaki benzerlikleri göz önüne alırken, diğerleri içerik analizi ve içerik benzerliklerini kullanmaktadır. Literatürde bu yöntemler genellikle işbirlikçi filtreleme, içerik tabanlı filtreleme ve hibrid olarak üçe ayrılmaktadır. Ancak öneri yöntemlerini daha farklı sınıflandırmak da mümkündür. Örneğin; (Burke, 2002)'de öneri teknikleri beş farklı grupta incelenmiştir. Bu beş grup işbirlikçi filtreleme, içerik tabanlı filtreleme, demografik, yardımcı tabanlı ve bilgi tabanlı

	Öğ 1	Öğ 2	Öğ 3	Öğ 4	Öğ 5	
Öğ 1	10	3	1	8	9	
Öğ 2	3	10	8	1	2	
Öğ 3	1	8	10	2	3	
Öğ 4	8	1	2	10	9	
Öğ 5	9	2	3	9	10	
.....						

Şekil 2.1: Öğeler arasındaki benzerlik oranı (1-10 arasında).

	Beğeni (1-5)
Öğ 1	5
Öğ 2	3
Öğ 3	?
Öğ 4	4
Öğ 5	?
.....	

Şekil 2.2: Kullanıcının öğeleri beğenme oranı.

olarak adlandırılmaktadır. Öneri tekniklerinin bu sınıflandırması çevrimiçi kullanıcı etkileşiminden elde edilen veri veya kullanıcı geribildirimi, öneri için kullanılan algoritma ve sistemde barındırılan arka plan verileri üzerine kurulmuştur.

Öneri tekniklerinin bahsedilen ilk sınıflandırması daha sık kullanıldığından bu kategoriler aşağıda daha yakından incelenmiştir. Ayrıca literatürde öneri sistemlerine anlamsal yaklaşım bulunmaktadır (Cantador et al., 2008). Bu yaklaşım, bu bölümde incelenecek olan öneri sistemleri yöntemlerinden farklı olup, bir başka bölümde detaylı olarak ele alınacaktır. Bu tez çalışması kapsamında yer verilen bir başka yöntem olan birliktelik kuralları, yukarıda söz edilen temel sınıflandırmada doğrudan yer almamaktadır. Ancak bu yöntemin detayları ve öneri sistemlerinde nasıl kullanıldığına dair literatürdeki çalışmalara da aşağıda yer verilmiştir.

2.1.1 İçerik tabanlı filtreleme

İçerik tabanlı öneri sistemlerinde bir öneride bulunmak için öğelerin özellikleri kullanılmaktadır. Bu tür sistemlerde kullanıcının geçmişte tercih ettiği öğelerle ortak özellikleri bulunan yeni öğeler kullanıcıya tavsiye edilir. Öğeler arasındaki benzerliklerin bulunması bu yöntem için oldukça önemlidir.

Örnek vermek gerekirse, içerik tabanlı filtreleme yöntemi ile kullanıcıya

Kullanıcı A		Kullanıcı B	
	Beğeni (1-5)		Beğeni (1-5)
Öğ 1	3	Öğ 1	
Öğ 2	5	Öğ 2	5
Öğ 3	1	Öğ 3	4
Öğ 4	5	Öğ 4	5
Öğ 5	4	Öğ 5	
....			

Şekil 2.3: İşbirlikçi filtreleme.

bir film önermek için kullanıcının geçmişte tercih ettiği filmleri ve bu filmlerin kullanıcının henüz izlemediği filmlerle olan benzerliklerini bilmemiz gerekir. Şekil 2.1’de 1-10 arasında bir derece ile öğeler arasındaki benzerliklere örnek verilmiştir. Buna göre Şekil 2.2’deki kullanıcının öğeleri beğeni oranı dikkate alınarak bu kullanıcıya önerilerde bulunulacaktır. (Adomavicius and Tuzhilin, 2005)’de detaylı olarak tartışıldığı gibi, bu yöntemin kökleri bilgi elde etme (information retrieval) ve bilgi filtreleme (information filtering) sistemlerine dayanmaktadır. Bazen bilgi filtreleme içerik tabanlı filtreleme anlamında kullanılmaktadır (Lee and Park, 2007).

2.1.2 İşbirlikçi filtreleme

Bu yöntemde öneriler kullanıcın diğer kişilerle olan benzer tercihleri kullanılarak yapılmaktadır. Örneğin; kullanıcının geçmişte izlediği ve beğendiği filmleri izleyip beğenen başka bir kullanıcı varsa bu iki kullanıcı arasında bağlantı kurulur, birisinin izleyip beğendiği filmler diğerine de önerilir. Şekil 2.3’de A ve B kullanıcıları arasındaki benzerlikler gösterilmiştir. İşbirlikçi filtreleme yöntemi kullanıldığında Öğ 5, B kullanıcısına önerilebilecektir.

İşbirlikçi filtreleme yöntemi kendi içinde üçe ayrılır: Hafıza tabanlı, model tabanlı ve hibrid. Hafıza tabanlı yöntem bazı kaynaklarda komşu tabanlı, kullanıcı tabanlı ya da buluşsal tabanlı olarak da kullanılmaktadır. Bu yöntemde kullanıcıdan alınan oylar, kullanıcılar ve öğeler arasında benzerliklerin hesaplanmasında kullanılır. İstatistiksel yöntemler kullanılarak hedeflenen kullanıcıya benzeyen diğer kullanıcılar bulunur. Benzer kullanıcılar bulunduktan sonra ise çeşitli algoritmalar kullanılarak öneriler yapılır. Model tabanlı (öğ 5 tabanlı da denmektedir) yöntemde her kullanıcı için bir model oluşturulur. Bu modellerin oluşturulmasında çeşitli veri madenciliği ve makine öğrenmesi algoritmaları kullanılmaktadır. Öneriler istatistiksel yöntemler kullanılarak yapılır. Bayes ağı ve kümeleme gibi yöntemler model tabanlı öneri sistemleri içinde yer almaktadır (Sarwar et al., 2001). İşbirlikçi

filtreleme yöntemi içindeki son yöntem olan hibrid yöntemler, model tabanlı ve hafıza tabanlı yöntemlerin birlikte kullanılması ile oluşturulur.

2.1.3 Hibrid yaklaşım

Bu yaklaşımda, içerik tabanlı ve işbirlikçi filtreleme yöntemlerinin birlikte kullanılır. Bu yaklaşımın amacı, tek bir yöntemin sahip olduğu dezavantajlardan mümkün olduğunca kurtulmak ve yöntemlerin avantajlarını birleştirmektir. İçerik tabanlı ve işbirlikçi filtreleme yöntemleri bir arada farklı şekillerde kullanılabilir. (Burke, 2002)'de hibridleştirme yöntemleri yedi ayrı grupta incelenmiş ve farklı yöntemlerin nasıl bir araya getirilebileceği detaylı olarak açıklanmıştır.

2.1.4 Anlamsal yaklaşım

Anlamsal web W3C tarafından öncülük edilen, web üzerindeki yapısal olmayan verinin yapısal hale getirilerek, birbiri ile ilişkili bir veri ağı kurulmasını hedefleyen bir harekettir. Buna göre web üzerindeki tüm veri makineler tarafından anlaşılabilir bir şekilde ifade edilmelidir. Bunun için W3C tarafından belirlenmiş çeşitli standartlar bulunmaktadır. RDF (Kaynak Tanımlama Çerçevesi / Resource Description Framework), XML (Genişletilebilir İşaretleme Dili / eXtensible Markup Language) ve OWL (Web Ontoloji Dili / Web Ontology Language) bu standartlar dahilinde geliştirilmiş, veri için özelleşmiş dillerdir. Bunlardan RDF çeşitli sözdizim modellerinde, örneğin özne-yüklem-nesne biçiminde, bilgiyi modellemek için kullanılır. XML de benzer şekilde veriyi hem insanlar tarafından okunabilir hem de makineler tarafından anlaşılabilir bilgiler olarak tanımlamaya yarayan bir dildir. OWL ise bilgiyi hiyerarşik olarak modellemeye yarayan ontolojileri oluşturmak için kullanılan bir dildir.

Ontolojiler bir konudaki veya alandaki kavramları hiyerarşik olarak modellemeye ve bu kavramlar arasındaki ilişkileri tanımlamaya yarar. Bu tanımlamalar ve ilişkilendirmeler sayesinde bir kavramın anlamsal olarak ne ifade ettiği anlaşılabilir. Anlamsal çıkarsama ontolojide tanımlanmış olan ve açıkça görülen kavramsal ilişkiler haricinde, saklı kalmış ilişkileri bulmak için kullanılır.

Anlamsal yaklaşım öneri sistemlerinde kullanılan yöntemlerden bir tanesidir. Anlamsal yaklaşımın öneri sistemlerinde kullanılmasının temel sebeplerinden bir tanesi içeriğin kültürel ve dilbilimsel arkaplan bilgisini kullanmayı mümkün kılmasıdır (Peis et al., 2008). Anlamsal yaklaşım, kelime tabanlı sistemlerle

ID	Öğeler
1	{Ekmek, Süt}
2	{Ekmek, Bebek bezi, Bira, Yumurta}
3	{Süt, Bebek bezi, Bira, Kola}
4	{Ekmek, Süt, Bebek bezi, Bira}
5	{Ekmek, Süt, Bebek bezi, Kola}

Çizelge 2.1: Birliktelik kuralları analizinde kullanılan örnek öğe listesi.

karşılaştırıldığında karmaşıklığı azalmakta, kavramların hiyerarşik olarak temsil edilmesine ve çıkarsama yapmaya olanak sağlamaktadır (Cantador and Castells, 2009).

2.1.5 Birliktelik kuralları

Birliktelik analizi çok büyük veritabanlarındaki veriler arasındaki tahmin edilemeyen ilişkileri bulmak için kullanılan bir veri madenciliği yöntemidir. Birliktelik analizi sonucunda öğeler arasındaki tahmin edilemeyen ilişkiler birliktelik kuralları şeklinde ifade edilir. Örneğin; bir marketin müşterilerinin tek bir seferde aldığı ürünlerin birliktelik kuralları ile analiz edilmesi oldukça çok kullanılan bir yöntemdir. Bu analiz sonucunda hangi ürünlerin birlikte daha çok satıldığına ilişkin bilgiler elde edilir ve marketin satış politikaları buna göre değiştirilir. Çizelge 2.1’de bu konuda oldukça sık verilen bir örnek görülmektedir. Çizelgedeki her bir satır herhangi bir müşterinin tek seferde satın aldığı öğelerden oluşmaktadır. Bu öğelerin analizi sonucu ortaya çıkan birliktelik kurallarından bazıları şöyledir:

$$Bebek\ bezi \rightarrow Bira$$

$$Süt, Ekmek \rightarrow Yumurta, Kola$$

$$Bira, Ekmek \rightarrow Süt$$

Bu kurallara göre bebek bezi ile bira arasında bir ilişki bulunmaktadır, bebek bezi alan müşterilerin aynı zamanda bira da almaları olasıdır.

Birliktelik kurallarının temel prensiplerini açıklamak için Çizelge 2.1’deki

ID	Ekmek	Süt	Bebek bezi	Bira	Yumurta	Kola
1	1	1	0	0	0	0
2	1	0	1	1	1	0
3	0	1	1	1	0	1
4	1	1	1	1	0	0
5	1	1	1	0	0	1

Çizelge 2.2: İkili ifade sistemi ile öge çizelgesi.

veriler Çizelge 2.2’de ikilik sistemde ifade edilmiştir. Ancak bu ifade şekli oldukça basittir ve satılan öge miktarı ya da alıcının o öge için ne kadar ödemeyi kabul ettiği gibi veriler yer almamaktadır. Bu verilerin de yer aldığı detaylı açıklamalar için, bu bölümün yazılmasında da kaynak olarak alınan (Tan and Kumar, 2005)’e bakılabilir.

Birliktelik kurallarının iki ana prensibi “öge kümesi sayısı” (itemset count) ve “destek sayısı” (support count)’dır. (Tan and Kumar, 2005)’te açıklandığı gibi, $I = \{i_1, i_2, \dots, i_d\}$ bir alışveriş sepeti verisindeki tüm öğelerin kümesi ve $T = \{t_1, t_2, \dots, t_N\}$ yapılan bütün alışverişlerin kümesi olsun. Her t_i alışverişi I kümesinden seçilen öğelerden oluşan bir altküme olacaktır. Bir öge kümesi k öğeden oluşuyorsa buna $k - \text{ögekümesi}$ denmektedir. Buna göre bir alışveriş üç öğeden oluşuyorsa, buna $3 - \text{ögekümesi}$ denmektedir. Destek sayısı ise belirli bir öge kümesini içeren alışverişlerin sayısı olarak tanımlanmaktadır. Buna göre, destek sayısı $\sigma(X)$ matematiksel olarak şöyle ifade edilmektedir:

$$\sigma(X) = \{ti | X \subseteq ti, ti \in T\} \quad (2.1)$$

Çizelge 2.1 ve 2.2’deki örneğe göre {Bebek bezi, bira, süt} öge kümesi, yapılan tüm alışverişlerde iki kez geçtiği için destek sayısı ikidir.

Birliktelik kuralları X ve Y ayrık öge kümeleri olmak üzere $X \rightarrow Y$ şeklinde ifade edilir ve X ile Y arasındaki ilişkiyi anlatır. Öğeler arasındaki ilişkinin ne kadar güçlü olduğunu anlamak için kullanılan iki sayı vardır. Bunlar “destek” (support) ve “güven” (confidence) sayılarıdır. Bu tez kapsamında çıkarılan birliktelik kurallarının analizinde de bu iki sayıdan faydalanılmıştır. Destek, bir kuralın verilen veri seti içinde ne sıklıkta bulunduğunu anlatır. Destek sayısı matematiksel olarak şöyle ifade edilir:

$$s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (2.2)$$

Çizelge 2.1’da verilen örneğe göre {Süt, Bebek bezi} $\rightarrow$ {Bira} kuralı ele

alındığında, destek sayısı $\frac{2}{5} = 0,4$ olmaktadır. Burada 2 rakamı {Süt, Bebek bezi} ikilisinin birlikte satın alındığı alışverişlerin sayısı, 5 rakamı ise toplam alışveriş sayısıdır.

Güven sayısı, kuraldaki Y ögesinin, X öge kümesi ile kaç kez birlikte geçtiğini anlatır. Güven sayısının matematiksel ifadesi şöyledir:

$$c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (2.3)$$

Bir önceki örnek olan {Süt, Bebek bezi} $\rightarrow$ {Bira} kuralı ele alındığında güven sayısı $\frac{2}{3} = 0,67$ olmaktadır. Çünkü Çizelge 2.1'e göre {Bira} ögesi, {Süt, Bebek bezi} öge kümesi ile birlikte iki kez, tek başına ise üç kez satın alınmıştır.

Destek ve güven sayıları birliktelik kuralları için oldukça önemlidir. Birliktelik kurallarının çıkarılmasında bu iki sayıya dikkat edilmelidir. Bir kuralda destek sayısı çok düşükse, bu o kuralın tesadüfen oluştuğu anlamına gelebilmektedir. Yukarıda ele alınan alışveriş örneğine göre eğer destek sayısı $\frac{1}{5} = 0,2$ olsaydı bu, {Süt, Bebek bezi} ikilisinin yalnızca bir kez birlikte alındığı anlamına gelecekti. Yapılan bu tek alışveriş ise bunun genelde böyle olduğunu söylemek için yeterli olmamaktadır. Güven sayısı ise bir kural tarafından yapılan çıkarsamanın güvenilirliği hakkında bir ölçüttür. Güven sayısı ne kadar yüksekse $X \rightarrow Y$ kuralı çerçevesinde Y 'nin X ile birlikte yer alma ihtimali o kadar yüksektir (Tan and Kumar, 2005).

Birliktelik kurallarının öneri sistemlerinde kullanılması konusunda literatürde yer alan çalışmalar bulunmaktadır. (Mobasher et al., 2001) tarafından birliktelik kurallarına dayalı bir kişiselleştirme çerçevesi önerilmiştir. Birliktelik kuralları daha güvenli öneri sistemleri geliştirilmesinde de kullanılmıştır. Örneğin (Sandvig et al., 2007)'de birliktelik kuralı tabanlı bir algoritma sunulmuştur. Bu algoritma öneri sistemlerinde profil enjeksiyon saldırılarını önlemek üzere geliştirilmiştir. (Lemdani et al., 2010) ise birliktelik kurallarını kullanarak bir işbirlikçi filtreleme yöntemi önermiştir. Benzer şekilde (Sun et al., 2005) nicel birliktelik kurallarını işbirlikçi filtreleme yönteminde kullanmış ve sonuçları geleneksel Pearson yöntemi ile karşılaştırmıştır. Pearson yöntemi iki değişken arasındaki ilişkinin gücünü ölçmeye yarayan bir yöntemdir.

2.2 Kullanıcı Geribildirimleri

Bir kişiye bir öneride bulunmak için öncelikle o kişinin beğenileri ve tercihleri hakkında bilgi sahibi olmak gerekir. Öneri sistemleri de kullanıcının daha önce incelediği, satın aldığı, izlediği vb. öğeler hakkındaki beğenilerini öğrenmeye çalışırlar. Bu sayede kullanıcıya ait bir profil oluşturulabilir ve kullanıcının hangi öğeleri beğenebileceği tahmin edilmeye çalışılır.

Öneri sistemini kullanan kişilerin kendilerine önerilen öğeler hakkındaki görüşlerini bir şekilde sisteme bildirmeleri gerekmektedir. Kullanıcı geribildirimleri, kullanıcının bilgisi dahilinde toplanan, öğeler hakkındaki beğeni ve tercihlerini anlamaya yönelik bilgilerdir. Öneri sisteminde kullanılan yöntemden bağımsız olarak kullanıcı geribildirimleri oldukça önemlidir. İçeriğin kullanıcılar arasındaki benzerliklerden daha öncelikli olarak ele alındığı içerik tabanlı sistemlerde bile kullanıcının geçmişte hangi öğeleri beğendiği bilgisi gereklidir. Kullanıcı geribildirimlerini açık ve örtük olmak üzere iki bölümde incelemek mümkündür.

2.2.1 Açık geribildirim

Açık geribildirimler kullanıcının el ile sisteme girdiği bildirimlerdir. Kullanıcı sistem tarafından sunulan öğe hakkında farklı şekillerde geribildirimde bulunabilir. En çok kullanılan açık geribildirim yöntemi kullanıcının öğeleri oylamasıdır. Oylama kullanıcı öğe hakkında belirli bir fikir edindikten sonra kullanıcıdan istenir. Oylamalar çoğunlukla 1 ile 5 arasındaki değerlerden oluşmakta ve kullanıcının öğeyi beğenip beğenmediği sorusuna yanıt istemektedir. Ancak daha farklı aralıklarda değerlerden oluşan ve kullanıcının birden fazla konudaki (içerik, kalite vb.) fikrini soran sistemler de bulunmaktadır.

Açık geribildirimler her zaman güvenilir olmamakla birlikte kimi zaman da kullanıcılar oy verme konusunda isteksizdir. Güvenilirlik sorunu kullanıcının bilinçli ya da bilinçsiz olarak yanlış oy vermesi anlamına gelebilmektedir. Örneğin; kullanıcılardan bir haber makalesine beğenileri doğrultusunda oy vermeleri istendiğinde, bazı kullanıcılar haber makalesinin kalitesini değerlendirmekte, bazıları ise haber içeriğini değerlendirmektedir. Bu olumsuzluklara rağmen Şekil2.4'te görülebileceği gibi açık geribildirimler öneri sistemlerinde en çok kullanılan geribildirim türüdür.

2.2.2 Örtük geribildirim

Örtük geribildirimler kullanıcının sistemi kullanımı sırasında toplanan arkaplan verileridir. Örneğin; kullanıcının sistemi ziyaret zamanı, hangi öğeye tıkladığı, sayfayı ne kadar süre görüntülediği gibi bilgiler örtük geribildirimdir. Bu tür geribildirimler kullanıcının herhangi bir müdahalesine gerek duymadan sistem tarafından otomatik olarak toplanmaktadır. Örtük geribildirimlerin çeşidi ve sayısı sistemden sisteme çok fazla değişiklik göstermektedir. Hangi verilerin toplanacağına ise sistem tasarımcıları karar vermektedir.

Örtük geribildirimler açık geribildirimlere göre daha güvenilirdir. Kullanıcının sistemi yanıltma şansı çok azdır. Diğer yandan bu tür geribildirimler kullanıcı gizliliği konusunda sorun oluşturabilmektedir. Örtük geribildirimler özellikle son yıllarda geliştirilen öneri sistemlerinde daha yaygın olarak görülmektedir.

2.3 Öneri Sistemlerinin Zorlukları

Bir öneri sistemi oluşturmak birçok kişi için ilk bakışta oldukça kolaydır. Kullanıcılara çeşitli öğeler önermek kulağa kolay bir işlem gibi gelmektedir. Ancak kişiye özel önerilecek olan uygun öğeyi bulmak, sadece kullanıcı hakkında değil, aynı zamanda öğeler ve genel bağlam hakkında da detaylı bilgi sahibi olmayı gerektiren karmaşık bir iştir. Kişisel tercihler ve ilgi alanları kullanıcının yaşına, kültür seviyesine, ait olduğu kültüre, cinsiyetine ve kişiliğine göre farklılık göstermektedir. Ayrıca ilgi alanları zaman içinde değişebilmektedir. Başarılı bir öneri sistemi, her biri ayrı bir araştırma alanı olabilecek birden fazla zorlukla başa çıkabilmelidir. Haber alanında ise öğelerin değişken özelliklerinden dolayı diğer alanlarla karşılaştırıldığında daha fazla zorluk bulunmakta veya bazı zorluklar daha fazla önem kazanmaktadır. Bu bölümde açıklanan zorluklarda tüm öneri sistemleri göz önüne alınmıştır. Bu sebeple söz edilen bazı zorluklar alan bağımlı olabilir ya da bazı alanlarda daha fazla öneme sahip olabilir.

2.3.1 Soğuk başlangıç sorunu

Soğuk başlangıç ya da ilk oylayıcı sorunu işbirlikçi filtreleme yöntemini kullanan öneri sistemlerinin en yaygın sorunlarından birisidir. İşbirlikçi filtreleme yönteminde öneriler diğer kullanıcıların öğeler üzerindeki oyları göz önüne alınarak yapıldığından, sisteme yeni bir öğe eklendiğinde daha önce hiçbir kullanıcı o öğeyi oylamadığı için öğenin tavsiye edilmesi imkansızdır. Benzer şekilde, sisteme

yeni bir kullanıcı eklendiğinde, o kullanıcıya ait herhangi bir oylama/beğeni verisi olmadığı için yeni kullanıcı ile sisteme daha önce kayıtlı kullanıcılar arasında bir benzerlik kurmak mümkün değildir. İçerik tabanlı filtreleme sistemlerinde ise kullanıcı ya da öğelerin oyları yerine öğelerin özellikleri göz önüne alındığından bu sorun söz konusu değildir.

2.3.2 Seyrek veri sorunu

Seyrek veri sorunu sistemde yeterince oy olmadığında işbirlikçi filtreleme yöntemi için kullanılan matrisler seyrek olurlar. Matrisin seyrek olmasının farklı sebepleri olabilir. Örneğin; kullanıcı sayısının az, öğe sayısının çok olması durumunda kullanıcıların tüm öğeleri oylaması mümkün değildir. Bu da kullanıcı – öğe matrisinin seyrek olmasına sebep olur. Benzer şekilde milyonlarca öğe ve kullanıcı barındıran sistemlerde seyrek veri sorunu belirgin hale gelmektedir. Matrisin seyrek olması sistemin performansının düşmesine sebep olmaktadır.

2.3.3 Anıdalık sorunu

Anıdalık haber öneri sistemlerinin en önemli zorluklarından bir tanesidir. Çoğu kullanıcı eski tarihli haberler yerine en yeni haberleri okumak ister. Yani bir haberin kullanıcılar için önemi zamanla azalmaktadır. Diğer yandan bazı haberler birbiri ile bağlantılı olabilir ve kullanıcı yeni bir haberle ilgili geçmiş makaleleri de okumak isteyebilir (Li et al., 2011). Bu yüzden eski makaleleri tamamen sistem dışı bırakmak ve sadece yeni makaleleri göz önüne almak iyi bir çözüm değildir.

Haber öneri alanında sisteme her saat yüzlerce yeni makale eklendiğini düşünürsek, anıdalık hem performans hem de öneri kalitesi açısından zorluk oluşturmaktadır.

2.3.4 Örtük kullanıcı geribildirim sorunu

Kullanıcı geribildirimleri daha geçerli ve uygun öneriler yapabilmek için oldukça önemlidir. Açık kullanıcı bildirimleri (oylama) sayesinde kullanıcının hangi öğeleri beğenip beğenmediğini ya da ne kadar beğendiğini anlamak mümkündür (Fortuna et al., 2010). Kimi zaman pek çok kullanıcı öğeler hakkında oy vermekten kaçınılmaktadırlar. Ayrıca açık geribildirimler için sistemin sürekli kullanıcı ile etkileşim halinde olması gerekmektedir. Bu özellikle mobil cihazlar üzerinde çalışan sistemlerde kullanıcı açısından zorluk yaratabilmektedir. Bu nedenlerden ötürü

kullanıcılardan daha geçerli geribildirimler alabilmek için örtük geribildirimlere ihtiyaç duyulmaktadır. Sistem kullanıcılardan etkin bir şekilde örtük geribildirimler alırken aynı zamanda da kullanıcı gizliliğini korumalıdır.

2.3.5 Kullanıcı beğeni ve ilgi alanlarının değişmesi

Öneri sistemlerindeki bir diğer önemli zorluk kullanıcıların beğeni ve ilgi alanlarının zamanla değişmesidir. Kullanıcının gelecekteki ilgi alanlarının tahminlenmesi gerekmektedir (Liu et al., 2010). Kitap, film ve müzik gibi bazı öneri sistemi alanları için bu nispeten kolaydır. Çünkü kullanıcıların bu alanlardaki beğenileri çok yavaş değişmektedir. Örneğin; bir kişinin çocukluğunda beğendiği bir filmi yetişkinliğinde beğenmemesi gibi. Benzer şekilde, bir kişinin beğendiği bir şarkıyı birkaç gün sonra beğenmemesi pek olası değildir. Haber alanında ise kişinin beğeni ve ilgi alanlarını takip etmek daha zordur. Birkaç gün önce kullanıcının ilgisini çeken bir haber o konuda daha sonra yaşanan gelişmeler nedeniyle kullanıcının ilgisini yitirebilir. Ya da kullanıcı bir haberi beğendiği için değil önemli bulduğu için okumak isteyebilir ancak bu, benzer diğer haberlerin de gelecekte kullanıcının ilgisini çekeceği veya kullanıcı için aynı önemi taşıyacağı anlamına gelmez.

2.3.6 Ölçeklenebilirlik

Öneri sistemleri pek çok kullanıcıya, hatta bazen milyonlarca kullanıcıya aynı anda hizmet vermeyi hedefler (Das et al., 2007). Ayrıca önerilebilecek olan öğelerin sayısı da çok yüksek olabilir. Gerçekten kullanışlı bir öneri sisteminin hızlı olması gerekmektedir. Milyonlarca kullanıcı ve öğenin bulunduğu bir sistemin, gerçek zamanlı önerilerde bulunması kolay değildir. Özellikle de anlık olarak sisteme eklenen ve sistemden çıkan öğelerin sayısı yüksekse verimli bir öneri sistemi inşa etmek daha da zorlaşmaktadır. Haber öneri sistemleri için durum böyledir. İnternetteki farklı haber kaynaklarında her saat onlarca yeni başlık bulmak mümkündür. Haber alanının bu dinamizminden dolayı, haber öneri sistemleri hızlı ve gerçek zamanlı işlem kapasitesine sahip olmalıdır (Li et al., 2011). Hangi öneri yönteminin kullanıldığından bağımsız olarak ölçeklenebilirlik öneri sistemlerinin en büyük zorluklarından bir tanesidir.

2.3.7 Düzensiz içerik

İçerik bilgisinin kullanılması gereken sistemlerde tüm içeriği analiz etmek zordur. Özellikle de içeriğin düzensiz ve yapısız olduğu durumlarda içeriğin sistem tarafından analiz edilip anlaşılması zorlaşmaktadır. Haber önerisi gibi metnin daha da önem kazandığı öneri alanlarında bu durum zorluk yaratmaktadır. Daha iyi haber önerileri için haber öğeleri yapısal ve makine tarafından okunabilir olmalıdır (Saranya.K.G and Sadhasivam, 2012).

2.3.8 Kullanıcı modelleme ve profilleme

Kullanıcı profilleme ya da kullanıcı tercihleri bilgisi edinme, öneri sistemlerinin önemli bir parçasıdır. Kişiselleştirilmiş öneriler için kullanıcının bir profilinin oluşturulması gerekmektedir. (Liu et al., 2010; Das et al., 2007; Saranya.K.G and Sadhasivam, 2012)'de belirtildiği gibi kullanıcı profili oluşturulması için çeşitli yöntemler vardır.

2.3.9 Gri koyun sorunu

Gri koyun sorunu işbirlikçi filtreleme yöntemi kullanıldığında ortaya çıkan sorunlardan bir tanesidir. İşbirlikçi filtreleme yönteminde kullanıcılar arasındaki benzerlikle göz önüne alınır ve bir kullanıcıya yapılan önerilerde o kişinin diğer kullanıcılarla olan ortak ilgi alanları kullanılır. Ancak bazı kullanıcılar hiçbir kullanıcı ya da kullanıcı grubu ile tutarlı ortak beğeni ve ilgi alanlarına sahip değildir. Bu durum gri koyun sorunu olarak adlandırılmaktadır (Su and Khoshgoftaar, 2009). Bu kişilere, diğer kullanıcılarla olan ortak beğeni ve ilgi alanları kullanılarak önerilerde bulunmak mümkün değildir. Sistemdeki toplam kullanıcı sayısı arttığında bu sorunun ortaya çıkma ihtimali azalmaktadır (Borges and Lorena, 2010).

2.3.10 Rastlantısallık sorunu

Rastlantısallık, aşırı uzmanlık ya da portföy sorunu olarak literatürde geçmektedir. Bu sorun sistemin daha önce önerilmiş öğelerle aynı veya çok benzer öğeleri tavsiye etmesidir. Örneğin; haber öneri sistemlerinde bir haber farklı kaynaklarda farklı şekillerde anlatılmış, farklı başlıklarla temsil edilmiş olabilir. Böyle bir durumda açıktır ki sistemin farklı kaynaklardan aynı haberi önermesi kullanıcıların hoşuna gitmeyecektir. Sistem aynı öğelerden kaçınarak her zaman

önerecek yeni öğeler keşfedebilmelidir. (Jaquinta et al., 2008)’da bu sorun içerik tabanlı filtreleme yöntemini kullanan sistemler için detaylı olarak tartışılmıştır. Ancak bu işbirlikçi filtreleme yöntemini kullanan sistemler için de zorluk oluşturmaktadır (Borges and Lorena, 2010).

2.3.11 Kullanıcı gizliliği sorunu

Bir kullanıcıya en uygun önerileri yapabilmek için sistem kullanıcının demografik özelliklerini, geçmiş tercihlerini, ilgi alanlarını ve hatta diğer insanlarla olan ilişkilerini bilmelidir. Bu kullanıcılar hakkında detaylı bilginin saklanması ve analiz edilmesini gerektirmektedir. Bu durum ise kullanıcı gizliliği konusunda sorunlara yol açabilir (Garcin et al., 2013). Öneri sistemlerinde kullanıcı gizliliği başlı başına bir çalışma konusudur. Ancak literatürde bu konuda henüz çok fazla çalışma bulunmamaktadır.

2.3.12 Komşu geçişkenliği sorunu

Veritabanı çok seyrek olduğunda ve kullanıcıların aynı öğeleri oylamadığı durumlarda komşu geçişkenliği durumundan söz edilebilir (Su and Khoshgoftaar, 2009). Bu durumda iki kullanıcı benzer tercih ve beğenilere sahip olsalar bile aynı öğeleri oylamadıkları için sistemin bunu bilmesi mümkün değildir. Komşu geçişkenliği sorunu gri koyun sorunu ile karıştırılmamalıdır. Gri koyun sorunu bir kullanıcının diğer herhangi bir kullanıcı ile tutarlı ortak beğenilerinin olmamasıdır. Sistemdeki kullanıcı sayısı arttıkça bu kullanıcıya benzer başka kullanıcılar bulma ihtimali artacaktır. Ancak komşu geçişkenliği sorunu öneri sisteminde kullanılan matrisin seyrek olması sonucu ortaya çıkar. Sorun benzer beğenilere sahip kullanıcıların olmaması değil, bu kullanıcıların oylarının farklı öğeler üzerinde kullanılmış olmasıdır.

2.3.13 Eş anlamlılık sorunu

Bazen aynı öğeler farklı kaynaklarda farklı isimlerle anılabilmektedir. Ancak makinelerin eş anlamlılığı algılaması oldukça güçtür. Buna bir örnek İngilizce “children’s movie” ile “children’s film” öğeleridir. Her iki öğe de “çocuk filmi” anlamına gelmesine rağmen farklı şekillerde ifade edilmiştir. Benzer şekilde, Türkçe “kupa” sözcüğü hem bir spor dalında kazanılan ödül, hem bir çeşit fincan, hem de kağıt oyunlarındaki bir simge anlamına gelmektedir. Bir öneri sistemi birbirinden çok farklı anlamlara sahip ya da aynı anlama gelen farklı öğeler arasındaki farkı

anlayabilmelidir.

2.4 Veri Setleri

Veri setleri geliştirilme aşamasındaki yeni sistemlerin eğitilmesi ve denenmesi amacıyla kullanılır. Bu nedenle veri setleri bilimsel araştırma ve geliştirmenin vazgeçilmez bir parçasıdır. Gerçek sistemler verilerle çalışırlar ve bir sistem pratik olarak kullanıcıların kullanımına açılmadan önce sistemin düzgün çalıştığının onaylanması çok önemlidir. Veri setleri olmadan yeni geliştirilen sistemlerin ne kadar doğru ve verimli çalıştığını anlamak mümkün değildir. Veri setinin kalitesi de sistemi test etmek için oldukça önemlidir.

Bilimsel yöntemler ölçümlerin sistematik kullanımını ve analizini temel almaktadır. (Dekker, 2006)'e göre veri setleri bilimsel araştırmada en az dört amaca hizmet ederler. 1. Yayınların onaylanması: Bilimsel yayınlar bir çalışmayı aynı veriyle tekrar ederek onaylanabilir. 2. Boylamsal (longitudinal) araştırma: Uzun süreli araştırmalar için veri setinin mevcut olması. 3. Verinin disiplinler arası kullanımı: Bir veri setinin farklı amaçlarla ve farklı disiplinlerde kullanılması yeni bilimsel bakış açılarına ve gelişmelere yol açabilir. 4. Veri seti sahipliği yeni projeler almaya yarayabilir. Ayrıca (Dekker, 2006)'e göre veri setleri akademik camiada kendi başına bir ürün olarak gittikçe değer kazanmaktadır.

2.4.1 Haber öneri sistemlerinde veri setleri

Veri setlerinin doğası hem uygulamanın çeşidine göre hem de uygulama alanına göre değişmektedir. Haber öneri sistemlerinde hem makine öğrenmesi teknikleri, hem de klasik arama teknolojileri kullanılabilir. Bütün bu yöntemler uygun veri setleri ile denenmelidir. Makine öğrenmesi teknikleri kullanılırken sistemin öncelikle eğitilmesi gerekmektedir. Kişisel farklılıkları da dikkate alan bu tür öneri sistemleri birçok parametreyi göz önüne almaktadır. Bu nedenle veri setinin özelliklerinin zengin olması beklenmektedir. Klasik arama teknolojileri ise daha dar bir alana odaklanmıştır ve kullanıcılar arasındaki farklılıkları göz önüne almazlar. Bu nedenle klasik arama yöntemi kullanılan öneri sistemlerinde veri setinin daha az özellikli olması yeterli olacaktır.

Haber alanının kendine has değişken ve tahmin edilemez özellikleri nedeniyle, haber öneri sistemlerini test etmekte kullanılan veri setlerini oluşturmak zordur. Oldukça hızlı değişen bir yapıya sahip olan haber alanı için sabit, kontrollü bir veri

seti kullanılmalıdır. Haber öneri sistemi veri setlerinin, kullanıcının özelliklerini ve tercihlerini gerçek haber bağlamında yansıtmaları beklenmektedir. Bu başarılması zor bir amaçtır ve belki de neden haber önerileri alanında sadece çok az küçük ölçekli veri seti olduğunu kısmen açıklamaktadır.

2.4.2 Öneri sistemi veri setlerinin karşılaştırılması

Bu bölümde literatürde sıkça yer alan, yeni geliştirilen öneri sistemlerinin denenmesi ve onaylanmasında kullanılan veri setleri açıklanmış ve karşılaştırılmıştır. Açıklama ve karşılaştırma için söz konusu veri setlerinin seçilme sebebi, öneri sistemleri araştırmalarında çoğunluk tarafından tanınması, kullanılması ve sıkça referans verilmesidir. Yukarıda söz edildiği gibi her öneri sistemi alanının (müzik, film, haber gibi) farklı özellikleri bulunmaktadır ve bu özelliklere uygun olan veri setlerinin kullanılması gerekmektedir. Bu bölümde karşılaştırılan öneri sistemi veri setleri seçilirken, veri setlerine daha geniş bir bakış açısı sağlaması ve haber öneri sistemlerinin veri seti özelliklerinin daha iyi anlaşılabilmesi amacıyla farklı alanlardan seçilmeye gayret edilmiştir. Bu bölümde yer alan tüm veri setleri açık olarak kullanıma sunulmuştur ve internetten ücretsiz olarak indirilebilmektedir. Şekil 2.4'te bu bölümde açıklanan tüm veri setlerinin özellikleri özet olarak görülebilmektedir.

- **MovieLens Veri Seti:** MovieLens Minnesota Üniversitesi GroupLens Araştırma Grubu (GroupLens, 2010) tarafından yürütülen bir film öneri sistemi projesidir. Bu proje kapsamında farklı zaman aralıklarında toplanmış üç adet veri seti yayınlanmıştır. Bu veri setleri ve içerdikleri kayıt sayıları Şekil 2.4'da görülebilir. Yayınlanan bu veri setleri için veriler MovieLens web sitesi üzerinden toplanmıştır. Buna ek olarak 100 000 ve 1 milyon boyutlu veri setleri kullanıcılara ait yaş, cinsiyet, meslek, adres gibi temel demografik bilgi içermektedir. 10 milyon boyutlu veri seti ise herhangi bir demografik bilgi içermemektedir.
- **Netflix Veri Seti:** Netflix(Netflix, 2009) dünya çapında hizmet veren bir çevrimiçi film kiralama servisidir. Netflix, 2 Ekim 2006 tarihinde film öneri sistemlerini geliştirmek amacıyla 1 milyon dolar ödülü olan bir yarışma başlatmıştır. Yarışmacıların geliştirdikleri sistemi test etmeleri için bir veri seti sağlanmıştır. Bu veri seti halka açık olarak sunulmuştur ve daha sonrasında da pek çok araştırmacı tarafından kullanılmıştır. Bu veri setinde 480.189 kullanıcı tarafından verilmiş 100.480.507 film oyu bulunmaktadır. Veri seti Netflix abonelerinden Aralık 1999 ve Aralık 2005 tarihleri arasında toplanmıştır. Büyük ödülü 2009 yılında, önerilerde %10,06 gelişme sağlayan "BellKor's

Pragmatic Chaos” adlı grup kazanmıştır. Kazanan algoritmanın detayları (Chaos, 2009)’da bulunabilir.

- MoviePilot Veri Seti:** MoviePilot veri seti 2011 yılında ACM RecSys konferansı kapsamında düzenlenen bağlam farkında film öneri sistemi yarışması (Context-Aware Movie Recommendation Challenge) (Said et al., 2011) sırasında yayınlanmıştır. Bu yarışma iki bölüm şeklinde düzenlenmiş ve her bölüm için veri setinin farklı bir bölümü kullanılmıştır. İlk bölümde belirli bir anda öneride bulunulacak uygun kullanıcıların seçilmesi, ikinci bölümde ise önerilerin yapılması hedeflenmiştir. Veri setinde kullanıcıların hangi filmi, ne zaman oyladığı ve oy değeri bulunmaktadır. Veri seti 4,536,891 oy, 171,670 kullanıcı ve 23,974 film den oluşmaktadır. Veri setinin bu bölümü %99,89 oranında seyrek tir. Bu veri seti yarışma sırasında geliştirilen sistemleri eğ itmek için kullanılmış tır. Daha sonrasında test etmek için ise 594 kullanıcı, 4,482 oy ve 811 film den oluşan daha küçük bir bölüm veri seti kullanılmış ve sistemlerin verimliliği ölçülmüştür.
- Last.fm Veri Seti:** Last.fm veri seti müzik öneri sistemi için geliştirilen en büyük veri setlerinden bir tanesidir (Celma, 2010). Toplamda 359,347 kullanıcı ve 17,559,530 satır “kullanıcı - sanatçı - dinlenme sayısı” bilgisi iç ermektedir. Bu veriler Last.fm API (Last.fm, 2008) kullanılarak, 2008 yılında Oscar Celma (Celma, 2008) tarafından toplanmıştır ve ticari olmayan kullanım için herkese açıktır. Veri setinde kullanıcılara ait cinsiyet, yaş, üyelik tarihi ve ülke gibi demografik bilgilerin yanı sıra, hangi kullanıcının hangi şarkıyı kaç kez dinlediği bilgisi de bulunmaktadır. Last.fm veri setinde sadece bir kullanıcı tarafından dinlenmiş olan şarkı ve sanatçı bilgileri bulunmaktadır. Şarkıların dinlenme sayılarına bakarak bir kullanıcı için en sevilen sanatçıyı ya da şarkıyı bulmak ve kullanıcılar arasındaki benzerlikleri tespit etmek mümkündür. Ancak şarkılar ya da sanatçılar arasındaki benzerlikleri bulmak mümkün değildir. Şarkılar ve sanatçılar hakkındaki metinsel bilgi bu veri setinde yoktur. Bu nedenle bu veri setinin işbirlikçi filtreleme yöntemi için daha uygun olduğunu söylemek mümkündür.
- Million Song Veri Seti:** Million Song veri seti (Dataset, 2011) müzik ses özelliklerini ve üst verilerini barındıran bir koleksiyondur. 1 milyon şarkıya ait şarkı başlığı, sanatçı, yayınlanma yılı, ses özellikleri ve çok daha fazlasını barındırmaktadır (McFee et al., 2012). Million Son veri setindeki veriler farklı kaynaklardan derlenmiştir. Örneğin; yeniden yorumlanan şarkılar için SecondHand, şarkı sözleri için musixmatch, şarkı benzerliği için Last.fm, kullanıcı verisi için de Taste Profile kaynakları kullanılmaktadır. Bütün bu veriler Million Song veri seti içinde derlenmiş ve daha büyük bir veri seti olarak kullanıma sunulmuştur.

- **Jester Veri Seti:** Jester kullanıcılara fıkra ya da şakalar tavsiye eden çevrimiçi bir sistemdir. Jester şaka öneri sistemi kullanıcılarının verileri kullanılarak üç farklı sürümü olan bir veri seti yayınlanmıştır. Bu veri seti işbirlikçi filtreleme için kullanılabilir bir veri sistemidir. Jester veri setinin ilk sürümü 73.421 kullanıcıdan toplanmış olan 4 milyon oydan oluşmaktadır. Sistemde toplamda 100 şaka bulunmaktadır. Veri setinin ilk sürümü Nisan 1999 - Mayıs 2003 arasında toplanmıştır. Jester veri setinin ikinci sürümü 1,7 milyonun üzerinde oy içermektedir. Sistemde toplam 59.132 kullanıcı ve 150 şaka bulunmaktadır. Bu sürüm Kasım 2006 - Mayıs 2009 arasında toplanmıştır. Veri setinin üçüncü sürümü ise ilk iki sürüme ek olarak 500.000 yeni oy ve toplam 79.681 kullanıcı barındırmaktadır. Jester veri setinin oyları sıkça kullanılan 1-5 aralığından farklı olup, -10 ve +10 arasında değişen ondalıklı sayılardan oluşmaktadır. Jester şaka öneri sistemi kullanıcı arayüzünde kullanıcının oy vermesi için bir kayan çubuk bulunmakta ve kullanıcı beğenisini rakamlarla değil görsel olarak ifade etmektedir. Veri seti toplam iki dosyadan oluşmaktadır. İlk dosyada öge numaraları ve şaka metinleri, ikinci dosyada ise kullanıcı numaraları, öge numaraları ve oylar bulunmaktadır. Bu veri seti işbirlikçi ve içerik tabanlı filtreleme sistemleri için kullanılabilir niteliktedir.
- **Book-Crossing Veri Seti:** Book-Crossing veri seti Cai-Nicolas Ziegler (Ziegler et al., 2005) tarafından 4 hafta süren bir çalışma ile Ağustos - Eylül 2004 arasında toplanmıştır. Bu veri seti kitap önerisi için oluşturulmuştur ve toplamda 278.858 kullanıcı, 271.379 kitap ve 1.149.780 açık ve örtük geribildirim içermektedir. Bu veri setinde kullanıcıların demografik özellikleri de bulunmaktadır ancak kullanıcı verisi gizliliği sebebi ile kullanıcılar anonimleştirilmiştir. Book-Crossing veri seti toplam 3 tablodan oluşmaktadır. BX-Users tablosunda kullanıcı numarası, bulunulan yer ve yaş, BX-Books tablosunda kitabın ISBN numarası, başlığı, yazarı, yayıncısı ve yayınlanma yılı, BX-Book-Ratings tablosunda ise 1-10 aralığında açık geribildirimler (oylar) ve örtük geribildirimler bulunmaktadır.
- **YOW Veri Seti:** YOW veri seti 25 kullanıcı, 7000'den fazla geribildirim ve 383 haber makalesi içeren bir haber öneri sistemi veri setidir. 2004 yılında Carnegie Mellon Üniversitesi'nde yürütülen bir proje kapsamında oluşturulmuştur. Hem kısaca haber metinlerini, hem de oldukça detaylı kullanıcı verilerini içermesi nedeniyle işbirlikçi ve içerik tabanlı filtreleme yöntemlerine uygundur. YOW veri seti bu her iki yöntemi de destekleyen, internette herkesin erişimine açık olan bulabildiğimiz tek haber öneri sistemi veri setidir. YOW veri seti 4.1.4'de detaylı olarak açıklanmıştır.

Veri Setinin Adı	Alan	Boyut			Geribildirim	
		Öge Sayısı	Kullanıcı Sayısı	Oy Sayısı	Açık	Örtük
MovieLens 100K	Film	1682 film	943	100	1-5 arası oy	Yok
MovieLens 1M	Film	3900 film	6040	1.000.209	1-5 arası oy	Yok
MovieLens 10M	Film	10.682 film	71567	10.000.054	1-5 arası oy	Yok
Netflix (Training)	Film	17.770 film	480.189	100.480.507	1-5 arası oy	Yok
MoviePilot	Film	23.974 film	171.67	4.536.891	Oy	Yok
Last.fm	Müzik	186.642 sanatçı	359.347	17.559.530	Oy	Yok
Million Song	Müzik	1.000.000 şarkı	0	0	0	0
Jester v1	Şaka	100 şaka	73.421	4.000.000	-10 / +10 arası oy	Yok
Jester v2	Şaka	150 şaka	59.132	1.700.000	-10 / +10 arası oy	Yok
BookCrossing	Kitap	271.379 kitap	278.858	1.149.780	1 - 10 arası oy	Var
YOW	Haber	383 makale	25	7000+	1 - 5 arası oy	Var

Şekil 2.4: Farklı öneri sistemi veri setlerinin karşılaştırılması ve özellikleri.

3 HABER ÖNERİ SİSTEMLERİ

Günümüzde kişiler haberleri internetten daha fazla takip etmeye başlamışlardır. Bazı görüşlere göre basılı gazete ve dergiler gelecekte yerini tamamen sayısal hallerine bırakacaktır. Son yıllarda basılı haber kaynaklarının tirajlarının düştüğü, sayısal haber kaynak aboneliklerinin ise arttığı gözlenmiştir. İnternet ortamının dinamikliği, güncel haberlere anında ulaşabilme imkanı pek çok kimse için oldukça önemlidir. Ancak bu dinamizmin olumsuz sonucu olarak çok fazla haber kaynağı ve haber başlığı bulunmaktadır. Sık sık güncellenen haber siteleri, bir saat içinde yüzlerce yeni makaleyi okuyuculara sunabilmektedir.

Pek çok okuyucu ilgisini çeken makaleleri bulabilmek için çokça zaman harcamaktadır. Kullanıcılar kimi zaman sadece haber başlıklarına bakarak ilgi çekici bir haber bulmaya çalışır, kimi zaman ise en çok okunan haberlere ya da son dakika gelişmelerine göz atarlar. Ancak bu yöntemlerin hiçbirisi tam olarak kullanıcının ilgisini çeken haberleri bulmayı garantilemez. Üstelik internetteki pek çok haber kaynağını göz önüne alırsak, okuyucular sadece haber okumak için çok fazla zaman harcayacaklardır. Haber öneri sistemleri, kullanıcılara kişisel ilgi ve tercihlerine göre en uygun makaleleri otomatik olarak sunmayı hedefler.

(Gulla et al., 2013)'da belirtildiği gibi U kullanıcılar kümesi, A haber makaleleri ve V kullanıcıların bir haber makalesi hakkındaki beğeni veya tercihleri kümesi olmak üzere haber öneri sistemleri matematiksel olarak şu şekilde gösterilebilir:

$$s : U \times A \rightarrow V \quad (3.1)$$

Buna göre bir haber öneri sisteminin amacı bir kullanıcı için aşağıdaki fonksiyonun en büyük değerini döndüren haber makalesini kullanıcıya önermektir.

$$\hat{\alpha} = \arg \max_{\alpha \in A} s(u, \alpha) \quad (3.2)$$

Haber önermek diğer öneri sistemleri ile karşılaştırıldığında gerçekleştirmesi en zor olanlardan bir tanesidir. Haber alanı diğer alanlardan birçok noktada ayrılır. Örneğin; makalelerin güncelliği ve beğenilirliği çok hızlı değişmektedir. Bir film öneri sisteminde, sistem 1970'li yıllardan bir film önerebilir ancak çoğu kullanıcı

sadece birkaç gün önceki haberi bile okumak istemez. Ya da beğenilen bir film birkaç hafta boyunca en sevilenler listesinde en üstte kalabilir ancak bir haber başlığı bu yerini birkaç saatten fazla koruyamaz. Haber alanı diğer tüm alanlara göre çok hızlı değişmektedir. Güncellik diğer alanlardan çok daha fazla önem arz etmektedir. Diğer yandan, bazı haberler birbiri ile bağlantılı olabilir. Kullanıcı bir yazı dizisinin geçen haftaki bölümünü okumak isteyebilir ya da bugün olan bir olay geçen ay olmuş bir başka olay ile bağlantılı olduğundan kullanıcı geçmiş makaleyi de okumak isteyebilir. Buna ek olarak, sadece kullanıcı tercihlerini öğrenmek de haber öneri sistemleri için yeterli olmayacaktır. Bazen kullanıcılar bir haber ilgi alanlarına girmese bile o haberi sadece önemli olduğunu düşündüklerinden dolayı okumak isteyeceklerdir. Örneğin; normalde politika haberleri kullanıcının ilgisini çekmese bile seçim döneminde bu tür haberleri takip etmek isteyebilir. Kullanıcılar çeşitli haber kategorilerini farklı zamanlarda ve belirli olaylara bağlı olarak okumayı tercih edebilirler. Tüm bu zorlukların yanı sıra her saat yayınlanan yüzlerce yeni haber başlığını da düşünürsek haber öneri sistemlerinin karmaşıklığı daha iyi anlaşılacaktır. Oysa bir film ya da bir müzik öneri sisteminde, sisteme yeni eklenen nesnelerin sıklığı çok daha düşüktür.

3.1 Haber Öneri Sistemi Zorlukları ve Literatürdeki Çözüm Önerileri

Bölüm 2.3'te her biri kısaca açıklanan öneri sistemlerinin zorluklarından bazıları bu bölümde haber öneri sistemleri açısından detaylı olarak ele alınmıştır. Aynı zamanda bu zorlukların literatürdeki çözüm önerileri kısaca açıklanmıştır. Bu bölümde yer verilen zorlukların bazıları sadece haber öneri sistemlerine özgüdür. Ancak bu zorlukların tamamı haber öneri sistemi araştırmalarında en çok ele alınan, önceliklendirilen ve önem verilen zorluklardır.

3.1.1 Soğuk başlangıç sorunu

İşbirlikçi filtreleme kullanıcılar arasındaki benzerlikleri bulup, farklı kullanıcıların benzer özellik ve tercihlerine göre önerilerde bulunur. Eğer sisteme yeni bir öğe eklenmişse, bu öğeyi daha önce kimse tercih etmediği için önerilmesi mümkün değildir. Benzer şekilde sisteme yeni bir kullanıcı eklendiğinde, bu kullanıcının geçmişte herhangi bir beğeni ya da tercihi olmadığı için bu kullanıcıya işbirlikçi filtreleme yöntemi ile bir şey tavsiye etmek mümkün değildir. Bu soruna soğuk başlangıç sorunu denmektedir. Soğuk başlangıç sorunu sadece işbirlikçi filtreleme yöntemini kullanan uygulamaların en yaygın sorunudur. (Liu et al.,

2010)'da belirtildiği gibi bazı araştırmacılar için bu sorun işbirlikçi filtrelemenin en önemli dezavantajıdır. Bu sorunu çözmek için (Liu et al., 2010; Fortuna et al., 2010) ve (Lin et al., 2012) birbirinden bağımsız ve biraz farklı şekillerde çalışmalarda hibrid yöntemi önermişlerdir. Hibrid yöntem, detayları Bölüm 2.1.3'de açıklandığı gibi işbirlikçi filtreleme ve içerik tabanlı filtreleme yöntemlerini birlikte kullanmaktadır. (Liu et al., 2010), yeni öğelerin önerilebilmesi için kullanıcıların özel ilgi alanlarına göre kişiselleştirilmiş profil oluşturulmasını önermiştir.

(Lin et al., 2012), soğuk başlangıç sorununun üstesinden gelebilmek için seçilmiş uzmanların beğenilerine yer veren bir sistem önermiştir. Bu çalışmada yapılan araştırmaya göre, sosyal ağları aktif olarak kullanan kişilerin yeni kullanıcılar üzerinde oldukça fazla etkisi bulunmaktadır. Buna göre, sisteme yeni eklenen bir kullanıcıya seçilmiş uzmanların tercihleri doğrultusunda öneride bulunulabilir. Ayrıca içerik tabanlı filtreleme yönteminin tekniklerinden birisi olan TF-IDF kullanılarak sisteme yeni eklenen öğeler analiz edilerek önerilebilir.

(Fortuna et al., 2010) ise kullanıcıları okudukları makale sayısına göre eski ve yeni kullanıcılar olarak gruplandırmayı önermiştir. Bu gruplandırmadan sonra her grup için en ilgi çekici haber kategorisini tahmin edebilmek için her gruba özel bir model eğitilmektedir. Sonuçta ortaya çıkan her bir kategorideki en üstteki makaleler kullanıcılara önerilebilmekte ve soğuk başlangıç sorunu oluşmamaktadır.

(Tavakolifard et al., 2013) kullanıcıların uzun ve kısa zamanlı tercihleri ile anlık bağlamı göz önüne alan bir mimari önermektedir. Böylece anlık bağlam bilgisi ilk öneriler için kullanılmakta ve soğuk başlangıç sorunu bu şekilde aşılabilmektedir.

(Lee and Park, 2007) de benzer şekilde kullanıcının yeni olup olmadığını kontrol ettikten sonra, eğer kullanıcı yeni ise onu demografik özelliklerine göre geçici olarak bir gruba eklemektedir. Daha sonra ise bu kullanıcıya ilk öneriler o grubun genel tercihleri göz önüne alınarak yapılmaktadır.

(Garcin et al., 2013), kimliği belirli olmayan, anonim kullanıcılar için çalışan bir sistem önermektedir. Kullanıcı bilgilerinin gizliliği konusuna da odaklanan bu çalışmada, kullanıcının sadece o anki oturumu boyunca toplanan bilgiler öneriler için kullanılmaktadır. Böylece kullanıcıların sisteme kayıt olması, oturum açması ve kişisel profil oluşturulmasına gerek kalmamaktadır. Önerilen bu sistemde bir kullanıcı haberleri okumaya başladığında, o anki oturum sona erene kadar kullanıcının bilgileri sisteme aktarılmakta, sistem de bu bilgilere göre bir model oluşturup oturum boyunca o modeli güncellemektedir. Yapılan öneriler oluşturulan ve güncellenen bu modele göre yapılmakta ve modeldeki her yeni güncellemeden

sonra sistem daha iyi önerilerde bulunabilmektedir.

3.1.2 Anındalık

Literatürdeki neredeyse bütün haber öneri sistemlerinde anındalığın öneminde değinilmiştir. (Yeung and Yang, 2010) anındalığı öneri sisteminin önemli bir özelliği olarak belirtmiş ve taşınabilir cihazlar için önetkin bir haber öneri sistemi sunmuşlardır. Taşınabilir cihazlar için çevre çok değişkendir. Hem bulunan fiziksel çevre hem de bağlam çok hızlı değişebilmektedir. Bu yüzden taşınabilir cihazlarda kullanıcıya hızlı ve o ana uygun bilgiyi ulaştırmak oldukça zordur. (Yeung and Yang, 2010)'ın önerdiği sistem hibrid bir P2P (peer-to-peer) sistemdir. Önerilen bu sistem sayesinde anlık bilgiler kullanıcının taşınabilir cihazına iletilmektedir. Saf bir P2P sistem diğer cihazlarla çok fazla iletişim gerektirdiğinden pil ömrü kısıtlı olan taşınabilir cihazlar için uygun değildir. Önerilen sistemde taşınabilir cihaz ağdaki sunuculardan bir tanesine bağlanarak bulunduğu bağlam bilgisini gönderir. Sunucu anlık haber makalelerini sürekli olarak RSS'ten alır. Öneri, kullanıcı profili, yer bilgisi, kullanım deseni ve kullanıcı oyları gibi bağlam bilgisi de hesaba katılarak sunucu tarafından üretilir.

Farklı bir yaklaşım olarak (Wen et al., 2012) öneri yöntemine zaman faktörünü de katmaktadır. Buna göre bir haber ögesini tavsiye etmek için o öğeye atanmış bir zaman katsayısı hesaba katılmakta ve kullanıcının ilgi alanları ve tercihleri de göz önüne alınarak hangi haber ögesinin tavsiye edileceğine karar verilmektedir. Benzer olarak (Lee and Park, 2007), haber makalelerine önem derecelerine ve anındalıklarına göre ağırlıklar atamaktadırlar. Öneri aşamasında ise ağırlığı fazla olan haber ögesi önerilerde daha öncelikli olarak yer almaktadır.

(Fortuna et al., 2010)'da öncelikli haber kategorileri her kullanıcı için ayrıca belirlenmektedir. Kullanıcıların tercihleri ve her kategorideki en yeni makaleye göre öneriler yapılmaktadır. Diğer yandan (Li et al., 2011) her haber makalesi için bir haber profili oluşturmaktadır. Bu profiller haberin yeniliği ve popülerliği ile orantılı olarak sürekli değişmektedir. (Liu et al., 2010) haber akımlarını göz önüne alan bir haber öneri sistemi önermiştir. Bu çalışmada akımların çoğunlukla en yeni haber başlıklarından oluştuğuna dikkat çekilmiş ve bu haber öneri sistemlerinin anındalık sorununa bir çözüm olarak önerilmiştir.

3.1.3 Örtük geribildirim

Bir kullanıcının gelecekteki tercihlerini tahminlemek ve uygun öneriler yapabilmek için sistemin kullanıcının geçmiş tercih ve ilgi alanlarını bilmesi gerekmektedir. Kullanıcının geçmiş tercihlerini öğrenebilmenin ise iki yolu vardır: Açık ve örtük geribildirimler. Açık geribildirimleri toplamak için sistemin kullanıcı ile sürekli etkileşim halinde olması gerekmektedir. Açık geribildirimler genellikle kullanıcıya yöneltilen sorular ya da kullanıcıdan istenen oylar sonucu elde edilir. Bu sorular ve oylar kullanıcının içeriği ya da içeriğe ait çeşitli detayları ne kadar beğendiğini anlamaya yöneliktir. Ancak sistemin sürekli olarak kullanıcı ile etkileşimde olması, kullanıcıdan el ile girilen çeşitli bilgiler girmesin istemesi taşınabilir cihazlar başta olmak üzere, hem kullanıcı hem de sistem açısından pratik değildir (Yeung and Yang, 2010). Ayrıca genel olarak kullanıcıların sorulan soruları yanıtlamama veya oy vermeme eğilimi bulunmaktadır. Sonuç olarak kullanıcı profili oluşturma ve filtreleme işlemlerinde kullanılmak üzere, kullanıcı bilgisinin, gizliliğin korunmasına da dikkat edilerek sistem tarafından otomatik olarak algılanması gerekmektedir. Bu da örtük geribildirimler sayesinde mümkün olmaktadır. (Garcin et al., 2013)'da belirtildiği gibi örtük geribildirimlerin toplanması genellikle kullanıcıların sisteme giriş yapmasını gerektirmektedir. Daha sonra ise kullanıcı hareketleri kayıt altına alınarak örtük geribildirim verisi olarak kullanılmaktadır. Bu da kullanıcı verisinin gizliliği konusunda sorunlar oluşturabilmektedir. (Garcin et al., 2013) kullanıcının sisteme giriş yapmasına gerek kalmadan sadece o anki oturum bilgilerini kullanarak önerilerde bulunan bir sistem önermiştir. (Liu et al., 2010) kullanıcıların gelecekteki tercihlerini tahmin etmek için kayıtlı kullanıcıların geçmiş hareketleri üzerinde analizler gerçekleştiren ve kullanıcının tercihlerinde bir değişiklik olup olmadığını izleyen bir sistem sunmuştur. Benzer şekilde (Wen et al., 2012; Tavakolifard et al., 2013; Lee and Park, 2007) da kullanıcı hareketlerini izleyen ve kullanıcı geçmişi üzerinde işlemler yaparak kullanıcının gelecek tercihlerini belirlemeye çalışan sistemler önermişlerdir.

3.1.4 Kullanıcı tercihlerinin değişmesi

Zaman içinde insanların tercihlerinin değişmesi doğal bir süreçtir. Bu değişimde etkili kişinin yaşı, kültür seviyesi ve çevresi gibi etkenler rol oynamaktadır. Örneğin bir kişinin gençliğinde beğendiği filmleri yetişkinliğinde ya da yaşlılığında beğenmemesi oldukça sık karşılaşılan bir durumdur. Ancak kişilerin beğendikleri ya da izlemeyi tercih ettikleri filmler kısa vadede genellikle çok büyük değişiklik göstermezler. Bu çeşitli öneri alanları için önemli bir ipucudur. Eğer bir kullanıcı macera filmlerinden hoşlanıyorsa sonraki aylar hatta yıllar boyunca

bundan vazgeçmeyecektir. Ancak haber öneri alanı için durum daha farklıdır. Bir kullanıcının haber okuma tercihleri kısa vadede pek çok şeyden etkilenebilir. Anlık gelişmeler, kullanıcının normalde ilgi alanına girmeyen ama önemli bulduğu olaylar, hatta kişinin anlık ruh hali bile kişinin haber makalesi seçiminde etkili olabilir. Dolayısı ile kullanıcının gelecek haber tercihlerini tahminlemek oldukça zor bir iştir. (Liu et al., 2010)'da bu zorluğa değinilmiş ve kullanıcının gelecek haber tercihlerini tahminlemek için bir mimari sunulmuştur. Bu mimaride, sisteme kayıtlı kullanıcının geçmiş hareketleri üzerinde geniş ölçekli analizler yapılarak gelecek tercihlerini tahmin etmek üzere hibrid bir öneri sistemi sunulmuştur. Bu çalışma kapsamında, hem birey hem de grup bazında farklı haber kategorileri üzerinde aylık olarak yapılan analizlerde kullanıcı tercihlerinin değiştiği gözlenmiştir. Kullanıcının gelecek tercihlerinin tahminlenmesinde Bayes çerçevesi kullanılmıştır. Daha sonra ise bu tahminler filtreleme yöntemlerinde kullanılmış ve daha uygun önerilerde bulunmak için işbirlikçi filtreleme yöntemi ile birleştirilmiştir. (Lee and Park, 2007) ise kullanıcının kısa ve uzun vadeli tercihlerini birbiri ile karşılaştırarak kullanıcı tercihlerinin değişimini gözlemlemiştir. Bu çalışmada, kullanıcının okumayı tercih ettiği haber kategorilerine ağırlıklar atanmış ve zaman içindeki okuma örüntülerine göre bu ağırlıklar değiştirilmiştir. Kullanıcıya yapılan öneriler, değişen kategori ağırlıkları ile orantılı olarak değişmekte ve böylece kullanıcı tercihlerinin zaman içindeki değişim sorununa bir çözüm olarak sunulmaktadır. Benzer şekilde (Wen et al., 2012) kullanıcının belirli bir konuya olan ilgi derecesini belirleyerek ve zaman içindeki okuma değişimlerini gözlemleyerek bu soruna bir çözüm olarak sunmuştur. Buna göre eğer bir kullanıcı sık sık belirli bir konu ile ilgili bir sayfayı ziyaret ediyorsa, bu o kullanıcının o konuya olan ilgisini göstermektedir. Eğer kullanıcı zaman içinde o sayfayı ziyaret etmeyi azaltır ya da bırakırsa, bu durum kullanıcının o konuya olan ilgisinin azaldığı şeklinde yorumlanmakta ve o konuyla ilgili daha az öneride bulunmaktadır. Kullanıcı tercihlerinin değişimini gözlemek için önerilen bir başka yöntem de (Saranya.K.G and Sadhasivam, 2012) tarafından sunulan iki farklı kullanıcı profili tanımlama yöntemidir. Bu yönteme göre durağan ve değişken olarak iki farklı kullanıcı profili tanımlanmaktadır. Durağan kullanıcı profili kullanıcının sisteme kayıt olurken verdiği kullanıcı adı, en çok ilgilendiği konular gibi bilgileri içerirken, değişken kullanıcı profili kullanıcının her oturumdaki örtük geribildirimleri ile oluşturulmuştur. Bu iki profil arasındaki farklar sisteme kullanıcı tercihlerinin değişimi ile ilgili bilgi vermektedir.

3.1.5 Ölçeklenebilirlik

Ölçeklenebilirlik bilgisayarla ilgili her sisteme etki eden bir kavramdır. Sistemin boyutları arttıkça onu idare etmek ve sistemin aynı etkinlikte çalışması

zorlaşmaktadır. Diğer tüm sistemlerde olduğu gibi öneri sistemlerinde de ölçeklenebilirlik önemli bir zorluktur. Gerçekten kullanılabilir ve sürdürülebilir bir öneri sisteminin ölçeklenebilir olması gerekmektedir. Haber öneri sistemleri alanında ise bu zorluk Bölüm 2.3'te söz edilen diğer zorluklarla birleşmekte ve son kullanıcılara açık, kullanışlı bir haber öneri sistemi geliştirilmesini daha da zorlaştırmaktadır. Ölçeklenebilirlik sorunu literatürde çokça ele alınmış ve farklı yaklaşımlarla çözülmeye çalışılmıştır. Ölçeklenebilir bir haber öneri sistemi için (Li et al., 2011)'de haber makalelerini gruplayan ve böylece gereksiz benzerlik hesaplamalarını ortadan kaldıran bir yöntem sunulmuştur. Bu sistemde anlık benzerlik hesaplamaları için daha az zaman harcandığından sistem kullanıcı isteklerine daha hızlı yanıt verebilmektedir. (Das et al., 2007) ise milyonlarca kullanıcıya aynı anda hizmet verebilmek için olasılıksal bir kümeleme yöntemi olan MinHash tabanlı yeni bir algoritma önermiştir. Ayrıca bu çalışmada işbirlikçi filtreleme yöntemi için kullanılan PLSI modeli yeniden düzenlenmiş ve büyük ölçekli verilerin işlenmesinde kullanılan MapReduce tekniğine uyarlanmıştır. MapReduce kümeler (cluster) üzerinde çalışan ve büyük ölçekli verilerin olduğu sistemlerin ölçeklenebilirliğini arttıran bir modeldir. Bu modelin işbirlikçi filtreleme yöntemine uyarlanması ile birlikte öneri sisteminin ölçeklenebilirliği artırılmıştır. (Saranya.K.G and Sadhasivam, 2012)'de da etkin öneri sistemleri için ölçeklenebilirliğin önemi ve gerekliliği tartışılmıştır. Bu çalışmada sistemlerin güvenilirliği ve ölçeklenebilirliğini arttıran Hadoop çerçevesi kullanılarak bu zorluk aşılmıştır. Öneri sistemlerine apayrı bir yaklaşım sunan (Garcin et al., 2013) de ölçeklenebilirlik sorununa değinmiştir. Bağlam ağaçlarının öneri sistemlerinde kullanan bu çalışmada, sistemin yapısı nedeniyle ölçeklenebilirlik büyük bir zorluk olmaktan çıkmıştır. Buna göre sistemde bir anda sadece tek bir ağaç bulunduğundan ve uygulanan bağlam sınırlamaları nedeniyle bu ağacın yapısı da oldukça sınırlı olduğundan, bu sistemlerde ölçeklenebilirlik sorunu bulunmamaktadır.

3.1.6 Veri seyrekliği

Veri seyrekliği işbirlikçi filtreleme yönteminin en önemli sorunlarından bir tanesi olmasına rağmen literatürdeki pek çok haber öneri sistemi çalışmasında bu sorundan söz edilmemiştir. Veri seyrekliği sorunu, öneri sisteminde kullanılan matriste pek çok boş alan olması şeklinde açıklanabilir. Bunun nedeni ise kullanıcı-öge matrisindeki boyutlardan bir tanesinin diğerinden daha az veri içermesidir. Örneğin; 500.000 öge bulunan bir sistemde sadece 20 kullanıcı varsa, oluşturulan kullanıcı-öge matrisi oldukça seyrek olacaktır. Bu durumda sistemin kullanım süresi arttıkça matrisin seyrekliğinin azalacağı söylenebilir. Ancak sisteme sürekli olarak yeni kullanıcılar ve öğeler eklendiği, hatta bazen de sistemden

çıkarıldığı unutulmamalıdır. Özellikle haber öneri sistemlerinde her saat sisteme yüzlerce yeni öge eklenebilmektedir. (Yeung and Yang, 2010)'da bu sorunun çözümü için Bayes ağlarının kullanımı tartışılmaktadır. Diğer yandan (Saranya.K.G and Sadhasivam, 2012)'da ilişkisel olmayan dağıtık bir veritabanı olan HBASE'i seyrek verinin hatasız saklanması için kullanılmıştır. Yine de bu yöntem veri seyrekliği sorununa tam bir çözüm önerisi getirmemektedir. (Lin et al., 2012) ise seyrek veri sorununun işbirlikçi ve içerik tabanlı filtreleme sistemlerinin birlikte kullanılarak, hibrid bir sistem sayesinde çözülebileceğini belirtmiştir. Hibrid sistemlerin kullanımı veri seyrekliği sorununun çözümü için yaygın olarak kullanılan yöntemlerden bir tanesidir.

3.2 Anlamsal Haber Öneri Sistemleri

Bölüm 2.1.4'de kısaca bahsedildiği gibi anlamsal yaklaşım öneri sistemlerinde kullanılan yöntemlerden bir tanesidir. Anlamsal yaklaşımın öneri sistemlerinde kullanılmasının temel sebeplerinden bir tanesi içeriğin kültürel ve dilbilimsel arkaplan bilgisini kullanmayı mümkün kılmasıdır (Peis et al., 2008). Anlamsal yaklaşım, kelime tabanlı sistemlerle karşılaştırıldığında karmaşıklığı azalmakta, kavramların hiyerarşik olarak temsil edilmesine ve çıkarsama yapmaya olanak sağlamaktadır (Cantador and Castells, 2009). (Lops et al., 2011)'de anlamsal öneri sistemleri kullandıkları farklı anlamsal yaklaşımlara göre gruplandırılmıştır. Farklı alanlardaki öneri sistemlerinin farklı zorlukları ve ihtiyaçları olduğundan, farklı alanlarda uygulanan farklı anlamsal yaklaşımlar bulunmaktadır. Bu bölümde literatürdeki anlamsal haber öneri sistemlerinden söz edilmiştir.

News@Hand yapılan önerilerde anlamsal teknolojileri kullanan bir haber öneri sistemidir (Cantador and Castells, 2009). Bu sistemde haber içeriklerinden elde edilen isim terimler kullanılarak ontolojiler oluşturulmuştur. Ontoloji sınıflarını oluşturmak için ise Wikipedia makaleleri kullanılmıştır. Bu çalışmada, öneri sistemlerindeki önemli problemlerden bir tanesi olan seyrek veri sorununun üstesinden gelebilmek için kullanıcı tercihlerinin genişletilmesi için bir mekanizma sunulmuştur. Bu sayede seyrek veri sorununun yol açtığı yeni öğelerin önerilmesi zorluğu, kullanıcıların tercih ve beğeni alanları genişlediği için azalmaktadır. News@Hand aynı zamanda öneri sistemlerinde bağlam konusuna da değinmiştir. Öneri sistemlerinde bağlamın farklı anlamlara gelebileceği belirtilmiş ve o anki bağlam çeşitli parametrelerle tanımlanmıştır. Zaman içerisinde bağlama ait kavramların önemleri azalmakta ve böylece sistemin dinamik olarak bağlam farkında olması sağlanmaktadır.

Hermes çerçevesi kişiselleştirilmiş haber servisi için anlamsal yöntemleri kullanmaktadır (Intema et al., 2010). Hermes'e ek olarak anlamsal haber önerileri için Athena adı verilen bir eklenti geliştirilmiştir. Öneriler için öncelikle kullanıcının geçmişte okuduğu haber makaleleri kullanılarak kişisel kullanıcı profili oluşturulmaktadır. Daha sonra ise çeşitli benzerlik ölçütleri ve Hermes çerçevesi tarafından oluşturulan ontoloji kullanarak öneriler yapılmaktadır. Athena'da uygulanan haber öneri yönteminin detayları (Goossen et al., 2011)'de yer almaktadır. Bu çalışmada, içerik tabanlı öneri sistemlerinde sıkça kullanılan TF-IDF yönteminin anlamsal öneri sistemlerine uyarlanmış biçimi olan CF-IDF yöntemi önerilmiştir. CF-IDF yöntemi bir metindeki sadece önemli kavramları dikkate alırken, TF-IDF yöntemi tüm terimleri dikkate alır ve analiz eder. Bu açıdan bakıldığında CF-IDF yöntemi daha verimli görünmektedir. Benzer şekilde (Capelle et al., 2012)'de SF-IDF ve SS adı verilen iki yöntem önerilmiştir. Bu yöntemler önceki çalışmalarla ilişkili bir şekilde, yukarıda sözü edilen Athena'ya eklenti olarak Ceryx adı ile uygulanmıştır. Bu yöntemlerden SF-IDF, yine anlamsal araştırma ve uygulamalarda sıkça kullanılan WordNet'in (WordNet, 1985) eş anlamlılar kümesini kullanırken, SS ögeleri arasındaki anlamsal benzerlikleri hesaplamaktadır. Bu iki yöntem birlikte kullanılarak haber ögelerini kullanıcı davranış profiline ve ögeleri arasındaki anlamsal benzerlik ölçütlerine göre önermektedir.

(Rao et al., 2013)'te ontoloji tabanlı bir benzerlik modeli önerilmiştir. Bu model bir anlamsal haber öneri sisteminde haber makaleleri ile kullanıcılar arasındaki benzerlikleri bulmak için kullanılmaktadır. Bu sistemde kullanılan ontolojiler İngilizce (Wikipedia, DBpedia, Freebase vb.) ve Çince (Hudong, Baidu vb.) çevrimiçi ansiklopediler kullanılarak üretilmiştir.

(Lašek and Vojtáš, 2011)'da haber filtreleme alanında bir anlamsal bilgi filtreleme iş akışı önerilmiştir. Bu iş akışında eleman tanımlanması, anlamsal veri taraması ve kullanıcı profili oluşturma gibi farklı adımlar bulunmaktadır. Bu çalışmada anlamsal arkaplan verilerinin kullanıcı profilleri ile birlikte kullanılması ve isim öge tanıması (name entity recognition) sonuçlarının kullanıcı geribildirimleri yardımı ile geliştirilmesi hedeflenmiştir.

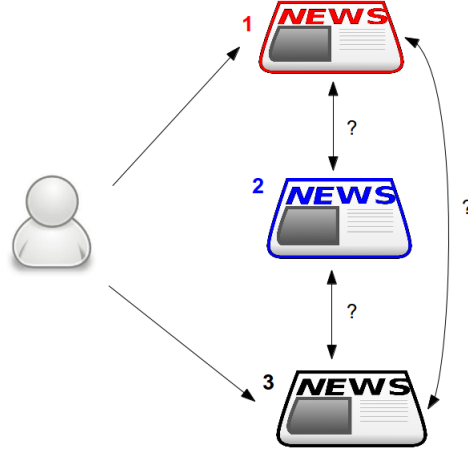
4 ANLAMSAL HABER KAYNAĞI ÖNERİ SİSTEMİ

Haber öneri sistemleri diğer öneri alanları ile karşılaştırıldığında birçok zorluk içermektedir. Bu zorluklardan Bölüm 3.1’de detaylı olarak söz edilmiştir. İnternette yayınlanan haberlerin sayısı çok hızlı artmakta, güncellikleri ve popülerlikleri de buna bağlı olarak çok hızlı değişmektedir. Bir haber farklı kaynaklarda farklı metinlerle yer alabilmektedir. Haber alanının bu dinamik yapısı içinde kullanıcıya farklı kaynaklardan aynı haberi önermemek, haber öneri sistemlerine özgü önemli zorluklarından bir tanesidir (Ozgobek et al., 2014).

Haber öneri sistemlerinde kullanılan haber kaynakları çoğunlukla geliştiriciler tarafından önceden belirlenmektedir. Kimi sistemlerde kullanıcıya seçim şansı tanınsa da kullanıcının daha önceden bilmediği kaynakları seçmek istememesi doğaldır. Ancak bu durum kullanıcının yeni haber kaynakları keşfetmesini engeller ve haber öneri sisteminin kapsamını daraltır. Literatürdeki çalışmalarda yapılan önerilerde kullanıcının kaynak tercihleri göz önüne alınmamaktadır. Öneri sistemindeki haberler belirlenen bu kaynaklardan otomatik olarak alınmakta ve öneri sistemi tarafından kullanılmaktadır. Ancak bu durum aynı haber metninin farklı kaynaklardan alınıp farklı haberlermiş gibi kullanıcıya önerilmesine yol açmaktadır. Doğaldır ki aynı haberi gün içinde defalarca öneren bir sistem kullanıcının hoşuna gitmeyecektir. Diğer yandan her haber kaynağının kendine özgü bir tarzı, görüşü ve anlatım biçimi vardır. Bu farklılıklar kullanıcılar açısından önem taşımaktadır. Farklı kaynaklardan alınan aynı habere ait metinler arasında bile kaynaktan kaynağa değişen anlatım farklılıkları bulunabilmektedir. Öneri sistemleri kullanıcının hangi anlatım tarzından hoşlandığından bağımsız olarak aynı içeriğe sahip haberlerden sadece bir tanesini seçmektedir.

Günümüzde internetten haber okuyan kişiler genellikle kendilerine uygun buldukları, hoşlandıkları sadece birkaç kaynağı takip etmektedir. (Ofcom, 2013)’un yaptığı araştırmaya göre İngiltere’de çevrimiçi haber okuyucularının %45’i haberleri sadece tek bir kaynaktan takip etmektedir. Benzer şekilde (Media, 2012)’de kullanıcılar tarafından takip edilen çevrimiçi haber kaynaklarının (web sayfaları ve uygulamalar dahil) sayısı demografik farklılıklara göre 1.4 ve 2.4 arasında değişmektedir. Toplamdaki ortalama ise 2.0’dır. Bu kullanıcılar açısından oldukça kısıtlayıcı ve kimi zaman da olaylara bakış açısını daraltıcıdır. Kullanıcıların beğenebilecekleri diğer kaynakları keşfetmeleri ise oldukça zordur ve çoğu zaman tesadüfe bağlı olarak bulunabilmektedir.

(Media, 2012)’de belirtilmiştir ki: “Nitel araştırmamız göstermiştir ki çok



Şekil 4.1: Bir kullanıcı ve farklı haber kaynakları arasındaki ilişki.

kaynaklı çevrimiçi haber takibi, çoğunlukla kişilerin farklı haber sağlayıcılar ve platformlardan bir haberin dengeli ve bütün bakışını edinmek istediklerinde yaptıkları aktif bir seçimdir.” Buna göre haberlerin farklı kaynaklardan edinilmesi kullanıcılara geniş, dengeli ve bütün bir bakış açısı sağlamaktadır. Bir kullanıcı bir kaynaktan alınan haberin ele alınış ve anlatım şeklini beğenmişse, o kaynaktan haber almaya devam etmek isteyecektir. Ayrıca bir seri halinde yayınlanan yazı dizilerinin aynı kaynaktan önerilmesi gerekmektedir. Farklı kaynaklardan aynı haberin önerilmemesi zorluğu da yapılan haber kaynağı önerisi sayesinde çözülmesi daha kolay bir sorun haline gelmektedir.

Literatürde pek çok kişiselleştirilmiş haber öneri sistemi olmasına rağmen, haber kaynakları arasındaki farkı dikkate alan ya da haber kaynağı önerisi üzerine odaklanan bir çalışma bulunmamaktadır.

Bu bölümde haber kaynağı önerisi için iki yöntem sunulmuş ve birbiri ile karşılaştırılmıştır. Şekil 4.1’de gösterildiği gibi bir kullanıcı farklı haber kaynaklarından haber okuyabilir. Ancak haber öneri sistemlerinde kullanıcının tercih ettiği farklı haber kaynakları arasında nasıl bir ilişki olduğu yapılan literatür taramalarında bulunamamıştır. Bu tez kapsamında sunulan yöntemlerden birisi anlamsal veri analizini içerirken diğeri okunan kaynaklar arasındaki örüntülerden, veri madenciliğinden faydalanmaktadır. Bu karşılaştırma sonucu “X haber kaynağından okuyan Y haber kaynağından da okumuş, öyleyse X ve Y anlamsal olarak ilişkili midir?” sorusunun yanıtı cevaplanacaktır. Bu sayede kullanıcılara en uygun haber kaynakları önerilebilecektir.

4.1 Kullanılan Araçlar

Bu tez kapsamında sunulan yöntemlerin geliştirilmesi, test edilmesi ve karşılaştırılması için çeşitli araçlar kullanılmıştır. Kullanılan araçların mümkün olduğunca açık ve özgür yazılımlar (FSF, 1986) olmasına ve ilerideki benzer çalışmalar için de yol gösterici olup daha da geliştirilebilmesine dikkat edilmiştir. Kullanılan araçların tümü herkes tarafından erişilebilir ve kullanılabilir araçlardır.

4.1.1 Protege

Protege ontoloji oluşturmak ve düzenlemek için geliştirilmiş, açık ve özgür bir yazılımdır (Protege, 2010). Diğer araçlarla yüksek uyumluluğu, birlikte kullanılabilirliği ve W3C standartları ile uyumluluğu nedeniyle Protege en yaygın kullanılan ontoloji düzenleyicilerden birisidir. Web ve masaüstü olmak üzere iki uygulaması olan Protege, OWL 2 web ontoloji dilini tam olarak desteklemekte ve çeşitli eklentilerle zenginleştirilebilmektedir. Protege masaüstü uygulaması kişisel bilgisayarlarda çalıştırılmakta ve yerel olarak kullanılmaktadır. Web uygulaması ise Protege geliştirici ekibi tarafından sunulan sunucu hizmeti kapsamında kullanılabilir ya da yerel olarak bir sunucu üzerinde çalıştırılabilmektedir.

Bu tez kapsamında Protege'nin masaüstü uygulamasının farklı sürümleri bir arada kullanılmıştır. Bunun nedeni oluşturulmak istenen ontoloji için gerekli olan eklentilerin sadece belirli Protege sürümleri ile çalışabilirliğidir. Örneğin; ontolojiye çok sayıda örneği aynı anda eklemek için geliştirmiş olan DataMaster (DataMaster, 2009) eklentisi sadece Protege 3.4.2 sürümü ile çalışmaktadır. Diğer yandan, bir sonraki bölümde detaylı olarak açıklanacak olan Termine eklentisi sadece Protege 4.0 ile çalışmaktadır. Ontolojideki sınıf, öge ve ilişkilerin görselleştirilmesi için Protege OntoGraf eklentisi kullanılmıştır.

4.1.2 Termine

Termine metin içindeki önemli terimleri bulmak için kullanılan bir metin analizi yazılımıdır (NaCTeM, 2011). Termine, metin analizi için dilbilimsel ve istatistiksel yöntemleri birleştiren "C-value/NC-value" yöntemini kullanmaktadır. C-value yöntemi bir sözcüğün metin içindeki kullanılma sayısına göre klasik istatistiksel yöntemleri kullanarak sözcüğün önemini belirlemektedir. NC-value ise bağlam kapsamında, terimlerle birlikte metinde bulunan sözcükleri bulmaktadır. Bu yöntemlerin detaylı açıklamaları (Frantzi et al., 2000)'da bulunabilir.

Termine web üzerinden bir servis olarak, Protege eklentisi olarak ya da geliştirilen uygulamalara eklemek üzere SOAP arayüzü ile kullanılabilir. Herkese açık olarak kullanılabilen web sürümü kısıtlı olup, büyük metin dosyaları üzerinde işlem yapılamamaktadır. Diğer sürümler ise akademik kullanım için ücretsiz olarak sunulmaktadır. Protege eklentisi hem büyük dosyalar üzerinde işlem yapmaya izin vermekte, hem de metinden çıkarılan terimleri OWL ontolojisine otomatik olarak aktarabilmektedir (Plugin, 2008). Metinden çıkarılan terimleri ontolojiye sınıf veya örnek olarak aktarmak mümkündür. Diğer yandan, Termine web sürümü daha günceldir ve daha iyi sonuçlar verdiği gözlenmiştir. Bu tez kapsamında haber içeriklerindeki önemli terimleri bulmak amacıyla Termine web servisi kullanılmıştır.

4.1.3 Orange

Orange veri analizi ve görselleştirilmesi için geliştirilmiş açık ve özgür bir yazılımdır (Orange, 2011). Aynı zamanda veri madenciliği ve makine öğrenmesi için araçlar barındırmaktadır. Orange ile sürükle bırak yöntemiyle görsel programlama yapılabilmekte, betik arayüzü kullanılarak Python betikleri yazılabilmekte ve veriler farklı grafiksel yöntemlerle görselleştirilebilmektedir. Özgür yazılım olması sayesinde pek çok şey eklenip değiştirilebilme imkanını sunan Orange, Python betikleri ve eklentilerle de istenilen şekilde genişletilebilmektedir. Orange hakkında daha detaylı bilgi (Demşar et al., 2013)'da bulunabilir. Bu tez kapsamında kullanılmış olan birliktelik kurallarını Orange sayesinde veri seti üzerinde çalıştırmak mümkün olmuştur.

4.1.4 Veri seti

Bölüm 2.4'te tartışıldığı gibi bilimsel araştırmalarda veri setlerinin önemi oldukça fazladır. Bu tez kapsamında sunulmuş olan haber kaynağı öneri sistemi yöntemleri de ancak uygun bir veri setinin varlığında test edilebilir ve birbiri ile karşılaştırılabilir. Maalesef haber öneri sistemleri alanında herkesin kullanımına açık ve internette mevcut olan bir veri seti bulmak çok zor olmaktadır. Haber öneri sistemlerinin değişken ve tahmin edilemez yapısı, kullanıcılardan alınan açık ve örtük geribildirimleri ile oluşturulacak veri setinin pek çok özelliğe sahip olması gerekmektedir.

Bu tez kapsamında, internette kullanıma açık olarak sunulan nadir haber öneri veri setlerinden birisi olan YOW veri seti kullanılmıştır. YOW veri seti

uzun zaman önce, 2004 yılında oluşturulmuş olmasına rağmen, bu tez kapsamında sunulan yöntemleri test etmek ve karşılaştırmak için gerekli olan pek çok özelliğe sahiptir. YOW veri seti ABD Carnegie Mellon Üniversitesi'nde yürütülen Yow-now haber filtreleme sistemi için oluşturulmuştur (Zhang, 2003). Yow-now projesi kullanıcılara farklı RSS kaynaklarından haber ulaştıran bir bilgi filtreleme sistemidir. Bu proje kapsamında bir ay boyunca yaklaşık 25 kullanıcıdan haber makaleleri hakkında 7000'den fazla geribildirim toplanmıştır. Oluşturulan veri setinde hem açık hem de örtük geribildirimler bulunmaktadır. Açık geribildirimler 1 ile 5 arasındaki oylardan oluşmaktadır. 1 kullanıcının ilgili makaleyi beğenmediği 5 ise çok beğendiği anlamına gelmektedir. Toplamda 383 haber makalesi her kullanıcı tarafından oylanmıştır. Örtük geribildirim olarak ise kullanıcıların bir haber makalesini okumak için o sayfada harcadığı zaman, fare, klavye ve ekran kaydırma gibi hareketleri kaydedilmiştir.

YOW veri setinin bir başka bölümünde, kullanıcılar tarafından oylanmış haber makalelerinin RSS besleme kayıtları da bulunmaktadır. Bu bölüm veri setinden sorumlu olan Yi Zhang'tan araştırma amaçlı olarak istenmiştir. Veri setinin bu kısmında her haber makalesinin RSS numarası (RSS_ID), başlığı, web adresi (URL) ve ilk 25-30 sözcükten oluşan giriş metni bulunmaktadır. Böylece YOW veri setinin iki bölümü birlikte bir kullanıcının haber makalelerini okuma geçmiş, geribildirimleri ve haber içeriklerini içeren bir bütün halindedir.

4.2 Anlamsal Haber Kaynağı Öneri Sistemi Detayları

Bu tez kapsamında bir anlamsal haber kaynağı öneri sistemi önerilmiştir. Yapılan literatür taramasında çeşitli anlamsal öneri sistemi yöntemleri ve örnekleri ile karşılaşılmışa rağmen, haber kaynağı öneri sistemine rastlanmamıştır. Bu bölümde detayları açıklanan sistemde kullanılan anlamsal çıkarsama ve birliktelik kuralları Bölüm 2.1.4 ve Bölüm 2.1.5'de detaylı olarak açıklanmıştır. Ayrıca anlamsal haber öneri sistemleri Bölüm 3.2'de açıklanmış, literatürden örnekler verilmiştir. Bu tez kapsamında tasarlanan sistemde haber kaynağı önerisi için iki farklı yöntem sunulmuş ve karşılaştırılmıştır. Bu yöntemlerden bir tanesi haber başlık ve özet metinleri üzerinde birliktelik kuralları ile yapılan analizler sonucu haber kaynakları arasındaki ilişkileri bulmak üzerine kurulmuştur. Diğer yöntem ise haber kaynakları ontolojisinin oluşturulması ve ontolojideki kaynaklar arasındaki benzerliklerin anlamsal ilişkiler kullanılarak bulunması üzerine kurulmuştur. Haber kaynakları arasındaki benzerliklerin bulunmasından sonra öneriler yapılabilir. Bu bölümde detayları açıklanan sistemde kullanılan anlamsal çıkarsama ve birliktelik kuralları Bölüm 2.1.4 ve Bölüm 2.1.5'de detaylı olarak açıklanmıştır. Ayrıca anlamsal haber öneri sistemleri Bölüm 3.2'de açıklanmış, literatürden örnekler verilmiştir. Bu tez kapsamında tasarlanan sistemde haber kaynağı önerisi için iki farklı yöntem sunulmuş ve karşılaştırılmıştır. Bu yöntemlerden bir tanesi haber başlık ve özet metinleri üzerinde birliktelik kuralları ile yapılan analizler sonucu haber kaynakları arasındaki ilişkileri bulmak üzerine kurulmuştur. Diğer yöntem ise haber kaynakları ontolojisinin oluşturulması ve ontolojideki kaynaklar arasındaki benzerliklerin anlamsal ilişkiler kullanılarak bulunması üzerine kurulmuştur. Haber kaynakları arasındaki benzerliklerin bulunmasından sonra öneriler yapılabilir.

NumOfUpArrow	MSecForPageDown	RSS_ID	readability	TimeOnHScroll	novelty	user_like	DOC_ID	TimeOnPage	Time/visit	user_id	NumOfDownArrow	authority
0	0	0	4	1	0	5	276101	1843	4/21/2004 10:25	63	0	1
1	351	4	4	1	0	5	276101	104330	4/21/2004 10:28	63	19	1
0	0	4	4	1	0	5	275841	7530	4/21/2004 10:29	63	0	1
1	0	4	4	1	0	3	275841	186778	4/21/2004 10:32	63	23	1
0	0	4	4	1	0	5	275840	6450	4/21/2004 10:32	63	0	1
0	0	4	4	1	0	5	275840	120585	4/21/2004 10:34	63	17	1
0	0	8	8	1	0	5	275877	12388	4/21/2004 10:36	63	0	1
0	0	8	8	1	0	5	275877	236711	4/21/2004 10:40	63	21	1
1	580	3	3	1	0	3	276097	10685	4/21/2004 10:42	63	0	1
0	0	2	2	1	0	5	276089	61258	4/21/2004 11:05	63	11	1
0	1332	2	2	1	0	3	276089	2343	4/21/2004 11:11	63	0	1
0	0	2	2	1	0	3	276091	149796	4/21/2004 11:22	63	30	1
2	1051	2	2	1	0	5	276091	8642	4/21/2004 11:22	63	0	1
0	0	4	4	1	0	5	276103	163305	4/21/2004 11:30	63	25	1
3	721	4	4	1	0	5	276103	8953	4/21/2004 11:30	63	0	1
0	0	7	7	1	0	5	276103	134514	4/21/2004 11:41	63	18	1
							276111	9243	4/21/2004 11:42	63	0	1

Şekil 4.2: YOW veri setinin ilk bölümü detaylı kullanıcı bilgilerini içeriyor.

doc ID	Title	RSS ID	Description	URL
6393	Catering to Small and Medium-Size Enterprises	6802	With opportunities in the large	newsfeed.technologyvalua
10280	Attacks in Mideast Raise Fear of More at Oil Installations	8197	With oil prices climbing to a 13	www.nytimes.com/2004/05/
11652	The Internet's Wilder Side	8209	With much of the Internet begi	www.nytimes.com/2004/05/
3497	First Napping Librarian Award winner!	1437	With much appreciation to the	www.ebn.weblogger.com/de
11917	UAE University and IBM hope to make roads safer (ITP.net)	8257	With more than one person be	us.rd.yahoo.com/dailynews/
1151	San Fran pictures	1690	With mom off staying with my	holycola.net/archives/0004?
5320	R&D innovations to drive growth in osteoporosis market	6609	With key R&D trends in the fie	www.pharmaceutical-busine
11647	Nets Abruptly Change Tone of Series	8208	With Jason Kidd struggling thr	www.nytimes.com/2004/05/
11776	As Inquiries Press Ahead, 2 Directors Quit at Azteca	8197	With its chairman under invest	www.nytimes.com/2004/05/
1512	When We're the Evildoers in Iraq	3412	With immoral U.S. leadership,	www.latimes.com/news/opi
86	What we can learn from Ikea blowing a \$350 sale	548	With any automated system,	www.healyourchurchwebsite
805	Microsoft premieres DRM technology	870	With an eye to capturing a larg	www.infoworld.com/cgi-bin/i
4715	What About the ITU?	2265	With all the talk about killing	www.snappingturtle.net/jmc
2286	Howie's	1821	With all the clout they carry, it	birdhouse.org/blog/archives
2858	Walking the Pitcher	1908	With a man on second, 1 out	www.baseballmusings.com/

Şekil 4.3: YOW veri setinin ikinci bölümü haber makalelerinin bilgilerini içeriyor.

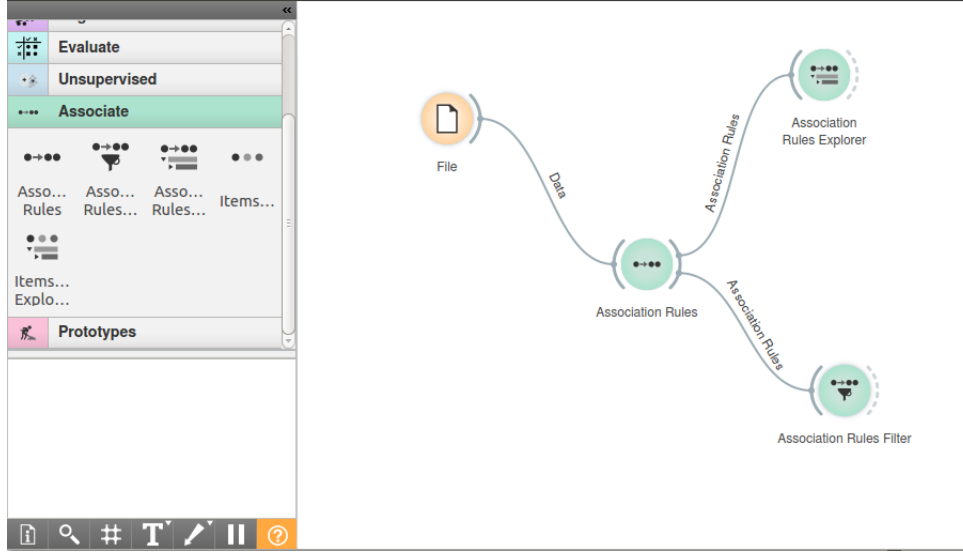
4.2.1 Birliktelik kuralları

Birliktelik kuralları ile ilgili tanımlamalar ve literatürdeki çalışmalara Bölüm 2.1.5’de yer verilmiştir. Bölüm 4.1’de söz edildiği gibi bu tez kapsamında birliktelik kurallarının çıkarılması ve analizinde Orange aracı kullanılmıştır. Şekil 4.4’te görülen, YOW veri setinden elde edilen veri dosyası Orange için girdi dosyası olarak kullanılmıştır. Bu veri dosyasının oluşturulmasında hesap tablosu (spreadsheet) kullanılmıştır. Bu veri dosyası kullanılarak birliktelik kuralları çıkarılmış ve farklı haber kaynakları arasındaki ilişkiler bulunmaya çalışılmıştır. Veri dosyasının ilk kolonu her biri farklı bir haber kaynağına ait RSS_ID değerini içermektedir. Diğer kolonlar toplam 24 kullanıcının bu haber kaynaklarından okudukları haber makalesi sayısını göstermektedir. Şekildeki son kolon ise her haber kaynağından okunan toplam haber makalesi sayısıdır. Bu örnek yukarıda söz edilen alışveriş listesi örneği ile karşılaştırılacak olursa, her RSS_ID bir alışveriş ögesi ve her kolon bir alışveriş olmaktadır. Birliktelik kuralları çıkarılırken sadece 26 satır göz önüne alınmıştır. Şekil 4.4’deki renklendirilmiş satırlar bu 26 satırdan bazılarıdır. Bunun nedeni bu veri setinin karşılaştırma sırasında Şekil 4.3’te örneği verilen haber içeriklerinin bulunduğu veri seti ile ilişkilendirilecek olmasıdır. Hem haber içeriklerinin analiz edilip ontolojiye aktarılacak kadar fazla, hem de kullanıcılar tarafından yeterince okunmuş olması istenmiştir.

Şekil 4.5’de Orange arayüzü kullanılarak birliktelik kurallarının nasıl çıkarıldığı gösterilmiştir. Buna göre “File” veri dosyasını içermektedir. Veri dosyasından okunan veriler birliktelik kurallarının çıkarılması için “Association Rules” (Birliktelik Kuralları) ögesine bağlanmıştır. Burada çıkarılan birliktelik kuralları, kuralların düzenli ve biçimli bir şekilde görüntülenebilmesi için “Association Rules Explorer” (Birliktelik Kuralları Araştırmacı) ögesine bağlanmıştır. Bu ögenin detayları Şekil 4.6’da görülmektedir. Burada çıkarılan birliktelik kurallarını görüntülemek ve detaylı olarak incelemek mümkündür. Örneğin; {8172 → 8155} kuralına göre, 8172 numaralı (RSS_ID’ye sahip) haber kaynağından haber okuyanlar genellikle 8155 numaralı kaynaktan da haber okumuşlardır. Bu kuralın destek sayısı 0,4 ve güven sayısı 0,9’dur. Bu arayüz kullanılarak birliktelik kurallarının kapsamı (coverage) ve dayanıklılığı (strength) gibi özellikler de görüntülenebilmekte, kuralların görüntülenme ayarları değiştirilebilmektedir.

Yine Şekil 4.5’te görülebileceği gibi kuralların analizi ve filtrelenmesi için çıkarılan kurallar “Association Rules Filter” (Birliktelik Kuralları Filtresi) ögesine bağlanmıştır. Bu ögenin detayları Şekil 4.7’de görülmektedir. Bu ekranda

Şekil 4.4: YOW veri seti kullanılarak elde edilen Orange girdi dosyası örneği.



Şekil 4.5: Orange arayüzü kullanılarak birliktelik kurallarının çıkarılması.

Association Rules Explorer

Info
Number of rules: 218
Selected rules:
Selected examples:
... matching:
... mismatching:

Options
Tree depth: 2
☒ Display whole rules

Send selection
☒ Purge attribute values/attributes
☒ Purge class attribute
☒ Send immediately
Send

Shown measures: ☒ Support ☒ Confidence ☐ Lift ☐ Leverage ☐ Strength ☐ Coverage

Rules	Supp	Conf
8172		
8172 -> 8155	0.4...	0.900
8172 -> 8154 8155	0.4...	0.900
8172 -> 8195	0.4...	0.900
8172 -> 8154 8195	0.4...	0.900
8172 -> 8154	0.4...	1.000
8154		
8155		
8195		
3616		
3616 -> 8154	0.4...	1.000
8195		
8195 -> 8155	0.5...	0.667
8195 -> 8156 8155	0.4...	0.556
8195 -> 8156 8154 8155	0.4...	0.556
8195 -> 8154 8155	0.5...	0.667
8195 -> 8154 8155 8153	0.4...	0.500
8195 -> 8155 8153	0.4...	0.500
8195 -> 8172	0.4...	0.500
8195 -> 8172 8154	0.4...	0.500
8195 -> 8157	0.4...	0.556
8195 -> 8157 8154	0.4...	0.500
8195 -> 8157 8154 8153	0.4...	0.500
8195 -> 8157 8153	0.4...	0.556

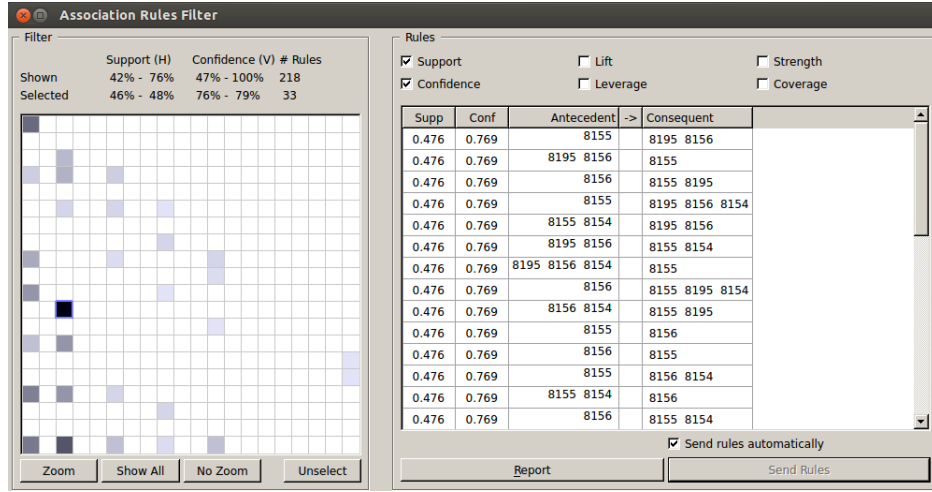
Report

Şekil 4.6: Orange arayüzünde birliktelik kurallarının görüntülenmesi.

kurallar destek ve güven sayılarına göre gruplanmış ve bir tabloda gösterilmiştir. Soldaki grafikte X ve Y koordinatları destek ve güven sayılarını temsil etmektedir. Renklendirilmiş her kare belirli bir destek ve güven sayısına sahip kurallardan oluşmaktadır. Karelerin renkleri ise o grupta bulunan kuralların sayısı arttıkça koyulaşmaktadır. Bu araç, çıkarılan birliktelik kurallarının destek ve güven sayılarına göre analiz edilmesinde kullanılmıştır.

4.2.2 Haber kaynağı ontolojisi

Haber alanı ontoloji oluşturmak için oldukça geniş bir alandır. Haberler her türlü konuyla ilgili olabilirler. Bu da sağlıktan spora, eğlenceden politikaya kadar



Şekil 4.7: Orange arayüzünde birliktelik kurallarının filtrelenmesi.

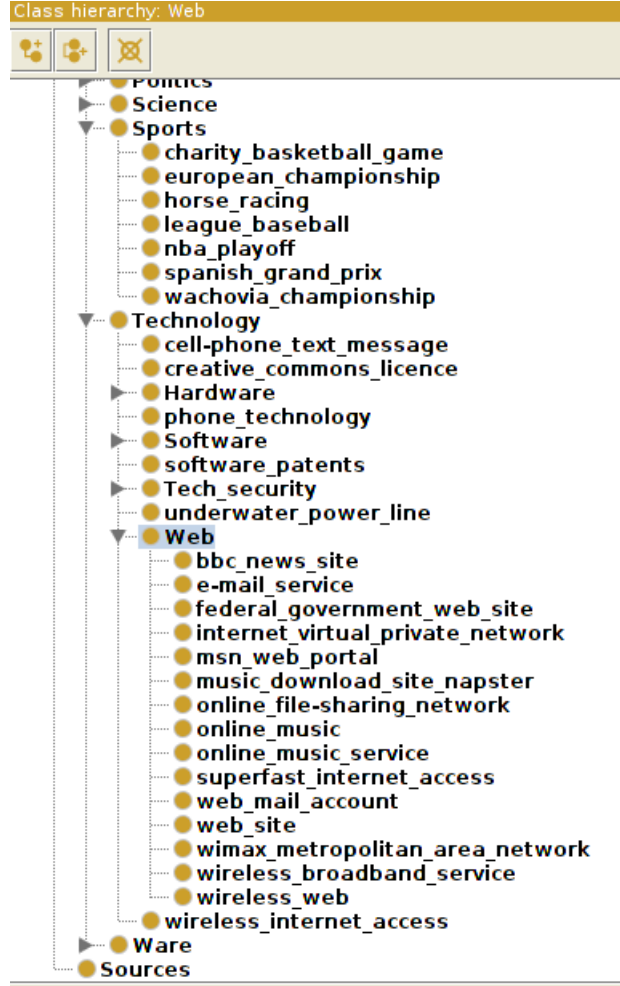
pek çok alanın modellenmesi anlamına gelmektedir. Tek başına sağlık ya da politika gibi alanların bile ontolojilerinin oluşturulması oldukça zorken bunların tamamı için durum daha da zorlaşmaktadır. Diğer yandan haber başlık ve metnilerindeki önemli terimlerin belirlenmesi, ait oldukları kaynaklarla ve birbirleri ile ilişkilendirilmesi öneri sistemleri açısından oldukça önemli katkılara sahiptir. Bu tez kapsamında YOW veri seti (Bölüm 4.1.4) kullanılarak küçük çaplı bir haber kaynağı ontolojisi oluşturulmuştur.

Ontoloji oluşturulmadan önce Bölüm 4.1.2’de açıklanan Termine kullanılarak her haber kaynağına ait metinlerin toplamından önemli terimler çıkarılmıştır. Önemli terimlerin çıkarılması için her haber kaynağının RSS beslemesinden toplanan haber başlıkları ve haber metninin giriş kısımları birlikte kullanılmıştır. Şekil 4.8’de Termine web arayüzünün kullanımından elde edilen sonuçlardan bir tanesi görülmektedir. Termine, verdiği sonuçları C-value değerine göre sıralamaktadır. C-value, dilbilimsel ve istatistiksel analizleri kullanan, otomatik terim tanıması için geliştirilmiş alan bağımsız bir yöntemdir. Sonuçlarda görünen “score” değeri çıkarsanan terimin metin içindeki önemini ifade etmektedir.

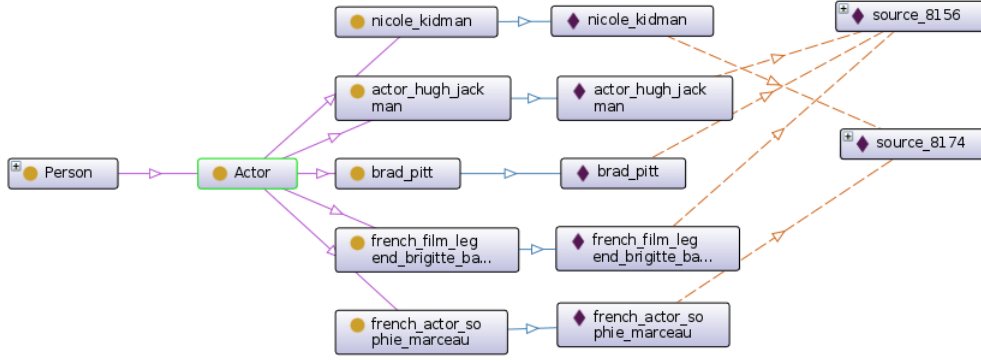
İkinci adımda Termine yardımı ile çıkarılan bu önemli terimler sınıflandırılarak ontolojiye aktarılmıştır. Ontolojinin sınıf hiyerarşisinin bir bölümü Şekil 4.9’da görülmektedir. Her haber kaynağı ontolojide bir eleman olarak yer almaktadır. Haber kaynakları ile haber kaynağından çıkarsanan önemli terimler arasında “has_source” (kaynağıdır) ilişkisi, bu ilişkinin tersi olarak da “includes” (içerir) ilişkisi tanımlanmıştır. Ontoloji oluşturulurken terimlerin c-value değerleri göz önüne alınmamıştır. Ancak bu değerlerin ontoloji ile ilişkilendirilmesi gelecek çalışmalarda değerlendirilecektir.

Rank	Term	Score
1	new york botanical garden	2
1	new york	2
1	new york city ballet	2
1	international fine art fair	2
1	actor writer director activist	2
6	grand central terminal	1.584962
6	red sox slugger	1.584962
6	irene diamond concert	1.584962
6	director lonny price	1.584962
6	sportive sexual habit	1.584962
6	19th-century american art	1.584962
12	commercial title	1
12	wynton marsalis	1
12	leigh montville	1
12	tough audience	1
12	anonymous buyer	1
12	united states	1
12	betty miller	1
12	english singer-songwriter	1
12	carnegie hall	1
12	character actor	1
12	southern california	1
12	smart-talking comedy	1
12	deputy director	1

Şekil 4.8: Termine web arayüzü kullanılarak elde edilmiş analiz sonucu örneği.



Şekil 4.9: Haber kaynağı ontolojisi sınıf hiyerarşisi örneği.



Şekil 4.10: Ontolojideki “Actor” (Oyuncu) sınıfının içerdiği haber kaynakları.

Ontolojideki Haber Kaynaklarının Benzerliklerinin Hesaplanması

Oluşturulan ontolojideki haber kaynaklarının birbiri ile benzerliği iki şekilde incelenmiştir. İlk yöntemde haber kaynaklarının içerdikleri benzer öğelerin sayısı dikkate alınmamıştır. İkinci yöntemde ise haber kaynakları arasındaki Jaccard benzerlik katsayısı hesaplanmıştır. Ayrıca anlamsal çıkarsama kullanılarak haber kaynakları arasındaki açıkça görülemeyen ilişkiler belirlenmiştir.

İlk yöntemde haber kaynakları arasındaki ilişkileri belirlemek için bir haber kaynağının belirli bir sınıf içerisinde yer alıp almadığı ve kaç kez yer aldığı bilgisi kullanılmıştır. Örneğin; Şekil 4.10’da görülen “Actor” (Oyuncu) sınıfındaki öğeler iki haber kaynağından (RSS_ID 8156 ve RSS_ID 8174) gelmektedir. Buna göre bu iki haber kaynağında oyuncuların söz edildiği anlaşılmaktadır. Her haber kaynağında oyuncularla ilgili birden fazla terimin bulunması, bunun sadece tesadüfi olmadığını, bu iki kaynağın da oyuncularla ilgili haberler içerdiği anlaşılmaktadır. Sonuç olarak, bu çıkarıma göre eğer bir kullanıcı 8156 numaralı haber kaynağındaki haberleri okumuşsa o kişiye 8174 numaralı haber kaynağını da önermek mümkündür.

Benzer olarak Şekil 4.15’de görüleceği gibi “Software” (Yazılım) sınıfında “computer worm” (bilgisayar solucanı) ile ilgili dört adet öğe bulunmaktadır. Bu öğelerden iki tanesi aynı haber kaynağından gelmekte, diğerleri ise farklı haber kaynaklarından gelmektedir. Buna göre bu üç haber kaynağının bilgisayarla ilgili benzer konular içerdiği söylenebilir. Ancak “Software” sınıfının haber kaynaklarının toplam sayısına ve dağılımına bakıldığında Sasser bilgisayar solucanından söz edilen kaynaklardan iki tanesinin (8152 ve 8155) bu sınıf içerisinde tekrarlamadığı görülmüştür. Farklı sınıflarda içerilen haber kaynakları ve dağılımları Şekil 4.16’de görülmektedir. Bu şekil incelendiğinde 8152 numaralı kaynağın “Web” sınıfında

birden fazla kere içerildiğini görmek mümkündür. Yani 8152 numaralı kaynak yazılım sınıfı ile çok ilişkili olmasa da web sınıfı ile oldukça ilişkilidir ve genel olarak bilgisayar teknolojileri konusunda ilgili olan kullanıcılara önerilebilir.

İkinci yöntemde ise ontolojide yer alan haber kaynaklarının benzerliklerinin hesaplanması için Jaccard benzerlik katsayısı kullanılmıştır. Bu yöntemin seçilme sebebi diğer benzerlik ölçütlerine göre daha uygun ve sade olması ve literatürde çokça kullanılmasıdır. Literatürde yer alan konuyla ilgili benzerlik ölçütlerinden bazıları şöyledir:

En Kısa Yol Yöntemi

Bu yöntem (Rada et al., 1989) tarafından önerilmiştir. Bu yönteme göre ontoloji hiyerarşisinde iki öge arasındaki anlamsal benzerlik, aralarındaki en kısa yoldur. c_1 ve c_2 ögeleri arasındaki benzerlik bu yöntemle şu şekilde hesaplanmaktadır:

$$Sim(c_1, c_2) = 2MAX - L \quad (4.1)$$

Buna göre c_1 ve c_2 aralarındaki benzerliğin hesaplanacağı iki öge, MAX hiyerarşideki en uzun yol, L ise en kısa yoldur. Bu yöntemin hesaplama karmaşıklığı düşük olduğundan avantaj sağlamaktadır.

Ortak Yol Yöntemi

Ortak yol (İng. Common Path) yöntemi iki terimin kavramsal benzerliğini bulmak için kullanılan yöntemlerden birisidir. (Wu and Palmer, 1994) tarafından önerilmiş olan bu yöntemde terimlerin en yakın ortak atasından köke kadar olan uzaklıkları kullanılır. Buna göre c_1 ve c_2 kavramları arasındaki benzerlik hesaplaması eşitlik 4.2’de gösterilmiştir. Bu eşitlikte D1 ve D2, c_1 ve c_2 kavramlarından, bu kavramların en yakın ortak atası olan c kavramına olan uzaklığı ifade etmektedir. H ise c kavramından köke kadar olan uzaklıktır.

$$Sim(c_1, c_2) = \frac{2H}{D1 + D2 + 2H} \quad (4.2)$$

Kavram Eşitliği Yöntemi

Ontolojideki iki kümenin benzerliğini bulmak için (Intema et al., 2010)'da tanımlanmıştır. Tanımlandığı çalışmada bir haber öneri sisteminde bir kullanıcı ve bir haber makalesinin benzerliğini bulmak için kullanılmıştır. Buna göre C ontolojideki kavramlar kümesi olmak üzere, U kümesi bir kullanıcının beğendiği kavramlar kümesi, A ise bir haber makalesinde geçen kavramlar kümesidir.

$$U = c_1^u, c_2^u, c_3^u, \dots, c_p^u, c_i^u \in C$$

$$A = c_1^a, c_2^a, c_3^a, \dots, c_p^a, c_j^a \in C$$

U ve A kümeleri arasındaki benzerlik ise bu iki kümenin kesişim eleman kümesi sıfırdan büyük ise 1, değilse 0 değerini alır. Bunun matematiksel ifadesi eşitlik 4.3'te görülmektedir. Bu yöntem ikilik sistemde, ya 0 ya da 1 olan bir benzerlik ölçütüdür.

$$Similarity(U, A) = \begin{cases} 1 & \text{if } |U \cap A| > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.3)$$

Jaccard Benzerlik Ölçütü

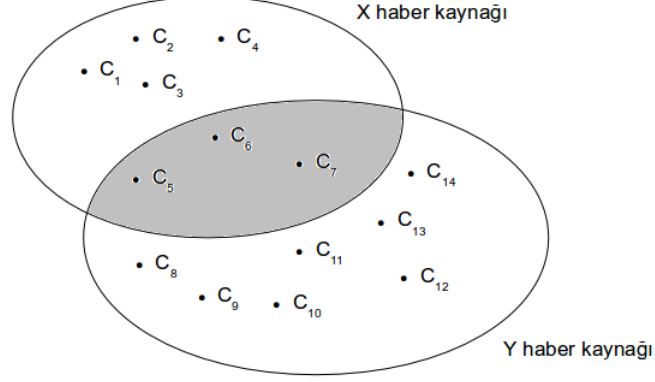
A ve B kümelerinin benzerlik oranını veren Jaccard katsayısı şöyle hesaplanır:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (4.4)$$

Buna göre A ve B iki sonlu küme olmak üzere Jaccard katsayısı, A ve B kümelerinin kesişim eleman sayısının birleşim eleman sayısına bölünmesi ile bulunur. Bu tez kapsamında geliştirilen haber kaynağı ontolojisindeki haber kaynakları, farklı haberlerden önemli terimleri içeren birer kümedir. Buna göre iki haber kaynağının içerdiği önemli terimlerin kesişim ve birleşim kümelerindeki eleman sayısının birbirine olan oranı, A ve B haber kaynakları arasındaki benzerlik oranını vermektedir. Bu oran 0 ve 1 arasında bir değerdir.



Şekil 4.11: Ontolojideki “Software” (Yazılım) sınıfındaki bir ögeyi içeren haber kaynakları.

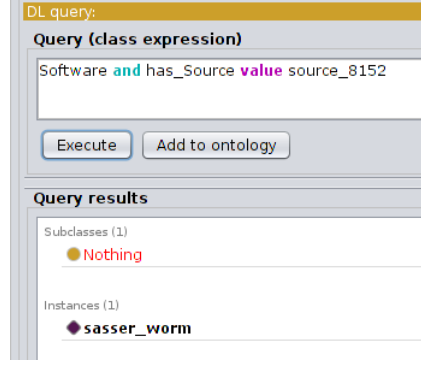


Şekil 4.12: X ve Y haber kaynaklarındaki öğelerin küme diyagramı.

$$0 \leq J(A, B) \leq 1 \quad (4.5)$$

Jaccard benzerlik ölçütü kullanılarak benzerliği hesaplanan iki haber kaynağı Şekil 4.11’de gösterilmiştir. Bu şekilde görüldüğü üzere ontolojideki “Software” (Yazılım) sınıfındaki bir öge (“sasser_worm” ögesi) iki farklı haber kaynağında yer almaktadır. Bu öge hem kaynak 8172’de hem de 8152’de yer almıştır. Buna göre bu iki kaynağın “Software” (Yazılım) sınıfındaki benzerliklerini hesaplamak istersek, bu iki kaynaktaki yer alan tüm kesişim ve birleşim öğelerinin sayılarına bakmamız gerekir. Şekil 4.11’de bu iki kaynağın kesişim kümesinden sadece 1 örnek gösterilmiştir. Bu şekilde ortak terimlere sahip tüm haber kaynakları çıkarılmış ve EK 3’te listelenmiştir. Şekil 4.12’de ise X ve Y haber kaynaklarındaki öğelerin kesişim kümesi bir diyagram ile gösterilmiştir. Bu diyagrama göre X ve Y haber kaynaklarının Jaccard benzerlik katsayısı $\frac{3}{14} = 0,214$ olmaktadır.

Ontolojide hangi sınıfta, hangi haber kaynağından toplam kaç öge olduğunu görebilmek için Protege’in “DL Query” özelliği kullanılarak “Class_name and has_Source value source_name” şeklinde sorgular yazılmıştır. Çıkarsama motoru olarak ise HermiT 1.3.8.3 kullanılmıştır. Bu sorgulardan ve sonuçlarından bir örnek Şekil 4.13’de görülmektedir. Buna göre ontolojide aynı ögeye sahip tüm haber kaynakları arasındaki benzerlikler Jaccard katsayısı kullanılarak hesaplanmıştır. Bir örnek vermek gerekirse, ontolojideki “Technology” (Teknoloji) sınıfı için 8152 ve



Şekil 4.13: Protege’de “Software” (Yazılım) sınıfı için bir DL sorgu örneği ve sonucu.

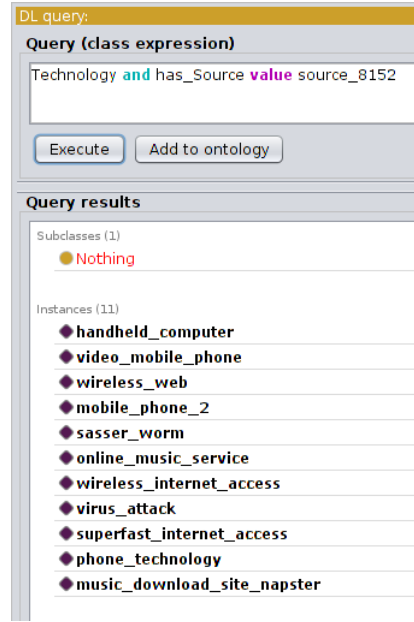
8172 numaralı kaynaklar arasındaki benzerlik şöyle hesaplanabilir:

$$J(8152, 8172) = \frac{|8152 \cap 8172|}{|8152 \cup 8172|} = \frac{3}{24} = 0,125 \quad (4.6)$$

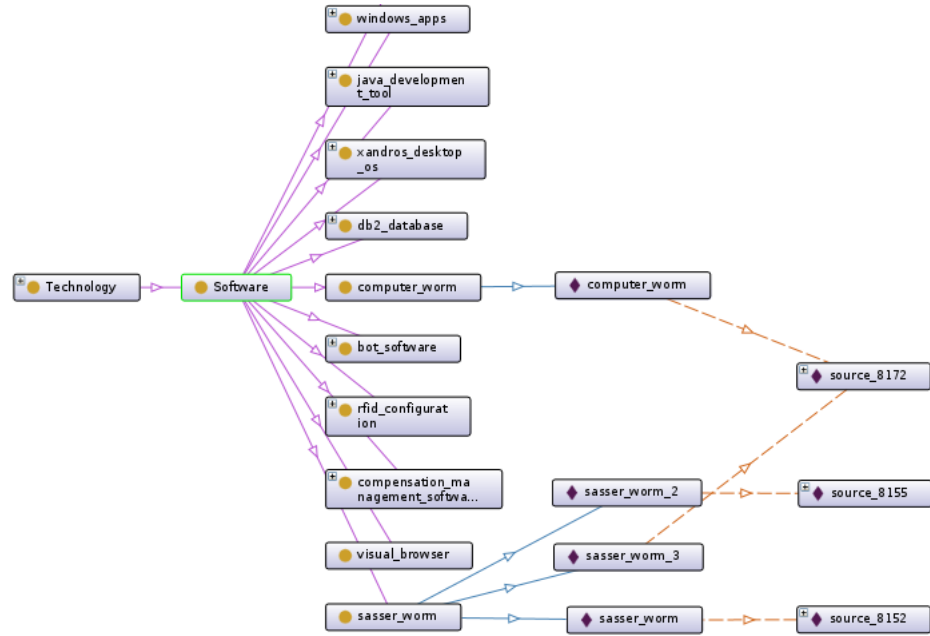
Yukarıdaki hesaplama göre 3, her iki haber kaynağında birden yer alan öğelerin sayısı, yani kesişim kümesindeki eleman sayısı, 24 ise her iki haber kaynağında yer alan öğelerin toplam sayısı, yani birleşim kümesindeki eleman sayısıdır. Bu şekilde hesaplanan tüm sonuçlar Ek-3’te listelenmiştir. Bu yöntemle hesaplanan benzerlik oranları her sınıf için ayrı ayrı hesaplanabilmektedir. Örneğin 8152 ve 8172 numaralı kaynakların benzerlikleri yukarıdaki örnekte olduğu gibi “Technology” (Teknoloji) sınıfına göre ya da “Technology” (Teknoloji) sınıfının bir alt sınıfı olan “Software” (Yazılım) sınıfına göre hesaplanabilir. Şekil ’da görülen sorgulama “Software” (Yazılım) sınıfında yapılmışken, Şekil ’da aynı haber kaynağı için “Technology” (Teknoloji) sınıfında sorgulama yapılmıştır. Şekillerde sorgu sonuçlarının farklılığı görülebilir. Yani her bir alt sınıf için tek bir haber kaynağının içerdiği öğe sayısı değişmektedir. Bu da Jaccard benzerlik sonucunu etkilemektedir. EK 3’te sunulan tabloda haber kaynaklarının benzerlikleri üst ve alt sınıflar için ayrı ayrı gösterilmiştir. “Software” (Yazılım) sınıfı için 8152 ve 8172 numaralı kaynaklar arasındaki benzerlik şöyledir:

$$J(8152, 8172) = \frac{|8152 \cap 8172|}{|8152 \cup 8172|} = \frac{1}{2} = 0,5 \quad (4.7)$$

Eşitlik 4.3 ev 4.4’te gösterildiği gibi iki haber kaynağının arasındaki benzerlik bir üst sınıfta (“Technology” (Teknoloji) sınıfında) 0,125 iken bir alt sınıf (“Software” (Yazılım)) için hesaplandığında 0,5 çıkmaktadır. Buna göre benzerlik oranları alt sınıflara inildikçe artmaktadır.



Şekil 4.14: Protege’de “Technology” (Teknoloji) sınıfı için bir DL sorgu örneği ve sonucu.



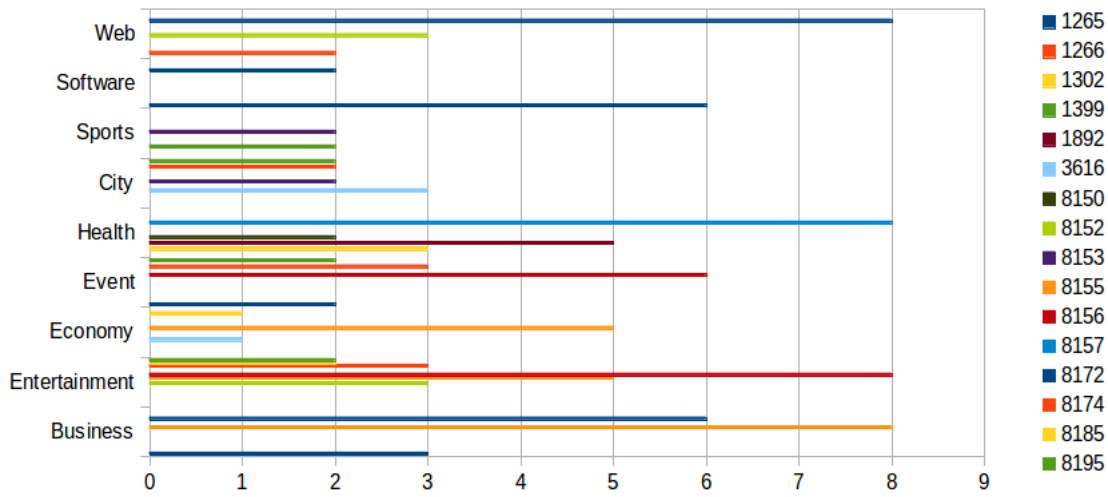
Şekil 4.15: Ontolojideki “Software” (Yazılım) sınıfının içerdiği haber kaynakları.

	1265	1266	1302	1399	1892	3616	8150	8152	8153	8155	8156	8157	8172	8174	8185	8195
Business	3									8			6			
Entertainment								3		5	8			3	2	2
Economy						1				5					1	
Event	2										6			3		2
Health			3		5		2					8				
City						3			2					2		2
Sports				2					2							
Software	6												2			
Web		2						3					8			

Şekil 4.16: Ontolojideki sınıflardan bazılarının sahip oldukları RSS_ID sayıları.

RSS_ID	Ontolojideki öge sayısı
1265	33
1266	7
1302	5
1399	7
1483	7
1892	7
1964	4
3616	15
3630	8
5580	14
8073	6
8150	7
8151	12
8152	19
8153	13
8154	16
8155	28
8156	31
8157	14
8172	29
8174	20
8185	10
8186	4
8187	4
8190	16
8195	13

Çizelge 4.1: RSS_ID'lerin ontolojide yer alma sıklığı.



Şekil 4.17: Ontolojideki kategorilerden bazılarının sahip oldukları RSS_ID sayıları grafiği.

	Sınıf Adı	RSS_ID (Bulunma Sayısı)					
1	Business	8155 (8)	8172 (6)	1265 (3)			
2	Entertainment	8156 (8)	8155 (5)	8152 (3)	8174 (3)	8195 (2)	8185 (2)
3	Health	8157 (8)	1892 (5)	1302 (3)	8150 (2)		
4	Event	8156 (6)	8174 (3)	8195 (2)	1265 (2)		
5	City	3616 (3)	8153 (2)	8195 (2)	8174 (2)		
6	Software	1265 (6)	8172 (2)				
7	Web	8172 (8)	8152 (3)	1266 (2)			

Çizelge 4.2: Ontolojiden elde edilen haber kaynakları arasındaki ilişkiler.

4.2.3 Ontoloji ve birliktelik kuralları sonuçlarının karşılaştırılması

Bu tez kapsamında kullanıcılara haber kaynağı önerisinde bulunmak için birliktelik kurallarının ve ontolojilerin kullanıldığı iki farklı yöntem sunulmuştur. Ontoloji ve anlamsal çıkarsama sonucunda bazı haber kaynakları arasındaki ilişkiler iki farklı şekilde tespit edilmiştir. Bunlardan ilki her haber kaynağının belirli sınıflar içindeki bulunma sayısıdır. Bu sonuçlar Şekil 4.16’de sayısal olarak ifade edilmiştir. Şekil 4.17’te ise bu sayılar ve ilişkiler grafik olarak görülmektedir. Buna göre elde edilen ilişkiler ise Çizelge 4.2’de görülmektedir. Bu çizelgede bir haber kaynağının her bir sınıfta bulunma sayısı, her RSS_ID’nin yanında parantez içinde ifade edilmiştir. Buna göre bulunma sayısı yüksek olan haber kaynaklarının birbiri ile ilişkili olduğu anlaşılmaktadır. Haber kaynakları arasındaki ilişkileri belirlemede kullanılan ikinci yöntem ise Jaccard benzerlik ölçütüdür. Buna göre haber kaynakları arasındaki Jaccard benzerlik katsayıları bir önceki bölümde anlatıldığı gibi bulunmuştur. İlişkili olan haber kaynakları, aralarındaki benzerlik katsayıları ve bu ilişkilerin birliktelik kurallarındaki destek ve güven sayıları EK 3’teki tabloda verilmiştir.

Elde edilen birliktelik kurallarından bazı örnekler Şekil 4.18, Şekil 4.20, Şekil 4.21, Şekil 4.22 ve Şekil 4.19’te görülmektedir. Buna göre 8172 ve 8155 numaralı haber kaynakları arasında 0,429 destek sayısı ve 0,9 güven sayısına sahip bir ilişki bulunmaktadır. Benzer şekilde 8156 ve 8154 numaralı haber kaynakları arasında 0,6 destek sayısı ve 1,0 güven sayısına sahip bir ilişki bulunmaktadır. Şekillerde yer alan birliktelik kurallarının özeti Çizelge 4.3’te verilmiştir. Ancak destek sayısı 0,5’in altında olan kuralların tesadüf sonucu oluşmuş olma olasılığı yüksektir. Bu nedenle ontolojiden elde edilen sonuçlar ile tutarlı olan ve destek sayıları 0,5’in üzerinde olan kurallar ayrılmıştır. Bu haber kaynakları arasındaki kurallardan ontolojiden çıkarılmış olan ilişkilerle tutarlı olanlar Çizelge 4.4’de gösterilmiştir. Çizelge 4.4’de özetlenen birliktelik kurallarının tamamının destek sayısı 0,5 ve üzerindedir.

Rules	Supp	Conf
8172		
8172 -> 8155	0.429	0.900
8172 -> 8155 8195	0.381	0.800
8172 -> 8156 8155 8195	0.333	0.700
8172 -> 8156 8154 8155 8195	0.333	0.700

Şekil 4.18: 8172 ve 8155 RSS_ID'lerine sahip haber kaynakları arasındaki birliktelik kuralı.

8156 -> 8157 8154 8153	0.381	0.615
8156 -> 8157 8153	0.381	0.615
8156 -> 8154	0.619	1.000
8156 -> 8154 8153	0.476	0.760

Şekil 4.19: 8156 ve 8154 RSS_ID'lerine sahip haber kaynakları arasındaki birliktelik kuralı.

Çıkarılan birliktelik kurallarının tamamı EK 2'de sunulmuştur.

Sonuç olarak ontolojiden (bulunma sayısı yöntemi ile) elde edilen sonuçlarla birliktelik kurallarının örtüştüğü kurallar (Çizelge 4.4) göz önüne alındığında, toplam 14 birliktelik kuralının 8 tanesinin tutarlı olduğu gözlemlenmiştir. Bu da $\frac{8}{14} \times 100 = 57$ yani yaklaşık %60 oranında bir tutarlılığa karşılık gelmektedir. Eğer destek sayısı 0,5'in üzerinde tutulmazsa bu oran daha da artacaktır. Bu karşılaştırma sonucu kullanıcıların ilgisini çeken, okudukları haber kaynaklarının benzer içeriklere sahip olduğu anlamına gelmektedir. Ontolojiden Jaccard katsayısı kullanılarak elde edilen sonuçlardan tıpkı bir önceki yöntemde olduğu gibi sadece üst sınıflar dikkate alındığında, toplam 33 tane kural çıkmaktadır. Birliktelik kurallarının destek sayısı 0,4'ün üzerinde tutulduğunda, bu kurallardan sadece 8 tanesi ontolojideki kurallar ile örtüşmektedir. Ancak destek sayısı 0,3 civarında tutulduğunda bu sayı 21'e çıkmaktadır. Eğer bu rakamları yüzde değerlerine çevirecek olursak, destek sayısı 0,3 civarında olduğunda, birliktelik kuralları ve ontoloji yaklaşık %64 uyumludur. Destek sayısı 0,4 değerinden büyük olduğunda ise yaklaşık %24 uyumludur. Buna göre destek sayısı sınırı düşürüldüğünde uyumluluk artmaktadır.

Ontoloji ve birliktelik kuralları sonuçlarının karşılaştırılmasında oluşturulan ontolojinin bir bölümü ele alınmıştır. Bunun sebebi oluşturulan ontolojinin küçük ölçekli olması sebebiyle birçok sınıfta bir haber kaynağına sadece tek bir bağlantı bulunmasıdır. Yapılan karşılaştırmanın daha anlamlı ve somut olması

8153 -> 8156 8152 8195	0.381	0.667
8153 -> 8154 8195	0.524	0.917
8153 -> 8154 8152 8195	0.381	0.667
8153 -> 8195	0.571	1.000
8153 -> 8152 8195	0.381	0.667
8153 -> 8156 8157 8154	0.381	0.667
8153 -> 8156 8157	0.381	0.667

Şekil 4.20: 8153 ve 8195 RSS_ID'lerine sahip haber kaynakları arasındaki birliktelik kuralı.

8156	->	8157 8154 8153 8195	0.381	0.615
8156	->	8157 8153 8195	0.381	0.615
8156	->	8195	0.619	1.000
8156	->	8154 8195	0.619	1.000
8156	->	8154 8153 8195	0.476	0.769
8156	->	8154 8152 8153 8195	0.381	0.615

Şekil 4.21: 8156 ve 8195 RSS_ID'lerine sahip haber kaynakları arasındaki birliktelik kuralı.

8172	->	8154	0.476	1.000
8172	->	8154 8152	0.333	0.700
8172	->	8152	0.333	0.700
8195				
8156				
8154				

Şekil 4.22: 8172 ve 8152 RSS_ID'lerine sahip haber kaynakları arasındaki birliktelik kuralı.

1	8195 → 8155
2	8195 → 8156
3	8195 → 8154
4	8195 → 8153
5	8195 → 8152
6	8155 → 8195
7	8155 → 8154
8	8152 → 8195
9	8153 → 8195
10	8153 → 8154
11	8156 → 8195
12	8154 → 8152
13	8154 → 8155
14	8154 → 8156

Çizelge 4.3: Birliktelik kurallarından elde edilen tüm tekli kurallar.

	Kaynak 1	Kaynak 2	Destek Sayısı	Güven Sayısı
1	8195	8155	0,571	0,667
2	8195	8156	0,619	0,722
3	8195	8154	0,762	0,889
4	8195	8153	0,571	0,667
5	8155	8195	0,571	0,923
6	8152	8195	0,524	1,000
7	8153	8195	0,571	1,000
8	8156	8195	0,619	1,000

Çizelge 4.4: Ontoloji ile tutarlı olan birliktelik kuralları.

için, bir sınıf içinde bir haber kaynağına birden fazla bağlantı varsa bu haber kaynağı karşılaştırmaya dahil edilmiştir. Diğer yandan oluşturulan ontolojideki sınıf hiyerarşisi çok özelleştirilmemiş, iki veya üç alt sınıf derinliğine inilmiştir. Bunun nedeni, veri setinde bulunan sınırlı sayıdaki içeriğin ve haber kaynağının olabildiğince ortak sınıflar altında toplanmasını sağlamaktır. Bu tez kapsamındaki örnekte ontolojideki sınıflar çok özelleştiğinde, o sınıf içindeki ortak haber kaynağı bağlantıları azalacak ve ilişkileri bulmak daha da zorlaşacaktır. Haber alanında bir ontoloji oluşturmak oldukça zordur. Haberler hayatın içindeki her şeyi konu edinebilecekleri için tam anlamıyla oluşturulan bir haber ontolojisi her şeyi kapsamak zorundadır. Bu da bu tez kapsamında oluşturulan ontolojinin küçük ölçekli ve çok özelleşmemiş sınıf hiyerarşisine sahip olmasının nedenlerinden bir tanesidir.

Bu tez kapsamındaki çalışmada birliktelik kuralları çıkarılırken kullanıcının sadece bir haber makalesini okuyup okumadığı dikkate alınmıştır. Eğer kullanıcı haber makalesini okuduysa, bu kuralların çıkarılması için kullanılmıştır. Ancak kullanıcının haber makalesini beğenip beğenmediği ya da onu okumak için ne kadar süre harcadığı dikkate alınmamıştır. Bu kavramların dikkate alınması sonucu ontolojiden elde edilen sonuçlarla olan tutarlılığın artması beklenmektedir.

4.3 Durum Çalışması

Bu bölümde veri setinden seçilen birkaç kullanıcı için haber kaynağı önerisinin birliktelik kuralları ve anlamsal çıkarsama kullanılarak nasıl yapıldığı gösterilmiş, her iki yöntemden elde edilen kullanıcılara özel sonuçlar karşılaştırılmış ve yorumlanmıştır. Şekil 4.23’de durum çalışması için veri setinden seçilen kullanıcılar ve hangi kaynaklardan kaç kez okudukları bilgisi yer almaktadır. Kullanıcılar veri setindeki daha aktif olan kullanıcılar arasından rastgele seçilmiştir.

Çizelge 4.5’de durum çalışması için seçilen dört kullanıcı için birliktelik kuralları ile tavsiye edilen, ontolojideki ilişkilerle (Bulunma sayısı yöntemi ile) tavsiye edilen ve her iki yöntemin ortak olarak tavsiye ettiği haber kaynaklarının numaraları görülmektedir. Buna göre en fazla haber kaynağı tavsiyesi ontolojideki anlamsal ilişkiler sonucu yapılmıştır. Bu çizelgenin oluşturulmasında 4.2, 4.3 ve 4.4 numaralı çizelgelerdeki veriler kullanılmıştır.

Kullanıcıların okuma geçmişleri ontolojideki kategoriler ile karşılaştırıldığında bazı kullanıcıların daha fazla tercih ettiği ya da hiç tercih etmediği alanlar da görünür hale gelmiştir. Örneğin; 63 numaralı kullanıcı

RSS ID / User ID	63	83	82	84
1265	2	17	1	0
1266	22	0	2	3
1302	0	1	0	0
1399	0	1	0	0
1483	0	0	0	0
1892	0	0	0	0
1964	2	0	0	0
3616	3	1	0	0
3630	0	0	0	0
5580	0	5	0	0
8073	3	2	3	0
8150	1	9	9	0
8151	0	0	0	0
8152	1	12	0	0
8153	0	57	0	0
8154	4	26	14	3
8155	10	22	1	2
8156	0	7	0	1
8157	0	7	0	0
8172	9	2	1	0
8174	0	0	0	1
8185	1	0	0	0
8186	0	0	0	0
8187	0	0	0	0
8190	1	1	0	0
8195	1	1	0	4

Şekil 4.23: Durum çalışması için seçilen kullanıcılar ve çeşitli kaynaklardan okuma bilgileri.

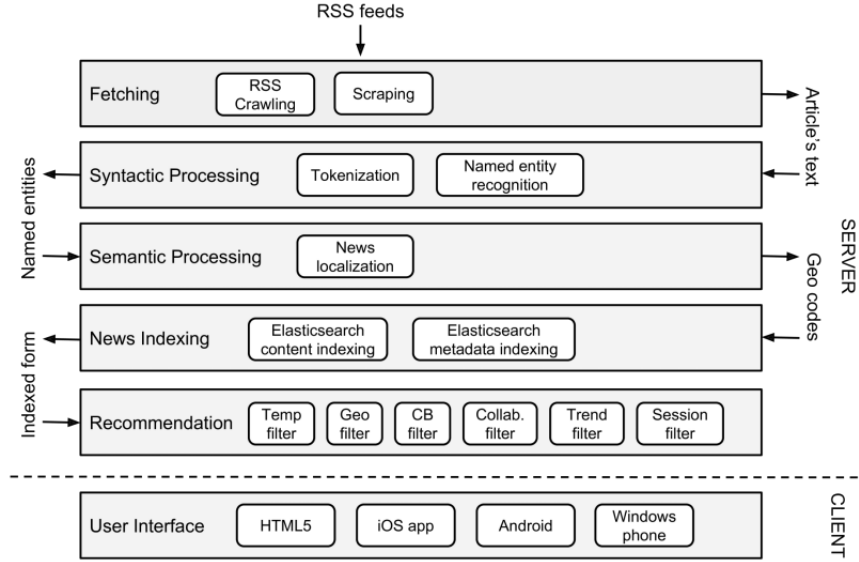
	Birliktelik Kuralları	Ontoloji (Bulunma sayısı yöntemi)	Ortak
Kullanıcı 63	8156	8174	8156
	8153	8156	8153
		8153	
Kullanıcı 82	8195	8152	8195
	8156	8157	8156
	8152	1892	8152
		1302	
		8156	
		8174	
		8195	
		8185	
Kullanıcı 83	-	8174	-
		1266	
		1892	
Kullanıcı 84	8153	1265	8153
	8152	8172	8152
		8152	
		8185	
		8153	
		3616	

Çizelge 4.5: Birliktelik kuralları ve ontoloji kullanılarak kullanıcılara tavsiye edilen haber kaynakları.

ontolojideki “Sağlık (Health)” kategorisinden, 82 numaralı kullanıcı “Şehir (City)” kategorisinden, 84 numaralı kullanıcı ise “Sağlık (Health)” ve “Yazılım (Software)” kategorilerine ait kaynaklardan hiç haber okumamış. Bunun tam tersi olarak da 82 numaralı kullanıcı tüm “İş Dünyası (Business)” ve “Yazılım (Software)” kategorilerine giren haber kaynaklarının hepsinden okumuş. Sonuç olarak, anlamsal yöntem sadece en fazla sayıda haber kaynağı tavsiyesinde bulunmakla kalmamış, aynı zamanda kullanıcıların genel ilgi alanları hakkında da yorum yapılabilmesini mümkün kılmıştır. Bu durum çalışması sadece ontolojideki bulunma sayısı yöntemi ile yapılmış, Jaccard katsayıları dikkate alınmamış ve bu iki yöntemle gerçekleştirilecek karşılaştırmalar gelecek çalışmalara bırakılmıştır.

4.4 SmartMedia Projesi

SmartMedia projesi (SmartMedia, 2012) Norveç Bilim ve Teknoloji Üniversitesi (Norges teknisk-naturvitenskapelige universitet - NTNU) Bilgisayar ve Bilgi Bilimleri Bölümü bünyesinde 2012 yılında Prof. Dr. Jon Atle Gulla yönetiminde başlamıştır. SmartMedia projesinin amacı taşınabilir cihazlar başta olmak üzere kaliteli ve uygun haber önerilerini kullanıcılara sunmaktır. Proje



Şekil 4.24: SmartMedia Projesi kapsamında geliştirilen haber öneri sistemi mimarisi.

kapsamında 2013-2014 yıllarında 12 kişi projenin kullanıcı arayüzü tasarımı, işbirlikçi filtreleme, içerik tabanlı filtreleme, konum tabanlı özelliklerin eklenmesi gibi farklı bölümlerinde görev almıştır. Şekil 4.24’de SmartMedia projesi kapsamında tasarlanan haber öneri sistemi mimarisi görülmektedir. Geliştirilen bu öneri sistemi mimarisi klasik bir sunucu - istemci mimarisidir. Sunucu günlük olarak geliştiriciler tarafından elle tanımlanmış Norveççe haber kaynaklarından RSS metinlerini almakta ve üst veriyi ayrıştırmaktadır. “Entity recognition” (öge tanıma) modülü yer ve kişi adları gibi öneri sistemi için oldukça önemli olabilecek bilgileri bulmaktadır. Öneri modülü, farklı öneri sistemi yöntemlerini kullanmaktadır. Şekil 4.24’da kullanıcı arayüzünde sadece taşınabilir cihazlar gösterilmekle birlikte, 2014 yılı itibariyle masaüstü sistemlerde de çalışan bir sistem mevcuttur.

SmartMedia Projesi kapsamındaki bir diğer önemli bölüm de kullanıcı profillemesidir. Kullanıcı profili seçilen bir zaman aralığı için oluşturulur ve eğer kullanıcıya ait başka bir profil daha önce oluşturulmuşsa, yeni profil daha ağırlıklı olmak üzere eski profile birleştirilir. Kullanıcı profillerinin bu şekilde oluşturulması, sisteme yeni kayıt olan kullanıcılara o oturum içinde oluşturulan profile öneri yapılabilmesini ve eski kullanıcılara her oturumda daha iyi öneriler yapılabilmesini sağlamaktadır. SmartMedia haber öneri sistemi ve kullanıcı profillemesi ile ilgili detaylar (Gulla et al., 2013)’de bulunabilir. Bu doktora çalışmasının bir kısmı SmartMedia Projesi ile birlikte yürütülmüştür.

5 SONUÇLAR

Haber kaynakları arasındaki benzerlik ilişkisi haber kaynağı öneri sistemi için oldukça önemlidir. Bu tez kapsamında haber kaynakları arasındaki benzerlik ontolojiler ve birliktelik kuralları ile belirlenmiş ve karşılaştırılmıştır. Bu karşılaştırma için öncelikle küçük çaplı bir haber kaynağı ontolojisi oluşturulmuştur. Birliktelik kuralları ise kullanıcıların haber okuma örüntüleri göz önüne alınarak çıkarılmıştır. Kullanılan YOW veri seti kullanıcıların hangi haber kaynağından gelen haberleri okuduğu ve bunları ne kadar beğendiği bilgisini de içermektedir. Ancak bu tez kapsamında kullanıcıların beğeni dereceleri göz önüne alınmamış, sadece okuma geçmişleri kullanılmıştır. Sonuç olarak haber kaynakları arasındaki bu benzerliklerden yola çıkarak, gerek işbirlikçi filtreleme gerekse de kullanıcının kendi okuma geçmişi ve beğenileri göz önüne alınarak haber kaynağı önerisi yapılması mümkündür. Ayrıca ontolojinin geliştirilmesi, öğeler arasında farklı ilişkiler tanımlanması, yeni haber kaynakları ile zenginleştirilmesi ve farklı sınıf hiyerarşileri ile oluşturulup sonuçların karşılaştırılması daha derin analizler yapmaya olanak sağlayacaktır. Elde edilen sonuçların ve geliştirilen sistemin Norveç Bilim ve Teknoloji Üniversitesi bünyesinde yürütülen ve detayları Bölüm 4.4'te açıklanan SmartMedia projesine entegre edilmesi planlanmıştır.

Bu tez kapsamında oluşturulan haber kaynağı ontolojisinden çıkarılan haber kaynağı benzerlikleri ile kullanıcıların haber okuma örüntülerinden çıkarılan birliktelik kurallarının gösterdiği benzerlikler arasında yaklaşık %60 benzerlik bulunmuştur. Bu benzerlik oranı kullanıcıların okumayı tercih ettikleri haber kaynakları arasında anlamsal olarak yaklaşık %60 benzerlik olduğu şeklinde yorumlanabilmektedir. Böylece, kullanıcının okuduğu tüm haber makalelerinden memnun olduğunu varsayarsak, sadece oluşturulan haber kaynağı ontolojisi kullanılarak yapılan önerilerde %60'a yakın bir kullanıcı memnuniyeti sağlanacaktır. Gelecek çalışmalarda yer verilen, kullanıcı beğenilerinin dikkate alınarak bu çalışmanın genişletilmesi sonucunda yapılacak önerilerde kullanıcı memnuniyetinin daha da artması beklenmektedir.

Durum çalışmasında veri setinden alınan gerçek kullanıcılar üzerinde, birliktelik kuralları ve ontoloji ile elde edilen haber kaynakları ilişkileri kullanılarak tavsiyelerde bulunulmuştur. Buna göre ontolojideki anlamsal çıkarsamalar kullanıldığında elde edilen haber kaynağı tavsiye sayısı, birliktelik kuralları ile elde edilenden daha fazladır. Ancak öneri sistemlerinde kullanıcı memnuniyetini ölçmek zor ve kullanıcılar üzerinde gerçek deneyler gerektiren bir iş olduğundan, yapılan durum çalışmasından elde edilen tavsiyelerin kullanıcı memnuniyetini nasıl

etkileyeceği açıklanamamıştır. Ancak durum çalışması göstermektedir ki anlamsal yöntem birliktelik kurallarından daha fazla sonuç vermektedir.

Bölüm 2.3’de açıklanan öneri sistemi zorluklarından rastlantısallık sorunun, bu tez kapsamında önerilen haber kaynağı önerisi sayesinde aşılması kolaylaşmaktadır. Rastlantısallık sorununda, farklı kaynaklardan gelen bir haber makalesinin aynı haber mi, yoksa aynı konuya ait bir başka haber mi olduğunun sistem tarafından anlaşılması oldukça zordur. Ancak haber kaynaklarının dikkate alındığı bir haber öneri sisteminde, eğer benzer iki haber metni aynı kaynaktan geliyorsa bunların ilişkili ancak aynı olmadığını söylemek oldukça kolaydır. Bu nedenle benzer haber makalelerinin sadece kullanıcının beğendiği bir haber kaynağından önerilmesi bu zorluğa oldukça etkin bir çözüm sağlamaktadır.

5.1 Kısıtlamalar

Bu tez kapsamında yürütülen çalışmalarda karşılaşılan çeşitli kısıtlamalar mevcuttur. Bilimsel çalışmaların temel dayanağını oluşturan veri, bu tez kapsamında da oldukça önemli bir yere sahiptir. Ancak yapılan çalışma ile ilgili verilere ulaşmak her zaman kolay olmamaktadır. Öneri sistemlerinde ise hem içeriğe ait veriler hem de kullanıcı verisi kullanıldığından, yapılan çalışmaya uygun veri seti bulabilmek oldukça zordur. Haber öneri sistemlerinde içerik haber makalelerinden oluşmaktadır. Haber makalelerinin tam metninin alınıp içerik analizlerinin yapılabilmesi için telif hakları ile ilgili kısıtlamaların aşılması gerekmektedir. Kullanıcı verilerinin bir veri seti içinde paylaşılması ise çoğunlukla kullanıcı gizliliği sebebiyle mümkün olmamaktadır. Öneri sistemlerinde gerçek kullanıcı verileri çoğunlukla ticari kullanıma açık sistemler tarafından elde edilebildiği için, yeterli büyüklükteki veri setleri bu hizmeti sunan şirketlerin elinde bulunmaktadır. Doğal olarak ticari kaygılar verilerin, kullanıcıların kişisel gizliliğinin korunması halinde bile açıkça paylaşılmasının önündeki bir başka kısıttır. Kullanıcı verilerini araştırma amaçlı olarak elde etmenin bir başka yöntemi ise yapılan kullanıcı çalışmalarıdır. Bu tez kapsamında kullanılan veri seti bu şekilde oluşturulmuş bir veri setidir. Ancak kullanıcı çalışmaları kısıtlı sayıda kullanıcının katıldığı ve çok uzun süreli olmayan deneyler şeklinde gerçekleştirilmektedir. Bu da sonuçta elde edilen veri setinin küçük ölçekli olmasına yol açmaktadır. Dolayısıyla bu tez çalışmasında kullanılan veri seti kullanıcı, haber makalesi ve haber kaynağı sayılarının gerçek dünya verilerine göre az olduğu küçük ölçekli bir veri setidir. Veri setinin küçük ölçekli olması, oluşturulan haber kaynağı ontolojisinin de küçük ölçekli olmasına sebep olmuş, elde edilen sonuçlar da bu nedenle küçük rakamlarla ifade edilmiştir.

Yapılan durum çalışması sonucu birliktelik kuralları ve ontoloji kullanılarak yapılan tavsiyelerde kullanıcı memnuniyetini ölçmek mümkün olmamıştır. Öneri sistemlerindeki önerilerin kalitesinin belirlenmesi ve kullanıcı memnuniyetinin ölçülmesi zor ve kullanıcı çalışması gerektiren bir iştir. Bu nedenle sonuçlar durum çalışmasından elde edilen haber kaynağı önerilerinin sayısı ile değerlendirilmiştir.

5.2 Tartışmalar

Bu bölümde bu tez çalışmasında elde edilen sonuçlarla ilgili detaylı tartışmalara ve açıklamalara yer verilmiştir.

5.2.1 Örtüşmeyen kurallar

Bu tez çalışmasında bir haber kaynağı öneri sistemi tasarlanmıştır. Tasarlanan bu sistemde ontolojiler ve birliktelik kuralları yöntemleri kullanılmış ve her birinden ayrı ayrı elde edilen sonuçlar karşılaştırılmıştır. Bölüm 4.2 (4.2.4)'de detaylı olarak açıklandığı gibi bu sonuçlara göre iki yöntem %60 oranında benzer sonuçlar vermiştir. Yani sadece ontolojiler ve anlamsal çıkarsama kullanılarak sistemin öneride bulunacağı haber kaynakları ve sadece birliktelik kuralları kullanılarak sistemin öneride bulunacağı haber kaynakları %60 oranında aynıdır. İki yöntem arasındaki bu benzerlik oranı, haber kaynağı ontolojisinden elde edilen tüm benzerliklerin birliktelik kurallarından elde edilen benzerliklerle olan örtüşme oranıdır. Açıkça görülmektedir ki bu örtüşmenin dışında kalan haber kaynakları olmuştur. Bu haber kaynakları iki şekilde incelenebilir. İlki haber kaynağı ontolojisinde aralarındaki ilişkileri görebildiğimiz ancak birliktelik kurallarında yer almayan kaynaklar, ikincisi ise birliktelik kurallarında aralarındaki ilişkiyi görebildiğimiz ancak ontolojide yer almayan kaynaklardır. İki yöntemden elde edilen sonuçların örtüşme oranlarını düşüren bu durumu yorumlamak gerekirse:

1. Bir kullanıcının okumayı tercih ettiği haber kaynakları arasında anlamsal olarak %100 yakınlık yoktur.
2. Aralarında anlamsal ilişki bulunan haber kaynaklarının tamamı kullanıcılar tarafından okunmamıştır.

Bu durum incelendiğinde aslında sonuçların oldukça doğal olduğu görülmektedir. Sonuçlara göre kullanıcıların haber okumayı tercih ettikleri haber kaynakları arasında %60 oranında anlamsal bir benzerlik vardır. Örneğin spor haberlerini

8153	->	8156 8157 8154	0.381	0.667
8153	->	8156 8157	0.381	0.667
8153	->	8157 8154	0.429	0.750
8153	->	8157	0.476	0.833
8153	->	8156 8154	0.476	0.833
8153	->	8156 8154 8152	0.381	0.667
8153	->	8156	0.476	0.833
8153	->	8156 8152	0.381	0.667

Şekil 5.1: Çoklu birliktelik kurallarına örnek - 1.

8195 8156	->	8155	0.476	0.769
8195 8156	->	8154 8155	0.476	0.769
8195 8156	->	8154	0.619	1.000
8195 8156	->	8154 8153	0.476	0.769
8195 8156	->	8154 8152	0.476	0.769
8195 8156	->	8153	0.476	0.769
8195 8156	->	8152	0.476	0.769
8195 8156 8155	->	8154	0.476	1.000
8195 8156 8154	->	8155	0.476	0.769
8195 8156 8154	->	8153	0.476	0.769

Şekil 5.2: Çoklu birliktelik kurallarına bir örnek - 2.

bir kaynaktan okuyan bir kullanıcı politika haberlerini başka bir haber kaynaktan okumuş olabilir. Bir kullanıcının spor haberleri ile ilgili tüm kaynakları okuması olanaklı değildir. Ancak kullanıcının çoğunlukla spor haberleri içerek kaynaklardan haber okumuş olması o kullanıcının spora olan ilgisini anlatmaktadır. Diğer yandan ikinci seçenekte tespit edilen durum kullanıcının henüz haberdar olmadığı haber kaynakları olması durumu olarak yorumlanabilir. Aralarında anlamsal ilişki olan ancak kullanıcılar tarafından okunmamış kaynaklar öneri sistemi tarafından önerilecektir. Sistem önerilerde bulundukça kullanıcıların haber kaynakları arasında anlamsal ilişki de gözlenerek hangi haber kaynaklarını tercih ettikleri daha net olarak ortaya çıkabilecektir.

5.2.2 Çoklu birliktelik kuralları

Bölüm 4'te detaylı olarak açıklandığı gibi veri setinin kullanıcı okuma bilgisini içeren bölümü kullanılarak birliktelik kuralları elde edilmiştir. Bu kurallar bize kullanıcıların hangi haber kaynaklarını birlikte okumayı tercih ettiğini anlatmaktadır. Örneğin; $X \rightarrow Y$ ifadesi X kaynağından haber okuyan bir kullanıcı Y kaynağından da okumuş kuralını ifade etmektedir. Ancak Figür 5.1 ve 5.2'de görüldüğü gibi bazı kurallar çokludur. Çoklu kurallardan $X \rightarrow Y, Z$ şeklinde olan kurallar X kaynağından okuyan kullanıcıların aynı zamanda Y ve Z kaynaklarından da okuduğunu, $X, T \rightarrow Y, Z$ şeklinde olan kurallar ise hem X hem de T kaynaklarından okuyan kullanıcıların aynı zamanda Y ve Z kaynaklarından da okuduğunu anlatır.

Bu tez çalışması kapsamında elde edilen birliktelik kurallarından sadece tekli yani $X \rightarrow Y$ şeklinde olan kurallar göz önüne alınmıştır. Bunun nedeni haber kaynakları arasında ontolojiden elde edilen anlamsal ilişkilerin çoklu kurallar tanımlamak için kullanılmasının mümkün olmamasıdır. Çizelge 4.2’de haber kaynakları arasındaki ontolojiden elde edilen ilişkiler görülmektedir. Bu tablodaki her bir satırdaki haber kaynakları birbiri ile ilişkilidir. Örneğin tablonun ilk satırında yer alan 8155, 8172 ve 1265 numaralı haber kaynakları birbiri ile anlamsal olarak ilişkilidir ancak aralarında birliktelik kurallarında olduğu gibi $1855 \rightarrow 8172, 1265$ ilişkisini kurmak mümkün değildir. Anlamsal ilişkiler ve birliktelik kuralları, haber kaynakları arasındaki ilişkileri birbirinden farklı tanımlamaktadır.

5.2.3 Duyarlılık - anma hesaplamaları

Duyarlılık (precision) ve anma (recall) bilgi elde etme sistemlerinde performans ölçümleri ve değerlendirmeler için sıkça kullanılır (Raghavan et al., 1989). Duyarlılık, elde edilen öğelerden uygun olanların oranı, anma ise uygun olan öğelerden elde edilenlerin oranı olarak tanımlanabilir. Buna göre;

A = Elde edilen öğelerin sayısı,

B = Elde edilemeyen öğelerin sayısı,

C = Elde edilen uygun olmayan öğelerin sayısı,

olmak üzere anma ve duyarlılık şöyle hesaplanır:

$$Anma = \frac{A}{A + B} \times 100\% \quad (5.1)$$

$$Duyarlılık = \frac{A}{A + C} \times 100\% \quad (5.2)$$

Anma ve duyarlılık değerleri birbiri ile ters orantılıdır, yani anma değeri arttığında duyarlılık değeri azalacaktır (Jizba, 2007). Ancak bu tez çalışması kapsamında duyarlılık ve anma değerlerinin hesaplanması mümkün olmamıştır. Bunun nedeni ise duyarlılık ve anma hesaplamalarının tek bir yöntem sonucunda elde edilen sonuçlar üzerinden yapılmasının mümkün olmasıdır. Bu tez kapsamındaki çalışmada ise iki farklı yöntemden elde edilen sonuçlar birbiri ile

karşılaştırılmıştır. İki farklı yöntemle elde edilen sonuçların birbirine göre duyarlılık ve anma değerleri hesaplanamamaktadır. Diğer yandan veri seti rastgele iki parçaya bölünüp her bir yöntem için ayrı ayrı duyarlılık ve anma hesaplanması mümkündür. Ancak hem veri setinin çok küçük olması, hem de her yöntemin kendi içindeki performansını ölçmenin tez kapsamında yapılmış olan çalışmaya bir katkısı olmaması sebebiyle bu hesaplamalar yapılmamıştır.

5.3 Edinimler

Kişisel verilerin gizliliğinin korunması gibi sebeplerle kullanıcı verisinin saklanması mümkün olmadığı durumlarda kişiselleştirilmiş öneri sistemlerinin performansı düşmektedir. Bu tez çalışması kapsamında tasarlanan sistemde ontolojiler ve birliktelik kuralları yöntemleri kullanılmıştır. Bu yöntemlerin her birinden ayrı ayrı elde edilen sonuçlar karşılaştırılmış ve sonuçlara göre iki yöntem %60 oranında benzer sonuçlar vermiştir. Sadece ontolojiler ve anlamsal çıkarsama kullanılarak sistemin öneride bulunacağı haber kaynakları ve sadece birliktelik kuralları kullanılarak sistemin öneride bulunacağı haber kaynakları %60 oranında aynıdır. Buna göre bu iki yöntemin birbiri yerine kullanıldığı durumlarda en az %60 tutarlılık sağlanacaktır. Bu oran detaylı kişisel verilerin saklanmadığı durumlarda, sadece haber kaynaklarının anlamsal analizi ve aralarındaki anlamsal ilişkilerin bulunması yöntemi ile kullanıcılara uygun haber kaynağını tavsiye etmek için uygun görünmektedir. Her iki yöntemin birlikte kullanılması durumunda ise öneri sisteminin daha uygun önerilerde bulunması beklenmektedir. Bölüm 5.2.1’de tartışıldığı gibi her iki yöntemde diğeri tarafından kapsanmayan kurallar, ilişkiler bulunmaktadır. Bu iki yöntemin birlikte kullanılması tek bir yöntemin olası eksikliklerinin tamamlanmasını sağlayacaktır.

Bu çalışmada anlamsal yöntemin seçilmesi haber kaynakları arasındaki anlamsal ilişkilerin, benzerlik ve farklılıkların bulunmasını sağlamaktadır. Eğer bir kullanıcı aynı haberi yayınlayan iki kaynaktan birisini seçiyorsa bu iki haber kaynağı arasındaki benzerlik ve farklılıkların bulunması kullanıcının tercihleri konusunda daha derin, anlamsal bir analiz yapmaya yarayacaktır. Bu sayede henüz kullanıcının keşfetmediği haber kaynaklarından kullanıcıya en uygun olanlar seçilebilecektir. Ayrıca bir kullanıcının tercih ettiği haber kaynakları arasındaki anlamsal farklılıklar kullanıcıların özellikleri hakkında da bilgi vermektedir. Örneğin; eğer bir kullanıcı çoğunlukla X ve Y kaynağından haber okuyorsa ve bu kaynaklar arasında anlamsal bir benzerlik varsa, kullanıcının benzer haberleri farklı kaynaklardan okumaktan hoşlandığını söylemek mümkündür. Tam tersi bir durumda ise kullanıcının farklı türde haberleri okumak için farklı kaynakları tercih ettiğini, örneğin spor haberlerini

1. İki yöntemin zorunlu durumlarda birbiri yerine kullanılabilmesi.
2. Anlamsal çıkarsamanın katkı ve getirileri.
3. Önerilerin nasıl oluşturulduğunun görülebilmesi.
4. Rastlantısallık sorununa çözüm getirilmesi veya sorunun hafifletilmesi.

Çizelge 5.1: Tez çalışması kapsamındaki edinimlerin özeti.

X kaynağından, eğlence haberlerini Y kaynağından vb. okuduğunu söylemek mümkündür. Bu tür kişisel özelliklere göre ise kişilere hem yeni haber kaynakları hem de haber kaynaklarını göz önüne alarak haber makaleleri önermek mümkündür. Buna ek olarak durum çalışmasında gösterildiği gibi anlamsal çıkarsama yöntemi birliktelik kurallarına göre daha fazla sayıda tavsiye üretmektedir.

Öneri sistemlerinde önerilen bir ögenin neden önerildiğinin öğrenilmesi önemlidir (Bilgic and Mooney, 2005). Kullanıcılar bir öneriyi neden aldıklarını öğrenmek, geliştiriciler ise öneri sisteminin nasıl sonuca vardığını görmek istemektedirler. Hem birliktelik kuralları hem de anlamsal analiz önerinin neden yapıldığının, sistemin nasıl bir çıkarsama sonucu öneride bulunduğunun anlaşılmasını sağlamaktadır. Ayrıca (Herlocker et al., 2000)'da açıklandığı gibi önerilerin nasıl yapıldığının gösterilmesi öneri sisteminin performansını arttırmaktadır.

Bölüm 2.3.10'da açıklanan rastlantısallık sorunu öneri sistemlerinde, özellikle de metinsel verinin çok olduğu haber öneri sistemleri gibi sistemlerde çözülmesi güç bir sorundur. Birbirine benzer içeriğe sahip iki haber metninin aynı olayı anlatan farklı metinler mi yoksa bir olayla ilgili devam haber metni mi olduğunun sistem tarafından anlaşılması çoğu zaman mümkün olmamaktadır. Haber kaynaklarının göz önüne alınması ile rastlantısallık sorununun çözümü ya da en azından iyileştirilmesi mümkün olmaktadır. Çünkü her bir haber kaynağında bir olayı anlatan sadece tek haber metni bulunmaktadır. Eğer birbirine çok benzeyen iki haber metni aynı kaynaktan geliyorsa, bu iki metnin birbiri ile alakalı olduğunu, ancak tamamen aynı olmadığını söylemek mümkündür. Diğer yandan, önceden yazılmış bir haber metninin güncellenmesi durumunda bu güncellenmenin sistem tarafından yeni bir haber makalesi olarak algılanması mümkündür. Bu da rastlantısallık sorununun çözümünü biraz daha güçleştirmektedir ancak farklı şekillerde önlemler alınarak üstesinden gelinmesi mümkün görünmektedir.

Bu bölümde açıklanan edinimlerin özeti Çizelge 5.1'de gösterilmiştir.

5.4 Katkıların Özeti

Bu tez kapsamında önerilen anlamsal haber öneri sisteminin katkıları şöyle özetlenebilir:

- Haber öneri sistemlerinde literatür taramasında örneğine rastlanmayan haber kaynakları arasındaki farklılıkların ve kullanıcı haber okuma örüntülerinin göz önüne alınarak zenginleştirilmiş haber ve haber kaynağı öneri yöntemi sunulması,
- Literatürde örneğine rastlanmayan küçük çaplı bir haber kaynağı ontolojisinin oluşturulması,
- Kullanıcıların haber okuma örüntülerinden çıkarsanan birliktelik kuralları ile haber kaynakları arasındaki benzerliklerin bulunması, haber kaynağı ontolojisi kullanılarak haber kaynakları arasındaki benzerliklerin bulunması ve bu iki yöntemin karşılaştırılması.

5.5 Gelecek Çalışmalar

Bu tez kapsamında yapılan çalışmaların devamı olarak önerilen gelecek çalışmalar şöyle özetlenebilir:

- Bu tez kapsamında haber kaynaklarının benzerliklerini bulmak için kullanıcıların okuma örüntüleri kullanılmış ancak kullanıcı geribildirimleri yani kullanıcıların beğeni dereceleri hesaba katılmamıştır. Ayrıca ontolojide yer alan ve haber metinlerinden çıkarılan önemli terimlerin önem değerleri (ağırlıkları) dikkate alınmamıştır. Kullanıcı geribildirimlerinin ve ontolojide yer alan öğelerin ağırlıklarının da hesaba katılması ile daha uygun öneriler yapılması,
- Bu tez kapsamında sunulan yöntemin Norveç Bilim ve Teknoloji Üniversitesi bünyesinde yürütülen SmartMedia (SmartMedia, 2012) haber öneri sistemi projesine entegre edilmesi,
- Ontoloji - öneri sistemi bütünleşik sistemi içinde ontolojiler kullanılarak yapılan önerilerin tersine çalışarak, öneri sisteminin ontolojiyi otomatik olarak geliştirip geliştiremeyeceğinin araştırılması,

- Anlamsal çıkarsama için kullanılan bulunma sayısı yöntemi ve Jaccard katsayısı yönteminin verimliliğinin karşılaştırılması,
- Daha geniş kapsamlı, güncel ve haber öneri sistemlerini tam destekleyen bir veri seti oluşturulması ya da medya kuruluşlarından edinilebilecek daha geniş kapsamlı bir veri seti üzerinde bu tez kapsamındaki çalışmaların tekrar edilmesi ve tutarlılığının gözlenerek daha derin analizler yapılması,
- Bu tez kapsamında tartışılan öneri sistemi zorluklarının daha etkin çözümleri üzerinde çalışılması.

KAYNAKLAR DİZİNİ

- Adomavicius, G. and Tuzhilin, A.**, 2005, Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, *Knowledge and Data Engineering, IEEE Transactions on* **17(6)**, 734
- Bilgic, M. and Mooney, R. J.**, 2005, Explaining recommendations: Satisfaction vs. promotion, in *Beyond Personalization Workshop, IUI*, Vol. 5
- Borges, H. L. and Lorena, A. C.**, 2010, A survey on recommender systems for news data, in *Smart Information and Knowledge Management*, pp 129–151, Springer
- Burke, R.**, 2002, Hybrid recommender systems: Survey and experiments, *User modeling and user-adapted interaction* **12(4)**, 331
- Cantador, I., Bellogín, A., and Castells, P.**, 2008, News@hand: A semantic web approach to recommending news, in *Proceedings of the 5th international conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, AH '08, pp 279–283, Springer-Verlag, Berlin, Heidelberg
- Cantador, I. and Castells, P.**, 2009, Semantic contextualisation in a news recommender system, in *Workshop on Context-Aware Recommender Systems (CARS 2009)*
- Capelle, M., Frasincar, F., Moerland, M., and Hogenboom, F.**, 2012, Semantics-based news recommendation, in *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*, p. 27, ACM
- Celma, O.**, 2008, *Last.fm Dataset Collection*, Internet: <http://ocelma.net/MusicRecommendationDataset/lastfm-360K.html>, Erişim tarihi: 11 Ağustos 2014
- Celma, O.**, 2010, *Music Recommendation and Discovery in the Long Tail*, Springer
- Chaos, B. P.**, 2009, *Netflix Prize Winning Algorithm*, Internet: <http://www.netflixprize.com/community/viewtopic.php?id=1537>, Erişim tarihi: 11 Ağustos 2014
- Das, A., Datar, M., Garg, A., and Rajaram, S.**, 2007, Google news personalization: Scalable online collaborative filtering, in *WWW '07 Proceedings of the 16th international conference on World Wide Web*, pp 271–280
- DataMaster**, 2009, *DataMaster 1.3.2*, Internet: http://protege.wiki.stanford.edu/wiki/DataMaster_1.3.2, Erişim tarihi: 11 Ağustos 2014

KAYNAKLAR DİZİNİ (devam)

- Dataset, M. S.**, 2011, *Million Song Dataset*, Internet: <http://labrosa.ee.columbia.edu/millionsong>, Erişim tarihi: 11 Ağustos 2014
- Dekker, R.**, 2006, The importance of having data-sets.
- Demšar, J., Curk, T., Erjavec, A., Črt Gorup, Hočevan, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M., and Zupan, B.**, 2013, Orange: Data mining toolbox in python, *Journal of Machine Learning Research* **14**, 2349
- Fortuna, B., Fortuna, C., and Mladenec, D.**, 2010, Real-time news recommender system, in *ECML PKDD'10 Proceedings of the 2010 European conference on Machine learning and knowledge discovery in databases: Part III*, pp 583–586
- Frantzi, K., Ananiadou, S., and Mima, H.**, 2000, Automatic recognition of multi-word terms: the c-value/nc-value method, *International Journal on Digital Libraries* **3**(2), 115
- FSF**, 1986, *Özgür Yazılım Tanımı*, Internet: <https://www.gnu.org/philosophy/free-sw.tr.html>, Erişim tarihi: 11 Temmuz 2014
- Garcin, F., Dimitrakakis, C., and Faltings, B.**, 2013, Personalized news recommendation with context trees, *CoRR* abs/1303.0665
- Goossen, F., IJntema, W., Frasincar, F., Hogenboom, F., and Kaymak, U.**, 2011, News personalization using the cf-idf semantic recommender, in *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, p. 10, ACM
- GroupLens**, 2010, *GroupLens*, Internet: <http://grouplens.org.edu>, Erişim tarihi: 11 Ağustos 2014
- Gulla, J. A., Ingvaldsen, J. E., Fidjestøl, A. D., Nilsen, J. E., Haugen, K. R., and Su, X.**, 2013, Learning user profiles in mobile news recommendation., pp 183–194
- Herlocker, J. L., Konstan, J. A., and Riedl, J.**, 2000, Explaining collaborative filtering recommendations, in *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pp 241–250, ACM
- Iaquinta, L., Gemmis, M. D., Lops, P., Semeraro, G., Filannino, M., and Molino, P.**, 2008, Introducing serendipity in a content-based recommender system, in *Hybrid Intelligent Systems, 2008. HIS'08. Eighth International Conference on*, pp 168–173, IEEE

KAYNAKLAR DİZİNİ (devam)

- IJntema, W., Goossen, F., Frasincar, F., and Hogenboom, F.**, 2010, Ontology-based news recommendation, in *Proceedings of the 2010 EDBT/ICDT Workshops*, p. 16, ACM
- Jizba, R.**, 2007, *Measuring Search Effectiveness*, Internet: <http://www.creighton.edu>, Erişim tarihi: 11 Eylül 2014
- Lašek, I. and Vojtáš, P.**, 2011, Semantic information filtering-beyond collaborative filtering, in *4th International Semantic Search Workshop*
- Last.fm**, 2008, *Last.fm*, Internet: <http://last.fm/>, Erişim tarihi: 11 Ağustos 2014
- Lee, H. and Park, S. J.**, 2007, Moners: A news recommender for the mobile web, *Expert Systems with Applications* **32(1)**, 143
- Lemdani, R., Bennacer, N., Polaillon, G., and Bourda, Y.**, 2010, A collaborative and semantic-based approach for recommender systems, in *Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on*, pp 469–476, IEEE
- Li, L., Wang, D., Li, T., Knox, D., and Padmanabhan, B.**, 2011, Scene : A scalable two-stage personalized news recommendation system, in *SIGIR '11 Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pp 125–134
- Lin, C., Xie, R., Li, L., Huang, Z., and Li, T.**, 2012, Premise: personalized news recommendation via implicit social experts, in *Proceedings of the 21st ACM international conference on Information and knowledge management*, pp 1607–1611, ACM
- Liu, J., Dolan, P., and Pedersen, E. R.**, 2010, Personalized news recommendation based on click behavior, in *IUI '10 Proceedings of the 15th international conference on Intelligent user interfaces*, pp 31–40
- Lops, P., De Gemmis, M., and Semeraro, G.**, 2011, Content-based recommender systems: State of the art and trends, in *Recommender systems handbook*, pp 73–105, Springer
- McFee, B., Bertin-Mahieux, T., Ellis, D. P., and Lanckriet, G. R.**, 2012, The million song dataset challenge, in *Proceedings of the 21st International Conference Companion on World Wide Web, WWW '12 Companion*, pp 909–916, ACM
- Media, K.**, 2012, Measuring news consumption and attitudes

KAYNAKLAR DİZİNİ (devam)

- Mobasher, B., Dai, H., Luo, T., and Nakagawa, M.**, 2001, Effective personalization based on association rule discovery from web usage data, in *Proceedings of the 3rd international workshop on Web information and data management*, pp 9–15, ACM
- NaCTeM**, 2011, *TerMine*, Internet: <http://www.nactem.ac.uk/software/termine>, Erişim tarihi: 11 Ağustos 2014
- Netflix**, 2009, *Netflix Prize*, Internet: <http://www.netflixprize.com>, Erişim tarihi: 11 Ağustos 2014
- Ofcom**, 2013, News consumption in the uk - 2013 report
- Orange**, 2011, *Orange*, Internet: <http://orange.biolab.si/>, Erişim tarihi: 11 Ağustos 2014
- Ozgobek, O., Gulla, J. A., and Erdur, R. C.**, 2014, A survey on challenges and methods in news recommendation, in *In Proceedings of the 10th International Conference on Web Information System and Technologies (WEBIST 2014)*
- Peis, E., del Castillo, J. M., and Delgado-López, J.**, 2008, Semantic recommender systems. analysis of the state of the topic, *Hipertext. net* **6**, 1
- Plugin, T. P.**, 2008, *TerMine Protege Plugin*, Internet: http://protegewiki.stanford.edu/wiki/TerMine_Plugin, Erişim tarihi: 11 Ağustos 2014
- Protege**, 2010, *Stanford Center for Biomedical Informatics Research*, Internet: <http://protege.stanford.edu>, Erişim tarihi: 11 Ağustos 2014
- Rada, R., Mili, H., Bicknell, E., and Blettner, M.**, 1989, Development and application of a metric on semantic nets, *Systems, Man and Cybernetics, IEEE Transactions on* **19(1)**, 17
- Raghavan, V., Bollmann, P., and Jung, G. S.**, 1989, A critical investigation of recall and precision as measures of retrieval system performance, *ACM Transactions on Information Systems (TOIS)* **7(3)**, 205
- Rao, J., Jia, A., Feng, Y., and Zhao, D.**, 2013, Personalized news recommendation using ontologies harvested from the web, in *Web-age information management*, pp 781–787, Springer
- Said, A., Berkovsky, S., and De Luca, E. W.**, 2011, Group recommendation in context, in *Proceedings of the 2nd Challenge on Context-Aware Movie Recommendation*, CAMRa '11, pp 2–4, ACM

KAYNAKLAR DİZİNİ (devam)

- Sandvig, J. J., Mobasher, B., and Burke, R.**, 2007, Robustness of collaborative recommendation based on association rule mining, in *Proceedings of the 2007 ACM conference on Recommender systems*, pp 105–112, ACM
- Saranya.K.G and Sadhasivam, G.**, 2012, A personalized online news recommendation system, *International Journal of Computer Applications* **57(18)**, 6
- Sarwar, B., Karypis, G., Konstan, J., and Riedl, J.**, 2001, Item-based collaborative filtering recommendation algorithms, in *Proceedings of the 10th international conference on World Wide Web*, pp 285–295, ACM
- SmartMedia**, 2012, *SmartMedia*, Internet: <http://research.idi.ntnu.no/SmartMedia>, Erişim tarihi: 11 Ağustos 2014
- Su, X. and Khoshgoftaar, T. M.**, 2009, A survey of collaborative filtering techniques, *Advances in artificial intelligence* **2009**, 4
- Sun, X., Kong, F., and Chen, H.**, 2005, Using quantitative association rules in collaborative filtering, in *Advances in Web-Age Information Management*, pp 822–827, Springer
- Tan, P.-N. and Kumar, V.**, 2005, Chapter 6. association analysis: Basic concepts and algorithms, *Introduction to Data Mining. Addison-Wesley. ISBN 321321367*
- Tavakolifard, M., Gulla, J. A., Almeroth, K. C., Ingvaldesn, J. E., Nygreen, G., and Berg, E.**, 2013, Tailored news in the palm of your hand: a multi-perspective transparent approach to news recommendation, in *Proceedings of the 22nd international conference on World Wide Web companion*, pp 305–308, International World Wide Web Conferences Steering Committee
- Wen, H., Fang, L., and Guan, L.**, 2012, A hybrid approach for personalized recommendation of news on the web, *Expert Systems with Applications* **39(5)**, 5806
- WordNet**, 1985, *WordNet*, Internet: <http://wordnet.princeton.edu>, Erişim tarihi: 11 Ağustos 2014
- Wu, Z. and Palmer, M.**, 1994, Verbs semantics and lexical selection, in *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, pp 133–138, Association for Computational Linguistics
- Yeung, K. F. and Yang, Y.**, 2010, A proactive personalized mobile news recommendation system, in *Developments in E-systems Engineering*, pp 207–212, IEEE

KAYNAKLAR DİZİNİ (devam)

- Zhang, Y.**, 2003, *YOW User Study*, Internet:
<http://users.soe.ucsc.edu/~yiz/papers/data/YOWStudy/>, Erişim tarihi: 11
Temmuz 2014
- Ziegler, C.-N., McNee, S. M., Konstan, J. A., and Lausen, G.**, 2005, Improving
recommendation lists through topic diversification, in *Proceedings of the 14th
international conference on World Wide Web*, pp 22–32, ACM

ÖZGEÇMİŞ

Adı Soyadı: Özlem ÖZGÖBEK
Doğum Tarihi: 1983
Doğum Yeri: İzmir
Uyruğu: T.C.

EĞİTİM

Lisans: Atılım Üniversitesi
 Bilgisayar Mühendisliği Bölümü, Ankara, 2005
Lise: Konak Anadolu Lisesi, İzmir, 2001

İLGİ ALANLARI

Öneri Sistemleri, Haber Öneri Sistemleri, Anlamsal Çıkarsama

YAYINLARI

- Ö. Özgöbek, L. Toker, “Üniversite Kampüslerinde 4. Nesil Mobil İletişim Ağlarına Geçişte İleri Bir Teknoloji Örneği: WiBro”, Akademik Bilişim 09, Harran Üniversitesi, Şanlıurfa, 2009.
- Ö. Özgöbek, R. C. Erdur, “Mobil Cihazlardan Web Servis Sunumu”, Akademik Bilişim 10, Muğla Üniversitesi, Muğla, 2010.
- Ö. Özgöbek, R. C. Erdur, “Akıllı Ortamlarda Sensör Kontrolüne Etmen Tabanlı Bir Yaklaşım: Bir Jadex Uygulaması”, Akademik Bilişim 11, İnönü Üniversitesi, Malatya, 2011.
- Ö. Özgöbek, “Açık Erişim ve Açık Erişim Sistemleri Altyapısı”, Akademik Bilişim 2012, Uşak Üniversitesi, Uşak, 2012.
- O.Ozgobek, J. A. Gulla, C. Erdur, “A Survey on Challenges and Methods in News Recommendation ”, Webist 2014, Barcelona, 2014.
- O. Ozgobek, N. Shabib, J.A. Gulla, “Data Sets and Recommendation”, NRA Workshop in UMAP 2014, Aalborg, Denmark, 2014.
- J.A. Gulla, V. Ollikainen, O. Ozgobek, N. Shabib, “2nd International Workshop on News Recommendation and Analytics (NRA2014)”, UMAP 2014, Aalborg, Denmark, 2014.
- Ö. Özgöbek, R.C. Erdur, “Öneri Sistemleri ve Bir Uygulama Alanı Olarak Haber Öneri Sistemleri", Akademik Bilişim 15, Anadolu Üniversitesi, Eskisehir, 2015.

EKLER

EK 1: Terimler Sözlüğü

EK 2: Çıkarılan Birliktelik Kuralları

EK 3: Haber Kaynakları Arasındaki İlişkiler ve Jaccard Katsayıları

EK 1: Terimler Sözlüğü

TÜRKÇE	İNGİLİZCE
Açık	Explicit
Ağ	Network
Akım	Trend
Anındalık	Recency
Anlamsal	Semantic
Anma	Recall
Arayüz	Interface
Arkaplan	Background
Aşırı uzmanlık	Over-specialization
Ayrıştırma	Parse
Bağlam ağacı	Context tree
Bayes ağı	Bayesian network
Benzerlik ölçütü	Similarity measure
Besleme	Feed
Betik	Script
Birliktelik kuralı	Association rule
Boylamsal	Longitudinal
Buluşsal	Heuristic
Çerçeve	Framework
Çevrimiçi	Online
Dağıtık	Distributed
Değerlendirme	Evaluation
Değişken	Dynamic
Dilbilimsel	Linguistic
Durağan	Static
Durum çalışması	Case study
Duyarlılık	Precision
Düzensiz	Unstructured
Eklenti	Plug-in
Elde etme	Retrieval
Eleman	Entity
Enjeksiyon	Injection
Eş anlamlılık	Synonymy
Filtreleme	Filtering
Geleneksel	Conventional
Geribildirim	Feedback
Girdi	Input

Gizlilik	Privacy
Gri koyun	Grey sheep
Güvenilirlik	Reliability
Hafıza tabanlı	Memory-based
Hesap tablosu	Spreadsheet
İçerik tabanlı	Content-based
İkili	Binary
İlgi alanı	Interest
İlişkisel olmayan	Non-relational
İş akışı	Workflow
İşbirlikçi	Collaborative
Karmaşıklık	Ambiguity
Kavram	Concept
Komşu geçişkenliği	Neighbor transitivity
Kümeleme	Clustering
Makine öğrenmesi	Machine learning
Masaüstü	Desktop
Model tabanlı	Model-based
Modelleme	Modelling
Nicel	Quantitative
Öge	Item
Ölçeklenebilirlik	Scalability
Öneri sistemi	Recommender system
Önetkin	Proactive
Örnek	Instance/Individual
Örtük	Implicit
Örüntü	Pattern
Oy	Rating
Portföy	Portfolio
Profil	Profile
Rastlantı	Serendipity
Ses	Audio
Seyreklik	Sparsity
Soğuk başlangıç	Cold-start
Sorun	Problem
Sunucu	Server
Tanıma	Recognition
Taşınabilir	Mobile
Tercih	Preference
Üst veri	Metadata

Veri madenciliđi	Data mining
Veri seti	Dataset
Veritabanı	Database
Yapısal	Structured
Yöntem	Method
Zorluk	Challenge

İNGİLİZCE

Ambiguity
Association rule
Audio
Background
Bayesian network
Binary
Case study
Challenge
Clustering
Cold-start
Collaborative
Concept
Content-based
Context tree
Conventional
Data mining
Database
Dataset
Desktop
Distributed
Dynamic
Entity
Evaluation
Explicit
Feed
Feedback
Filtering
Framework
Grey sheep
Heuristic
Implicit
Injection
Input
Instance/Individual
Interest
Interface
Item
Linguistic
Longitudinal

TÜRKÇE

Karmaşıklık
Birliktelik kuralı
Ses
Arkaplan
Bayes ağı
İkili
Durum çalışması
Zorluk
Kümeleme
Soğuk başlangıç
İşbirlikçi
Kavram
İçerik tabanlı
Bağlam ağacı
Geleneksel
Veri madenciliği
Veritabanı
Veri seti
Masaüstü
Dağıtık
Değişken
Eleman
Değerlendirme
Açık
Besleme
Geribildirim
Filtreleme
Çerçeve
Gri koyun
Buluşsal
Örtük
Enjeksiyon
Girdi
Örnek
İlgi alanı
Arayüz
Öge
Dilbilimsel
Boylamsal

Machine learning	Makine öğrenmesi
Memory-based	Hafıza tabanlı
Metadata	Üst veri
Method	Yöntem
Mobile	Taşınabilir
Model-based	Model tabanlı
Modelling	Modelleme
Neighbor transitivity	Komşu geçişkenliği
Network	Ağ
Non-relational	İlişkisel olmayan
Online	Çevrimiçi
Over-specialization	Aşırı uzmanlık
Parse	Ayrıştırma
Pattern	Örüntü
Plug-in	Eklenti
Portfolio	Portföy
Precision	Duyarlılık
Preference	Tercih
Privacy	Gizlilik
Proactive	Önetkin
Problem	Sorun
Profile	Profil
Quantitative	Nicel
Rating	Oy
Recall	Anma
Recency	Anındalık
Recognition	Tanıma
Recommender system	Öneri sistemi
Reliability	Güvenilirlik
Retrieval	Elde etme
Scalability	Ölçeklenebilirlik
Script	Betik
Semantic	Anlamsal
Serendipity	Rastlantı
Server	Sunucu
Similarity measure	Benzerlik ölçütü
Sparsity	Seyreklik
Spreadsheet	Hesap tablosu
Static	Durağan
Structured	Yapısal

Synonymy	Eş anlamlılık
Trend	Akım
Unstructured	Düzensiz
Workflow	İş akışı

EK 2: Çıkarılan Birliktelik Kuralları

Rules	Supp	Conf
8172		
8172 -> 8155	0.429	0.900
8172 -> 8154 8155	0.429	0.900
8172 -> 8195	0.429	0.900
8172 -> 8154 8195	0.429	0.900
8172 -> 8154	0.476	1.000
8154		
8172 8154 -> 8155	0.429	0.900
8172 8154 -> 8195	0.429	0.900
8155		
8172 8155 -> 8154	0.429	1.000
8195		
8172 8195 -> 8154	0.429	1.000
3616		
3616 -> 8154	0.429	1.000
8195		
8195 -> 8155	0.571	0.667
8195 -> 8156 8155	0.476	0.556
8195 -> 8156 8154 8155	0.476	0.556
8195 -> 8154 8155	0.571	0.667
8195 -> 8154 8155 8153	0.429	0.500
8195 -> 8155 8153	0.429	0.500
8195 -> 8172	0.429	0.500
8195 -> 8172 8154	0.429	0.500
8195 -> 8157	0.476	0.556
8195 -> 8157 8154	0.429	0.500
8195 -> 8157 8154 8153	0.429	0.500
8195 -> 8157 8153	0.476	0.556
8195 -> 8156	0.619	0.722
8195 -> 8156 8154	0.619	0.722
8195 -> 8156 8154 8153	0.476	0.556
8195 -> 8156 8154 8152	0.476	0.556
8195 -> 8156 8153	0.476	0.556
8195 -> 8156 8152	0.476	0.556
8195 -> 8154	0.762	0.889
8195 -> 8154 8153	0.524	0.611
8195 -> 8154 8152	0.524	0.611
8195 -> 8153	0.571	0.667
8195 -> 8152	0.524	0.611

8172

8195 8172 -> 8154 0.429 1.000

8156

8195 8156 -> 8155 0.476 0.769

8195 8156 -> 8154 8155 0.476 0.769

8195 8156 -> 8154 0.619 1.000

8195 8156 -> 8154 8153 0.476 0.769

8195 8156 -> 8154 8152 0.476 0.769

8195 8156 -> 8153 0.476 0.769

8195 8156 -> 8152 0.476 0.769

8195 8156 8155 -> 8154 0.476 1.000

8195 8156 8154 -> 8155 0.476 0.769

8195 8156 8154 -> 8153 0.476 0.769

8195 8156 8153 -> 8154 0.476 1.000

8195 8156 8154 -> 8152 0.476 0.769

8195 8156 8152 -> 8154 0.476 1.000

8157

8195 8157 -> 8154 0.429 0.900

8195 8157 -> 8154 8153 0.429 0.900

8195 8157 -> 8153 0.476 1.000

8195 8157 8154 -> 8153 0.429 1.000

8195 8157 8153 -> 8154 0.429 0.900

8154

8195 8154 -> 8156 8155 0.476 0.625

8195 8154 -> 8155 0.571 0.750

8195 8154 -> 8155 8153 0.429 0.562

8195 8154 -> 8172 0.429 0.562

8195 8154 -> 8157 0.429 0.562

8195 8154 -> 8157 8153 0.429 0.562

8195 8154 -> 8156 0.619 0.812

8195 8154 -> 8156 8153 0.476 0.625

8195 8154 -> 8156 8152 0.476 0.625

8195 8154 -> 8153 0.524 0.688

8195 8154 -> 8152 0.524 0.688

8195 8154 8155 -> 8156 0.476 0.833

8195 8154 8156 -> 8155 0.476 0.769

8195 8154 8155 -> 8153 0.429 0.750

8195 8154 8153 -> 8155 0.429 0.818

8195 8154 8157 -> 8153 0.429 1.000

8195 8154 8153 -> 8157 0.429 0.818

8195 8154 8156 -> 8153 0.476 0.769
 8195 8154 8153 -> 8156 0.476 0.909
 8195 8154 8156 -> 8152 0.476 0.769
 8195 8154 8152 -> 8156 0.476 0.909

8155

8195 8155 -> 8156 0.476 0.833
 8195 8155 -> 8156 8154 0.476 0.833
 8195 8155 -> 8154 0.571 1.000
 8195 8155 -> 8154 8153 0.429 0.750
 8195 8155 -> 8153 0.429 0.750
 8195 8155 8156 -> 8154 0.476 1.000
 8195 8155 8154 -> 8156 0.476 0.833
 8195 8155 8154 -> 8153 0.429 0.750
 8195 8155 8153 -> 8154 0.429 1.000

8152

8195 8152 -> 8156 8154 0.476 0.909
 8195 8152 -> 8156 0.476 0.909
 8195 8152 -> 8154 0.524 1.000
 8195 8152 8156 -> 8154 0.476 1.000
 8195 8152 8154 -> 8156 0.476 0.909

8153

8195 8153 -> 8154 8155 0.429 0.750
 8195 8153 -> 8155 0.429 0.750
 8195 8153 -> 8157 8154 0.429 0.750
 8195 8153 -> 8157 0.476 0.833
 8195 8153 -> 8156 8154 0.476 0.833
 8195 8153 -> 8156 0.476 0.833
 8195 8153 -> 8154 0.524 0.917
 8195 8153 8155 -> 8154 0.429 1.000
 8195 8153 8154 -> 8155 0.429 0.818
 8195 8153 8157 -> 8154 0.429 0.900
 8195 8153 8154 -> 8157 0.429 0.818
 8195 8153 8156 -> 8154 0.476 1.000
 8195 8153 8154 -> 8156 0.476 0.909

8156

8156 -> 8155 8195 0.476 0.769
 8156 -> 8154 8155 8195 0.476 0.769
 8156 -> 8155 0.476 0.769
 8156 -> 8154 8155 0.476 0.769
 8156 -> 8195 0.619 1.000

8156 -> 8154 8195	0.619 1.000
8156 -> 8154 8153 8195	0.476 0.769
8156 -> 8154 8152 8195	0.476 0.769
8156 -> 8153 8195	0.476 0.769
8156 -> 8152 8195	0.476 0.769
8156 -> 8154	0.619 1.000
8156 -> 8154 8153	0.476 0.769
8156 -> 8154 8152	0.476 0.769
8156 -> 8153	0.476 0.769
8156 -> 8152	0.476 0.769
8195	
8156 8195 -> 8155	0.476 0.769
8156 8195 -> 8154 8155	0.476 0.769
8156 8195 -> 8154	0.619 1.000
8156 8195 -> 8154 8153	0.476 0.769
8156 8195 -> 8154 8152	0.476 0.769
8156 8195 -> 8153	0.476 0.769
8156 8195 -> 8152	0.476 0.769
8156 8195 8155 -> 8154	0.476 1.000
8156 8195 8154 -> 8155	0.476 0.769
8156 8195 8154 -> 8153	0.476 0.769
8156 8195 8153 -> 8154	0.476 1.000
8156 8195 8154 -> 8152	0.476 0.769
8156 8195 8152 -> 8154	0.476 1.000
8154	
8156 8154 -> 8155 8195	0.476 0.769
8156 8154 -> 8155	0.476 0.769
8156 8154 -> 8195	0.619 1.000
8156 8154 -> 8153 8195	0.476 0.769
8156 8154 -> 8152 8195	0.476 0.769
8156 8154 -> 8153	0.476 0.769
8156 8154 -> 8152	0.476 0.769
8156 8154 8155 -> 8195	0.476 1.000
8156 8154 8195 -> 8155	0.476 0.769
8156 8154 8195 -> 8153	0.476 0.769
8156 8154 8153 -> 8195	0.476 1.000
8156 8154 8195 -> 8152	0.476 0.769
8156 8154 8152 -> 8195	0.476 1.000
8155	
8156 8155 -> 8195	0.476 1.000

8156 8155 -> 8154 8195	0.476	1.000
8156 8155 -> 8154	0.476	1.000
8156 8155 8195 -> 8154	0.476	1.000
8156 8155 8154 -> 8195	0.476	1.000
8152		
8156 8152 -> 8154 8195	0.476	1.000
8156 8152 -> 8195	0.476	1.000
8156 8152 -> 8154	0.476	1.000
8156 8152 8195 -> 8154	0.476	1.000
8156 8152 8154 -> 8195	0.476	1.000
8153		
8156 8153 -> 8154 8195	0.476	1.000
8156 8153 -> 8195	0.476	1.000
8156 8153 -> 8154	0.476	1.000
8156 8153 8195 -> 8154	0.476	1.000
8156 8153 8154 -> 8195	0.476	1.000
8157		
8157 -> 8195	0.476	1.000
8157 -> 8154 8195	0.429	0.900
8157 -> 8154 8153 8195	0.429	0.900
8157 -> 8153 8195	0.476	1.000
8157 -> 8154	0.429	0.900
8157 -> 8154 8153	0.429	0.900
8157 -> 8153	0.476	1.000
8154		
8157 8154 -> 8195	0.429	1.000
8157 8154 -> 8153 8195	0.429	1.000
8157 8154 -> 8153	0.429	1.000
8157 8154 8195 -> 8153	0.429	1.000
8157 8154 8153 -> 8195	0.429	1.000
8153		
8157 8153 -> 8154 8195	0.429	0.900
8157 8153 -> 8195	0.476	1.000
8157 8153 -> 8154	0.429	0.900
8157 8153 8195 -> 8154	0.429	0.900
8157 8153 8154 -> 8195	0.429	1.000
8195		
8157 8195 -> 8154	0.429	0.900
8157 8195 -> 8154 8153	0.429	0.900
8157 8195 -> 8153	0.476	1.000

8157 8195 8154 -> 8153	0.429	1.000
8157 8195 8153 -> 8154	0.429	0.900
8073		
8073 -> 8154	0.429	1.000
8155		
8155 -> 8172	0.429	0.692
8155 -> 8172 8154	0.429	0.692
8155 -> 8195	0.571	0.923
8155 -> 8156 8195	0.476	0.769
8155 -> 8156 8154 8195	0.476	0.769
8155 -> 8154 8195	0.571	0.923
8155 -> 8154 8153 8195	0.429	0.692
8155 -> 8153 8195	0.429	0.692
8155 -> 8156	0.476	0.769
8155 -> 8156 8154	0.476	0.769
8155 -> 8154	0.619	1.000
8155 -> 8154 8153	0.429	0.692
8155 -> 8153	0.429	0.692
8172		
8155 8172 -> 8154	0.429	1.000
8195		
8155 8195 -> 8156	0.476	0.833
8155 8195 -> 8156 8154	0.476	0.833
8155 8195 -> 8154	0.571	1.000
8155 8195 -> 8154 8153	0.429	0.750
8155 8195 -> 8153	0.429	0.750
8155 8195 8156 -> 8154	0.476	1.000
8155 8195 8154 -> 8156	0.476	0.833
8155 8195 8154 -> 8153	0.429	0.750
8155 8195 8153 -> 8154	0.429	1.000
8156		
8155 8156 -> 8195	0.476	1.000
8155 8156 -> 8154 8195	0.476	1.000
8155 8156 -> 8154	0.476	1.000
8155 8156 8195 -> 8154	0.476	1.000
8155 8156 8154 -> 8195	0.476	1.000
8154		
8155 8154 -> 8172	0.429	0.692
8155 8154 -> 8156 8195	0.476	0.769
8155 8154 -> 8195	0.571	0.923

8155 8154 -> 8153 8195 0.429 0.692
 8155 8154 -> 8156 0.476 0.769
 8155 8154 -> 8153 0.429 0.692
 8155 8154 8195 -> 8156 0.476 0.833
 8155 8154 8156 -> 8195 0.476 1.000
 8155 8154 8195 -> 8153 0.429 0.750
 8155 8154 8153 -> 8195 0.429 1.000
 8153
 8155 8153 -> 8154 8195 0.429 1.000
 8155 8153 -> 8195 0.429 1.000
 8155 8153 -> 8154 0.429 1.000
 8155 8153 8195 -> 8154 0.429 1.000
 8155 8153 8154 -> 8195 0.429 1.000
 8152
 8152 -> 8156 8154 8195 0.476 0.909
 8152 -> 8156 8195 0.476 0.909
 8152 -> 8154 8195 0.524 1.000
 8152 -> 8195 0.524 1.000
 8152 -> 8156 8154 0.476 0.909
 8152 -> 8156 0.476 0.909
 8152 -> 8154 0.524 1.000
 8156
 8152 8156 -> 8154 8195 0.476 1.000
 8152 8156 -> 8195 0.476 1.000
 8152 8156 -> 8154 0.476 1.000
 8152 8156 8195 -> 8154 0.476 1.000
 8152 8156 8154 -> 8195 0.476 1.000
 8154
 8152 8154 -> 8156 8195 0.476 0.909
 8152 8154 -> 8195 0.524 1.000
 8152 8154 -> 8156 0.476 0.909
 8152 8154 8195 -> 8156 0.476 0.909
 8152 8154 8156 -> 8195 0.476 1.000
 8195
 8152 8195 -> 8156 8154 0.476 0.909
 8152 8195 -> 8156 0.476 0.909
 8152 8195 -> 8154 0.524 1.000
 8152 8195 8156 -> 8154 0.476 1.000
 8152 8195 8154 -> 8156 0.476 0.909

1266 -> 8154	0.429 1.000
8150	
8150 -> 8154	0.429 1.000
8153	
8153 -> 8154 8155 8195	0.429 0.750
8153 -> 8155 8195	0.429 0.750
8153 -> 8154 8155	0.429 0.750
8153 -> 8155	0.429 0.750
8153 -> 8157 8154 8195	0.429 0.750
8153 -> 8157 8195	0.476 0.833
8153 -> 8156 8154 8195	0.476 0.833
8153 -> 8156 8195	0.476 0.833
8153 -> 8154 8195	0.524 0.917
8153 -> 8195	0.571 1.000
8153 -> 8157 8154	0.429 0.750
8153 -> 8157	0.476 0.833
8153 -> 8156 8154	0.476 0.833
8153 -> 8156	0.476 0.833
8153 -> 8154	0.524 0.917
8195	
8153 8195 -> 8154 8155	0.429 0.750
8153 8195 -> 8155	0.429 0.750
8153 8195 -> 8157 8154	0.429 0.750
8153 8195 -> 8157	0.476 0.833
8153 8195 -> 8156 8154	0.476 0.833
8153 8195 -> 8156	0.476 0.833
8153 8195 -> 8154	0.524 0.917
8153 8195 8155 -> 8154	0.429 1.000
8153 8195 8154 -> 8155	0.429 0.818
8153 8195 8157 -> 8154	0.429 0.900
8153 8195 8154 -> 8157	0.429 0.818
8153 8195 8156 -> 8154	0.476 1.000
8153 8195 8154 -> 8156	0.476 0.909
8156	
8153 8156 -> 8154 8195	0.476 1.000
8153 8156 -> 8195	0.476 1.000
8153 8156 -> 8154	0.476 1.000
8153 8156 8195 -> 8154	0.476 1.000
8153 8156 8154 -> 8195	0.476 1.000
8157	

8153 8157 -> 8154 8195 0.429 0.900
 8153 8157 -> 8195 0.476 1.000
 8153 8157 -> 8154 0.429 0.900
 8153 8157 8195 -> 8154 0.429 0.900
 8153 8157 8154 -> 8195 0.429 1.000

8154

8153 8154 -> 8155 8195 0.429 0.818
 8153 8154 -> 8155 0.429 0.818
 8153 8154 -> 8157 8195 0.429 0.818
 8153 8154 -> 8156 8195 0.476 0.909
 8153 8154 -> 8195 0.524 1.000
 8153 8154 -> 8157 0.429 0.818
 8153 8154 -> 8156 0.476 0.909
 8153 8154 8155 -> 8195 0.429 1.000
 8153 8154 8195 -> 8155 0.429 0.818
 8153 8154 8195 -> 8157 0.429 0.818
 8153 8154 8157 -> 8195 0.429 1.000
 8153 8154 8195 -> 8156 0.476 0.909
 8153 8154 8156 -> 8195 0.476 1.000

8155

8153 8155 -> 8154 8195 0.429 1.000
 8153 8155 -> 8195 0.429 1.000
 8153 8155 -> 8154 0.429 1.000
 8153 8155 8195 -> 8154 0.429 1.000
 8153 8155 8154 -> 8195 0.429 1.000

8154

8154 -> 8073 0.429 0.474
 8154 -> 8172 8155 0.429 0.474
 8154 -> 8156 8155 8195 0.476 0.526
 8154 -> 8155 8195 0.571 0.632
 8154 -> 8155 8153 8195 0.429 0.474
 8154 -> 8156 8155 0.476 0.526
 8154 -> 8155 0.619 0.684
 8154 -> 8155 8153 0.429 0.474
 8154 -> 8172 8195 0.429 0.474
 8154 -> 8172 0.476 0.526
 8154 -> 1266 0.429 0.474
 8154 -> 8157 8195 0.429 0.474
 8154 -> 8157 8153 8195 0.429 0.474
 8154 -> 8156 8195 0.619 0.684

8154 -> 8156 8153 8195	0.476 0.526
8154 -> 8156 8152 8195	0.476 0.526
8154 -> 8195	0.762 0.842
8154 -> 8153 8195	0.524 0.579
8154 -> 8152 8195	0.524 0.579
8154 -> 8157	0.429 0.474
8154 -> 8157 8153	0.429 0.474
8154 -> 8156	0.619 0.684
8154 -> 8156 8153	0.476 0.526
8154 -> 8156 8152	0.476 0.526
8154 -> 8153	0.524 0.579
8154 -> 8152	0.524 0.579
8154 -> 8150	0.429 0.474
8154 -> 3616	0.429 0.474
8172	
8154 8172 -> 8155	0.429 0.900
8154 8172 -> 8195	0.429 0.900
8195	
8154 8195 -> 8156 8155	0.476 0.625
8154 8195 -> 8155	0.571 0.750
8154 8195 -> 8155 8153	0.429 0.562
8154 8195 -> 8172	0.429 0.562
8154 8195 -> 8157	0.429 0.562
8154 8195 -> 8157 8153	0.429 0.562
8154 8195 -> 8156	0.619 0.812
8154 8195 -> 8156 8153	0.476 0.625
8154 8195 -> 8156 8152	0.476 0.625
8154 8195 -> 8153	0.524 0.688
8154 8195 -> 8152	0.524 0.688
8154 8195 8155 -> 8156	0.476 0.833
8154 8195 8156 -> 8155	0.476 0.769
8154 8195 8155 -> 8153	0.429 0.750
8154 8195 8153 -> 8155	0.429 0.818
8154 8195 8157 -> 8153	0.429 1.000
8154 8195 8153 -> 8157	0.429 0.818
8154 8195 8156 -> 8153	0.476 0.769
8154 8195 8153 -> 8156	0.476 0.909
8154 8195 8156 -> 8152	0.476 0.769
8154 8195 8152 -> 8156	0.476 0.909
8156	

8154 8156 -> 8155 8195 0.476 0.769
8154 8156 -> 8155 0.476 0.769
8154 8156 -> 8195 0.619 1.000
8154 8156 -> 8153 8195 0.476 0.769
8154 8156 -> 8152 8195 0.476 0.769
8154 8156 -> 8153 0.476 0.769
8154 8156 -> 8152 0.476 0.769
8154 8156 8155 -> 8195 0.476 1.000
8154 8156 8195 -> 8155 0.476 0.769
8154 8156 8195 -> 8153 0.476 0.769
8154 8156 8153 -> 8195 0.476 1.000
8154 8156 8195 -> 8152 0.476 0.769
8154 8156 8152 -> 8195 0.476 1.000

8157

8154 8157 -> 8195 0.429 1.000
8154 8157 -> 8153 8195 0.429 1.000
8154 8157 -> 8153 0.429 1.000
8154 8157 8195 -> 8153 0.429 1.000
8154 8157 8153 -> 8195 0.429 1.000

8155

8154 8155 -> 8172 0.429 0.692
8154 8155 -> 8156 8195 0.476 0.769
8154 8155 -> 8195 0.571 0.923
8154 8155 -> 8153 8195 0.429 0.692
8154 8155 -> 8156 0.476 0.769
8154 8155 -> 8153 0.429 0.692
8154 8155 8195 -> 8156 0.476 0.833
8154 8155 8156 -> 8195 0.476 1.000
8154 8155 8195 -> 8153 0.429 0.750
8154 8155 8153 -> 8195 0.429 1.000

8152

8154 8152 -> 8156 8195 0.476 0.909
8154 8152 -> 8195 0.524 1.000
8154 8152 -> 8156 0.476 0.909
8154 8152 8195 -> 8156 0.476 0.909
8154 8152 8156 -> 8195 0.476 1.000

8153

8154 8153 -> 8155 8195 0.429 0.818
8154 8153 -> 8155 0.429 0.818
8154 8153 -> 8157 8195 0.429 0.818

8154 8153 -> 8156 8195 0.476 0.909
8154 8153 -> 8195 0.524 1.000
8154 8153 -> 8157 0.429 0.818
8154 8153 -> 8156 0.476 0.909
8154 8153 8155 -> 8195 0.429 1.000
8154 8153 8195 -> 8155 0.429 0.818
8154 8153 8195 -> 8157 0.429 0.818
8154 8153 8157 -> 8195 0.429 1.000
8154 8153 8195 -> 8156 0.476 0.909
8154 8153 8156 -> 8195 0.476 1.000

EK 3: Haber Kaynakları Arasındaki İlişkiler ve Jaccard Katsayıları

Haber Kaynakları (RSS_ID)	Jaccard Katsayısı	Birlikte Kuralları	
Sınıf adı: “Business”		Destek	Güven
(8172, 1266)	0.125	0.286	0.667
Sınıf adı: “Entertainment”			
(8155, 8156)	0.231	0.476	0.769
(8156, 8185)	0.100	-	-
(8156, 8174)	0.091	0.286	0.857
(8156, 8186)	0.111	N/A	N/A
(8185, 8174)	0.250	-	-
(8185, 8186)	0.500	N/A	N/A
(8174, 8186)	0.333	N/A	N/A
Movie (“Entertainment” alt sınıfı)			
(8155, 8156)	0.375	0.476	0.769
(8156, 8185)	0.167	-	-
(8156, 8174)	0.143	0.286	0.857
(8156, 8186)	0.167	N/A	N/A
(8185, 8174)	0.500	-	-
(8185, 8186)	1.000	N/A	N/A
(8174, 8186)	0.500	N/A	N/A
Sınıf adı: “Newspaper”			
(8190, 1266)	0.333	-	-
Sınıf adı: “Organization”			
(8185, 8186)	1.000	N/A	N/A
Sınıf adı: “Person”			
(8154, 5580)	0.125	0.333	1
(8185, 8174)	0.111	-	-
(8185, 8151)	0.125	N/A	N/A
(8185, 8150)	0.200	-	-
(8174, 8151)	0.100	N/A	N/A
(8174, 8150)	0.143	-	-
(8151, 8150)	0.167	N/A	N/A
(8153, 5580)	0.077	0.333	1
(1265, 1483)	0.111	N/A	N/A
(8154, 8153)	0.083	0.524	0.917

Haber Kaynakları (RSS_ID)	Jaccard Katsayısı	Birliktelik Kuralları	
Sınıf adı: “Person” (devam)		Destek	Güven
(8154, 8172)	0.100	0.476	1
(8154, 8185)	0.091	0.286	1
(8172, 5580)	0.091	-	-
(8172, 8185)	0.167	-	-
(5580, 8185)	0.083	-	-
Sınıf adı: “Place”			
(8153, 8150)	0.167	0.286	0.667
(8172, 8155)	0.500	0.429	0.9
(8172, 8154)	0.250	0.476	1
(8155, 8154)	0.200	0.619	1
(1964, 8195)	0.200	N/A	N/A
(1964, 3616)	0.167	N/A	N/A
(8195, 3616)	0.200	0.381	0.444
(8174, 8195)	0.167	0.286	0.857
(8174, 3616)	0.143	-	-
(1399, 3616)	0.143	N/A	N/A
(1399, 8151)	0.333	N/A	N/A
(3616, 8151)	0.143	N/A	N/A
(8174, 1265)	0.250	-	-
City (“Place” alt sınıfı)			
(1964, 8195)	0.500	N/A	N/A
(1964, 3616)	0.333	N/A	N/A
(8195, 3616)	0.500	0.381	0.444
(8174, 8195)	0.333	0.286	0.857
(8174, 3616)	0.250	-	-
(1399, 3616)	0.333	N/A	N/A
(1399, 8151)	1.000	N/A	N/A
(3616, 8151)	0.333	N/A	N/A
(8174, 1265)	1.000	-	-
Sınıf adı: “Politics”			
(8154, 8155)	1.000	0.619	0.684
Sınıf adı: “Sports”			
(8186, 8187)	1.000	N/A	N/A

Haber Kaynakları (RSS_ID)	Jaccard Katsayısı	Birliktelik Kuralları	
Sınıf adı: “Technology”		Destek	Güven
(8152, 8172)	0.125	0.333	0.7
(8152, 8073)	1.000	0.286	0.667
(8152, 8155)	0.167	0.381	0.615
(8155, 8172)	0.067	0.429	0.9
(1266, 8172)	0.063	0.286	0.667
(1266, 1265)	0.067	-	-
(8172, 1265)	0.038	0.286	0.75
Software (“Technology” alt sınıfı)			
(8152, 8155)	1.000	0.381	0.615
(8152, 8172)	0.500	0.333	0.7
(8155, 8172)	0.500	0.429	0.9
Web (“Technology” alt sınıfı)			
(8152, 8172)	0.091	0.333	0.7
(1266, 8172)	0.111	0.286	0.667
(1266, 1265)	0.500	-	-
(8172, 1265)	0.125	0.286	0.75
(8152, 8155)	0.250	0.381	0.615