

UZMAN BULMA PROBLEMİNE SEMANTİK WEB YAKLAŞIMI

ONUR DEMİRCAN

**YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**TOBB EKONOMİ VE TEKNOLOJİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

EYLÜL 2011

Fen Bilimleri Enstitü onayı

Prof. Dr. Ünver KAYNAK
Müdür

Bu tezin Yüksek Lisans derecesinin tüm gereksinimlerini sağladığını onaylarım.

Doç. Dr. Erdoğan DOĞDU
Anabilim Dalı Başkanı

Onur Demircan tarafından hazırlanan UZMAN BULMA PROBLEMİNE SEMANTİK-
WEB YAKLAŞIMI adlı bu tezin Yüksek Lisans tezi olarak uygun olduğunu onaylarım.

Doç. Dr. Erdoğan DOĞDU
Tez Danışmanı

Tez Jüri Üyeleri

Başkan: Yrd. Doç. Dr. Tansel ÖZYER

Üye : Doç Dr. Erdoğan DOĞDU

Üye : Doç. Dr. Bülent TAVLI

Tez Bildirimi

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada orijinal olmayan her türlü kaynağa eksiksiz atıf yapıldığını bildiririm.

Onur DEMİRCAN

Üniversitesi : TOBB Ekonomi ve Teknoloji Üniversitesi
Enstitüsü : Fen Bilimleri Enstitüsü
Anabilim Dalı : Bilgisayar Mühendisliği
Tez Danışmanı : Doç. Dr. Erdoğan Dođdu
Tez Türü ve Tarihi : Yüksek Lisans – Eylül 2011

Onur DEMİRCAN

UZMAN BULMA PROBLEMİNE SEMANTİK-WEB YAKLAŞIMI

ÖZET

Uzman Bulma problemi bir gerçek hayat problemidir. Özel bir konu üzerinde profesyonelleşmiş uzmanlara hem akademik hem de endüstriyel dünyada olan ihtiyaç yadsınamaz. Bugüne kadar uzman bulma problemine çözüm bulmak isteyen birçok sistem ortaya konulmuştur. Bu sistemler büyük ölçüde insan gücü ile toplanmış verilerden elle oluşturulmuş veritabanları kullanılarak gerçekleştirilmişlerdir. Veritabanlarının sadece belirli bir büyüklüğe kadar oluşturulabilmesi ve hem akademisyenlerin yeni çalışmaları ile güncellenememesi hem de güncel konuların kapsamaması sebebiyle güncelliğini ve etkinliğini zamanla yitirmektedir. İnternette toplanan veriler ile oluşturulmuş veritabanları üzerine geliştirilen yöntemlerin yerine, semantik web yaklaşımlarıyla oluşturulan, yaşayabilen sistemler hem güncelliğini korumakta hem de veriye ulaşımında daha kolay ve etkin bir yol olabilmektedirler. Bu sebeple semantik web yaklaşımının uzman bulma probleminin çözümünde kullanılabileceğini düşünerek çalışmalarımıza bu şekilde yön verdik. Çalışmamızda uzman bulma problemi için kullanılan diğer modelleri ve daha önceki çalışmaları incelerken, veri toplama, yorumlama ve anlamlandırma için bu çözümlerin üzerine kendi modellerimizi geliştirdik. Çalışmamızda akademisyenlerin profilleri ve uzmanlık alanlarının tespiti için DBLP datasını ve diğer semantik web verilerini kullandık. Bunun haricinde modellerimiz için Google Scholar üzerine web arayıcısı (crawler) yazılımını geliştirdik. Modellerimizin Java ve C# altyapısını kullanarak uygulamasını geliştirdik. Çıkan sonuçların performans değerlerini karşılaştırabilmek için de belirli bir küme setinde uzmanlık alanlarını bildiğimiz akademisyenlere ait sonuçlarla karşılaştırdık. Geliştirdiğimiz modelleri temel alarak geliştirdiğimiz hibrit modelle belirli bir konudaki en iyi (top-k) uzmanı bulma problemi için bir yaklaşım geliştirdik. En iyi uzmanı bulma problemine olan çözümümüzün etkinliğini değerlendirmek için daha önceden tanımlı veri kümelerini kullandık. Geliştirdiğimiz hibrit çözümün önceki çalışmalardan daha iyi sonuçlar verdiğini gösterdik.

Anahtar Kelimeler: Uzman bulma, semantik web.

Üniversitesi : TOBB Ekonomi ve Teknoloji Üniversitesi
Enstitüsü : Fen Bilimleri Enstitüsü
Anabilim Dalı : Bilgisayar Mühendisliği
Tez Danışmanı : Doç. Dr. Erdoğan Dođdu
Tez Türü ve Tarihi : Yüksek Lisans – Eylül 2011

Onur DEMİRCAN

SEMANTIC APPROACH FOR EXPERT FINDING PROBLEM

Abstract

Expert Finding is a real life problem. The need for finding experts on a specific topic in both academic and industrial world is undeniable. Many systems and models have been performed in order to find a solution to this problem. These systems have used the data that has been collected manually in order to arrange an expert database. Since the collection of data is and keeping the data up to date is a big concern, we refer to Semantic Web data in order to collect, organize and analyze the data. We have used DBLP data in order to obtain academicians' profiles and their articles. Moreover we have developed a Google Scholar crawler in order to get citation values of articles. We have examined the previous methods and models that have been implemented to find experts. We have introduced our models and then implemented them with C# and Java programming languages. After the implementation we have tested and verified our solutions with the predefined sets of academicians and their area of interests. With the integration of our models, we have established a hybrid model in order to find top-k experts in a specific area. In order to measure our hybrid model's performance, we have compared our results with the predefined top-k experts that have been verified.

Keywords: Expert finding, semantic web.

TEŐEKKÜR

Deęerli bilgilerini benimle paylaőan ve hiębir konuda benden yardımını esirgemeyen tez danıőmanım Doę.Dr. Erdoğan DOęDU'ya, asistanlık ve öğrencilik yaptığım süre boyunca desteklerini esirgemeyen TOBB ETÜ Bilgisayar Mühendislięi Bölümü öğretim üyelerine teşekkür ederim. Öğrenim hayatım boyunca beni destekleyen anne ve babama, sevgili eşime, Eventseer [17] sisteminin kurucusu ve yöneticisi Thomas Brox Røst'a ęalıőmalarım boyunca bana verdikleri destek için teşekkür ederim.

İÇİNDEKİLER

1. GİRİŞ	12
2. UZMAN BULMA PROBLEMİ.....	15
2.1. Karşılaşılan Problemler	16
2.1.1. Profil Çıkarma ve İsim Benzerliği Problemi (Profile Extraction and Name Disambiguation).....	16
3. İLGİLİ ÇALIŞMALAR	18
3.1. İzlenen Yöntemler ve Modeller	18
3.1.1. Olasılıksal (Probabilistic) Model	18
3.1.2. Basit Dil Modeli (Basic Language Model).....	18
3.1.3. Ağırlıklı Dil Modeli (Weighted Language Model).....	20
3.2. Sorgu Bağımlı ve Sorgu Bağımsız Modeller (Query Dependent and Query Independent Approaches).....	21
4. SEMANTİK WEB YAKLAŞIMI ile UZMAN BULMA.....	23
4.1. Neden Semantik Yaklaşım?.....	23
4.2. Semantik Web ve İlgili Teknolojiler	23
4.3. Linked Data, DBLP ve Diğer Veri Web Kaynakları.....	24
4.3.1. DBLP	24
4.3.2. Linked Data.....	24
4.4. Semantik Web-tabanlı Uzman Bulma Yöntemleri.....	25
4.4.1. Başlık Tabanlı Arama Modeli.....	25
4.4.2. Linked Data Tabanlı Semantik Web Modeli	30
4.4.3. Belirli bir Konudaki En iyi Uzmanları Bulma.....	38
5. GERÇEKLEŞTİRİM	43
5.1. Kullanılan Geliştirme Araçları	43
5.2. Kullanılan Çerçeve Programlar (Frameworks).....	44
5.3. Veri Modeli	44
6. PERFORMANS ANALİZİ	46
6.1. Test Ortamı.....	47
6.2. Performans Metrikleri.....	48
6.3. Test Sonuçları ve Analiz.....	48

6.3.1. Başlık Tabanlı Model ve Linked Data tabanlı Semantik Web Modeli Sonuçları	48
6.3.2. En iyi (Top-k) Uzmanı Bulma Algoritması Sonuçları.....	53
7. SONUÇ ve GELECEK ÇALIŞMALAR	62
8. REFERANSLAR	64
Ek 1. Veri seti için elde edilen sonuçlar.....	66
Ek 2. Aktör Ontolojisi, http://www.aktors.org/ontology/portal	85

Tablolar

Tablo 4-1 DBLP İstatistikleri.....	24
Tablo 4-2 DBLP verisinde Erdoğan Doğdu'ya ait olan makaleler	26
Tablo 4-3 Başlık Tabanlı Modele göre Erdoğan Doğdu'ya ait ilgi alanları.....	28
Tablo 4-4 Veri setimizde Erdoğan Doğdu'ya ait ilgi alanları.....	28
Tablo 4-5 Yazarlarımıza ait toplam makale sayısı.....	29
Tablo 4-6 DBLP'de makalesi bulunan yazar sayısı.....	29
Tablo 4-7 DBLP verisi için ortalama ilgi alanı değerleri.....	29
Tablo 4-8 Veri setimiz için DBLP ilgi alanı istatistiği.....	29
Tablo 4-9 Linked Data tabanlı Semantik Web modelimize göre yazarın ilgi alanları.....	32
Tablo 4-10 Yazarlarımızın ACM ve IEEE'de bulunan makale sayısı	33
Tablo 4-11 ACM ve IEEE'de makalesi bulunan yazar sayısı.....	33
Tablo 4-12 ACM ve IEEE verisi için ortalama ilgi alanı değerleri	34
Tablo 4-13 Linked Data tabanlı Semantik Web modelimize (DBLP++ ile) göre yazarın ilgi alanları.....	35
Tablo 4-14 Yazarlarımızın DBLP++'da bulunan makale sayısı.....	37
Tablo 4-15 DBLP++'da makalesi bulunan yazar sayısı	37
Tablo 4-16 DBLP++ verisi için ortalama ilgi alanı değerleri	37
Tablo 4-17 Konular için benzer ve alt konular	40
Tablo 5-1 Kullanılan Frameworkler.....	44
Tablo 6-1 Başlık tabanlı modelimiz için bilinen ilgi alanlarını bulma oranı	48
Tablo 6-2 Linked Data tabanlı modelimiz için ACM ve IEEE ile bilinen ilgi alanlarını bulma oranı.....	48
Tablo 6-3 Linked Data tabanlı modelimiz için DBLP++ ile bilinen ilgi alanlarını bulma oranı.....	49
Tablo 6-4 Bilinen ilgi alanı ile DBLP++'da bulunan ilgi alanı karşılaştırması	50
Tablo 6-5 Linked Data tabanlı modelimiz için hem ACM, IEEE hem de DBLP++ verisi kullanılarak bilinen ilgi alanlarını bulma oranı	50
Tablo 6-6 ACM, IEEE kütüphanesinde DBLP++ kütüphanesine göre daha fazla makalesi bulunan akademisyenlere ait makale sayıları.....	51
Tablo 6-7 Erdoğan Doğdu'ya ait ACM ve IEEE kütüphanesinde bulunan makaleler	51
Tablo 6-8 Erdoğan Doğdu'ya ait DBLP++ kütüphanesinde bulunan makaleler	52
Tablo 6-9 Modeller için farklı veri setleri kullanılarak karşılaştırmalı sonuçları	52
Tablo 6-10 DBLP++ ve Benzer İlgi alanı Bulma Algoritması kullanarak "Support Vector Machine" konusunda bulunan ilk 5.....	53
Tablo 6-11 Deng'e göre "Support Vector Machine" başlığı için ilk 5 akademisyen	54
Tablo 6-12 "Vladimir Vapnik" için DBLP++ veri setinde bulunan makaleler	55
Tablo 6-13 "Vladimir Vapnik" için ACM ve IEEE veri setinde bulunan makaleler.....	55
Tablo 6-14 "Vladimir Vapnik" için DBLP veri setinde bulunan makaleler	56

Tablo 6-15 Hibrit modelimizin bütün veri setleri ile çalışması sonucunda “Vladimir Vapnik” için alınan sonuçlar	57
Tablo 6-16 Makale alıntılanma değerleri ve yazar sıralaması	58
Tablo 6-17 Deng’in çalışmasında information extraction için sonuçlar	58
Tablo 6-18 Bizim çalışmamızın benzer ilgi alanı bulma algoritması olmadan Information extraction için bulduğu sonuçlar	58
Tablo 6-19 Bizim çalışmamızın benzer ilgi alanı bulma algoritması ile Information extraction için bulduğu sonuçlar	59
Tablo 6-20 Deng’in çalışmasında Semantik Web için sonuçlar	59
Tablo 6-21 Bizim çalışmamızın benzer ilgi alanı bulma algoritması olmadan Semantik Web için bulduğu sonuçlar	60
Tablo 6-22 Bizim çalışmamızın benzer ilgi alanı bulma algoritması kullanarak Semantik Web için bulduğu sonuçlar.....	60
Tablo 6-23 Semantik Web alt başlıkları.....	60
Tablo 6-24 Deng’in veri setindeki konular ile ilk 5 ve bizim sonuçlarımız için uyuşan yazar sayıları ve oranlar	61

Şekiller

Şekil 3-1 Ağırlıklı Dil Modeli.....	21
Şekil 4-1 Başlık tabanlı model Veri Transferi Şeması.....	27
Şekil 4-2 ACM, IEEE ile Linked Data tabanlı Semantik Web Algoritması Veri Transferi Şeması	31
Şekil 4-3 Faceted DBLP ile Linked Data tabanlı Semantik Web Yaklaşımı Veri Transferi Şeması	36
Şekil 4-4 Dokümanlar ve sorgular [5].....	39
Şekil 4-5 Google Scholar ile alıntılanma sayısı tespiti	39
Şekil 5-1 Kullandığımız veri tabanındaki veri modeli.....	44

1. GİRİŞ

Hem endüstri hem de akademik dünyada önemli problemlerden birisi, özel bir başlık için uzman bulma problemidir. Kullanılabilirliği ve geçerliliği ispatlanmış ve en iyi k sayıdaki (top k) uzmanları bulabilecek bir sistem çok kullanışlı olabilir.

Bu sistem bir kişinin akademik uzmanlık konularının ortaya çıkarılması veya belirli bir konuyla ilgili en iyileri bulmak için kullanılabilir. Örneğin yeni bir araştırma alanında danışmanlık hizmeti verecek kişilerin tespit edilmesi, devlet tarafından verilecek hibelerin değerlendirilmesi için hangi hakemlerin atanacağı [1] , akran değerlendirme sürecinde makalelerin değerlendirilmesi için yapılan atamaların otomatikleştirilmesi gibi ana başlıklar kullanım alanı olarak sunulabilir [2] [3] .

Uzman bulma, “bilirkişi keşfi” (Who knows what?) olarak da adlandırılabilir.

Geleneksel yaklaşımlardan bir tanesi elle bir uzmanlık veritabanı oluşturmak ve bu veri tabanındaki verileri elle düzenlemek ve kontrol etmek şeklindedir [6] . Akademik ve endüstriyel alanda sürekli gelişen ve değişen konuların olması ve uzmanların uzmanlıklarının göstergesi olan verilerin sürekli değişmesi yaşayan bir veri kümesine olan ihtiyacı açık bir şekilde ortaya koymaktadır. Yaşayan bir veri tabanı kümesi oluşturmak için semantik web verisi çok ciddi bir girdi olabilir.

Bu tezde akademi ve endüstri dünyası için bu şekilde güvenilir ve sürekli yaşayan bir veri seti ve tabanı oluşturabilmek için web de bulunan Semantik Web verisini sistemimizde kullandık.

Semantik olmayan grafik tabanlı yaklaşımlar yerine, SPARQL sorgularını kullanarak verileri toplamaya, topladığımız verileri anlamlandırmaya ve sınıflandırmaya çalıştık. Bir akademisyenin uzmanlık alanının en büyük göstergesi olan yazdığı makale bilgilerini, konferans bilgilerini ve ayrıca akademisyenlerin bilgilerini içeren DBLP verisini kullandık.

Kullandığımız DBLP veri seti 474,235 akademisyen ve bunlara ait 673,789 makaleyi içermektedir.

DBLP verisi makalelerin hangi akademisyen tarafından yazıldığı ve makalelerin başlıklarını barındırmasına rağmen makaleler ile ilgili anahtar kelimeleri ve özetleri içermemektedir. DBLP veri setindeki bu eksikliği ACM ve IEEE çevrimiçi kütüphanelerinden SPARQL sorguları ile makalelere ait olan ilgi alanlarını (area of interest) sorgulayarak giderdik.

ACM ve IEEE kütüphanelerinde bulunan verilerin de performans olarak yetersiz kalması bizi başka arayışlara yöneltti. Daha sonra Faceted DBLP [19] isimli çalışmayı projemize veri kaynağı olarak dâhil edip daha etkin sonuçlar almaya çalıştık.

Başlık tabanlı yaklaşım ve Linked Data tabanlı Semantik Web yaklaşımı şeklinde iki farklı algoritma kullanarak akademisyenlerin yazdığı makalelerden yola çıkarak uzmanlıklarını tespit etmeye çalıştık.

Başlık tabanlı ilk modelimizde DBLP verisinde bulunan akademisyenlerin yine DBLP verisinde bulunan makale başlıklarını ele alarak, bilgisayar bilimi için özelleştirilmiş bir konu havuzu ile eşleştirmesini yapmaya çalıştık. Bilgisayar bilimine ait güncel, güvenilir ve kullanılabilir bütün konuları içeren bir veri tabanı oluşturma süreci oldukça zahmetli bir süreç olmasına rağmen Eventseer [17] isimli sistemden bilgisayar bilimi ile alakalı, bütün popüler ve güncel başlıkları içeren 5,282 başlık bulduk ve onu kullandık.

Linked Data tabanlı Semantik Web yaklaşımımız için ise ACM ve IEEE kütüphaneleri tarafından kullanılan aktör ontolojisinde bulunan generic area of interest property'sini kullanarak makalenin hangi ilgi alanına ait olduğunun tespitinden yola çıkarak geliştirdik.

Linked Data tabanlı Semantik- Web yaklaşımımız için ek olarak Faceted DBLP projesi ve bünyesinde bulunan Dublin Core ontolojisi ve “subject” özelliğini (property) kullanarak akademisyenlere ait ilgi alanı kümemizi genişlettik.

Sonrasında da ilgi alanları noktasında en iyi (top-k) akademisyenleri bulmak için Başlık Tabanlı modelimizle Linked Data tabanlı Semantik Web yaklaşımımızı bütünleştirerek hibrit bir model oluşturduk.

Deney sonuçlarımızla Semantik Linked Data tabanlı yaklaşımımızın Başlık tabanlı modelimize göre daha iyi performans göstermiş olduğunu gördük. Sonuçları değerlendirebilmek için 111 kişiden oluşan bir veri seti hazırladık ve bu veri seti akademisyenlerin web sayfalarında kendilerini tanımlamış oldukları uzmanlık alanlarını içermektedir. Sonrasında bu veri setindeki değerler ile kendi algoritmalarımızda ortaya çıkan sonuçları karşılaştırdık. Bu karşılaştırmaları hem Google arama ara yüzünü (search api) kullanarak hem de elle kontrol ederek yaptık.

Uzmanlara ait web sayfalarını DBLP veri setinden SPARQL sorguları neticesinde elde ettik.

2. Bölümde uzman bulma problemi açıklanmaktadır. 3. Bölümde daha önce yapılan çalışmalar incelenmiştir. 4. Bölümde önerdiğimiz Semantik Web yaklaşımı ile uzman bulma yaklaşımı açıklanmıştır. 5. Bölümde sistem gerçekleştirimi anlatılmıştır. 6. Bölümde ise makalelerden tespit etmiş olduğumuz uzmanlık alanlarının akademisyenlerin kendini tanımladıkları uzmanlık alanları ile uyduğu sonucu detaylı olarak değerlendirilmiştir. 7. Bölümde ise sonuçlar ve gelecek çalışmalara değinilmiştir.

2. UZMAN BULMA PROBLEMİ

Uzman bulma yöntemi verilen başlığa göre (query topic) verilen akademisyenin o konu üzerinde uzman olabilme olasılığını hesaplamaya çalışmaktadır. Bu hesaplama yapılırken uzmanlar ile ilgili dokümanlar kullanılmakta, dokümanların ilgili oldukları alan ilgili yazarların uzmanlık alanı olarak değerlendirilmektedir.

Uzman Bulma konusunun TREC Enterprise Track [13] içine dâhil edilmesi bu konu üzerine yapılan çalışmaları oldukça artırmıştır [5] .

Uzman Bulma konusu ile ilgili yapılan çalışmalar iki ana model üzerinden ilerlemektedir. Sorgu Bağımlı ve Sorgu Bağımsız olarak iki ayrı model altında incelenen bu çalışma bir sonraki başlıkta ayrıntılı olarak ele alınmıştır [5] .

Balog [8] bu iki yöntemi incelemiş ve performanslarını karşılaştırmıştır. Uzman bulma probleminde önce sorgu ile alakalı dokümanları tarayıp sonra da yazarları ile sorguyu ilişkilendiren sorgu bağımlı modelin, aday profili üzerinden giderek uzmanlık tespit eden sorgu bağımsız modele göre daha iyi performans ortaya koyduğunu ispatlamıştır. Bu iki modeli de ayrıntılı olarak bir sonraki bölümde ele alacağız (3.2).

Cao [15] iki aşamalı bir yöntem izlemiş relevance (uygunluk) ve co-occurrence (eş-oluşum) modellerini birleştirmiştir. Relevance (uygunluk) modeli verilen dokümanın sorgu ile ilişkili olup olmadığını ortaya koymaktayken, co-occurrence (eş-oluşum) modeli verilen sorgunun kişi ile ilişkili olup olmadığını doküman üzerinden ortaya koyar.

Macdonald ve Ounis [16] yeni bir bakış açısıyla voting (oylama) modelini ortaya çıkarmışlardır.

Petkova ve Croft [14] hiyerarşik bir yöntem izlemişler ve Sorgu Bağımlı ve Bağımsız yöntemleri kombinasyon şeklinde kullanmışlardır.

Uzman bulma konusunda bu ve buna benzer birçok araştırma yapılmasına rağmen akademik dünya için özelleşmiş bir uzman bulma çalışması çok yapılmamıştır [5] . Ayrıca bu çalışmaların hepsinde internet veya intranetten toplanan veriler kullanılmış fakat Semantik Web'den faydalanılmamıştır [12]

Milan Stankovic ve Jelena Jovanovic Linked Data verisini uzman bulma problemi için kullanmış ve farklı uzmanlık alanları için farklı kütüphaneleri sorgulamak gerektiğini düşünmüşler ve uygulamışlardır [20] .

Daniel M. Herzig and Hristina Taneva ise Wikipedia'yı veri kaynağı olarak kullanmış ve buradaki semantik verilerden hareketle uzman profillerini çıkarmaya ve uzmanlık alanlarını bulmaya çalışmışlardır [21] .

2.1. Karşılaşılan Problemler

2.1.1. Profil Çıkarma ve İsim Benzerliği Problemi (Profile Extraction and Name Disambiguation)

Uzman bulma probleminde önemli aşamalardan bir tanesi ilgili akademisyenlerin veri setinin ortaya çıkarılabilmesidir. Bu veri setini elde etme aşamasında veri kaynağı olarak DBLP kullanılmıştır. DBLP verisinin içerisindeki bütün akademisyenlerin bilgisine ulaşabilmek için Java da yazılmış bir kütüphane olan SwetoDBLP [10] parseri kullanılmıştır. Akademisyen veri seti oluşturmak için web'i tarayan (crawl) eden araştırmalara göre bir aşama önde başlanmıştır. Deney sonuçları DBLP datasının profil çıkarma ve isim benzerliği problemini minimize ettiği ve uzman bulma algoritmalarını güçlendirdiğini ortaya koymuştur [4] . Diğer taraftan DBLP verisinde bile isim ile alakalı problemler gözlemlenebilmektedir. Örnelemek gerekirse Nihan Kesim Çiçekli ve Nihan K. Çiçekli aynı kişi olmasına rağmen DBLP verisinde Nihan Kesim Çiçekliye ait 14, Nihan K. Çiçekli'ye ait 2 farklı makale yer almaktadır. Nihan K. Çiçekli ve Nihan Kesim Çiçekli aynı kişiler olmasına rağmen farklı kişi gibi yazar listesinde yer almıştır. En iyi uzman

bulma (top -k)problemimizin daha verimli çalışabilmesi için Nihan Kesim Çiçekli ve Nihan K. Çiçeklinin aynı kişiler olduğunu geliştirdiğimiz sistemin anlaması gerekmektedir.

Bu problemin çözümü için birçok yöntem takip edilebilir. Kendi sistemimizde bunu yaparken takip etmiş olduğumuz metot, yazarlar için unique identifier olan web adresi veya çalıştığı iş yeri verisinden yararlanmaktır. Özellikle birden çok ismi olan yazarlar için aynı kişi olup olmadığı bu veriler kontrol edilerek sağlanmıştır. Deney sonuçlarımız bu yaklaşımın en iyi (top-k) uzmanı bulma probleminin çözümünü iletmiş olduğunu göstermektedir.

3. İLGİLİ ÇALIŞMALAR

Uzman bulma problemi ile alakalı ikinci bölümde de bahsettiğimiz üzere birçok yöntem izlenmiş ve birçok yaklaşım benimsenerek uygulanmıştır. İncelediğimiz makalelerde uygulanan yöntemler aşağıda özet olarak açıklanmıştır.

3.1. İzlenen Yöntemler ve Modeller

3.1.1. Olasılıksal (Probabilistic) Model

Dengin makalesinde [5] değindiği üzere belirli bir sorguya göre uzman bulma problemini olasılıksal (probabilistic) modele göre formülize etmek gerekirse:

$p(ca, q)$ = Akademisyen ca 'nın verilen q sorgusuna göre uzman olma olasılığı [8] .

Buradaki q herhangi bir terimi veya kavramı içerebilir. Örneğin “semantik web” sorgusu semantik web konusunda çalışma yapmış uzman akademisyenleri aramak için yapılmış bir sorgudur.

3.1.2. Basit Dil Modeli (Basic Language Model)

Dengin makalesinde [5] değindiği şekilde basit dil modelini (basic language model) formülleştirme işlemini inceleyebiliriz:

“Bayes Teoremi” kullanılarak bu sorgu şu şekilde formülize edilebilir.

$$p(ca|q) = \frac{p(ca, q)}{p(q)} \quad (1)$$

$p(ca, q)$ Akademisyen ve sorgu için ortak (joint) olasılıktır.

$p(q)$ Sorgu için olasılıktır. $p(q)$ Sabit bir değer olduğu için sıralama yaparken hesaplama dışı bırakılabilir. Bu şekilde $p(ca|q)$ 'nın hesaplanmasının $p(ca, q)$ ortak olasılığının (joint probability) hesaplanması ile doğru orantılı olduğu görülebilir.

$p(ca, q)$ Şu şekilde hesaplanabilir.

$$\begin{aligned} p(ca, q) &= \sum_{d \in D} p(d) p(ca, q|d) \\ &= \sum_{d \in D} p(d) p(q, d) p(ca|d, q) \end{aligned} \quad (2)$$

$p(d)$ Dokümanın prior (öncelik) olasılığıdır. Eğer bir akademisyen bir dokümanın yazarı ise dokümana ait konularla ilgili bilgi sahibidir. Buradaki destekleyici dokümanlar D , sorgu q ve akademisyen ca arasındaki bir köprü vazifesi görmektedir. Bu model ile uzman bulma işlemi şu şekilde yürümektedir:

Verilen sorgu q 'ya uygun olan ve puanlama yapılarak sıralanmış dokümanlar, sorgu q ile ilgi derecesine göre incelendikten sonra, dokümanlarla ilgili akademisyenler tespit edilir. Buradaki doküman inceleme işleminin ne kadar fazla olduğu ve aşırı bir süreç içerdiği ortadadır.

Verilen bir dokümana göre, verilen sorgu q 'nun olasılığını hesaplamak için doküman dil modelini (document language model) θd 'yi kullanabiliriz.

$$p(q|\theta d) = \prod_{t \in q} p(t|\theta d)^{n(t,q)} \quad (3)$$

Buradaki $p(t|\theta d)$, t teriminin d dokümanında maksimum olabilirlik tahminidir ve $n(t|q)$ ise t teriminin q sorgusu içinde kaç defa bulunduğudur

Bir q sorgusunun belli sayıdaki t terimi için d dokümanında bulunma olasılığı Jelinek-Mercer smoothing'e [11] göre şu şekildedir.

$$p(t|\theta d) = (1 - \lambda) p(t|d) + \lambda p(t) . \quad (4)$$

Balog 'a göre $\lambda = 0,5$ olarak değerlendirilmiştir.

Deng [5] bir varsayım yaparak akademisyen ca 'nın doküman ile ilgili q sorgusundan bağımsız olduğunu varsaymıştır. Bu varsayıma göre;

$$p(ca, q|d) = p(ca, d) \text{ şeklinde bir eşitlik çıkabilir.} \quad (5)$$

Eğer bir akademisyen bir dokümanın yazarı ise dokümana ait konularla ilgili bilgi sahibidir. Birden çok yazara sahip dokümanlar için bir akademisyenin tek başına sahip olduğu konu başlığı hakkındaki bilgi sahibi olma olasılığı $p(ca, d)$ ' dan düşük olmalıdır [5]. Deng'e göre eğer bir doküman birden çok yazara sahipse, bir yazarın doküman ile ilgili konu üzerinde bilgi sahibi olma olasılığı şu şekilde hesaplanmaktadır.

$$p(ca, d) = \begin{cases} \frac{1}{nd}, & ca \text{ } d \text{ dokümanının yazarıdır} \\ 0, & \text{diğer koşulda} \end{cases} \quad (6)$$

Burada nd o dokümana ait yazar sayısıdır ve $p(ca, d)$ doküman ile yazarın ilişki derecesini ortaya koymaktadır [5].

Doküman önceliği $p(d)$ burada uniform olarak değerlendirildiğinden dolayı sonucu etkilemeyecektir. Basit Dil Modeli için en son hesaplama 2 numaralı eşitlikte $p(q|d)$ yerine $p(q|\theta d)$ konularak yapılmaktadır [5].

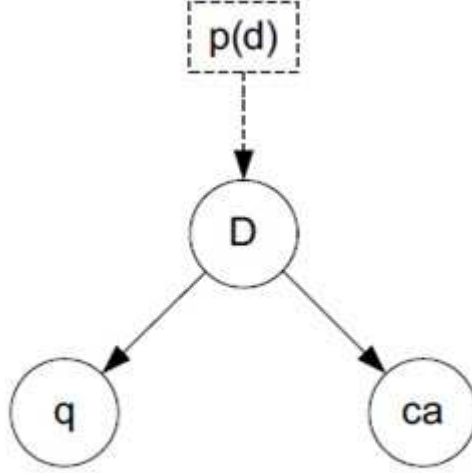
$$p(ca, q|d) = \sum_{d \in D} \{ \prod_{t \in q} (p(t|\theta d))^{n(t,q)} \} p(ca|d) \quad (7)$$

Uzman bulma probleminde bu model basit dil modeli olarak isimlendirilmektedir [5].

3.1.3. Ağırlıklı Dil Modeli (Weighted Language Model)

Basit dil modeli sorgu ve doküman arasındaki ilişkiyi bulabilmesine rağmen, dokümanların önceliği ve önemi hakkında herhangi bir değerlendirme yapmamaktadır. Şekil1'de görüleceği üzere Ağırlıklı Dil Modelinde [5], Basit Dil Modelinden farklı olarak

dokümanın önceliğini belirten $p(d)$ katsayısı uniform olarak değerlendirilmemiş ve hesaplamaya ait bir parametre olarak kullanılmıştır. Eğer bir yazar diğer bir yazara göre daha çok öneme sahip bir makale yazdıysa, bu onun sıralamada öne geçmesini sağlayacaktır. Bu model bizim en iyi uzmanı bulma algoritmamızda da kullanılmıştır. Daha ayrıntılı bilgi için “4.4.3” başlığı incelenebilir.



(a) Weighted model

Şekil 3-1 Ağırlıklı Dil Modeli

3.2. Sorgu Bağımlı ve Sorgu Bağımsız Modeller (Query Dependent and Query Independent Approaches)

Sorgu Bağımlı ve Sorgu Bağımsız Modeller uzman bulma problemi için iki önemli yaklaşımdır. İki yaklaşım içinde geçerli olan durum, sistem uzman bulurken ilgili dokümanları incelemekte, değerlendirmekte ve ortaya çıkardığı sonuçlarla o dokümanları hazırlayan kişilerin konu üzerindeki uzmanlık olasılığını ortaya çıkarmaktadır.

Sorgu bağımsız (query independent) yöntemi doğrudan aday ile ilgili tüm belgelere dayalı bir aday profili (sanal bir "belge") modeller ve uzman bulma sorgusuna yanıtı bu profile göre sıralama puanı oluşturarak geri döner. Öte yandan sorgu bağımlı model (query

dependent) korpus dâhilinde bulunan dokümanları sıralayıp derecelendirdikten sonra ilgili konularla alakalı dokümanların yazarlarına dokuman alt kümesini kullanarak ulaşmaktadır [5].

Balog yazdığı makalede [8] iki dil tabanlı modeli araştırmalarında kullanmış ve ikisinin performanslarını karşılaştırmıştır. Birinci modellerinde sorguyu uzmanların yazdıkları makaleler için kullanmış, ikinci modellerinde ise ilgili sorguya göre dokümanları bulmuş daha sonra da dokümanların yazarlarından hareketle ilgili uzmanları bulmaya çalışmıştır. Deney sonuçları da ikinci modellerinin birinci modele göre çok daha iyi performans ortaya çıkardığını kanıtlamıştır. Sonuçlara göre sorgu bağımsız (query independent) modelin, sorgu bağımlı (query dependent) modele göre çok daha iyi performans verdiğini ortaya koymuştur.

4. SEMANTİK WEB YAKLAŞIMI ile UZMAN BULMA

Uzman bulma konusundaki bugüne kadar yapılan çalışmalarda Semantik Web verilerinde çok yararlanılmamıştır. Biz tezimizde Semantik Web verilerini kullanarak uzman bulma problemine çözüm bulunabileceğini göstermekteyiz.

4.1. Neden Semantik Yaklaşım?

Bugüne kadar uzman bulma problemi ile ilgili farklı modeller geliştirilmiş ve yapılan bütün çalışmalarda internetten toplanan veriler kullanılmıştır [12]. Semantik Web verisinin kullanılmasıyla daha güncel verilere, büyük veri kümeleri ile uğraşmak zorunda kalınmadan, daha etkin bir şekilde ulaşılabileceği düşünüldüğü için Semantik Web yaklaşımı kullanılmıştır.

4.2. Semantik Web ve İlgili Teknolojiler

Semantik Web internet ortamında bilgiyi sadece insanların anlayabileceği doğal dillerle değil aynı zamanda ilgili yazılımlar tarafından anlaşılabilir, erişilebilir ve sorgulanabilir şekilde ifade edildiği ve ulaşılmak istenen bilginin daha kolay ve zahmetsiz kullanılabilirdiği internet tabanlı bir teknolojidir.

İnternetin kurucusu olarak adlandırılan Tim-Berners Lee Semantik web'i şu şekilde tanımlamaktadır:

"Web için bir hayalim var, öyle ki bilgisayarlar web üzerindeki bütün veriyi, içerikler, linkler ve insanlarla bilgisayarlar arasındaki bütün işlemler gibi, analiz etmeye muktedir olacaklar. Henüz ortaya çıkmamış olsa da, ortaya çıktığı zaman anlamsal ağ ticaretin günlük mekanizmaları, bürokrasi ve günlük yaşamlarımız birbiri ile konuşan makineler tarafından yürütülecek. İnsanlığın asırlardır konuşup durduğu "akıllı ajanlar" nihayet gerçekleşecek." [22]

4.3. Linked Data, DBLP ve Diğer Veri Web Kaynakları

4.3.1. DBLP

DBLP1 Dijital Kaynakça ve Kütüphane Projesinin (Digital Bibliography & Library Project) kısaltılmış adıdır. Öncelikle Bilgisayar Bilimleri için oluşturulan DBLP Kütüphanesi sonradan diğer disiplinleri de içine alacak şekilde genişlemiştir. DBLP şuanda bir kaynakça (bibliography) sunucusu olarak hizmet vermektedir. 900.000'nin üzerinde bibliyografik verinin bulunduğu DBLP verisi akademisyenlerin profilleri, makaleleri, tezleri, kitapları ve dergileri hakkında bilgileri bünyesinde barındırmaktadır.

DBLP (1 Temmuz 2011 verisi) ile ilgili SwetoDBLP [10] ayrıştırıcısı kullanarak aşağıdaki istatistikleri elde ettik:

Yayın Türü	Sayı
Makale (Articles)	673789
Kitap (Books)	7915
Dergi (InCollection)	16624
Konferans (InProceedings)	944245
Doktora Tezi (PhdThesis)	813
Yayıncı (Publishers)	22833
Bağlantılar (WWW)	954460

Tablo 4-1 DBLP İstatistikleri

DBLP bünyesinde bulunan verilere topluca indirerek veya çevrimiçi olarak sorgularla (SPARQL) erişebilmek mümkündür.

4.3.2. Linked Data

Linked Data² veriyi insanların anlayabileceği metinlerden ziyade bilgisayarlarında otomatik olarak anlayabileceği formatta paylaşır. Linked Data sayesinde veriler birbirleriyle ilişkili olarak sunulurken verilerin arasındaki bariyerler ortadan kalkmaktadır. Linked Data üzerindeki verilere anlamsallık kazandırarak anlamsal sorgulara cevap dönebilir.

¹ <http://www.dblp.org/search/>

² Linked Data, <http://linkeddata.org/>

Linked Data verisi içinde yukarıda bahsettiğimiz DBLP setine ek olarak ACM, IEEE ve DBLP++ veri setlerini kullanıyoruz.

DBLP++ 3 veri setinde DBLP’de bulunan 870.000 doküman kapsamakta ve bunların 95.000 tanesinden daha fazlası ve onlara ait 513.000’den fazla anahtar kelime ve özet verisi bulunmaktadır.

ACM veri setinde 15 milyon ve IEEE setinde 120 bin civarında triple bulunmaktadır.

ACM ve IEEE için sorgu ara yüzü (end point) olarak <http://acm.rkbexplorer.com/sparql/> ve <http://ieee.rkbexplorer.com/sparql/> adresleri kullanılmaktadır.

DBLP++ verisi için ise <http://dblp.l3s.de/d2r/sparql/> sorgu ara yüzü kullanılmıştır.

4.4. Semantik Web-tabanlı Uzman Bulma Yöntemleri

4.4.1. Başlık Tabanlı Arama Modeli

Yazar bilgilerinin erişim için DBLP datasını kullandığımız araştırma da yazarın ismini, web sayfasını ve çalıştığı yer bilgisini içeren profil datası SwetoDBLP [10] çerçevesi kullanılarak çıkarılmıştır. <http://dblp.rkbexplorer.com/sparql/> adresinde koştığımız SPARQL sorguları ile de yazarların yazdığı bütün makale bilgilerine erişilmiştir.

Erdoğan Doğdu’ya ait makaleleri output olarak veren SPARQL sorgusu bu şekildedir:

```
SELECT DISTINCT ?title
WHERE {
    ?x akt:has-title ?title
    ?x akt:has-author ?me.
    ?me akt:full-name "Erdogan Dogdu".
}
```

Bu sorgu Erdoğan Doğdu’ya ait 24 farklı makale dönmüştür. Java altyapısı kullanılarak otomatik olarak programımızın içinden çekilen bu sorgular için Jena Framework4 kullanılmıştır.

³DBLP++, <http://dblp.l3s.de/dblp++.php>

Sorgumuz için aldığımız cevap Tablo 4-2’de verilmiştir.

1	Scheduling Adaptive Transactions in Real-Time Databases.
2	A Java API for global querying and updates for a system of databases.
3	A Small Listener for Heterogeneous Mobile Devices: A Service Enabler with a Uniform Web Object View.
4	An Agent Module for a System on Mobile Devices.
5	A Framework for Constraint-Based Collaborative Web Service Applications and A Travel Application Case Study.
6	SyD: A Middleware Testbed for Collaborative Applications over Small Heterogeneous Devices and Data Stores.
7	Enforcing Interdependencies and Executing Transactions Atomically over Autonomous Mobile Data Stores Using SyD Link Technology.
8	Implementation of a Calendar Application Based on SyD Coordination Links.
9	Efficient Scheduling Strategies for Web Services-Based E-Business Transactions.
10	Scheduling Web Services Transactions.
11	Enhancements to Securing XML Web Services.
12	Health Level-7 compliant clinical patient records system.
13	Real-Time Transactions with Execution Histories: Priority Assignment and Load Control.
14	Scheduling Nested Transactions in Real-Time Database System.
15	User Control and Dynamic Reorganization of Multimedia Presentations During Payout.
16	Design and Implementation of a Middleware for Mobile Networked Devices and Applications.
17	A User Management System for Federated Databases Using Web Services.
18	A Generic Database Web Service.
19	Utilization of execution histories in scheduling real-time database transactions.
20	An Extended Web Services Framework.
21	Error-Constraint COUNT Query Evaluation in Relational Databases.
22	Semantic web in eHealth.
23	Automatic Detection of Erythematous-Squamous Diseases Using k-Means Clustering.
24	Dynamic Selection of Web Services Using QoS Constraints in BPEL.

Tablo 4-2 DBLP verisinde Erdoğan Dođdu’ya ait olan makaleler

⁴ Jena Framework, <http://jena.sourceforge.net/>

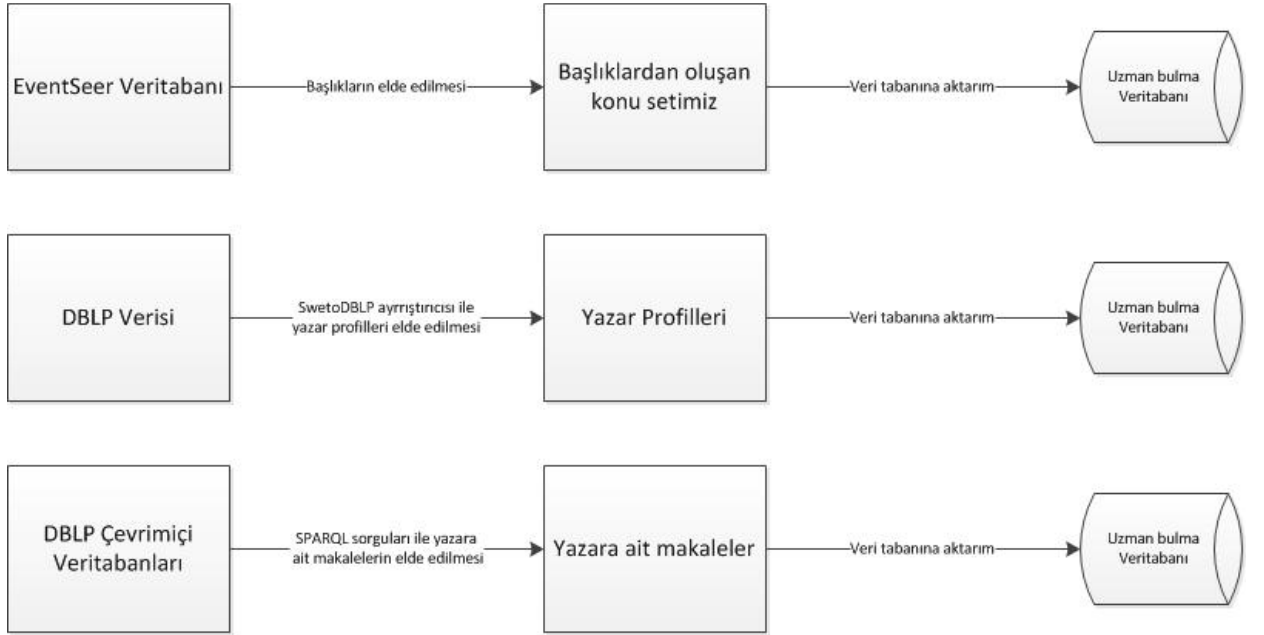
DBLP verisinin makalelere ait olan anahtar kelime ve özet verilerine sahip olmamasından dolayı yazarlara ait ilgi alanlarının bulunması için ilgili makalelerin başlıklarında Eventseer'den [17] aldığımız 5,282 konudan oluşan konu setimizi kullanarak tarama işlemi yapıyoruz:

Başlık Tabanlı İlgi Alanı Bulma Algoritması:

Girdi: Eventseer'den alınmış tanımlı konu setimiz, Y yazar listesi

Çıktı: Yazarlar ve ilgi alanları listesi.

- 1: Her bir $y \in Y$ için: DBLP veri setinde yazarın bulunması ($y_i \in \text{DBLP yazar seti}$)
- 2: Yazarlara ait makalelerin DBLP çevrimiçi (online) data setine SPARQL sorguları çekilerek elde edilmesi
- 3: Bütün tanımlı konu setimizin makale başlıklarında aranması
- 4: Konu makale başlığında bulunduysa akademisyenin ilgi alanı olarak o yazarın ilgi alanına eklenmesi.



Şekil 4-1 Başlık tabanlı model Veri Transferi Şeması

Erdoğan Dođdu için bulunan başlıklar:

Area of Interest	Title
databases	A Java API for global querying and updates for a system of databases.
database	A Java API for global querying and updates for a system of databases
database	Utilization of execution histories in scheduling real-time database transactions.
real-time data	Utilization of execution histories in scheduling real-time database transactions.
real-time database	Utilization of execution histories in scheduling real-time database transactions.

Tablo 4-3 Başlık Tabanlı Modele göre Erdoğan Dođdu'ya ait ilgi alanları

Veri setimizde yer alan Erdoğan Dođdu'ya ait bilinen ilgi alanları aşağıdaki şekildedir:

1	Database systems	6	Middleware
2	Transaction processing	7	Web technologies
3	Web-based databases	8	Mobile and context-aware computing
4	Semantic web	9	Software engineering
5	Web services	10	Object-oriented design and programming

Tablo 4-4 Veri setimizde Erdoğan Dođdu'ya ait ilgi alanları

Erdoğan Dođdu'nun kendi ilgi alanlarını incelediğimizde “Database systems, transaction processing, web-based databases” bizim algoritmamızın bulduğu sonuçlarla eşleşen ilgi alanlarıdır.

Basit bir hesaplamayla Erdoğan Dođdu'nun ilgi alanlarının başlık tabanlı modelimiz tarafından karşılanma oranının %30 olduğunu görmekteyiz. Bu oranlar Performans Bölümünde (6) daha detaylı olarak deney setimiz dâhilinde incelenecektir.

Başlık tabanlı modelimizin verimini artırmak için konu setimiz kendi içinde sınıflandırılabilir. Bunun pozitif etkisi olacağı muhakkaktır. Yalnız bu çalışmayı gelecek dönemde yapılabilecek çalışmalar içerisinde ele almaktayız. Semantik Web verisi kullanarak bu çalışmanın temelleri benzer ilgi alanı bulma algoritması “4.4.3” konu başlığı altında kısmen uygulanmıştır.

111 akademisyen ve ilgi alanlarından oluşan deney setimizle yaptığımız çalışma istatistik sonuçları aşağıdadır:

Toplam Yazar Sayısı	DBLP deki toplam makale sayısı
111	1221

Tablo 4-5 Yazarlarımıza ait toplam makale sayısı

Deney setimizdeki 111 yazara ait olan 1221 makaledeki bütün başlıklar modelimiz için kullanılmıştır.

Toplam Yazar Sayısı	DBLP’de makalesi bulunan yazar sayısı	DBLP’de makalesi bulunmayan yazar sayısı
111	64	47

Tablo 4-6 DBLP’de makalesi bulunan yazar sayısı

Deney setimizdeki 111 akademisyenin 64’nün DBLP veri setinde en az bir tane makalesi bulunmaktadır. Bu oran deney setimizdeki akademisyenlerin %70’ine denk gelmektedir.

Akademisyen başına düşen ortalama makale sayısı	11
Akademisyen başına düşen ilgi alanı	1.96
Makale başına düşen ortalama ilgi alanı	0.2

Tablo 4-7 DBLP verisi için ortalama ilgi alanı değerleri

DBLP’de makalesi bulunan toplam yazar sayısı	DBLP’de makalesi ve ilgi alanı bulunan yazar sayısı	DBLP’de makalesi bulunan ve ilgi alanı bulunmayan yazar sayısı
64	51	13

Tablo 4-8 Veri setimiz için DBLP ilgi alanı istatistiği

Deney setimizde bulunan akademisyenlerden DBLP verisinde makalesi bulunan 64 akademisyen için yapılan başlık taramasında 13 akademisyenimizin makaleleri ile ilgili Eventseer’den [17] aldığımız konu setimizle ilişkili ilgi alanı bulunamamıştır. Bu %20

oranında akademisyen için DBLP setinde makalesi bulunduğu halde ilgi alanı bulunamadığı sonucunu doğurmuştur.

Daha ayrıntılı analiz Performans (6) bölümünde yapılacaktır.

4.4.2. Linked Data Tabanlı Semantik Web Modeli

DBLP yazarlar ve yazarlarla ilgili makaleler için çok iyi bir veri kaynağı olmasına rağmen, makalelerin hangi konularla ilgili olduğunu ortaya çıkaran anahtar sözcük ve özeti içermemektedir.

Bu darboğazı ortadan kaldırmak için ACM ve IEEE veri tabanlarına semantik sorgular göndererek hangi makalenin hangi konu başlığı ile alakalı olduğunu yöntemini bulmaya çalıştık. İki veritabanının da ortak olarak kullandığı aktör ontolojisinde [23] (actors ontology) [Ek 2. Aktör Ontolojisi, <http://www.actors.org/ontology/portal>] makalenin ilgili olduğu alanı gösteren “**akt: addresses-generic-area-of-interest**” şeklinde bir alan olduğunu tespit ettik. Uzman bulma problemi ve sosyal network analizi ile alakalı birçok makaleyi taradığımız halde bu tarz bir yaklaşımın daha önce kullanılmadığını gördük. Bize Uzman Bulma problemimizde oldukça yararlı olacak makalenin hangi konu başlığı ile alakalı olduğunu gösteren bu alanı bu modelimiz için kullandık.

DBLP verisini ayrıştırarak elde ettiğimiz her bir makalenin konu başlığını ACM ve IEEE kütüphanelerinden aşağıdaki SPARQL sorgusu sayesinde elde ettik:

```
SELECT ?title ?areofinterest
WHERE {
  ?x akt:addresses-generic-area-of-interest ?areofinterest
  ?x akt:has-title ?title
  ?x akt:has-author ?me.
  ?me akt:full-name "Erdogan Dogdu".
}
```

Bu yapıyı otomatize edebilmek içinde yine aynı şekilde Jena Frameworkunu kullandık.

Modelimizin algoritması da aşağıdaki şekildedir:

Linked Data Tabanlı Semantik Web Algoritması:

Girdi: DBLP den alınmış yazar bilgileri ve yazara ait makale başlıkları

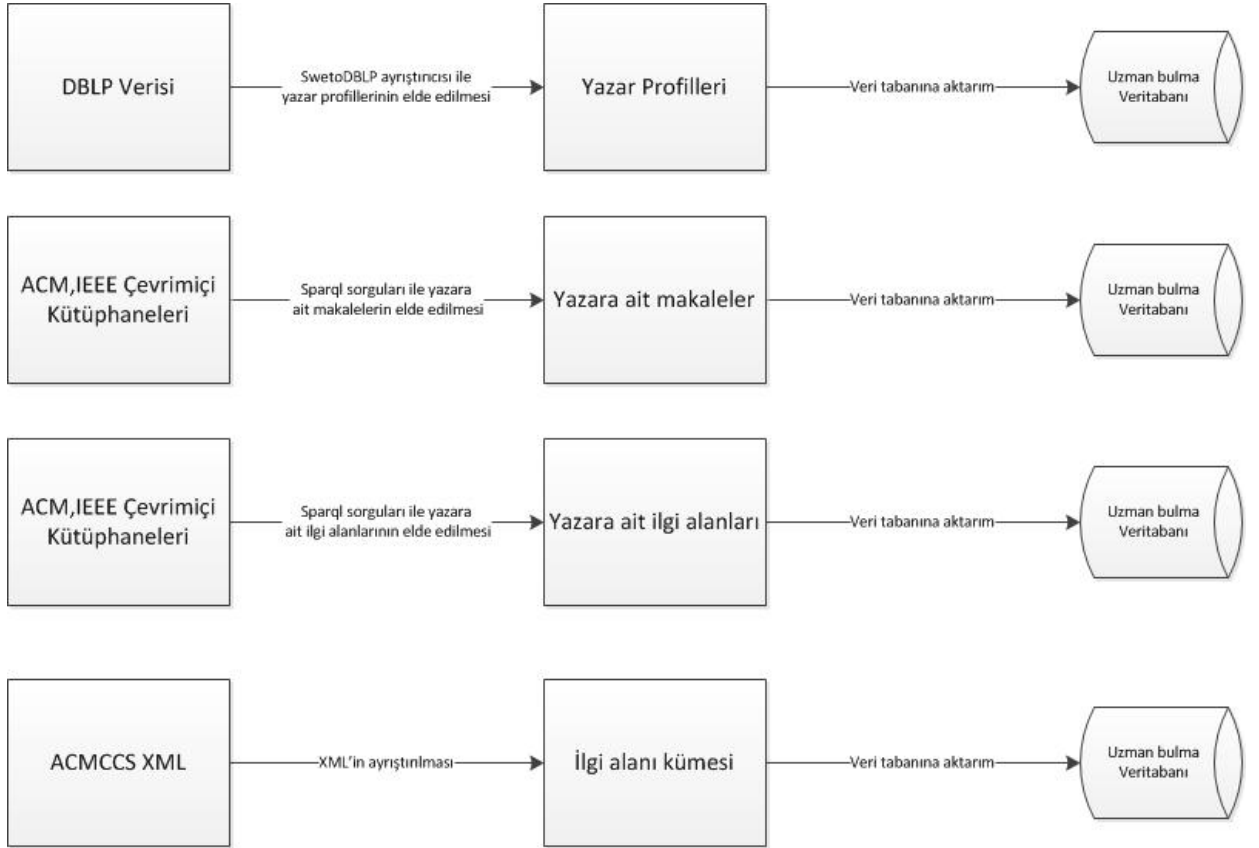
Çıktı: Yazara ait ilgi alanları listesi

1: Bütün a yazarları için

2: a yazarına ait olan makaleleri ACM ve IEEE database'inde bul ve S kümesine ekle

3: S kümesi içinde bulunan bütün p makaleleri için

4: İlgi alanı kümesini p için bul ve a'nın ilgi alanı kümesine ekle.



Şekil 4-2 ACM, IEEE ile Linked Data tabanlı Semantik Web Algoritması Veri Transferi Şeması

Algoritmanın “Erdogan Dogdu” sorgusu için bize verdiği sonuçlar aşağıda yer almaktadır:

Title	Area of Interest
Error-constrained COUNT query evaluation in relational databases	Logical Design, Query processing, Data models, SPECIAL-PURPOSE AND APPLICATION-BASED SYSTEMS
Enforcing Interdependencies and Executing Transactions Atomically over Autonomous Mobile Data Stores Using SyD Link Technology	Distributed Systems, Other Architecture Styles, Office Automation
Implementation of a Calendar Application Based on SyD Coordination Links	Distributed Systems, Other Architecture Styles, User Interfaces, Distributed Artificial Intelligence

Tablo 4-9 Linked Data tabanlı Semantik Web modelimize göre yazarın ilgi alanları

Aktör ontolojisinde sorgumuza dönen cevaplar ACM Computing Classification isimli xml dosyasında yer alan konu bilgilerinin ID’leri şeklindedir. Dönen bu ID’leri konulara çevirebilmek için ACMCCS isimli xml dosyasını ayrıştırdık. Bu xml dosyası konuları sınıflandırmak için University of Minho tarafından oluşturulmuş ve ACM tarafından kullanılmaktadır.

Bu xml incelendiği zaman konu başlıkları arasında “isRelated” şeklinde bir alanın olduğu görülmüştür. Algoritmamıza bir parametre olarak bu da girmiştir. Örnek vererek açıklamak gerekirse H.5.3 kodlu “User Interfaces” konu başlığı D.2.2 kodlu “Design Tools and Techniques”, H.1.2 kodlu “User/Machine Systems” ve I.3.6 kodlu “Methodology ve Techniques” konu başlıkları ile ilgilidir. “User Interfaces” konu başlığını ilgi alanı olarak içeren bir makale aynı zamanda ikincil olarak “Design Tools and Techniques”, “User/Machine Systems” ve “Methodology ve Techniques” konuları ile de ilgilidir.

Aldığımız sonuçları karşılaştırdığımızda Tablo 4-4 Veri setimizde Erdoğan Doğdu’ya ait ilgi alanları’de yer alan “Erdoğan Doğdu’ya” ait gerçek ilgi alanlarının karşılama oranının Linked Data tabanlı Semantik web modelimizle daha çok olduğunu görmekteyiz.

Modelimizin verdiği sonuçlara göre gerçek ilgi alanlarını karşılama oranı %40'a çıkmıştır. “Database systems”, “transaction processing”, “web-based databases” ve “Software engineering” bizim bulduğumuz ilgi alanlarına karşılık gelmektedir.

Detaylı analiz için Performans (6) bölüm incelenebilir.

Bu metot bize çevrimiçi veritabanlarının ve SPARQL sorgularının bir yazarın ilgi alanını bulmak için etkili bir şekilde kullanılabilir olduğunu göstermiştir. Dokümanları parse etmek, kendi veri setini oluşturmak ve tarama (crawl) algoritmalarını kullanmadan daha güncel bir veri tabanına ulaşım imkânı sağlamaktadır.

Bu metodun başarısı yazarın yazdığı makalelerin ACM ve IEEE'deki sayılarının toplam makalelerine oranıyla doğru orantılıdır. Maalesef DBLP veri setinde bulunan bütün makaleler ACM ve IEEE veri tabanlarında bulunmamaktadır.

Kullandığımız setimiz için ACM ve IEEE'de bulunan makale sayılarımız ve oranlarımız şu şekildedir:

Toplam Yazar Sayısı	DBLP deki toplam makale sayısı	ACM ve IEEE'de bulunan makale sayısı
111	1221	221

Tablo 4-10 Yazarlarımızın ACM ve IEEE'de bulunan makale sayısı

Tablo 4-10'dan göreceğimiz üzere ACM ve IEEE deki makalelerin DBLP'deki makale sayısına oranı %20 civarındadır.

Toplam Yazar Sayısı	ACM ve IEEE'de makalesi bulunan yazar sayısı	ACM ve IEEE'de makalesi bulunmayan yazar sayısı
111	47	64

Tablo 4-11 ACM ve IEEE'de makalesi bulunan yazar sayısı

Tablo 4-11'den göreceğimiz üzere %42 oranında setimizdeki yazarların en az 1 tane ACM, IEEE setinde makalesi bulunmaktadır.

Akademisyen başına düşen ortalama makale sayısı	1.9
Akademisyen başına düşen ilgi alanı	5.27
Makale başına düşen ortalama ilgi alanı	2.8

Tablo 4-12 ACM ve IEEE verisi için ortalama ilgi alanı değerleri

Yaptığımız çalışmadaki akademisyenlerin yazdığı makalelerin anahtar kelime ve özetlerine ulaşmadaki sıkıntı modelimizin performansını etkileyebileceğini düşündüğümüzden veri setindeki yazarlarımız için daha çok makale bilgisinin anahtar kelime ve özetine ulaşmamız gerekmektedir.

Burada da farklı bir kütüphanenin yardımına başvuruyorduk. Araştırmalarımız bize DBLP++ projesi veya diğer adıyla DBLP Faceted projesini kullanarak bu sorunu aşabileceğimiz göstermiştir. DBLP++, DBLP ontolojisindeki bu eksikliği görmüş ve makaleler için anahtar kelime değerlerine yer vermiştir. Bu anahtar kelime değerlerini internet verisinden toparlamışlar ve Dublin Core ontolojisi ile semantik-web kullanımına sunmuşlardır. Kullandığımız sorgu ara yüzü (endpoint) ve SPARQL sorguları aşağıdaki şekildedir:

Sorgu ara yüzü : <http://dblp.l3s.de/d2r/sparql/>

```
SELECT ?article ?areaofinterest
WHERE {
  ?article dc:subject ?areaofinterest
  ?article foaf:maker ?author
  ?author foaf:name "Erdogan Dogdu".
  OPTIONAL { ?article dc:subject ?areaofinterest }
}
```

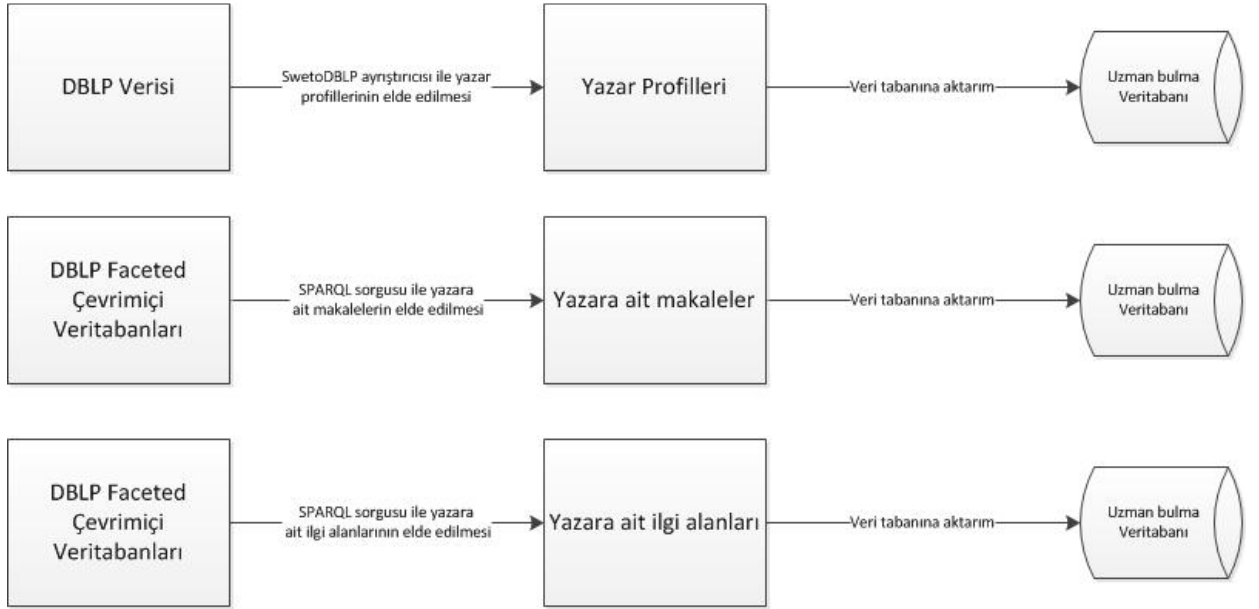
Title	Area of Interest
A Java API for global querying and updates for a system of databases.	distributed information retrieval, global queries/updates, multidatabases, system of databases
Semantic web in eHealth	eHealth, medical information systems, semantic web
An Agent Module for a System on Mobile Devices.	Agent Middleware for Mobile Devices, Agent Based Execution Engine, Mobile Agents, System on Mobile Devices Middleware
Implementation of a Calendar Application Based on SyD Coordination Links	Middleware for mobile devices, Coordination links, Peer-to-peer distributed applications, Calendar of Meetings
SyD: A Middleware Testbed for Collaborative Applications over Small Heterogeneous Devices and Data Stores.	Mobile Servers, SyD Coordination Bonds, Object and Web Service Coordination, Atomic Transactions, Application-Level QoS
Health Level-7 compliant clinical patient records system.	HL7, XML, clinical document architecture, medical informatics, schema, universal patient records, web-based application

Tablo 4-13 Linked Data tabanlı Semantik Web modelimize (DBLP++ ile) göre yazarın ilgi alanları

Tablo 4-13'te yukarıdaki SPARQL sorgusunun sonuçlarını göstermektedir. Sorgu Erdoğan Doğdu'nun makalelerinin (varsa) ilgi alanlarını listelemektedir. Anlaşılabileceği üzere Dublin Core ontolojisinin kullanılması ve DBLP Faceted bünyesinde bulunan verilerin sonuçlarımıza dâhil edilmesi Erdoğan Doğdu'nun yazmış olduğu makalelerin daha fazlasının ilgi alanlarına ulaşmamızı sağlamıştır. Bu ilgi alanlarının artması Erdoğan Doğdu'nun gerçek ilgi alanları ile kapsama alanımızı ve dolayısıyla modelimizin verimliliğini artırmıştır.

Tablo 4-4’de yer alan gerçek ilgi alanlarının DBLP Faceted kütüphanesinde modelimiz tarafından bulunan ilgi alanlarının karşılanma oranı şimdi %70’in üzerindedir. Web-based databases, middleware, web technologies, web services, database systems, semantic web, mobile computing ve transaction processing konuları DBLP Faceted verisinin içerisinde çıkarılabilen ve Erdoğan Doğdu’ya ait alanları karşılayan ilgi alanlarıdır. Modelimizde iki veri kaynağı ACM, IEEE ve DBLP Faceted kütüphanelerinden gelen verileri dâhil ettiğimizde karşılanma oranının Erdoğan Doğdu için %80’in üzerine çıkmış olduğunu görüyoruz. Elde edilen bir kısım örnek sonuçlar Ek 1. Veri seti için elde edilen sonuçlar’dan görüntülenebilir.

Bu da modelimizin Erdoğan Doğdu için çalıştığında yeterli bir sonuç verdiğini ortaya koyuyor. Daha ayrıntılı değerlendirme performans analizi (6. Bölüm) kısmında yapılacaktır.



Şekil 4-3 Faceted DBLP ile Linked Data tabanlı Semantik Web Yaklaşımı Veri Transferi Şeması

DBLP Faceted ile geliştirilmiş Linked Data Tabanlı Semantik Web Algoritması:

Girdi: Yazar listesi

Çıktı: Yazarlara ait ilgi alanları listesi

- 1: Bütün a yazarları için
- 2: a yazarına ait olan makaleleri ACM, IEEE ve Faceted DBLP database'inde bul ve S kümesine ekle
- 3: S kümesi içinde bulunan bütün p makaleleri için:
- 4: İlgi alanı kümesini p için bul ve I kümesine ekle
- 5: I kümesi içinde bulunan bütün p ilgi alanları için:
- 6: p yi yazarın ilgi alanı listesine ekle

Toplam Yazar Sayısı	DBLP deki toplam makale sayısı	DBLP Faceted de bulunan makale sayısı
111	1221	339

Tablo 4-14 Yazarlarımızın DBLP++'da bulunan makale sayısı

Tablo 4-14'den göreceğimiz üzere veri setimiz için DBLP++'daki makalelerin DBLP'deki makale sayısına oranı yani birbirini karşılama oranı %30 civarındadır.

Toplam Yazar Sayısı	DBLP Faceted'de makalesi bulunan yazar sayısı	DBLP Faceted'de makalesi bulunmayan yazar sayısı
111	78	33

Tablo 4-15 DBLP++'da makalesi bulunan yazar sayısı

Tablo 4-15'den göreceğimiz üzere setimizdeki yazarların %70'inin en az 1 tane DBLP++ setinde makalesi bulunmaktadır.

Akademisyen başına düşen ortalama makale sayısı	3.05
Akademisyen başına düşen ilgi alanı	13.7
Makale başına düşen ortalama ilgi alanı	4.5

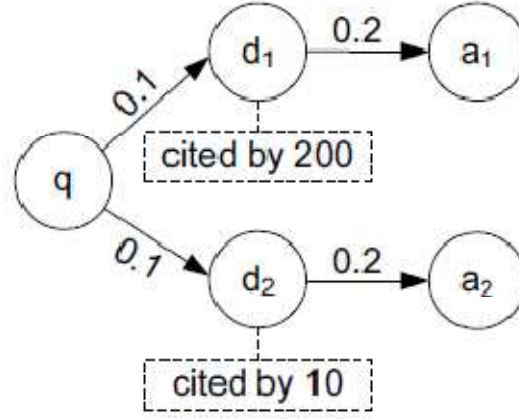
Tablo 4-16 DBLP++ verisi için ortalama ilgi alanı değerleri

Daha ayrıntılı analizler performans analizi (6. Bölüm) bölümünde ele alınacaktır.

4.4.3. Belirli bir Konudaki En iyi Uzmanları Bulma

Daha önceki yaklaşımlarda, TREC Enterprise Track'te [13] uzman bulma problemi information retrieval (bilgi toplama) konu başlığı altında incelendi. Bu yaklaşımlardaki modelde genelde statistical language model (istatistiksel dil modeli) kullanarak uzmanları derecelendirme şeklinde olmaktadır. Temel dil modeli (basic language model) sorgu ile doküman arasındaki benzerliği ölçüp sonradan bu dokümanlardan akademisyenlerin bilgi seviyesinin ve ilgi alanının belirlenmesi şeklindedir. Temel dil modeline göre Şekil 4-4'deki sorgu (query) q için dokümanlar d1 ve d2 ye göre a1 ve a2 akademisyenleri için uzmanlık alanları ve seviyeleri aynı durumdadır.

Diğer taraftan alıntı (citation) değeri basit dil modelinde kullanılmamaktadır. Ancak dokümanların önemlilik derecesine göre sıralanması algoritmanın bulacağı sonuçların daha etkin olmasını sağlayabilir. Daha fazla kişi tarafından alıntılanmış bir doküman daha az kişi tarafından alıntılanmış bir dokümana göre daha önemli bir dokümandır. Alıntılanma sayılarını hesaba kattığımız zaman Şekil 4-4'deki d1 dokümanı, d2 dokümanına göre daha önemli bir dokümandır. Buradan hareketle q sorgusu için a1 akademisyeni a2 akademisyenine göre daha uzmandır. Alıntılanma değerlerini hesaba katarak yapılan bu değerlendirme weighted language model (ağırlıklı dil modeli) olarak isimlendirilmektedir [5]. Yapılan deneylere göre basit dil modeline göre çok daha iyi sonuçlar verdiği gözlemlenmiştir [5]



Şekil 4-4 Dokümanlar ve sorgular [5]

Biz de bu yaklaşımları çalışmalarımızda kullandık ve yayınlanan bir makalenin alıntılanma değerini bulmak için Google Scholar'dan yararlandık. Yazılan web arayıcısı (crawler) sayesinde Şekil 4-5 Google Scholar ile alıntılanma sayısı tespiti'ten görülebileceği gibi makalenin başlığına göre makalenin toplam alıntılanma sayısı web arayıcısı sayesinde elde ettik.



Şekil 4-5 Google Scholar ile alıntılanma sayısı tespiti

Alıntı sayılarını da hesaba katarak hibrit bir model geliştirdik. Ayrıca hibrit modelimizde hem Başlık tabanlı yaklaşımımızı hem de Linked Data tabanlı Semantik Web yaklaşımımızı hesaba kattık. Ayrıca algoritmamızın etkinliğini artırmak için Deng'in makalesinde yer alan Ağırlıklı Dil Modeli ve Benzer ilgi alanı bulma yöntemini kullandık. Benzer ilgi alanı bulma yöntemine Semantik Web olarak bir çözüm bulduk.

Benzer İlgi Alanı Bulma Yöntemi

En iyi uzman bulma probleminde bir makale için olan ilgi alanı eğer sadece kendisiyle ilişkili veya kendisinin alt kümesi olabilecek konuları ilgi alanı olarak içeriyorsa mevcut algoritmamız tarafından değerlendirmeye katılmamaktaydı. Örnek vererek açıklamak gerekirse “Semantik Web” sorgumuz için sadece ilgi alanı olarak “Semantik Web” kelimesi aranmaktaydı. Fakat bu ilgi alanı ile doğrudan ilişkili “Ontoloji, Ontolojiler, Semantik, RDF, OWL, SPARQL ” gibi ilgi alanları aranmamaktaydı. Benzer ve alt kümesi olabilecek konuların yapılan sorguya katılması en iyiyi bulma algoritmamızı iyileştireceği düşünülmektedir.

Çalışmalarımız içerisinde bir örnek vermek gerekirse “SPARQ2L: towards support for subgraph extraction queries in rdf databases.” makalesi “Amit P. Sheth” tarafında yazılmış olmasına ve Semantik Web konusu ile alakalı olmasına rağmen, ilgi alanları kısmında “Semantik Web” bulunmadığı için şu anda algoritmamızın değerlendirme kriterlerine girememektedir.

Benzer konuların ve alt başlıkların bulunabilmesi için “Benzer ilgi alanı bulma yöntemi” en iyiyi bulma algoritmamıza eklenmiştir.

Bunu yaparken de sorgumuza aldığımız cevap için dönen ilgi alanlarından en fazla sayıda bulunan ilgi alanlarından n tanesi seçilerek algoritmamıza katılmaktadır. Bu örnekten hareketle “information extraction” ve “support vector machine” sorgularımız için bu ilgi alanlarının yer aldığı ve kendilerinin haricinde görünen ilgi alanlarında en çok yer alan 5 başlık Tablo 4-17’de gösterilmektedir:

Information retrieval	Support Vector Machine
Information extraction	Classification, Regression
Data mining	Inductive Logic Programming
Text mining	Classifier
Machine learning	Genetic Algorithms
Extraction	Learning

Tablo 4-17 Konular için benzer ve alt konular

Aşağıda benzer ilgi alanı bulma algoritmamız sunulmuştur:

Benzer İlgi Alanı Bulma Algoritması

Girdi: Bir ilgi alanı ve yazar kümesi ve yazarlara ait makaleler

Çıktı: İlgi alanına benzer diğer ilgi alanları

Aradığımız sorguyu içeren makalelerin sahip olduğu diğer ilgi alanlarında en çok yer alan “n” kadarını benzer olarak kabul ettik.

Benzer ilgi alanı bulma algoritmamız şu şekilde çalışmaktadır:

Algoritma:

- 1: Veri setinde bulunan bütün a yazarları için
- 2: a yazarına ait olan makalelerin aradığımız ilgi alanını barındıran makalelerini bul ve M kümesine ekle
- 3: M kümesi içinde bulunan bütün p makaleleri için
- 4: İlgi alanları kümesini p için bul ve ilgi alanları kümesi I ye ekle
- 5: I ilgi alanları kümesi için
- 6: En çok yer alan n kadar ilgi alanını bul ve benzer ilgi alanları olarak B kümesine ekle

En iyi Uzmanı Bulma Algoritması:

Girdi: İlgi alanı (q sorgusu) ve akademisyen veri seti

Çıktı: Veri setinde bulunan akademisyenlerin uzmanlık derecesine göre sıralaması

- Veri setindeki yazarlar için Linked Data tabanlı Semantik Web algoritmamızı uygula ve **q** sorgusu için olan sonuçları listele.
- Veri setindeki yazarlar için Başlık Tabanlı modelimizi uygula ve **q** sorgusu için sonuçları listele.
- Benzer ilgi alanlarını bulma algoritmamızı **q** sorgusu için uygula.
- Benzer ilgi alanı bulma algoritmamızdan dönen benzer ilgi alanları için Linked Data tabanlı Semantik Web yaklaşımımızı tekrar uygula.
- İki modelden dönen sonuçları (yazarlara ait makaleler ve ilgi alanları) birleştir.

- Bulunan makaleler için Google Scholar web arayıcısından (crawler) alıntılanma (citation) deęerlerini al.
- Aęırlıklı dil modeli (3.1.3) uygulayarak uzmanları sırala.

Uygulama ve deęerlendirme Performans (6.3.2) bölümünde ayrıntılarıyla işlenmiştir.

5. GERÇEKLEŐTİRİM

Java ve C# altyapısı kullanılarak yukarda bahsedilen bütün algoritmalar ve modeller gerekleŐtirilmiŐtir. AŐađıda bahsedilen geliŐtirme araları ve ereveler kullanılmıŐtır.

5.1. Kullanılan GeliŐtirme Araları

- Visual Studio 2010 Ultimate
<http://www.microsoft.com/visualstudio/en-us/products/2010-editions/ultimate/overview>
- Eclipse-jee-helios-SR2-win32-x86_64
<http://www.eclipse.org/downloads/>
- Microsoft Visio 2010
<http://office.microsoft.com/en-us/visio/>
- Google Chrome
<http://www.google.com/chrome>
- Microsoft SQL Server 2008 R2
<http://www.microsoft.com/sqlserver>
- Microsoft SQL Server Management Studio
<http://www.microsoft.com/download/en/details.aspx?id=7593>
- Microsoft Excel 2007
<http://office.microsoft.com/en-us/excel/>
- Microsoft Word 2007
<http://office.microsoft.com/en-us/word/>

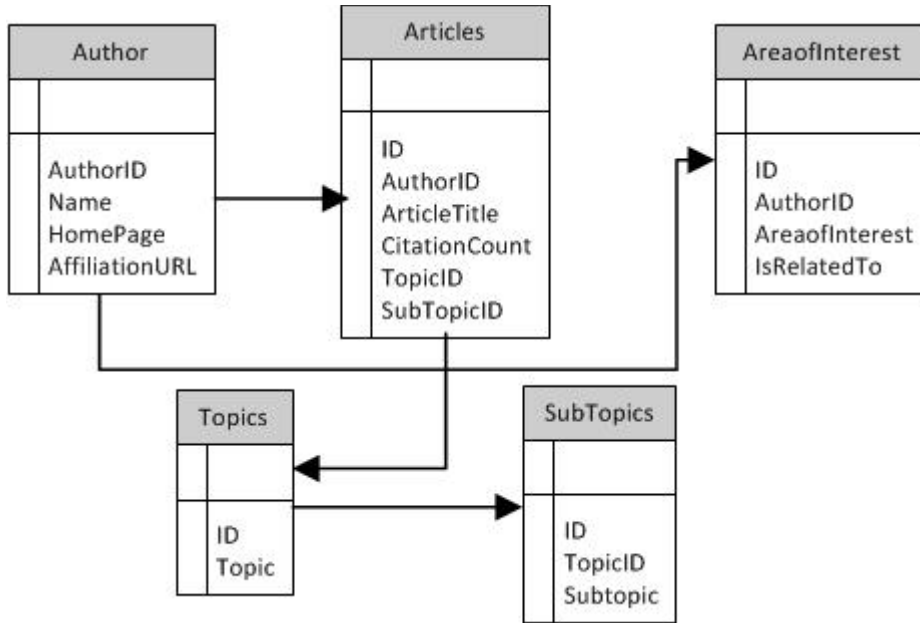
5.2. Kullanılan Çerçeve Programlar (Frameworks)

İsmi	Versiyonu	Amaç
ARQ	2.8.0	Jena için SPARQL sorgu motoru.
Jena	2.3	Semantik Web uygulamaları için kullanılan çerçeve.
SemdisAPI	3.0	RDF ayrıştırıcı ve işletici.
Commons-logging	1.1.1	Kayıt tutmak için kullanılan çerçeve.
Xerces-J	2.11.0	XML ayrıştırıcı (parser)
SemdisImpl	0.5	RDF ayrıştırıcı ve işletici.
Slf4j	1.6.1	Kayıt tutmak için kullanılan çerçeve.
.Net Framework	4.0	Web sistemi tasarımı için kullanılan çerçeve.
Stax-Api	1.0	XML ayrıştırıcı (parser)

Tablo 5-1 Kullanılan Frameworkler

5.3. Veri Modeli

Sistemimizde Expert Finder veritabanı için kullanılan veri modeli aşağıdaki gibidir:



Şekil 5-1 Kullandığımız veri tabanındaki veri modeli

Veri setimizde Şekil 5-1'den anlaşılacağı üzere yazarların adı, soyadı, web sayfası ve çalıştıkları yerin web adresi bilgisi vardır. Ayrıca her makalenin bir yazarı ve en az bir tane

de ilgili olduđu bir konu bařlıđı vardır. Konular alt konulardan oluşabilir ve ilgi alanları birbirleriyle ilişkili olabilir.

6. PERFORMANS ANALİZİ

Bu çalışmada Uzman Bulma problemine Semantik Web açısından bir yaklaşım sergilenmiş, veri toplama ve analiz etme süreçlerinde Semantik Web verisinin kullanımının ne kadar mümkün olduğu araştırılmıştır. Bu verinin kullanımının mümkün olduğu anlaşıldıktan sonra etkinliği yönünde bir çalışma yapılmıştır.

Benzerlerinden farklı olarak toplanan veri seti internet verisinden taranarak (crawl) değil Semantik Web ontolojilerinden ve kütüphanelerinden faydalanarak oluşturulmuştur.

Ortaya 3 ayrı yöntem, 3 ayrı model konulmuştur. Başlık tabanlı arama modeli, Linked Data Tabanlı Semantik Web Yaklaşımı ve bu iki modeli entegre ederek oluşturulmuş Hibrit Model bu 3 modeldir.

Kısaca özetlemek gerekirse Başlık tabanlı arama modelinde DBLP veri setindeki deney kümemizde olan bütün akademisyenlere ait makalelerin başlıklarında Eventseer'den [17] oluşturulmuş konu setindeki konular aranmakta ve ilişki bulunan konular akademisyenin ilgi alanı olarak değerlendirilmektedir.

Linked Data Tabanlı Semantik Web yaklaşımında ise ontolojiler kullanılarak doğrudan ilgi alanı olarak kayıt edilmiş semantik web bilgisi varsa, o değerlendirilmekte ve akademisyenin ilgi alanı olarak kayıt edilmektedir. Üç farklı kütüphane (ACM, IEEE ve Faceted DBLP) ve iki farklı ontolojiden (Actors ve Dublin Core) makalelere ait ilgi alanlarını karşılayan özellikler (property) bulunmuş ve akademisyenlerin ilgi alanları olarak değerlendirilmiştir.

Hibrit modelimizde ise Başlık Tabanlı arama modelimiz ve Linked Data Tabanlı Semantik Web yaklaşımımız entegre edilmiştir. Başlık tabanlı modelimizin avantajı daha çok makale ve daha çok akademisyeni kapsama alanına almış olmasıyken, Linked Data Tabanlı Semantik Web yaklaşımımızın avantajı herhangi bir arabirim yazılımına gerek duymadan

ve daha önceden analiz edilmiş, doğruya yakın verileri SPARQL sorguları ile kolay bir şekilde ulaşılabilmesidir.

Başlık tabanlı arama modelimizdeki konu setimizin avantajı oldukça güncel olması ve bilgisayar bilimleri alanındaki bütün konu başlıklarını hemen hemen kapsamasıdır.

Hibrit modelimizde iki modelimizin avantajlarını bir araya getirerek, alıntılanma (citation) değerlerini de kullanarak belli bir sorguya ait en iyi k sayıdaki (top-k) akademisyeni bulmamıza fayda sağlamasıdır.

Aşağıda ilk önce Başlık tabanlı arama modelimizle, Linked Data Tabanlı Semantik Web yaklaşımımızın performansları karşılaştırılacaktır. Daha sonra Linked Data Tabanlı Semantik Web yaklaşımımızın DBLP++ kütüphanesinin kullanımıyla performansının nasıl arttığı ortaya konacak neden üç farklı kütüphaneye ihtiyacımızın olduğu incelenecektir. Hibrit modelimizde ise iki farklı arama modelinin avantajları ortaya konularak alanından en iyi uzmanı bulma problemine bulunan çözüm citation (alıntılanma) değerlerinin de katılımı ile değerlendirilecektir. Ayrıca benzer ilgi alanı bulma yöntemimizin sonuçları ne ölçüde geliştirdiği değerlendirilecektir.

6.1. Test Ortamı

Testleri üzerinde koştığımız sistemin donanım özellikleri 4 GB Ram'e sahip 2 GHz işlemciye sahiptir. Donanım üzerinde Windows 7 işletim sistemi kuruludur.

Testleri koşturmak için 111 akademisyen seçilmiş, bu akademisyenler ile ilgili gerçek ilgi alanları verisi toplanmıştır. Bu verilerin toplanmasında internet verisi kullanılmış ve veri tabanı insan gücü ile oluşturulmuştur. Veri tabanında kullanılan ilgi alanları kişilerin kendilerini tanımladıkları web sayfalarının yanı sıra o akademisyenleri tanıyan kişiler yardımıyla oluşturulmuştur. Ayrıca www.arnetminer.org [18] sisteminden de faydalanılmıştır. Bu sisteme girilen akademisyenler için ilgi alanları sağlanmaktadır.

En iyi uzmanı bulma yöntemimiz için ise Deng'in [5] makalesinden ve deney setinden yararlanılmıştır.

6.2. Performans Metrikleri

Sistemin performansı oluşturulan veri setindeki akademisyenlere ait olan ilgi alanlarının ne kadarının modellerimiz tarafından karşılandığı ile doğru orantılıdır. Örneğin gerçek veri setimizde bir akademisyene ait 10 ayrı ilgi alanı varsa ve bu ilgi alanlarından 4'ü birinci modelimiz tarafından kapsanıyorsa, ikinci modelimiz tarafından ise 4'ten fazla ilgi alanını kapsanıyorsa, ikinci modelimiz birinci modelimize göre daha iyi bir performans gösteriyordur.

En iyi uzman bulma algoritmamız için de değerlendirme kriterimiz Deng'in bulduğu sonuçlarla eşleşme oranıdır.

6.3. Test Sonuçları ve Analiz

6.3.1. Başlık Tabanlı Model ve Linked Data tabanlı Semantik Web Modeli Sonuçları

Başlık Tabanlı Modelimiz için karşılaştırma sonuçları:

Koşul	Akademisyen Sayısı	Bilinen ilgi alanlarını bulma oranı
Akademisyenin hepsi katılarak	111	%19
Akademisyenlerden yalnızca en az bir makalesi olanlar katılarak	73	%28

Tablo 6-1 Başlık tabanlı modelimiz için bilinen ilgi alanlarını bulma oranı

Linked Data tabanlı Semantik Web Modelimiz için sadece ACM ve IEEE kütüphaneleri kullanılarak alınan sonuçlar:

Koşul	Akademisyen Sayısı	Bilinen ilgi alanlarını bulma oranı
Akademisyenin hepsi katılarak	111	%22
Akademisyenlerden sadece en az bir makalesi olanlar katılarak	47	%53

Tablo 6-2 Linked Data tabanlı modelimiz için ACM ve IEEE ile bilinen ilgi alanlarını bulma oranı

Yukarıdaki Tabloları (Tablo 6-1 ve Tablo 6-2) incelediğimizde Başlık Tabanlı arama modelimize göre Linked Data tabanlı Semantik Web modelimizin hem akademisyenlerin hepsi hesaba katılarak yapılan analizde hem de sistemde sadece en az bir tane makalesi bulunan akademisyenlerin katıldığı iki durumda da, daha iyi bir performans ortaya koyduğunu görmekteyiz.

Fakat iki modelinde bütün akademisyenler için karşılama oranında %20 olarak gerçekleşmesi kötü bir performans olarak değerlendirilebilir.

Diğer yandan sadece en az bir makalesi olan akademisyenlerle hesaplama yapıldığında Linked Data tabanlı Semantik Web modelimizin hesaplamaya giren akademisyenlerden iki uzmanlık alanından en az bir tanesini bulması iyi bir performans olarak değerlendirilebilir.

Linked Data tabanlı Semantik Web modelimizi DBLP++ verisi ile genişletmek ve güçlendirmek isterken aldığımız sonuçlara bakmak gerekirse:

Sadece DBLP++ kullanılarak elde edilen sonuçlar:

Koşul	Akademisyen Sayısı	Sonuç
Akademisyenin hepsi katılarak	111	%31
Akademisyenlerden sadece en az bir makalesi olanlar katılarak	78	%61

Tablo 6-3 Linked Data tabanlı modelimiz için DBLP++ ile bilinen ilgi alanlarını bulma oranı

Tablo 6-2 ve Tablo 6-3'ten gözlemleyebileceğimiz gibi bütün akademisyenler hesaba katılarak ve en az bir tane makalesi olan akademisyenler hesaba katılarak yapılan hesaplamalar sadece DBLP++ verisi kullanılarak oluşturulmuş veri setinin daha başarılı olduğunu göstermektedir.

DBLP++ veri seti kullanılarak yapılan hesaplamalarda, hesaplamaya giren her akademisyenlerin kendisini tanımladığını her 3 ilgi alanından 2 tanesi DBLP++ verisi kullanılarak bulunmuştur.

Analizler esnasında dikkate değer bazı ilgi alanları eşleşme verisi gözlemlenmiştir. Bu ilgi alanlarının DBLP++ verisi eklenerek, Linked Data tabanlı Semantik Web modeli ile bulunması sistemin verimliliğini göstermesi açısından önemlidir:

Akademisyenin İsmi	Gerçek İlgi Alanı	DBLP++'da Bulunan İlgi Alanı
Nihan Kesim Çiçekli	Event Calculus	Event Calculus
Ahmet Oğuz Akyüz	High – Dynamic Range Imaging	High – Dynamic Range Imaging
Halit Oğuztüzün	Distributed simulation	Distributed Interactive Simulation

Tablo 6-4 Bilinen ilgi alanı ile DBLP++'da bulunan ilgi alanı karşılaştırması

Hem ACM, IEEE hem de DBLP++ kütüphaneleri kullanılarak alınan sonuçlar:

Koşul	Akademisyen Sayısı	Sonuç
Akademisyenin hepsi katılarak	111	%34
Akademisyenlerden sadece en az bir makalesi olanlar katılarak	88	%66

Tablo 6-5 Linked Data tabanlı modelimiz için hem ACM, IEEE hem de DBLP++ verisi kullanılarak bilinen ilgi alanlarını bulma oranı

Tablo 6-5'dan anlaşılacağı üzere Linked Data tabanlı Semantik Web yaklaşımımızın en iyi performansın hem ACM, IEEE hem de DBLP++ kütüphanelerinden gelen verileri kullanarak elde edildiği görülmektedir.

DBLP++ kütüphanesinin makale ve akademisyen kapsama sayılarının ACM ve IEEE'dekinden fazla olmasına rağmen ikisinin birlikte kullanıldığında modelin performansının artmasının sebebi DBLP++'da bulunmayan makalelerin ACM ve IEEE kütüphaneleri tarafından kapsanmasıdır. Bu durum iki şekilde olabilmektedir.

Birincisi Tablo 6-6'den görülebileceği üzere akademisyenlerin %30'una yakınının ACM ve IEEE kütüphanelerinde bulunan makale sayısı DBLP++'da bulunan makale sayısından fazladır. Burada bulunan fazla sayıların DBLP++ kütüphanesi tarafından kapsanmadığı açıktır.

Akademisyenin İsmi	ACM, IEEE'deki makale sayıları	DBLP++'daki makale sayıları
Volkan Atalay	8	4
Cem Bozsahin	2	0
Muslim Bozyigit	2	2
Tolga Can	3	2
Ahmet Cosar	6	2
Ayşe Kiper	2	0
Sibel Tarı	3	1
İsmail Hakkı Toroslu	7	1
Fatos Yarman	2	0
Adnan Yazıcı	14	8
Oguz Ergin	9	8
Selim Aksoy	3	2
Fazli Can	21	11
Tolga Capin	6	0
Tugrul Dayar	4	3
Hayri Sever	5	3
Ethem Alpaydın	10	3
A.Taylan Cemgil	2	0
Cem Ersoy	5	3

Tablo 6-6 ACM, IEEE kütüphanesinde DBLP++ kütüphanesine göre daha fazla makalesi bulunan akademisyenlere ait makale sayıları

İkinci olarak da DBLP++'da bulunan makale sayısı ACM ve IEEE'deki makalelerden fazla olan akademisyenler için ACM ve IEEE'de, DBLP++'da bulunmayan farklı makalelerin yer alma durumudur.

Bu durumu örnekle görmek istersek:

ACM ve IEEE:

- 1 Error-constrained COUNT query evaluation in relational databases
- 2 Enforcing Interdependencies and Executing Transactions Atomically over Autonomous Mobile Data Stores Using SyD Link Technology
- 3 Implementation of a Calendar Application Based on SyD Coordination Links

Tablo 6-7 Erdoğan Doğdu'ya ait ACM ve IEEE kütüphanesinde bulunan makaleler

DBLP++ :

- 1 A Java API for global querying and updates for a system of databases.
- 2 Semantic web in eHealth.
- 3 An Agent Module for a System on Mobile Devices.
- 4 Implementation of a Calendar Application Based on SyD Coordination Links.
- 5 SyD: A Middleware Testbed for Collaborative Applications over Small Heterogeneous Devices and Data Stores.
- 6 Health Level-7 compliant clinical patient records system.

Tablo 6-8 Erdoğan Dođdu'ya ait DBLP++ kütüphanesinde bulunan makaleler

Erdoğan Dođdu'ya ait DBLP veri setinden toplam 24 makale vardır. Tablo 6-8'dan anlaşılacağı üzere sadece DBLP++'da bulunan makalelerin toplam makaleleri karşılama oranı %25'tir. Eğer Tablo 6-7'deki verileri de beraber değerlendirecek olursak bu oran %33'e çıkmaktadır. Bu artış oranı kapsanan ilgi alanları sayısını da artırmaktadır.

Yöntem	Veri Seti	Koşul	Akademisyen Sayısı	Bilinen ilgi alanlarını bulma oranı
Başlık Model	Tabanlı DBLP	Akademisyenin hepsi katılarak	111	%19
Başlık Model	Tabanlı DBLP	En az bir makalesi olanlar katılarak	73	%28
Linked Tabanlı Model	Data ACM, IEEE	Akademisyenin hepsi katılarak	111	%22
Linked Tabanlı Model	Data ACM, IEEE	En az bir makalesi olanlar katılarak	47	%53
Linked Tabanlı Model	Data DBLP++	Akademisyenin hepsi katılarak	111	%31
Linked Tabanlı Model	Data DBLP++	En az bir makalesi olanlar katılarak	78	%61
Linked Tabanlı Model	Data ACM, IEEE, DBLP++	Akademisyenin hepsi katılarak	111	%34
Linked Tabanlı Model	Data ACM, IEEE, DBLP++	En az bir makalesi olanlar katılarak	88	%66

Tablo 6-9 Modeller için farklı veri setleri kullanılarak karşılaştırmalı sonuçları

Tablo 6-9 Modeller için farklı veri setleri kullanılarak karşılaştırmalı sonuçlar'dan anlaşılacağı üzere sadece bir veri seti kullanıldığında en iyi sonucu Linked Data tabanlı Model ve DBLP++ veri seti ile alıyoruz. Eğer veri setimiz de akademisyene ait bir makale varsa, akademisyenin her 3 ilgi alanından yaklaşık 2 tanesini, yani %61 oranında bilinen ilgi alanlarını tespit edebiliyoruz.

Eğer DBLP++ verisinin yanında ACM, IEEE kütüphanelerine ait veri setini de dâhil edersek de bu oranın %66'ya kadar çıkmış olduğunu görüyoruz. Ayrıca ACM, IEEE kütüphanelerine ait veri setini DBLP++ veri seti ile beraber kullandığımız zaman bilinen ilgi alanlarını karşılama oranının artışı yanında en az bir makalesi olan akademisyenin sayısının da %70'ten (78/111), %79'a (88/111) çıkmış olduğunu görüyoruz.

6.3.2. En iyi (Top-k) Uzmanı Bulma Algoritması Sonuçları

Deng [5]'in yaptığı çalışmada kullanılan sorgu kümesi (query set) En iyi Uzmanı Bulma algoritmamızın değerlendirilmesinde kullanılmıştır. Deng çalışmasında en iyi uzmanları insan gücü kullanarak bulmuş daha sonra da algoritmasından çıkan sonuçlarla karşılaştırmıştır. Deng'in çalışmasında temel aldığı veri seti ve sonuçlarını kendi modelimizin değerlendirmesinde kullanacağız.

Sadece DBLP++ verisini ve Benzer ilgi alanı bulma algoritmasını kullandığımızda "Support Vector Machine" konusu için yapılan sıralamada hibrit modelimizin bulduğu ilk beş akademisyen aşağıdadır:

Akademisyenin İsmi	Makale Sayısı	Citation Değerleri Toplamı
Massimiliano Pontil	10	13590
Ingo Steinwart	2	8544
Bernhard Scholkopf	4	4907
Edward Y. Chang	19	1643
Thorsten Joachims	8	1091

Tablo 6-10 DBLP++ ve Benzer İlgi alanı Bulma Algoritması kullanarak "Support Vector Machine" konusunda bulunan ilk 5.

Deng'in algoritması ile "Support Vector Machine" başlığı için ilk beş akademisyen:

Akademisyenin sıralaması ve ismi	
1	Bernhard Scholkopf
2	Vladimir Vapnik
3	Alex J. Smola
4	Ingo Steinwart
5	Thorsten Joachims

Tablo 6-11 Deng'e göre "Support Vector Machine" başlığı için ilk 5 akademisyen

Tablo 6-10 ve Tablo 6-11'i incelediğimiz de sadece DBLP++ ve benzer ilgi alanı bulma algoritmasını kullandığımızda ilk 5 içerisindeki 3 en iyi uzmanın bizim algoritmamız tarafından da bulunduğunu görüyoruz.

Bulunan isimler "Bernhard Scholkopf", "Ingo Steinwart" ve "Thorsten Joachims" bize algoritmamızın %60 oranında belirlenmiş en iyi uzmanları tespit edebildiğini göstermektedir.

Bulunamayan isimlerden "Vladimir Vapnik"i incelediğimizde DBLP++'da kendine ait sadece bir makale bulunduğunu ve bu makalenin alıntılanma değerinin 8 olduğunu görüyoruz. Bu sebeple "Vladimir Vapnik" için sadece DBLP++ verisinin değerlendirildiği ve Benzer ilgi alanı bulma algoritmamızın kullanıldığı durumda başarısız bir değerlendirme yapılmıştır.

Fakat daha sonra Başlık Tabanlı modelimiz ve Linked Data tabanlı Semantik Web modelimiz uygulandığında "Vladimir Vapnik" için başarı oranının arttığı tespit edilmektedir. Başlık tabanlı modelimiz de ilgi alanları alt kümesi için Tablo 4-17 kontrol edilebilir.

DBLP++ verisi ile "Vladimir Vapnik" için "Support Vector Machines" için bulunan sonuçlar:

	Makale Başlığı	İlgi Alanı	Alıntılanma
1	Large margin vs. large volume in transductive learning	Learning	8

Tablo 6-12 “Vladimir Vapnik” için DBLP++ veri setinde bulunan makaleler

ACM ve IEEE verisi ile “Vladimir Vapnik” için “Support Vector Machines” için Benzer ilgi alanı bulma algoritması kullanılarak bulunan sonuçlar:

1	Makale Başlığı	İlgi Alanı	Alıntılanma
2	Prior knowledge in support vector kernels	Learning	224
3	A training algorithm for optimal margin classifiers	Learning	3463
4	Measuring the VC-dimension of a learning machine	Learning	193
5	Support-Vector Networks	Learning	7105
6	Support vector density estimation	Learning	81
7	Support vector regression with ANOVA decomposition kernels	Learning	61
8	Using support vector machines for time series prediction	Learning	132
9	Discovering informative patterns and data cleaning	Learning	179
10	Local learning algorithms	Learning	331
11	Three remarks on the support vector method of function estimation	Learning	2401

Tablo 6-13 “Vladimir Vapnik” için ACM ve IEEE veri setinde bulunan makaleler

Başlık tabanlı Modelimiz ile “Vladimir Vapnik” için “Support Vector Machines” için Benzer ilgi alanı bulma algoritması kullanılarak bulunan sonuçlar:

	Makale Başlığı	İlgi Alanı	Alıntılanma
1	Incorporating Invariances in Support Vector Learning Machines.	Support Vector	184
2	Support Vector Regression Machines.	Support Vector	583
3	Support Vector Method for Function Approximation, Regression Estimation and Signal Processing.	Support Vector	992
4	Parallel Support Vector Machines: The Cascade SVM.	Support Vector	138
5	Model Selection for Support Vector Machines.	Support Vector	219
6	Support Vector Method for Multivariate Density Estimation.	Support Vector	78
7	Support vector machines for spam categorization.	Support Vector	732
8	Support vector machines for histogram-based image classification.	Support Vector	680
9	Predicting Time Series with Support Vector Machines.	Support Vector	511

	Makale Başlığı	İlgi Alanı	Alıntılanma
10	The Support Vector Method.	Support Vector	19
11	Prior Knowledge in Support Vector Kernels	Support Vector	224
12	Support Vector Clustering.	Support Vector	577
13	Learning with Rigorous Support Vector Machines.	Support Vector	14
14	A Support Vector Clustering Method.	Support Vector	59
15	Bounds on Error Expectation for Support Vector Machines.	Support Vector	322
16	Choosing Multiple Parameters for Support Vector Machines.	Support Vector	1147
17	Gene Selection for Cancer Classification using Support Vector Machines.	Support Vector	2295
18	Principles of Risk Minimization for Learning Theory.	Learning	185
19	An overview of statistical learning theory.	Learning	1521
20	Learning Curves: Asymptotic Values and Rate of Convergence.	Learning	38
21	Boosting and Other Machine Learning Algorithms	Learning	3955
22	Learning using hidden information (Learning with teacher).	Learning	4
23	A new learning paradigm: Learning using privileged information	Learning	14
24	Local Learning Algorithms.	Learning	331
25	Large Margin vs. Large Volume in Transductive Learning.	Learning	8
26	Learning hidden information: SVM+.	Learning	258
27	Measuring the VC-Dimension of a Learning Machine	Learning	193
28	A Training Algorithm for Optimal Margin Classifiers	classifier	3463
29	Automatic Capacity Tuning of Very Large VC-Dimension Classifiers.	classifier	193
30	Support vector machines for histogram-based image classification.	classification	680
31	Gene Selection for Cancer Classification using Support Vector Machines.	classification	2295

Tablo 6-14 “Vladimir Vapnik” için DBLP veri setinde bulunan makaleler

Yukarıdaki tabloları incelediğimizde (Tablo 6-12,Tablo 6-13 ve Tablo 6-14) sadece Başlık Tabanlı modelimizi veya sadece Linked Data tabanlı Semantik Web modelimizi kullandığımızda akademisyenlere ait tüm verileri kullanamadığımız ve değerlendirmemizin sağlıklı olmadığını ortaya koyan örneklerden bir tanesini görmüş oluyoruz. Hibrit modelimizin verimliliği işte bu aşamada kendini gösteriyor.

Hibrit modelimiz kullanıldığında sıralamamız değişmesine rağmen ilk 5 ismin yine 3 tanesinin algoritmamız tarafından tespit edilmiş olduğunu görüyoruz. Deng’in ilk 5’inde yer alıp bizim bulduğumuz ilk 5’e giremeyen kişilerin sadece DBLP++ kullanılarak sıralaması bulunan “Vladimir Vapnik (29/34) ve Alex J.Smola (22/34) ’dan” farklı olarak

sıralamanın en sonlarında değil hemen 9. ve 10. olarak sıralamada yer aldıklarını görüyoruz.

Akademisyenin İsmi	Citation Değerleri Toplamı	Sıralama
Vladimir Vapnik	37351	1
Massimiliano Pontil	37330	2
Bernhard Scholkopf	30002	3
Thorsten Joachims	28906	4
John Shawe-Taylor	28856	5
Ingo Steinwart	18438	9
Alex J. Smola	14906	10

Tablo 6-15 Hibrit modelimizin bütün veri setleri ile çalışması sonucunda “Vladimir Vapnik” için alınan sonuçlar

Tablo 6-15’den da görüleceği üzere Deng’in 34 akademisyen içerisinde bulunduğu “Support Vector Machines” konusunda en iyi 5 akademisyenin 3’ü de bizim ilk 5’imizde yer almıştır. Deng’in bulunduğu ilk 5 akademisyenin hepsi bizim sıralamamızda da ilk 10’a girme başarısını göstermiştir.

Bu ilk beş için %60’lık bir başarıyken, 5 akademisyenin hepsinin ilk 10 da yer alması %50 başarı olarak yorumlanabilir. Bu da Semantik Web tabanlı verilerden yola çıkarak uzmanların ilgi alanlarının rahatlıkla bulunabileceğini ve bu uzmanlık alanlarından yola çıkılarak sıralama yapılabileceğini göstermiştir.

Alıntılanma değerleri açısından baktığımızda da bazı yazarların çok sayıda makale yazdıkları ama bu makalelerden fazla sayıda alıntı yapılmadığından dolayı algoritmamızda geri sıralarda yer aldığı görülmektedir.

Model	Yazarın İsmi	Makale Sayısı	Alıntılan Değeri	Sıralaması
Başlık Tabanlı Model	Edward Chang	Y. 24	1497	24
Linked Data tabanlı Semantik Model	Edward Chang	Y. 24	3506	24
Başlık Tabanlı Model	Ingo Steinwart	5	18438	9
Linked Data tabanlı Semantik Model	Ingo Steinwart	14	9410	9

Tablo 6-16 Makale alıntılanma değerleri ve yazar sıralaması

Tablo 6-16'dan görüleceği üzere Edward Y. Chang (24/34), Ingo Steinwart'tan (9/34) fazla sayıda makale yazmasına rağmen Ağırlıklı Dil Modelimizden dolayı sıralama da Steinwart'ın gerisinde yer almaktadır.

Diğer sorguları da koştığımız zaman benzer sonuçlarla karşılaşmaktayız:

Deng'in "Information extraction" başlığı için koştığı sonuçları Tablo 6-17 Deng'in çalışmasında information extraction'den görmekteyiz.

Akademisyenin sıralaması ve ismi	
1	Ellen Riloff
2	Dayne Freitag
3	Stephen Soderland
4	Raymond J. Mooney
5	Andrew McCallum

Tablo 6-17 Deng'in çalışmasında information extraction için sonuçlar

Akademisyenin sıralaması ve ismi	Makale Sayısı	Citation Toplamı
1 Ellen Riloff	2	1375
2 Daniel S. Weld	4	628
3 Oren Etzioni	5	548
4 Stephen Soderland	1	420
5 Andrew McCallum	4	278

Tablo 6-18 Bizim çalışmamızın benzer ilgi alanı bulma algoritması olmadan Information extraction için bulduğu sonuçlar

Information Extraction konusu için benzer ilgi alanları bulma algoritması olmadan Deng'in sonuçları ile bizim sonuçlarımız %60 oranında uyuşmaktadır. Yani Deng'in sıralamasında 20 akademisyen arasından ilk 5'e giren akademisyenlerden 3 tanesi bizim sıralamamızda da ilk 5 'e girmiştir.

Akademisyenin sıralaması ve ismi	
1	Ronen Feldman
2	Ellen Riloff
3	Dayne Freitag
4	Daniel S. Weld
5	Raymond J. Mooney

Tablo 6-19 Bizim çalışmamızın benzer ilgi alanı bulma algoritması ile Information extraction için bulduğu sonuçlar

Benzer ilgi alanı bulma algoritmamızla seçilen başlıkları Tablo 4-17'den görülebilir.

Information Extraction konusu için Benzer ilgi alanı bulma algoritması uyguladığımız zaman Deng'in sonuçlarıyla yine %60 oranında uyuşmuş olduğumuzu görüyoruz. Fakat bir önceki örneğimizde olduğu gibi Benzer ilgi alanı bulma algoritması uyguladığımızda Tablo 6-17'da Deng'in listesinde olup bizim listemizde yer almayan kişilerin sıralamalarının 3 kişinin içinde olmayan Raymond J. Mooney (19/20) ve Dayne Freitag (18/20)'in çok gerilerde yer aldığını görüyoruz. Ancak Benzer ilgi alanı bulma algoritması kullanılarak listeye giremeyen iki kişinin Andrew McCallum (8/20) ve Stephen Soderland (9/20) daha mantıklı ve makul olan ön sıralarda yer aldığını görüyoruz.

Akademisyenin sıralaması ve ismi
Deiter Fensel
James A. Hendler
Katia P. Sycara
Amit P. Sheth
Ian Horrocks

Tablo 6-20 Deng'in çalışmasında Semantik Web için sonuçlar

Akademisyenin sıralaması ve ismi	Makale Sayısı	Alıntılanma Toplamı
Stefan Decker	16	3196
Ian Horrocks	8	2581
Sheila A. McIlraith	4	2517
Amit P. Sheth	21	1892
Frank van Harmelen	4	1711

Tablo 6-21 Bizim çalışmamızın benzer ilgi alanı bulma algoritması olmadan Semantik Web için bulduğu sonuçlar

Semantik Web konusu için benzer ilgi alanı bulma algoritması olmadan Deng'in sonuçları ile bizim sonuçlarımız %40 oranında uyuşmaktadır. Yani Deng'in sıralamasında 20 akademisyen arasından ilk 5'e giren akademisyenlerden 2 tanesi bizim sıralamamızda da ilk 5'e girmiştir.

Akademisyenin sıralaması ve ismi	Makale Sayısı	Alıntılanma Toplamı
Stefan Decker	20	3666
Ian Horrocks	10	2643
Katia P. Sycara	5	2554
Steffen Staab	22	2394
Amit P. Sheth	32	2198

Tablo 6-22 Bizim çalışmamızın benzer ilgi alanı bulma algoritması kullanarak Semantik Web için bulduğu sonuçlar

Benzer ilgi alanı bulma algoritmamızla seçilen başlıklar:

Semantic Web
Ontology
Ontologies
RDF
OWL
Sparql

Tablo 6-23 Semantik Web alt başlıkları

Konular	Sıralamalarda uyuşan kişi sayısı	Oran	Benzer ilgi alanı bulma algoritması ile uyuşan kişi sayısı	Oran
Semantic Web	2	%40	3	%60
Boosting	3	%60	4	%80
Information Extraction	3	%60	3	%60
Computer Vision	3	%60	3	%60

Tablo 6-24 Deng'in veri setindeki konular ile ilk 5 ve bizim sonuçlarımız için uyuşan yazar sayıları ve oranlar

Semantik Web konusu için benzer ilgi alanı bulma algoritması uyguladığımız zaman da, yukarıdaki örneklerde de gördüğümüz gibi Deng'in deney kümesini kullanarak Deng'in sonuçlarıyla %60'ın üzerinde bir oranla uyuşmuş olduğumuzu görüyoruz.

Ayrıca benzer ilgi alanı bulma algoritmamızın sonuçlara katkısının pozitif yönde olduğu burada da görülebilmektedir.

Buradaki örneklerden de ve Tablo 6-24'den anlaşılacağı üzere Deng'in makalesinde yer verdiği akademisyenlerin sıralaması Benzer ilgi alanı bulma algoritması ve Hibrit modelimiz sayesinde %60'ın üzerinde bir oranla yakalanabilmektedir.

Bu da bize Tezimizin başında savunduğumuz Uzman Bulma problemine Semantik Web yaklaşımıyla yeterli ve yerinde sonuçlar alabildiğini göstermiştir.

7. SONUÇ ve GELECEK ÇALIŞMALAR

Önceki yapılan çalışmalarda temel alınarak uzman bulma problemi ile ilgili Semantik Web yaklaşımı kullanarak Başlık tabanlı model ve Linked Data tabanlı Semantik Web modeli olmak üzere iki farklı model geliştirilmiştir.

Bu modeller için veri toplamak üzere farklı Semantik Web kütüphaneleri çevrimiçi olarak kullanılmış ve bu iki farklı modelin gerçekleştirimi sağlanarak performansları karşılaştırılmıştır.

Linked Data tabanlı Semantik Web modelinin sonuçlarının daha iyi performans gösterdiği tespit edilmiştir.

Daha sonra bu iki model temel alınarak en iyi uzman bulma problemi için hibrit bir yaklaşım ortaya konulmuş, bu yaklaşımın başarısını artırmak için benzer ilgi alanı bulma algoritması uygulanmıştır. Benzer ilgi alanı bulma algoritmasının verimliliği artırdığı daha önceki yapılan çalışmalardaki alınan sonuçlarla elde edilen sonuçlar karşılaştırılarak gösterilmiştir.

Çalışmalarımızdan ortaya çıkan modellerin performans değerlendirmeleri göstermiştir ki; uzman bulma problemi için Semantik Web verileri kullanılabilir ve bu veriler internetten klasik yöntemlerle toplanan verilere göre daha etkin ve verimlidir.

İlerde yapılacak çalışmalar içinde dokümanlar ve sorgular üzerine kurulmuş modellerin haricinde yazar ilişkilerini ortaya çıkaran SPARQL sorguları da uzman bulma problemine dâhil edilebilir. Birlikte yazılan makalelerde yazarlar arası ilişki (Co-authorship) ile aynı konular üzerinde çalışan akademisyenler tespit edilebilir. Bunun için FOAF ontolojisinde bulunan “relationship:worksWith” alanı kullanılabilir.

Hirsh (h) index [24] bir akademisyenin üretkenliğini ve verimliliğini ölçmek için kullanılan bir index olmakla beraber, yazarın en çok atıf yapılan makaleleri ve o makalelerin alıntı

sayısı ile hesaplanmaktadır. Weighted Language (Ağırlıklı Dil) modelimizdeki alıntı yaklaşım sayısı gibi bu da bir parametre olarak hesaplama algoritmamıza girebilir. Google h-index hesaplaması için bir ara yüz (API) sunmaktadır.

8. REFERANSLAR

- [1] S. Hettich and M. J. Pazzani. Mining for proposal reviewers: lessons learned at the national science foundation. In *KDD*, pages 862–871, 2006.
- [2] D. M. Mimno and A. McCallum. Expertise modeling for matching papers with reviewers. In *KDD*, pages 500–509, 2007.
- [3] M. A. Rodriguez and J. Bollen. An algorithm to determine peer-reviewers. *CoRR*, abs/cs/0605112, 2006.
- [4] Social Network Extraction of Academic Researchers ICDM '07 Proceedings of the 2007 Seventh IEEE International Conference on Data Mining.
- [5] Formal Models for Expert Finding on DBLP Bibliography Data 2008 Eighth IEEE International Conference on Data Mining.
- [6] M. S. Ackerman. Augmenting organizational memory: a field study of answer garden. *ACM Transactions on Information Systems (TOIS)*, 16(3):203–224, 1998.
- [7] Desislava Petkova and W. Bruce Croft. 2006. Hierarchical Language Models for Expert Finding in Enterprise Corpora. In *Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI '06)*. IEEE Computer Society.
- [8] K. Balog, L. Azzopardi, and M. de Rijke. Formal models for expert finding in enterprise corpora. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference*, 2006.
- [9] Krisztian Balog and Maarten De Rijke. 2007. Determining expert profiles (with an application to expert finding). In *Proceedings of the 20th international joint conference on Artificial intelligence (IJCAI'07)*, Rajeev Sangal, Harish Mehta, and R. K. Bagga (Eds.). Morgan Kaufmann.
- [10] B. Aleman-Meza, F. Hakimpour, I.B. Arpinar, A.P. Sheth: SwetoDblp Ontology of Computer Science Publications In: *Web Semantics: Science, Services and Agents on the World Wide Web*, volume 5, issue 3, pages 151-155, 2007.
- [11] C. Zhai and J. D. Lafferty. A study of smoothing methods for language models applied to information retrieval. *ACM Trans. Inf. Syst.*, 22(2):179–214, 2004.
- [12] W3C. The W3C test collection, 2005. URL: <http://research.microsoft.com/users/nickcr/w3c-summary.html>.

- [13] TREC. Enterprise track, 2005. URL: <http://www.ins.cwi.nl/projects/trec-ent/wiki/>.
- [14] D. Petkova and W. B. Croft. Hierarchical language models for expert finding in enterprise corpora. In ICTAI, pages 599–608, 2006.
- [15] Y. Cao, J. Liu, S. Bao, and H. Li. Research on expert search at enterprise track of trec 2005. In Proceedings of TREC 2005, 2005.
- [16] C. Macdonald and I. Ounis. Voting for candidates: adapting data fusion techniques for an expert search task. In CIKM, pages 387–396, 2006.
- [17] Eventseer.net. URL:<http://eventseer.net/topic/>.
- [18] Arnetminer.org. URL: <http://arnetminer.org/>.
- [19] Faceted DBLP. URL: <http://dblp.l3s.de>.
- [20] Milan Stankovic, Jelena Jovanovic, Philippe Laublet: Linked Data Metrics for Flexible Expert Search on the Open Web. ESWC 2011:108-123.
- [21] Daniel Herzig, Hristina Taneva: Multilingual Expert Search using Linked Open Data as Interlingual Representation. CLEF (Notebook Papers/LABs/Workshops) 2010.
- [22] Semantik Web Tanımı URL: tr.wikipedia.org/wiki/Anlamsal_ağ
- [23] Aktör Ontolojisi: URL: <http://www.aktors.org/ontology/portal>
- [24] Hirsh index : <http://en.wikipedia.org/wiki/H-index>

Ek 1. Veri seti için elde edilen sonuçlar

Örnek Sonuçlar

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
Adnan Yazıcı	Intelligent Database Systems, Multimedia Databases, Fuzzy Database Modelling, Design and Analysis of Algorithms, Wireless Sensor Networks	Logical Design, General, Deduction and Theorem Proving, Information Search and Retrieval, Systems, Database Applications, GENERAL, Vision and Scene Understanding, Knowledge Representation Formalisms and Methods, Multimedia Information Systems, Content Analysis and Indexing, Mathematical Logic, SOCIAL AND BEHAVIORAL SCIENCES, Clustering, Logical Design, Similarity measures, Object-oriented databases, Data models, DISCRETE MATHEMATICS, Applications, Languages, Algorithms, Systems and Information Theory, Physical Design, PERFORMANCE OF SYSTEMS, Network Architecture and Design, Systems and Software,	classification, constraints, data model, data modeling, database applications, databases, database, education, genetic algorithm, genetic algorithms, index structure, knowledge-based system, model, object-oriented data, object-oriented database, object-oriented databases, semantic, reasoning, similarity, spatial data, spatial databases, video data, video database, news video,	OWL, RDF, content-based retrieval, event extraction, fuzziness, genetic algorithm, object extraction, ontology, video model, audio retrieval; news broadcasts; fuzzy classes; query-by-example, information extraction; named entity recognition; Turkish, Data mining; fuzzy association rules; fuzzy spatio-temporal data cube; association rule mining; association rule mining comparison criteria, fuzzy object-oriented data model, fuzzy set theory, uncertain information, computer engineering education, online education, online information technologies certificate program, multimedia databases; fuzzy object-oriented data model; multimedia data modeling, Object-oriented databases, knowledge-base systems, uncertainty, fuzzy set theory, flexible querying.,
Ahmet Coşar	Relational Databases, Software Engineering, Computer Networks	Applications and Expert Systems, Problem Solving, Control Methods, and Search, Information Search and Retrieval, Systems, Graph and tree search strategies, Query	database, distributed data, distributed database, genetic algorithm, graph, mining, optimization, optimization problem, query	Web mining, web usage mining, session reconstruction, agent simulator and web topology, graph mining, map/reduce, parallel data mining, web usage mining,

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
		formulation, Query processing, Electronic Commerce, Languages, Document Preparation, Nonnumerical Algorithms and Problems, General, Optimization,	optimization, web usage	web user modeling
Ali Hikmet Doğru	Component based Software engineering	Design Tools and Techniques, Software Management, Management	service, design environment, service-oriented, system design	Axiomatic Design Theory, COSEML, Collaboration Diagrams., Requirements specification; rapid prototyping; integrated environments, component-oriented development, methodologies, object-oriented methodologies, process models,
Anne Adams	Privacy, Multi-Media communications	User Interfaces, User/Machine Systems, MANAGEMENT OF COMPUTING AND INFORMATION SYSTEMS, INFORMATION INTERFACES AND PRESENTATION (e.g., HCI), MODELS AND PRINCIPLES, Public Policy Issues, Security and Protection, Multimedia Information Systems, Privacy, Information Search and Retrieval, INFORMATION STORAGE AND RETRIEVAL, Miscellaneous, Sociology, Human factors, hebele, OPERATING SYSTEMS, SOCIAL AND BEHAVIORAL SCIENCES, DATA ENCRYPTION, Group and Organization Interfaces, Organizational Impacts, Digital Libraries, Systems and Information Theory,	trust, compliance, digital libraries, health informatics, multimedia communication, privacy, user needs, user perspectives, multimedia, security	Social exclusion, communities of practice, digital libraries, grounded theory, social empowerment, Digital libraries; User communities; Social structures; Usability; Grounded Theory, HCI, academia, clinical, context of use, digital libraries, digital library intermediaries, grounded theory, health, user communities, Digital library design; Educational digital libraries; African Context of Use; Cross-cultural usability, MMORPG, MUVE, attention, gaming, immersion, social worlds, e-commerce, online banking, situational awareness, collaboration, cooperation, novel mechanisms, trust, grounded theory, multicasting,

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
				multimedia communications, privacy, trust
Asuman Doğaç	eHealth, B2B Interoperability, Semantic Web	Network Protocols,Distributed Systems,Document Preparation,Other Architecture Styles,User Interfaces,COMPUTER-COMMUNICATION NETWORKS,Process Management,Systems,Physical Design,Online Information Services,Electronic Commerce,Graph Theory,Distributed databases,Graph algorithms,Query processing,Database Machines,Heterogeneous Databases,DATABASE MANAGEMENT,Language Classifications,Languages,Database (persistent) programming languages,COMPUTER SOCIETY,DOCUMENT AND TEXT PROCESSING	service, content management, e-commerce, interoperability, management system, management systems, metadata, mobile commerce, personalization, semantic, semantic interoperability, standards,	distributed processing; distributed workflow enactment service; coordinated task execution; processing entities; distributed heterogeneous environments; commercial workflow systems; workflow schedulers; failure resiliency; performance; block-structured workflow specification language; process instance distributed scheduling; system testing; system debugging; execution efficiency; messages,OODBMS kernel implementation; query optimization in OODBMSs; dynamic function linker; Graphical User Interface,Workflow, Agent, Web, Internet,XML, XML-QL, mobile E-commerce, personalization,DDL, DML, E/R model query languages, GDBMS, RAP database machine, associative processors, data models, database machines, operational transformations, schemas, structural transformations,Electronic Healthcare Record standards, eHealth, interoperability,semantics; ebXML registry;

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
				<p>Web Ontology Language (OWL); semantic web service discovery, Universal Business Language (UBL); eBusiness interoperability; Semantics; Business context; Context ontology, Electronic marketplace; Workflow; Agent; Extensible Markup Language (XML); Common Business Library (CBL), graphical user interfaces, object-oriented databases, database design; Generalized Expert System for Database Design; ESGM; requirement specification phase; logical schema; hierarchical data model; network data model; relational data model; GESDD; menu-driven system; database management systems; expert systems; software tools., concurrent transaction execution; distributed scheduler; distributed databases; concurrency control; distributed database management systems; ordering by serialization numbers; OSN method; certifier model; time-interval techniques; short-term locks; serializability; deadlocks; standard transaction execution policy; concurrency; log classification; timestamp ordering; two-phase locking; complexity analysis; concurrency control; distributed</p>

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
				databases, optimistic locking technique; concurrency control; distributed databases; optimistic method; dummy locks; write sets; validated transactions; validation test; short-term locks; ODL; strict two-phase locking; low conflict cases; strict 2PL; concurr, personalization; user profiles; query indexing; querying XML documents; information delivery to mobile devices,
Ben Adida	security, privacy, transparency	Document Preparation, Document Preparation, Systems, Design Tools and Techniques, Group and Organization Interfaces, General, Computer and Information Science Education, Online Information Services, Web-based interaction, Computer science education	authentication	web security, barcode, cryptographic voting, homomorphic tallying, paillier cryptosystem, paper ballot, scratch surface, phishing, two-factor authentication, web security"

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
Cem Ersoy	Computer Networks; Wireless Networks; Sensor Networks	Local and Wide-Area Networks,PROBABILITY AND STATISTICS,Network Architecture and Design,PERFORMANCE OF SYSTEMS,Other Architecture Styles,Models,COMPUTERS IN OTHER SYSTEMS,PHYSICAL SCIENCES AND ENGINEERING,Network Protocols,Distributed Systems,Optimization,Nonnumerical Algorithms and Problems,Simulated annealing,Pattern matching,Centralized networks	planning, service, admission control, assignment problem, cellular networks, control, evaluation, integer programming, mobile ad, mobile ad hoc networks, mobile communication, mobile network, model, multicast routing, network topologies, networking, network, path routing, performance evaluation, quality of service, routing algorithm, routing protocol, sensor network, sensor networks, tabu search, wireless sensor, wireless sensor network, ad hoc network, ad hoc networks, wireless, wireless sensor networks	Video Surveillance Sensor Networks; Fairness; Queue Management,reliable and secure data dissemination; wireless hot spots; wireless LANs,call admission control; channel reservation; DS-CDMA systems,Satellite; Routing; Minimum flow; Routing set; LEO; Distributed flow; Congestion,Surveillance; wireless sensor networks; deployment quality.,
Duo Zhang	unified framework, name disambiguation, topic modeling	Distributed Artificial Intelligence,Feature Measurement,Design Methodology	classification, databases, database, graph, its, language model, model, optimization, network, probabilistic framework, texture classification, unified framework,	graph-based regularization, social networks, statistical topic models,digital library, name disambiguation, probabilistic model,Language modeling, document and word graph, graph structure, smoothing,digital library, name disambiguation, semi-supervised clustering, social network analysis,advisor-advisee prediction, coauthor network,

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
				relationship mining, time-constrained factor graph
Ethem Alpaydın	Machine Learning	Simulation Output Analysis,PROBABILITY AND STATISTICS,Learning,Applications,Design Methodology,Modes of Computation,Problem Solving, Control Methods, and Search,Clustering,Simulation Output Analysis,Model Validation and Analysis,PERFORMANCE OF SYSTEMS,Other Architecture Styles,Types and Design Styles,Digitization and Image Capture,Deduction and Theorem Proving,MATHEMATICAL SOFTWARE,Models,Natural Language Processing,Scene Analysis,	color image, color images, decision tree, decision trees, error diffusion, face recognition, vector quantization	image colour analysis; fuzzy error diffusion; color images; quantization errors; dithering operation; error diffusion dithering; color cluster; closest codebook vector; error accumulation; attraction-repulsion schema; fuzzy membership function; mean squared error values,selective attention, Markov models, feature integration, face recognition, handwritten digit recognition,Machine learning, classifier design and evaluation, experimental design.,

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
Faruk Polat	Autonomous Agents and Multi-Agent Systems, Artificial Intelligence	Learning, Distributed Artificial Intelligence, Learning, Systems, Relational databases, Object-oriented databases, Logical Design, Reusable Software, DATABASE MANAGEMENT, SOFTWARE ENGINEERING, Data models, Distributed Systems, Organization and Design, Information Search and Retrieval,	planning, case-based reasoning, control, cooperation, coordination, databases, database, dynamic environment, dynamic environments, evaluation, function approximation, gene regulatory networks, information system, multi-agent reinforcement, model, object-oriented data, object-oriented database, object-oriented databases, optimization, network, pattern generation, problem solving, query optimization, regulatory networks, reinforcement learning, schema evolution, reasoning,	applications of autonomous agents and multi-agent systems, cooperation and coordination among agents, multi-agent planning, grid world, heuristic search, path planning, real-time search, grid world, moving target search, path planning, predator, probabilistic boolean networks, optimal control, Markov decision processes, monitoring, closure, database system, message expression, object algebra expression, object-oriented data model, object-oriented query language, query model, total instances, Algorithms; Base Classes; Modification Lists; Object-Oriented Databases; Views, association rules, CURE clustering algorithm, fuzzy sets, data mining, quantitative attributes, learning (artificial intelligence); multi-agent systems; function approximation; function approximation; multi-agent reinforcement learning; multi-agent based domain independent coordination mechanisms; coordination information; state transitions; reward distribution; region-wide joint rewards; Adversarial Food-Collecting World; multi-agent environments, Path

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
				Planning; Real-Time Search; Moving Target Search, PaGE, differentially expressed genes, microarray, pattern generation, q-values, algorithms, data migration, data re-engineering, information systems, object-oriented databases, relational databases, unifying databases, algorithms, data migration, forward engineering, object-oriented databases, reengineering of legacy databases, relational databases., Real-time search; Path planning, Reinforcement learning; Hierarchical reinforcement learning; Multiagent learning, q-values, significantly differentially expressed genes, microarray, pattern generation, PaGE., Class hierarchy; Object-oriented databases; Reusability; Schema evolution; Views,

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
Ilyas Cicekli	Machine Translation, Natural Language Processing, Computational Linguistics, Logic Programming, Artificial Intelligence	Programming Languages and Software, Processors, Deduction and Theorem Proving, Problem Solving, Control Methods, and Search, Inference engines, Backtracking, Natural Language Processing, MATHEMATICAL SOFTWARE, Learning, GENERAL,	constraints, databases, database, text summarization, video data, video database	text summarization; lexical cohesion; lexical chains, Event Calculus; Web Service Composition; Planning, Inductive logic programming; Machine learning; String generalization, Example-based MT; Machine learning,
Ismail Hakkı Toroslu	Databases, Algorithms, Logic	Systems and Software, Mathematical Logic, Information Search and Retrieval, Object-oriented Programming, Deduction and Theorem Proving, Language Classifications, Inference engines, Lambda calculus and related systems, PROGRAMMING TECHNIQUES, PROGRAMMING LANGUAGES, Applications and Expert Systems, MATHEMATICAL SOFTWARE, Systems, GENERAL, Requirements/ Specifications, Office Automation, Problem Solving, Control Methods, and Search, Nonnumerical Algorithms and Problems, Graph Theory	assignment problem, constraints, data mining, databases, database, dynamic programming, its, mining, multi-relational data mining, optimization, optimization problem, programming algorithm, protein structure, protein structures, query optimization, relational data, relational database, relational databases, resource allocation, sequence data, similarity, similarity search, web usage,	Knowledge bases; deductive databases; recursive rules; recursive query processing; transitive closure queries

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
Kemal Bicakci	Systems Biology, Machine Learning, Bioinformatics, Chemoinformatics, Data mining, Security	General, Combinatorics, DATA STORAGE REPRESENTATIONS, Security and Protection	trust, cooperation, network, sensor network, sensor networks, wireless network, wireless networks, wireless sensor, wireless sensor network, wireless, wireless sensor networks,	server-assisted signature, one-time signature, digital signature, pervasive computing, digital signature, PMI, attribute certificate, revocation, binding signature semantics, Real-time communication security, TCP/IP security, security management, network security, PKI, public key certificates, revocation, trust metrics, authentication, digital signatures, hash functions, hash chains, electronic commerce, network security, digital signatures,
Lada A. Adamic	social and information networks, expertise sharing, online communities	Systems and Software	knowledge sharing, network, question answering, social network, social networks	Viral marketing, e-commerce, long tail, network analysis, recommender systems, word-of-mouth, CSCW, expertise finding, expertise location, social networks, online interactions, social media, social networks, expertise finding, help seeking, knowledge sharing, online communities, question answering, social network analysis, expertise finding, expertise locators, help seeking, online communities, simulation, social network analysis, online communities, social networks, graph compression, graph synopsis, social influence; diffusion of innovations; virtual worlds; social networks, E-

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
				commerce, recommender systems, viral marketing, crowdsourcing, knowledge market, learning, online communities, question-answer sites, virtual communities, witkey, reputation, social networks, online community, question-answering
Lale Akarun	Computer Vision, Image Processing, Computer Graphics	Scene Analysis	belief functions, classification, color image, color images, color quantization, error diffusion, expert system, expert systems, face model, face recognition, feature extraction, fusion approach, graph, language recognition, model, network, sensor network, sensor networks, shape recognition, sign language, wireless sensor, wireless sensor network, wireless, wireless sensor networks,	Sign language recognition; hand gestures; head gestures; non-manual signals; hidden Markov models; belief functions, image colour analysis; fuzzy error diffusion; color images; quantization errors; dithering operation; error diffusion dithering; color cluster; closest codebook vector; error accumulation; attraction-repulsion schema; fuzzy membership function; mean squared error values, selective attention, Markov models, feature integration, face recognition, handwritten digit recognition, Multimodal, biometrics, database, evaluation, performance, benchmark, face, voice, speaker, signature, fingerprint, hand, iris

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
Mete Kabatepe	Data networks and telecommunications. Cellular networks	PERFORMANCE OF SYSTEMS,Network Protocols,PERFORMANCE OF SYSTEMS,Network Architecture and Design,Modeling techniques,Network topology,Protocol architecture (OSI model),PROBABILITY AND STATISTICS,GENERAL	network	fairness, high-speed networks, medium access control protocols, metropolitan-area networks, optical fiber networks, scalability
Michel Buffa	Mobile Robot, semantic web, Delaunay Triangulation, Robotics	Robotics,Vision and Scene Understanding,Scene Analysis,Problem Solving, Control Methods, and Search,	knowledge sharing, ontologies, semantic, semantic web,	stereo image processing; computerised navigation; navigation; image matching; mobile robots; path planning; robot vision; calibration; weakly-calibrated stereo perception; vision system; autonomous navigation; stereo perception; rover navigation; weakly calibrated stereo images; epipolar geometry; image rectification; point matching; relative point elevation; reference plane; projection images; vision module; complete navigation system; shape indicator; appropriate steering directions; unstructured outdoor environments; wheeled rover,Web 2.0, Wiki, ontology, semantic web, social tagging,semantic web; social network analysis; graph representations

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
Michel E. Adiba	Database Systems, Object-Oriented Database Systems, Database Management Systems	Office Automation, Database Applications, Systems, Logical Design, Parallel Architectures, Applications and Expert Systems, Semantics of Programming Languages, Language Classifications, Programming Languages and Software, Deduction and Theorem Proving, Languages, Mathematical Logic, Formal Definitions and Theory,	cooperation, data base, data bases, data warehouse, database system, database, distributed data, heterogeneous data, management system, management systems, medical data, model, application development,	mobile transactions; databases; mobility; transaction execution; atomicity; consistency; isolation; durability
Nihan Kesim Çiçekli	Semantic Web, Event Calculus, Multimedia databases	Vision and Scene Understanding, Knowledge Representation Formalisms and Methods, Multimedia Information Systems, Content Analysis and Indexing	service, databases, database, monitoring system, network, semantic, semantic web, semantic web services, video data, video database, web service, web services, boundary detection	Event Calculus; Web Service Composition; Planning, data mining, k-means clustering, power quality, power quality event, Inductive logic programming; Machine learning; String generalization

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
Oguz Ergin	processors,computer architecture,microarchitecture,performance of systems	PERFORMANCE OF SYSTEMS,GENERAL,Types and Design Styles,Other Architecture Styles,Design Aids,Microprogram Design Aids,Design,Single Data Stream Architectures,Process Management,General,Performance Analysis and Design Aids,High-Speed Arithmetic,Nonnumerical Algorithms and Problems,Multiple Data Stream Architectures (Multiprocessors),	dynamic scheduling, instruction scheduling	error correcting codes (ecc), narrow values, parity generation,low-complexity datapath, low-power design, reorder buffer, short-lived values,low-power comparators, superscalar datapath,instruction packing, issue queue, low power,variation-tolerant design, processor architectures, performance and reliability, impact of VLSI on system design,Low-power design, compiler, energy efficiency, microarchitecture, register file,Issue queue, instruction packing, low power,Superscalar processors, register file optimization, precise interrupts
Osman Abul	Data Mining,Bioinformatics, Machine Learning	Distributed Artificial Intelligence,Machine Learning	clustering, databases, database, evaluation, function approximation, moving object, moving objects, multi-agent reinforcement, model, pattern generation, performance evaluation, reinforcement learning,	probabilistic boolean networks, optimal control, Markov decision processes, monitoring,learning (artificial intelligence); multi-agent systems; function approximation; function approximation; multi-agent reinforcement learning; multi-agent based domain independent coordination mechanisms; coordination information; state transitions; reward distribution; region-wide joint rewards; Adversarial Food-Collecting World; multi-agent environments,PaGE, differentially expressed genes,

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
				microarray, pattern generation, q-values, q-values, significantly differentially expressed genes, microarray, pattern generation
Peter Aczel	Symbolic Logic, Dependent Type Theory, general topology	MISCELLANEOUS, Miscellaneous, Graph Theory, Mathematical Logic, Specifying and Verifying and Reasoning about Programs, Software/Program Verification, Semantics of Programming Languages, Deduction and Theorem Proving, Studies of Program Constructs	set theory	Constructive set theory; Exponentiation axiom, Final Universe; Process; Coalgebra; Labelled; Transition System; CCS; CSP
Selim Aksoy	Computer vision, Statistical and structural pattern recognition, Machine learning, Data mining	Segmentation, Database Applications, Design Methodology, Information Search and Retrieval, Clustering, Scene Analysis	evaluation, image retrieval, mining, object detection, remote sensing, similarity, similarity measure, similarity measures, training	Graph theoretic clustering; Content based image retrieval; Image clustering; Image grouping, data fusion, decision tree classifiers, land cover analysis, missing data, remote sensing,

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
Sibel Adalı	multimedia database systems, information integration. semantic web	Simulation Support Systems, Systems, Types of Simulation, Multimedia Information Systems, Online Information Services, General, Languages and Systems, Database Applications, Nonprocedural languages, INFORMATION INTERFACES AND PRESENTATION (e.g., HCI), INFORMATION STORAGE AND RETRIEVAL, Multimedia databases, DISCRETE MATHEMATICS, Clustering, Content Analysis and Indexing, Information Search and Retrieval, PATTERN RECOGNITION, Query formulation, Processors, Network Protocols, Group and Organization Interfaces,	trust, government, information system, multimedia presentation, multimedia presentations, network, parallel simulation, social network, social networks, storage system, multimedia,	Multimedia databases; Multimedia presentations, information integration; meta-search; relational algebra; early returns, temporal constraint reasoning, multimedia presentations, collaborative quality assessment, user type detection, user type model, city government, county government, digital government, information system, youth services, city government, co-design, collaboration, communication, cooperative design, county government, digital government, information system, participatory design, user-centered design, youth services
Sibel Tari	Computer vision, Differential Equations, Variational methods and axis morphology	Applications, Design Methodology, Segmentation, Vision and Scene Understanding, Feature Measurement, Miscellaneous, Multimedia Information Systems	classification, context, model, similarity	level set methods; PDEs and variational methods; curve evolution; shape representation; shape decomposition

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
Sumalatha Adabala	Computer Systems, Computational Grids, Data Placement, Security	PERFORMANCE OF SYSTEMS, Large and Medium ("Mainframe") Computers, Local and Wide-Area Networks, SOFTWARE ENGINEERING, Super (very large) computers, COMPUTER-COMMUNICATION NETWORKS, Security and Protection, Systems and Software, Information Search and Retrieval,	computing, security issues, security	access control, grid environments, grid security, Unix access model
Tolga Can	Bioinformatics, Statistical Methods	Nonnumerical Algorithms and Problems, Information Search and Retrieval, LIFE AND MEDICAL SCIENCES, Design Methodology	classification, context, marketing, network, random walk, random walks, similarity, video sequences	protein classification; SCOP; decision trees, bag-of-features, copy detection, keyframe extraction, story and media tracking, tree approach
Tugrul Dayar	Performance modelling and analysis, scientific computing, bioinformatics, computer networks.	Numerical Algorithms and Problems, Modes of Computation, PROBABILITY AND STATISTICS, Numerical Linear Algebra, Local and Wide-Area Networks	service, network, queueing network, queueing networks	Markov chains; Kronecker products; reordering; grouping; lumping; block iterative methods; multilevel methods; preconditioned projection methods; Multilevel methods; multigrid; aggregation-disaggregation; Markov chains; Kronecker-based numerical techniques, Closed queueing network; Phase-type service distribution; Kronecker representation; Network

Akademisyenin İsmi	Bilinen İlgili Alanları	Linked Data Tabanlı Semantik Web, ACM ve IEEE ile	Başlık Tabanlı Arama	Linked Data Tabanlı Semantik Web, DBLP ++ ile
				decomposition; Fixed-point iteration; Multilevel method,
Zehra Cataltepe	Machine Learning	Learning, General, Knowledge Representation Formalisms and Methods	classification, model	Feature Selection; Stable Feature Selection; Stability; MRMR (Minimum Redundancy Maximum Relevance), Co-training; Random Subspace Methods; RASCO; mRMR, Biological Sequence Analysis; Optimal and Information Theoretic Syntactic Classification; Peptide Classification; Sequence Processing; Syntactic Pattern Recognition

Ek 2. Aktör Ontolojisi,

<http://www.aktors.org/ontology/portal>

Aşağıda aktör ontolojisi için yararlanılan özellikler (property) bulunmaktadır:

RESEARCH AREAS

```
<owl:Class rdf:ID="Generic-Area-Of-Interest">
  <rdfs:comment>A generic class to specify generic areas for research
or business initiatives. For instance, the area in which a project is
situated</rdfs:comment>
  <rdfs:subClassOf rdf:resource="&support;Intangible-Thing"/>
  <rdfs:isDefinedBy rdf:resource="&base;"/>
</owl:Class>
```

AUTHOR

```
<owl:ObjectProperty rdf:ID="has-author">
  <rdfs:domain>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <owl:Class rdf:about="#Information-Bearing-Object"/>
        <owl:Class rdf:about="#Publication-Reference"/>
        <owl:Class rdf:about="#Technology"/>
        <owl:Class rdf:about="#Method"/>
      </owl:unionOf>
    </owl:Class>
  </rdfs:domain>
  <rdfs:range rdf:resource="#Generic-Agent"/>
  <rdfs:isDefinedBy rdf:resource="&base;"/>
</owl:ObjectProperty>
```

TITLE

```
<owl:DatatypeProperty rdf:ID="has-title">
  <rdfs:subPropertyOf rdf:resource="&support;has-pretty-name"/>
  <rdfs:isDefinedBy rdf:resource="&base;"/>
</owl:DatatypeProperty>
```

FULL NAME

```
<owl:DatatypeProperty rdf:ID="full-name">
  <rdfs:domain rdf:resource="#Person"/>
```