

T.C.
PAMUKKALE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

DAĞITIK VERİ KAYNAKLARI İÇİN BİRLEŞİK
SORGULARIN OLUŞTURULMASI

YÜKSEK LİSANS TEZİ

AFRA ADAK

DENİZLİ, TEMMUZ - 2024

T.C.
PAMUKKALE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI



DAĞITIK VERİ KAYNAKLARI İÇİN BİRLEŞİK
SORGULARIN OLUŞTURULMASI

YÜKSEK LİSANS TEZİ

AFRA ADAK

DENİZLİ, TEMMUZ - 2024

Bu tezin tasarımı, hazırlanması, yürütülmesi, araştırmalarının yapılması ve bulgularının analizlerinde bilimsel etiğe ve akademik kurallara özenle riayet edildiğini; bu çalışmanın doğrudan birincil ürünü olmayan bulguların, verilerin ve materyallerin bilimsel etiğe uygun olarak kaynak gösterildiğini ve alıntı yapılan çalışmalara atfedildiğine beyan ederim.

AFRA ADAK

ÖZET

DAĞITIK VERİ KAYNAKLARI İÇİN BİRLEŞİK SORGULARIN OLUŞTURULMASI

YÜKSEK LİSANS TEZİ

AFRA ADAK

PAMUKKALE ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

(TEZ DANIŞMANI:DR. ÖGR. ÜY. FATMANA ŞENTÜRK)

DENİZLİ, TEMMUZ - 2024

Ontolojiler, bir alan ile ilgili özellikleri, bu alana ait örnek verileri ve bu özellikler arasındaki ilişkileri açıklayan üst verilerdir. Bilgi gösterimi, doğal dil işleme, bilgi getirme, veritabanları, bilgi yönetimi, veritabanı entegrasyonu, bilgi dönüşümü, vb. gibi birçok alan için, otomatik ya da yarı otomatik olarak çeşitli ontolojiler oluşturulup kullanılmaktadırlar. Farklı alanlar için oluşturulan ontolojilerin birlikte çalışılabilir olması, veri esnekliğinin sağlanması gibi pek çok problem ortaya çıkmaktadır. Bu problemlerin ortadan kaldırılması amacıyla, dağıtık bir hiyerarşide bulunan ontolojilerin (veriler dâhil) birlikte çalışması için dağıtık sorgulama mimarisinin geliştirilmesi hedeflenmiştir. Bu mimari için, dağıtık ontolojilerin ya da veri kaynaklarının birbirleri ile entegre çalışması ve optimum sonuncun son kullanıcıya iletilmesi amacıyla dağıtık sorgular yazılmıştır. Dağıtık sorgular, DbPedia, Wikidata ve Drugbank gibi farklı veri platformları üzerinde test edilmiştir. Böylece her birinden verilerin çekilip, sonrasında istenilen raporlar için birleştirme ve filtreleme algoritmaları kullanılması yerine, tek bir sorgu ile istenilen raporlara ulaşılması sağlanmıştır. Geliştirilen sorgular sayesinde hem veri trafiği azaltılmış hem verilerin ayrı ayrı filtreler uygulanarak birleştirilmesi maliyetinden kaçınılmıştır. Ek olarak dağıtık sorgular aracılığıyla hızlı sonuç elde edilmiştir.

ANAHTAR KELİMELER: Anlamsal Web, Dağıtık Sorgu Motorları, Ontoloji, SPARQL

ABSTRACT

GENERATING FEDERATED QUERIES FOR DISTRIBUTED DATA SOURCES

MSC THESIS

AFRA ADAK

**PAMUKKALE UNIVERSITY INSTITUTE OF SCIENCE
COMPUTER ENGINEERING**

(SUPERVISOR:ASST. PROF. FATMANA ŞENTÜRK)

DENİZLİ, JULY 2024

Ontologies are metadata that describe features related to a field, sample data belonging to this field, and the relationships between these features. They are used in information representation, natural language processing, information retrieval, databases, information management, database integration, information transformation, and other fields. Various ontologies are created and used automatically or semi-automatically for many fields. However, there are many problems that arise, such as the interoperability of ontologies created for different fields and ensuring data flexibility. In order to eliminate these problems, it is aimed to develop a distributed query architecture for ontologies (including data) in a distributed hierarchy to work together. This architecture employs distributed queries, which integrate distributed ontologies or data sources with each other and deliver the optimum result to the end user. Distributed queries have been tested on different data platforms such as DbPedia, Wikidata, and Drugbank. Consequently, instead of extracting data from each source and then applying merging and filtering algorithms to generate the desired reports, the desired reports can be accessed with a single query. The development of these queries has resulted in a reduction in data traffic and the avoidance of the cost associated with applying separate filters to combine data. Furthermore, the generation of results has been expedited through the use of distributed queries.

KEYWORDS: Semantic Web, Distributed Query Engines, Ontology, SPARQL

İÇİNDEKİLER

Sayfa

ÖZET.....	i
ABSTRACT	ii
İÇİNDEKİLER	iii
ŞEKİL LİSTESİ.....	v
SEMBOL LİSTESİ	vi
ÖNSÖZ.....	vii
1. GİRİŞ.....	1
2. ONTOLOJİ.....	4
2.1 Bilgi Entegrasyonu	5
2.2 Bilgi Çıkarma ve Anlamlandırma	6
2.3 Sağlık Bilgi Sistemleri.....	7
2.4 E-ticaret ve İş Zekası	8
2.5 Robotik ve Otomasyon	9
2.6 Büyük Veri Yönetimi	10
3. ONTOLOJİLERİN TEMSİL EDİLDİĞİ DİLLER	13
3.1 URI	14
3.2 XML	14
3.3 RDF	14
3.4 RDFS	16
3.5 OWL	17
3.6 OWL Web Ontoloji Dilleri.....	18
3.6.1 OWL-Lite.....	18
3.6.2 OWL-DL.....	18
3.6.3 OWL-Full.....	18
3.7 Ontoloji Uygulamaları.....	19
3.7.1 Eşleme.....	19
3.7.2 Birleştirme	21
3.7.2.1 Nested Loop Join	21
3.7.2.2 Hash Join.....	21
3.7.2.3 Bind Join	22
4. MİMARİ.....	23
4.1 SPARQL.....	23
4.2 Dağıtık Sorgu	26
4.2.1 Dağıtık Sorgulama Motorlarının Mimari Yapısı	26
4.2.2 Dağıtık Sorgulama Motoru Seçerken Dikkate Almanız Gereken Faktörler.....	26
4.3 Dağıtık Sorgulama Motorları	27
4.3.1 DAW (Duplicated Aware)	27
4.3.2 FEDX	27
4.3.3 DARQ	28
4.3.4 Splendid	28
4.3.5 Ontario	28
4.3.6 Metafactory.....	29
4.3.7 ColChain	29
4.3.8 Comunica	30

4.4	Dağıtık Sorgu Yapmanın Adımları	30
4.5	Dağıtık Sorgulama Motorlarının Hızı	30
5.	TEST.....	32
5.1	Data Set	32
5.2	Dağıtık Sorgu Örnekleri	33
5.2.1	Sorgu 1	33
5.2.2	Sorgu 2	36
5.2.3	Sorgu 3	40
5.2.4	Sorgu 4	45
5.2.5	Sorgu 5	48
6.	SONUÇ VE ÖNERİLER	52
7.	KAYNAKLAR.....	54
8.	ÖZGEÇMİŞ.....	60

ŞEKİL LİSTESİ

Sayfa

Şekil 3.1: Tim Berners-Lee'nin Anlamsal Web Katmanları (Akyokuş 2024).	13
Şekil 3.2: Örnek Bir XML Dokümanı (W3Schools 2024).	15
Şekil 3.3: Örnek Rdf Dokümanı (Wikidata query service 2024).	16
Şekil 3.4: Örnek Rdfs Dokümanı (W3Schools 2024).	17
Şekil 4.1: Örnek Sparql Sorgusu (Dbpedia 2024).	24
Şekil 4.2: Distinct İfadesinin Kullanımı.	25
Şekil 4.3: Filter İfadesinin Kullanımı.	25
Şekil 5.1: Belirli İlacın Ticari Adını ve Atılım Yolunu Getiren Sorgu.	33
Şekil 5.2: Sorgu 1 İçin Veri İlişkileri Diyagramı.	35
Şekil 5.3: Sorgu 1 Sonucunda Elde Edilen Veriler.	36
Şekil 5.4: Obesite Hastalarında Covid-19 İle Birlikte Kullanılan İlaçların Bilgisini Getiren Sorgu.	37
Şekil 5.5: Sorgu 2 İçin Veri İlişkileri Diyagramı.	40
Şekil 5.6: Sorgu 2 Sonucunda Elde Edilen Veriler.	40
Şekil 5.7: Belirli Bir Hastalıkta Covid-19 ile Beraber Uygulanan Tedaviler ve Kullanılan İlaçların İçeriği ve Diğer İlaçlar ile Etkileşimini Getiren Sorgu.	41
Şekil 5.8: Sorgu 3 İçin Veri İlişkileri Diyagramı.	44
Şekil 5.9: Sorgu 3 Sonucunda Elde Edilen Veriler.	44
Şekil 5.10: Herhangi Bir İlacın Drugbank Bilgisi ve Ticari Bilgilerini Görüntüleyen Sorgu.	45
Şekil 5.11: Sorgu 4 İçin Veri İlişkileri Diyagramı.	47
Şekil 5.12: Sorgu 4 Sonucunda Elde Edilen Veriler.	48
Şekil 5.13: İlaçları ve Bu İlaçların Ticari İsimlerini Birlikte Getiren Sorgu.	49
Şekil 5.14: Sorgu 5 İçin Veri İlişkileri Diyagramı.	50
Şekil 5.15: Sorgu 5 Sonucunda Elde Edilen Veriler.	51

SEMBOL LİSTESİ

CSV	:	Comma Separated Values
HTTP	:	Hypertext Transfer Protocol
IOT	:	Internet of Things
JSON	:	JavaScript Object Notation
OWL	:	Web Ontology Language
R2RL	:	RDB to RDF Mapping Language
RDF	:	Resource Description Framework
RDFS	:	RDF Schema
RML	:	Redline Markup Language
SPARQL	:	Anlamsal web için bir sorgulama dili
UML	:	Unified Modeling Language
URI	:	Uniform Resource Identifier
URL	:	Uniform Resource Loader
W3C	:	World Wide Web Consortium
WWW	:	World Wide Web
XML	:	Extensible Markup Language
XSD	:	XML Schema Definition

ÖNSÖZ

Tez çalışmam boyunca bilgi ve tecrübelerini benimle paylaşan, bana her konuda sonsuz yardımcı olan değerli danışmanım Dr.Öğr. Üyesi Fatmana ŞENTÜRK'e ve Pamukkale Üniversitesi Bilgisayar mühendisliğindeki hocalarıma,

Tez yazma sürecimde bana her konuda kolaylık sağlayan okul müdürüme ve destek olan öğretmen arkadaşlarıma,

Bundan önce olduğu gibi tez yazma sürecimde de bana destek olan, dostum Funda KIRCA AKSEN'e,

Hayatım boyunca hep yanımda oldukları için annem Zeynep ŞAHİN, babam Şahabettin ŞAHİN ve abim Akın ŞAHİN'e,

Ve bu sürece yakından şahitlik eden sevgili eşim Fatih ADAK, ilham kaynaklarım, kızlarım Zeynep Mira ADAK ve Eylül ADAK'a gönülden teşekkür ederim.

1. GİRİŞ

İnternet üzerindeki veri miktarı günden güne artarken, bu veri ancak insanlar tarafından anlaşılıp, yorumlanmaktadır. İnternet üzerindeki verilerin çoğalması nedeniyle, verinin sadece insanlar tarafından değil makineler tarafından da hızlı şekilde anlaşılması ve yorumlanabilmesi ihtiyacı ortaya çıkmıştır. Bunun için de bu verileri anlayabilen ve yorumlayabilen modellerin geliştirilmesi gereklidir. Tüm bu ihtiyacı karşılayan yeni bir model, 1999 yılında Tim Barners-Lee tarafından ortaya atılmış ve bu modele “Anlamsal Web” adı verilmiştir (Breitman ve diğ. 2007).

Tim Barners-Lee’ye göre, anlamsal web yeni bir web olmayıp, verilerin üçlü yapılarla tanımlandığı insan ve bilgisayarların her ikisinin de anlamlandırıldığı güncel web’in eklentisi olarak tanımlanmıştır. Tasarlanan semantic web modeli, bilgisayarların verileri anlaması, yorumlaması ve kullanılabilmesi mantığına dayalı olan bir modeldir (Breitman ve diğ. 2007).

Gruber (1993) ontolojiyi “kavramsallaştırmanın açıkça belirtilmesidir.” şeklinde tanımlamış ve bu tanımlama literatürde yaygın olarak kabul görmüştür. Bir ontoloji ilgili alan içerisindeki kavramları ve bunların arasındaki ilişkilerini içeren bir yapıdır (Öztürk 2004).

Anlamsal web’in temelini RDF (Research Description Framework) standardı oluşturur. Bu standart veriyi üçlü (triple) bir yapı olan, özne - yüklem - nesne (Subject - object - predicate) şeklinde tanımlar. Bu yapıda her düğüm özne - yüklem - nesne olarak görev alabilir. Birbiriyle ilişki içinde olan üçlüler arasında, bağlantılar oluşturularak çizge (graph) denilen bir yapı oluşturulmaktadır. Bu şekilde düğümler arası bağlantı tanımlanmış olur.

Anlamsal web, içerisinde bağlı veri yapılarının olduğu bir modeldir. Bu veriler birbirleriyle çeşitli özellikler yönünden ilişkili olan üst verilerdir. Bu üst verileri sağlayan yapı ise ontolojilerdir. Ontolojiler farklı alanların belirlenmiş özelliklerini incelemek için sınıflandırılmıştır. Ontolojiler üst seviye ontolojiler, orta seviye ontolojiler ve alan ontolojisi olarak sınıflandırılabilir. Üst seviye ontolojiler

daha genel kavramları tanımlar, orta seviye ontolojiler belirli bir alandaki kavramları ve ilişkileri tanımlarken alan ontolojisi daha özel konularla ilgili kavramları tanımlamaktadır.

Ontolojilerin bu denli çeşitli olması ve geniş kapsamlı kullanımı farklı zorlukları da beraberinde getirmektedir. Örneğin, bir ontolojide bulunan bir veri, başka bir ontolojide de bulunabilmektedir. Birbirleriyle ilişkili olan ancak farklı ontolojilerde yer alan bu verilerin birbirine bütünleştirilmesi ve birlikte çalışması büyük önem taşımaktadır. Bu kapsamda bu verilerin sorgulanması için, bileşik sorgu motorları kullanılabilir (Schwarte ve diğ. 2011) Bileşik sorgu motorları, ontolojilerin modellendiği graflardaki düğümler üzerinden (bağlı veri) bilgiyi sorgulayıp ve sonucu döndürebilir.

Birleşik sorgu işleme algoritmaları ve bileşik sorgu motorları, kullanıcıların kullandıkları verileri tek bir merkezde toplamadan yani verileri çoğaltmadan dağıtılmış bir şekilde bulunan verilerin kaynaklarına erişim sağlar ve bu veriler üzerinde istenilen sorgulamaların yapılabilmesine olanak tanır. Birleşik sorgu işleme algoritmaları ve motorları, gerçek zamanlı veri toplamayı destekledikleri gibi verileri kendilerine kopyalamadıkları için de veri gizliliğini desteklemektedir. Ayrıca veri yönetimi konusunda da kullanıcılara esneklik sağlar. Ancak bu tür işlemlerin kullanıcılar için erişilebilir ve kullanılabilir olması zorlukları da beraberinde getirmektedir. Bu zorlukların en başında sorgulama yaparken karmaşık sorguların yazılması işlemi gelmektedir. İkinci bir problem, kullanılacak olan veri kaynaklarının istenilen işlemi yaparken hangi kısımlarının kullanılacağını tespit edilmesidir. Üçüncü olarak ise, kullanılacak veri kaynaklarının dağıtık veri sorgulama işlemi sırasında uyumudur. Bu amaç doğrultusunda bu zorlukların üstesinden gelmek için bir yöntem geliştirilmiştir.

Tezin kapsamında dağıtık veri kaynakları kullanılarak birleşik bir sorguların geliştirilmesi amaç edinilmiştir. Bu kapsamda, Dbpedia, Wikidata, Drugbank ...vb gibi farklı fiziksel alanlardaki veri kaynaklarının birbirleriyle sorunsuz bir şekilde haberleşmesi ve yazılan filtrelerle göre verileri süzüp getirmesi hedeflenmiştir. Bu kapsamda, öncelikli olarak tüm bu farklı veri kaynaklarına erişim sağlanmasında ve sorguların çalıştırılmasını sağlayacak olan dağıtık sorgu motoru belirlenmiştir. Bu bağlamda, Sakor ve diğ. (2023) tarafından tanımlanan web ortamında sorgulama

imkânı sađlayan dađıtık sorgulama motoru sečilmiřtir. Diđer benzer alıřmalar ise, veri setinin sabit olması, verilerin boyutlarının az olması ya da sadece kendileri tarafından belirlenen sorguların yazılması gibi gerekelerle sečilmemiřtir. Dađıtık sorgu motoru seçimi yapıldıktan sonraki aşamada SPARQL sorguları oluřturulmuřtur. Oluřturulan sorgularda, Dbpedia, Wikidata, Drugbank, gibi farklı veri kaynaklarından veriler kullanılmıřtır. Yazılan dađıtık sorgular sayesinde, farklı kaynaklardaki verilerin tek tek sorgulanıp, sonrasında bu verilerin birleřtirilmesi iřleminden kaınılmıřtır. Dađıtık sorgulama ile ilgili kaynaktaki sadece önemli veriler ekilmiř, gereksiz olan veriler dâhil edilmemiřtir. Böylece zamanlama ve ek birleřtirme maliyetlerinden kaınılmıřtır.

Tezin geri kalan kısmında, 2. Bölümde Ontolojiler ve uygulama alanları ile ilgili bilgilendirme yapılmıř olup, 3. Bölümde ise; Ontolojilerin yapısı ile ilgili detaylara yer verilmiřtir. 4. Bölümde Dađıtık Sorgulama Mimarilerine yer verilmiř olup, 5. Bölümde geliřtirilen sorgular ele alınmıřtır. Son bölümde ise; bulgulara yer verilmiřtir.

2. ONTOLOJİ

Bilgisayarların hayatımızda yoğun bir şekilde kullanılmasıyla birlikte, elektronik ortamdaki verinin boyutunda ciddi artışlar meydana gelmiştir. Bu durum verinin nasıl saklanacağı sorusuyla beraber, var olan verinin tekrar kullanılmak gerektiğinde erişimin doğru ve hatasız olabilmesini sağlayan gereksinimleri de beraberinde getirmiştir. Zaman içerisinde bu gereksinimlerin ontolojiler aracılığıyla karşılanabileceği görülmüştür.

Ontoloji, felsefede “Varlık Bilimi” anlamına gelmektedir. Ontoloji varlığı soyut ve somut olarak inceler ve sınıflandırır. Bilgisayar bilimlerinde kullanılan ontoloji tanımını (Gruber 1993) “Kavramsallaştırmanın açıkça belirtilmesidir.” olarak tanımlamaktadır. Ontoloji, alandaki kavramları ve kavramların birbiriyle ilişkisini kapsamaktadır.

Ontolojiler 1990’ların başından itibaren bilgisayar bilimleri alanında merak konusu olmaya başlamış ve bu alanda araştırmalar yapılmaya başlanmıştır. Ontolojiler, belirli bir alandaki bilginin bütün özelliklerini ve bu bilgilerin birbirleriyle olan ilişkilerini tanımlar. Ontolojiler (Fisseha 2003);

- Verinin anlamsal gösterimini sağlamaktadır.
- Veri edinme ve veri yönetimi için kavramsal diziler oluşturarak, insanlar ve makineler arasında iletişimi sağlamaktadırlar.
- Tekrar kullanılabilir yapılardır. Bu sayede diğer ontolojileri oluşturmak için veriler kullanılabilir.
- Paylaşılabilir. Bu sayede bilimsel çalışmaların geliştirilmesine destek sağlanmış olur.
- Son zamanlarda ontolojiler akıllı sistemler, elektronik ticaret, doğal dil işleme ve bilgi işleme konularında sıklıkla kullanılmaktadır.

Akıllı sistemlerde kullanılan makine öğrenmesi elimizdeki verilerden tahmin yürütür, ontolojiler ise çıkarım yapar. Örneğin; bir makine öğrenmesi elimizde bulunan bir dizi veri kümesini kullanarak kötü amaçlı e-postaların ortak özelliklerini

belirleyebilir ve yeni gelen e-postanın tehlikeli olup olmayacağını tahmin edebilir. Siber güvenlik alanında kullanılan makine öğrenmesine ontolojileri dâhil ederek daha güçlü ve güvenli sistemler elde edebilir. Ontolojiler yapılarında veri kümesini barındırmaktadır ancak herhangi bir tahminde bulunmazlar. Ontolojiler yapıları gereği, kendi içinde tanımlı verileri ve bu verilerin bağlantılarını kullanarak çıkarımda bulunabilirler. Ontolojiler, bu özelliği ile akıllı sistemlere fayda sağlar. Akıllı sistemlerde kullanılan makine öğrenmesi sayesinde eğitilen veriyi sınıflandırır.

Ontolojiler soyut kavramlar gibi görünmesine rağmen aslında günlük hayatımızda sürekli karşımıza çıkmaktadır. İnternette yaptığımız aramaların anlamlı hale gelip bize bir bütün halinde çıkarılan sonuçlar aslında ontolojilerin sayesinde ortaya çıkan anlamsal aramaların bir sonucudur. Bunun yanında günümüzde sosyal medya uygulamalarında da yine ontolojilerden faydalanılır. Örneğin, Lu ve diğ. (2014), sosyal olarak etiketlenen görsel ve yazılı kelimeler arasındaki korelasyon matrisini öğrenmek için, anlamsal bir analiz yöntemi sunmuşlardır. Benzer şekilde, Sridhar ve Sanagavarapu (2020), önerdikleri anlamsal model sayesinde kullanıcıların tweet içeriklerinin benzerliğine dayalı bir sosyal etkileşim grafiği oluşturmayı amaçlamışlardır.

Akıllı sistemlerde ontolojiler, bilgiyi yapılandırmak, paylaşmak ve anlamak için kullanılan önemli araçlardır. Ontolojiler, bir konu alanındaki kavramları, ilişkileri ve özellikleri tanımlayan biçimsel bir yapıdır ve farklı amaçlar için kullanılabilirler. Kullanıldıkları farklı uygulama alanları bu bölümün kalanında verilmiştir.

2.1 Bilgi Entegrasyonu

Farklı kaynaklardan gelen bilgileri birleştirmek ve entegre etmek için ontolojiler kullanılabilir. Farklı veri formatlarından gelen bilgileri ontoloji üzerinde birleştirmek, veri entegrasyonunu kolaylaştırabilir.

Apaydin ve diğ. (2010), farklı alanlardaki bilgileri içeren anlamsal tabanlı bir model önermişlerdir ve bu modeli deneyimler arasındaki görüntü ilişkilerini

araştırmak için kullanmışlardır. Önerdikleri modele sosyal ağlarda bilgi paylaşımı için Semantic Image Retrieval Model (SIRM) ismini vermişlerdir.

Çevrimiçi platformların eğitimde kullanımı COVID-19'un ortaya çıkmasıyla artmıştır. Bu durum akademik kurumlara, eğitim platformlarına yönetilmesi gereken büyük miktarda veri üretilmesine sebep olmuştur. Ancak birden fazla heterojen kaynaktan gelen, bazen anlamsal tutarsızlıkları olan veriyi yönetmek bazı zorluklar yaratabilir. Bu amaçla semantik web teknolojileri çok kaynaklı e-öğrenme verilerinin semantik entegrasyonu için önemli bir çerçeve olarak ortaya çıkar. Bu bağlamda, ilk kez e-LION (e-Öğrenme Entegrasyon Ontolojisi) semantik modeli, farklı e-öğrenme bilgi tabanlarının veri konsolidasyon yaklaşımı olarak çalışmak üzere önerilmiştir, dolayısıyla üst düzey analizi zenginleştirmeye yol açmıştır (Paneque ve diğ. 2023).

Borowicc ve Souza (2024) çalışmalarında heterojen veri bütünleştirmesinde uygulanan ana teknikleri vurgulayan bir literatür kapsamı incelemesi sunar. İncelemede tekniklerin eksiksiz bir bütünleştirme sürecinde nasıl kullanılabileceğine dair net bir anlayış yerine, çoğunlukla bütünleştirme sürecinin belli bir amacına ve bir kısmına odaklanan çözümler sundukları görülmüştür. Bu nedenle çalışmada teknikleri işlevselliklerine göre düzenleyen bir veri bütünleştirme sürecinin bütün halini gösterir. Ayrıca eşleme ve eşleştirme görevleri gibi semantik uygulamalara odaklanan kaynaklarla ve tekniklerle bir iş akışı sunar.

2.2 Bilgi Çıkarma ve Anlamlandırma

Ontolojiler, bilgisayarların metinleri anlamasına yardımcı olabilir. Metin madenciliği ve doğal dil işleme uygulamalarında, ontolojiler kullanılarak belirli kavramlar, ilişkiler ve bağlamlar ortaya çıkarılabilir.

Doğal dil sorularının doğal dil işleme yöntemleri ile anlamlandırılması ve anlamlandırılan soruların SPARQL sorgu dili ile sorgularının iletilmesi ve DBpedia servisi üzerinden SPARQL sorgularının işletilmesi ve sonuçların web arayüzünde gösterilmesi üzerine bir çalışma yapılmıştır (Kabakuş ve Doğdu 2003).

Birçok sektörde yaşanan kritik sorunlardan birisi büyük ölçekte üretilen veri miktarı ve bu verilerin entegrasyonu problemidir. Veriler farklı formatlar ve yazılımlar kullanılarak kalıcı hale getirilir. Bu amaçla (Almeida ve diğ. 2023) veri entegrasyonu için semantik iş birliğine dayanan INSIDE’i önerirler. INSIDE her bir hizmetin bir ontoloji kullanılarak tanımlandığı veri hizmetleri kavramları aracılığıyla bir veya daha fazla veritabanına yapılan sorguları temsil eder.

İlaç belgelerindeki yapılandırılmamış veriden otomatik bilgi çıkarmak, ilaç keşfi ve geliştirilmesi açısından önemlidir. Bu amaçla özel olarak geliştirilmiş bir ilaç ontolojisi, bağlamsallaştırma algoritmaları ve bir BioBERT modeline uygun uçtan uca bir bilgi çıkarma çerçevesi önermişlerdir (Mann ve diğ. 2023). Önerdikleri çerçeve SUSIE (Şema Tabanlı Denetimsiz Anlamsal Bilgi Çıkarımı) yapılandırılmamış metinden önemli varlıkları ve ilişkileri belirlemek aynı zamanda yapılandırılmış bir biçimde önemli bilgileri temsil eden bilgi grafiklerin otomatik olarak oluşturmak için ICH(Uluslararası Uyum Konferansı) belgeleri üzerinde eğitildi. Çalışmada belirli görevler için eğitim veri kümelerinin manuel olarak düzenlenmesine gerek kalmadan alan, otomatik olarak bilgi çıkarma çerçevesi sunmaktadır.

2.3 Sağlık Bilgi Sistemleri

Sağlık sektöründe, ontolojiler hastalıklar, ilaçlar, hastalar ve tedavi yöntemleri gibi konuları tanımlayarak bilgi paylaşımını ve anlamını kolaylaştırabilir. Bu, hastaların sağlık verilerinin etkili bir şekilde yönetilmesine ve paylaşılmasına yardımcı olabilir. Hastaneler ve klinikler verileri çeşitli formatlarda veritabanlarında saklamaktadır. Bu veritabanları ve ontolojilerin birlikte kullanımı sağlık sektörü açısından hızlı ve anlamlı çözümler üretilmesine yardımcı olabilir.

Kiong ve diğ. geliştirdikleri sağlık ontolojisi sisteminde bu verileri kavramlar ve semantiklerle, makinelerin bunları yorumlamasını sağlamak için ortak bir kümülatif ontoloji şeklinde entegre etmek için bir araç sağlar.(Kiong ve diğ. 2011)

Sakor ve diğ.(2023) COVID-19 ile ilgili net bir şekilde belirtilmemiş tedavileri tanımlayan bir platform(framework) oluşturmuşlardır. Bu platformda

COVID-19 tedavileri boyunca kullanılan ilaçların yan etiklerini (ilaç-ilaç etkileşimlerini) bulmak için DrugBank, CORD-19, vb. gibi farklı kaynakları birbirine entegre etmişlerdir. Entegrasyon sırasında doğal dil işleme yöntemleri ve makine öğrenmesi yöntemlerini kullanarak KG (Knowledge Graph) modelini oluşturmuşlar ve bu model üzerinden ilaç- ilaç etkileşimlerini bulmayı amaçlamışlardır. Son olarak, bu etkileşimleri kullanarak muhtemel tedavileri belirlemişlerdir (Sakor ve diğ. 2023).

Tıbbi alanda her bir doktor hastalarıyla ilgili verileri okuyup yazarak kapsamlı bilgiye sahip olmalıdır. Bu kapsamda her gün artan çok sayıda belge nedeniyle ilaçlar, hastalıklar ve hastalar ile ilgili tüm bilgileri okumak çok yorucudur. Sonuç olarak çok sayıda hata meydana gelmektedir. Bu sorunu ele alabilecek alan, bilgi çıkarmadır. Bu alanda, doğal dilde yazılmış yapılandırılmamış metinden önemli ve istenen bilgileri çıkarmak için birkaç yöntem vardır. Ana yöntem, ilgili bilgileri çıkararak metni yapılandırabildikleri için varlık tanıma ve ilişki çıkarma olarak adlandırılır. Bu kapsamda yararlı bilgi ve özellikleri çıkarsamak için bazı yöntemleri kullanmışlardır. (Landolsi ve diğ. 2023).

BT tabanlı sağlık uygulamalarının kullanımının yaygınlaşması ile birlikte büyük hacimlerde çeşitli sağlık verisi biçimleri oluşturulmaktadır. Bu veriler yapılandırılmamış, yarı yapılandırılmış, yapılandırılmış veriler olmakla beraber büyük verinin tüm özelliklerine sahiptir. Bu veriler NoSQL veritabanları vasıtasıyla depolanmaktadır. İlişkisel veritabanlarının aksine, NoSQL veritabanı tasarımı için standart yöntemler veya araçlar yoktur. Sen ve Mukherjee (2023) ontoloji tabanlı bir şema tasarım yaklaşımı benimsimişlerdir. Bu yaklaşımda birincil sağlık hizmetleri için bir ontoloji açıklamışlardır.

2.4 E-ticaret ve İş Zekası

E-ticaret platformlarında, ontolojiler ürün kategorileri, özellikleri ve ilişkileri tanımlayarak daha iyi ürün önerileri sunmaya yardımcı olabilir. İş zekası uygulamalarında ise ontolojiler, iş verilerini daha anlamlı bir şekilde organize etmek ve analiz etmek için kullanılır.

Yücebaş ve arkadaşları e-ticaret sitelerinde kullanıcıların kendi sorgularını girebilecekleri bir ortam hazırlamayı amaçlamışlar, bu anlamda girilen sözcüklerin metin analizi yöntemiyle analiz edilmesi için ontolojilerden faydalanmışlardır (Yücebaş ve diğ. 2018).

Akıllı şehir uygulamaları geliştirmesi yaygınlaşan bir kavram olmakla birlikte birçok zorlukla karşılaşılan bir alandır. Bunlar çeşitli veri kaynaklarına entegre edilmesi, fiziksel kentsel çevreyi yansıtan coğrafi verilerin dahil edilmesi gibi karmaşık gereksinimlerin ele alınması gerekir. Akıllı şehirler için tasarlanan platformlar, geliştiriciler tarafından yeniden kullanılabilen üst düzey hizmet sundukları için uygulamaları somutlaştırmada önemli bir role sahiptir.

Bu amaçla Lympers ve Goumopoulos (2023) etkili veri analizi ve entegrasyonu için semantik olarak zenginleştirilmiş bir veri modelini desteklemek üzere coğrafi bilgiler de dâhil olmak üzere çeşitli veri kaynaklarına dayalı akıllı uygulamalar geliştirmek için bir platform olan SEDIA tanıtmışlardır. Platform toplanan verileri anlamsal olarak açıklamak için ontoloji sınıflarını ve özelliklerini kullanır. Bu araştırma ayrıca SEDIA'nın performans değerlendirmesini gösteren deneysel veriler sunar.

2.5 Robotik ve Otomasyon

Ontolojiler, robotların çevrelerini daha iyi anlamalarına ve insanlarla etkileşim kurmalarına yardımcı olabilir. Bu, endüstriyel otomasyon, akıllı ev sistemleri veya hizmet robotları gibi alanlarda kullanılır. Bayraktar ve diğerleri şehirleşmenin çok arttığı günümüzde farklı kaynaklardan üretilen verinin kontrolünü sağlamak önem arz ettiğini ifade etmişlerdir. Şehirler her geçen gün artan altyapı, konut, temizlik, su, sağlık, istihdam, güvenlik ve ulaşım gibi konularda ortaya çıkacak talepler ile başa çıkmak zorundadır. Tüm bunların çözümü olarak akıllı şehirler fikri ortaya çıkmıştır. Farklı kaynaklardan gelen verinin birbiriyle ilişkisi tespit edilip aralarında bağlantı kurmak yoluyla büyük veri yönetimine ihtiyaç duyulacağını ifade etmişlerdir (Bayraktar ve diğ. 2012).

Son yıllarda mimari açıdan karmaşık binaların sayısının artması nedeniyle mobil bina denetim sistemlerinin giderek daha karmaşık ortamlarda çalışması gerekiyor. Bina denetim robotlarının tasarımına ilişkin mevcut literatür, robot donanım tasarım gereksinimlerini tanımlamada bina ortamlarının ve bina kusurlarının etkisinden dolayı olarak bahsedilmiştir. Ancak belirli bir donanım gereksinimini tanımlamak için hangi bilgilerin gerekli olduğu açık bir şekilde gösterilmesi gerekmektedir. Bu gereksinimi karşılamak amacıyla Leyuan ve Timo (2023) robot donanım tasarım gereksinimlerinin belirlenmesini etkileyen bina ve denetim alanı nesnelere genel bir bakış sağlayan bir ontoloji sunmaktadır.

2.6 Büyük Veri Yönetimi

Ontolojiler, büyük veri setlerini düzenlemek, sorgulamak ve anlamak için kullanılabilir. Veri entegrasyonunu kolaylaştırmak ve büyük veri analitiği yapmak için ontolojiler önemli bir rol oynar. Bağlantılı Veri Teknolojileri web üzerinde yapılmış veriyi yayınlamak için RDF dilini ve HTTP protokolünü kullanır. Birçok alanda kullanılan verilerin Bağlantılı Veri İlkelerine göre yayınlanmasının yanı sıra çok sayıda kitle kaynaklı girişim gerçekleştirilmiştir. Bunların başında DBpedia gelir. (DBpedia 2024) DBpedia, Wikidata'dan yapısal bilginin çıkarılmasıyla oluşturulan bağlantılı veri setlerinden biridir. Bir diğer bağlantılı veri girişimi ise, GeoNames(GeoNames 2024)'dir. GeoNames, içerisinde yer adlarını bulunduran bir dizindir. GeoNames, 8 milyon konuma ait bağlantılı veriyi yayınlamış olan açık lisanslı konumsal veri tabanıdır.

Veri kalitesi, başarılı bir analiz için veriye ihtiyaç duyulduğu gerçeğini düşündüğümüzde yapay zekâ çalışmalarının merkezinde yer almalıdır. Kalite düzeyi ve değerlendirilmesi veriyi kullanılmasına karar verilmesi açısından önemlidir. Bu önemin farkında olan bilim ve akademi son yıllarda veri kalitesi ölçümleri ve değerlendirme çerçeveleri önermiştir. Fakat veri kalitesi terimleri için ortak ve paylaşılan bir kelime dağarcığı yoktur. Bu nedenle yapay zeka görevlerini gerçekleştirmek için bir veri iş akışına, veri kalitesi analizini entegre etmek zor ve zaman alıcıdır. Bunun başlıca nedenlerinden biri daha az sayıda öneri dışında sunulan kelime dağarcıklarının ne makine tarafından ne de işlenebilir olması ve

sisteme entegre edilmesi için insana ihtiyaç duymasdır. González ve diğ. (2024) yaptıkları çalışmada birleşik bir veri kalitesi ölçümü ve değerlendirme bilgi modeli önermektedir. Bu modelde birlikte çalışabilir ve makine tarafından okunabilir hale getirmek için ontoloji olarak geliştirilmiştir.

Kulkami ve diğ.(2023) yaptıkları çalışmada heterojen veri kaynaklarından veri almak ve yönetmek için bir yöntem geliştirmişlerdir. Bangalore'un çeşitli hükümet sektörlerindeki verileri barındırmak ve analiz etmek için semantik tabanlı bir veri gölü çerçevesi oluşturmaya katkısını ele almışlardır. Çalışmada ontoloji tabanlı bir büyük veri yönetim sistemi, dağıtılmış heterojen bir veri depolaması ile aramanın aldığı şey arasındaki boşluğu kapatmak için bir belge alma çözümünü açıklamaktadırlar.

Arora ve Sharma (2023) yapay zekânın ve büyük verinin günümüz dünyasında ne kadar kritik ve hayati olduğunu ifade etmişlerdir. Çalışmalarında ayrıca dünya çapındaki işletmelerin çok sektörlü senaryolarında yapay zekâ ve büyük veri uygulamalarının önemini vurgulamak için nitel ve teorik perspektifleri keşfetmeyi amaçlamaktadırlar. Bu amaçla modern işletmelerde yapay zekâ ile birleştirilen büyük verilerin faydalarını keşfetmeyi ve yapay zekâ uygulamalarının günümüzdeki işletmelerin çok sektörlü uygulamalarındaki önemini vurgulamışlardır. NVivo yazılımıyla birlikte bibliyometrik analiz kullanarak duygu analizi yapılan bu çalışmada, yapay zekâ ve büyük veri işletmelerde birleştirildiğinde ne olacağına dair bir anlayış geliştirmeye çalışmaktadır.

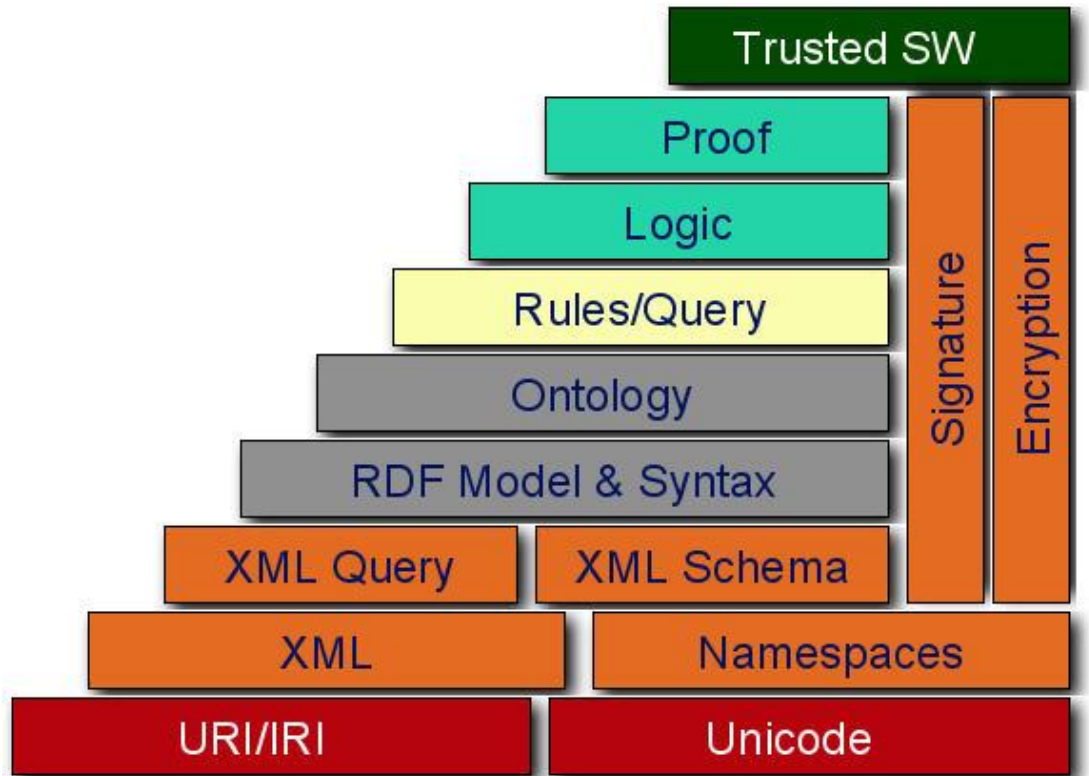
Anlamsal zekâ teknolojileri, makinaların kaynakları bağlamsal ve akıllı bir şekilde oldukları için yapay zekâlı bilgi tabanlı sistemlerin geliştirilmesinde en önemli bileşendir. Dağıtılmış, kurumsal ve web tabanlı bilgi sistemlerinde anlamsal zekâyı gizemden arındırmayı amaçlamaktadır. Son 20 yılda bilişim teknolojilerindeki ilerlemeyle IOT cihazlarından gelen veri akışını ele almak için yeni yaklaşımlar geliştirildi. Yazılım endüstrisi, IOT ağları tarafından üretilen veri hacmi ve veri depolama yöntemleri, veri hızı ve işleme hızı ve farklı protokoller aracılığıyla iletilen (yapılandırılmamış) veri çeşitliliği ve standartların benimsenmesi ihtiyacı dâhil olmak üzere, IOT ağlarından üretilen verilerin farklı yönleri (boyutları) ile ilgili sorunları ele alan çözümler bulmakla ilgilenmektedir. Bu amaçla büyük veri ve IOT

teknolojilerinin birleřtirilmesi ve makine öğrenmesindeki gelişmelerle birlikte karmaşık sistemler için hizmet geliştirme fırsatları yarattı (Janev 2021).

3. ONTOLOJİLERİN TEMSİL EDİLDİĞİ DİLLER

Ontolojiler, anlamsal web'in temelini oluşturur. Web 3.0 olarak ifade ettiğimiz çağ anlamsal web çağı olarak kabul edilir. Web 2.0'dan sonra, internetteki veri miktarı artmış, arama motorları sayesinde aranılan bir bilgi rahatlıkla bulunur hale gelmiştir. Ancak bu veri insanın anlamlandırabildiği, makinelerin anlamlandıramadığı bir veridir. Verilerin makineler tarafından anlamlandırılabilmesi için anlamsal ağ yapısı ortaya çıkmıştır (Borner-Lee ve Hendler 2001).

Ontolojiler, Semantic web uygulamalarının temelini oluşturmaktadır ve RDF (Resource Description Framework) ve OWL (Web Ontology Language), XML gibi standartlar kullanılarak ontolojiler tanımlanabilmektedir. Bu tanımlama katmanlı bir mimariye sahiptir ve en tabanında URI ve Unicode bulunmaktadır. Sonrasında biraz daha gelişmiş versiyonu olan XML yapısı bu katmanlı yapıya yapı taşı olarak eklenir. Üzerine RDF, RDFS ve OWL gibi farklı yapı taşları eklenerek semantic web mimarisi oluşmaktadır. Tüm bu katmanlı mimari Şekil 3.1'de gösterilmiştir.



Şekil 3.1: Tim Berners-Lee'nin Anlamsal Web Katmanları (Akyokuş 2024).

3.1 URI

Uniform Resource Identifier'in kısaltılmış hali olan URI, internet üzerinde bir kaynağın tam yerine işaret eden (resim veya belge) standart formata uygun bir karakter dizisidir. Kısaca bir URL'nin altında bulunan kaynağın tam yoluna işaret eder.

3.2 XML

Genişletilebilir işaretleme dili (XML) verileri paylaşılabilir bir şekilde tanımlamamız, saklamamız ve transferi için olanak tanımaktadır. W3C tarafından tanımlanmış bir standarttır. Bir programlama dili değildir. Herhangi bir algoritmaya dayanmamaktadır.

XML'in avantajları;

- Herhangi bir bilgisayar sisteminde kullanılabilir
- Belgeleri kolayca okuyabilir.
- Unicodu'yu destekler.
- Genişletilebilirdir.
- Verileri taşır.
- Platform değiştirme sürecini kolaylaştırır.

Şekil 3.2'de bir kitap evinde çocuk kategorisinde ve web kategorisinde bulunan kitapların özellikleri (kategorisi, kitabın adı, yazarı, basım yılı, ücreti) XML dilinde nasıl yazıldığı görülmektedir.

3.3 RDF

RDF, XML üzerine oturtulmuş bir standarttır. World Wide Web (WWW) tarafından geliştirilmiştir (Koivunen ve Miller 2001). Özne- Yükleme- Nesne üçlülerinden oluşan RDF belgeleri XML kullanılarak ifade edilir. (Chong ve diğ. 2005).

RDF verileri aynı zamanda üç sütunlu (özne-yüklem-nesne) bir tablo olarak da düşünülebilir. RDF'deki konu belirli bir nesneye ilişkin veri öğelerinin birden fazla sütuna yerleştirildiği, benzersiz alanlardan oluşan bir veritabanına benzer. RDF'deki bu alanlar aynı konuyu, genellikle aynı benzersiz anahtarı paylaşan ayrı yüklem/nesne satırları olarak temsil edilir; yüklem sütun adına, nesne ise gerçek verilere benzer.

```
<?xml version="1.0" encoding="UTF-8"?>

<bookstore>

  <book category="children">

    <title lang="en">Harry Potter</title>

    <author>J K. Rowling</author>

    <year>2005</year>

    <price>29.99</price>

  </book>

  <book category="web">

    <title lang="en">Learning XML</title>

    <author>Erik T. Ray</author>

    <year>2003</year>

    <price>39.95</price>

  </book>

</bookstore>
```

Şekil 3.2: Örnek Bir XML Dokümanı (W3Schools 2024).

RML, RDF haritalama dili olarak ifade edilir. CSV, XML, JSON ve ilişkisel tablolar gibi çeşitli formatlarda temsil edilen heterojen veri kaynaklarını yönetmek için W3C standardı eşleme dili R2RML'yi genişletir. RML üçlü haritaları olarak adlandırılan bu kurallar, RDF sınıflarının örneklerini ve özelliklerini mantıksal bir kaynak açısından tanımlar (Schwarteve diğ. 2001).

Şekil 3.3'te Tony Benn ile ilgili bir Wikipedia makalesi RDF yapısında ifade edilmiştir.

```
<rdf:RDF

xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntaxns#"

xmlns:dc="http://purl.org/dc/elements/1.1/">

<rdf: Description

rdf: about="http://en.wikipedia.org/wiki/Tony_Benn">

    <dc: title>Tony Benn</dc: title>

    <dc: publisher>Wikipedia</dc: publisher>

</rdf: Description>
```

Şekil 3.3: Örnek Rdf Dokümanı (Wikidata query service 2024).

3.4 RDFS

RDFS, veri modellemek için kullanılır. RDF veri modelini genişleterek, alanda kullanılacak sözcük kümesini nesneler, nesneler arası ilişkiler, özellikler ve özelliklerin alabileceği değerler açısından tanımlamaktadır.

```
<?xml version="1.0" ?>

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xml:base="http://www.cicekler.fake/cicekler#" > <rdf:Description rdf:ID="Çiçek" >

<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class" />

</rdf:Description>

<rdf:Description rdf:ID="Papatya" >

<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class" />

<rdfs:subClassOf rdf:resource="#Çiçek" />

</rdf:Description> </rdf:RDF>
```

Şekil 3.4: Örnek Rdfs Dokümanı (W3Schools 2024).

Örnekte “Papatya, bir Çiçek’tir” ifadesi modellenmiştir. Bu ifadede Çiçek ve Papatya varlık adları olup, bu varlıklar arasında kalıtım (is-a) ilişkisi vardır (Keskin ve Sezer 2009).

3.5 OWL

Anlamsal web temelini ontolojiler oluşturmaktadır, OWL yani Web Ontoloji Dili, tanımlamalar ve veriler arasında iletişim için kullanılır (Çankaya 2007). OWL, RDF/XML sözdizim temeline dayanan bir dildir (Pascal ve diğ. 2012).

OWL, özellikleri ve sınıfları tanımlayabileceğimiz daha geniş bir alt yapı sağlamaktadır. Sınıfların özellikleri tanımlanarak terimlerin açıklamalarını içerir. Özellikler, sınıflara ait bileşenlerin birbiri ile olan ilişkilerini belirtmek için kullanılmaktadır. Ontolojilerde çeşitli özellikler ve bu kavramlar arasında anlamsal ilişkiler tanımlanmaktadır. OWL, XML tabanlı olduğu için, OWL bilgisi kolayca farklı işletim sistemlerine sahip bilgisayarlar ve farklı programlama dilleri arasında iletimi sağlanabilir. OWL, değer yönetimi çerçevesi, büyük ölçekli entegrasyon ve ağ üzerinde veri paylaşımı için standart yaratmada kullanılmaktadır.

3.6 OWL Web Ontoloji Dilleri

OWL, W3C tarafından tanımlanmış dil olup, gösterim şekline göre OWL Lite, OWL DL ve OWL Full olmak üzere üç türü vardır (Davies ve diğ. 2004).

3.6.1 OWL-Lite

OWL ailesine ait bir alt dildir. OWL'nin ifadesel olarak basit ve sınırlı sürümüdür. Sadece basit sınıf (kavram) sıra düzenleri, daha az karmaşık ontoloji yapıları, basit kısıtların ihtiyaç duyulduğu durumlarda kullanılabilir.

Sınırlı ifade edilmesi sayesinde daha hızlı ve kolayca işlenebilir. Bu özelliği sayesinde verimin önemli olduğu olduğu uygulamalarda kolaylık sağlar.

3.6.2 OWL-DL

OWL-DL (Web Ontology Language Description Logic) OWL dilinin mantıksal kuralları üzerine geliştirdiği dilidir. Üst seviye anlamsallığa sahip olan OWL-DL çıkarımları hesaplayabilir ve belli bir zamanda bitebilir olması yönünden OWL-Lite göre daha güçlüdür. Bu yönüyle semantik uygulamalarda daha fazla kullanılır. Karmaşık web ontolojilerinde, akademik çalışmalarda, semantik web uygulamalarında kullanılır.

3.6.3 OWL-Full

OWL dil ailesinin ifadesel olarak en güçlü ve esnek sürümüdür. RDF ve RDFS ile uyumludur. OWL -Lite ve OWL-DL'ye göre daha fazla ifade gücü bulunur. Bu da daha güçlü bir dil olmasını sağlar ancak işlem sonunda hesaplanabilirlik seviyesini zorlaştırır. Bu OWL-Full'un en önemli dezavantajıdır.

Karmaşık ontolojiler, esneklik gerektiren ontolojiler, araştırma projelerinde sıklıkla kullanılır.

RDF üzerindeki veriyi OWL katmanı üzerindeki sınıfları ve yüklemeleri sorgulamak için SPARQL dili geliştirilmiştir.

Bunlar üzerine bilgisayarların anlamlandırma, yorumlama yapabildiği Mantık (Logic) katmanı bulunmaktadır. Bu katmanların hepsini dikey olarak kesen İmzalama (Signature) ve Şifreleme (Encription) katmanları ile güvenlik unsurları yer almaktadır.

3.7 Ontoloji Uygulamaları

Ontolojiler içinde bir alan ile örnek verileri, bu alan ilgili özellikleri(öznitelikleri), ve bu özellikler arasındaki bağlantıları açıklayan yapılardır. Ontolojiler, doğal dil işleme, bilginin gösterilmesi, bilginin getirilmesi, bilginin yönetilmesi, dönüşümü vb. gibi pek çok farklı alanda kullanılmaktadır (Kolli 2008). Bu bölümünün devamında ontolojilerin farklı kullanım amaçları irdelenmiştir.

3.7.1 Eşleme

Kullanım alanlarının geniş olması sebebiyle çok farklı amaçlar için ontolojiler oluşturulmuştur. Ancak, ontolojilerin oluşturulması işlemi zaman zaman tekrarlanabilmekte olup, yani aynı işlem için farklı ontolojiler tanımlanması gibi bir durum ortaya çıkmıştır. Hem kullanıcılar hem de yazılım geliştiricileri için zaman kaybına sebep olabilecek bu durumu çözmek için ontolojilerin birbirleri ile benzer olup olmadığının ya da ne kadar benzer olduğunun belirlenmesi gereklidir (Huang ve diğ. 2012). Bir ontolojideki kavramların veya örneklerin, başka bir ontolojide hangi yapıya karşılık geldiğini bulmak için ontoloji eşleme algoritmaları kullanılmaktadır. Ontoloji eşleme algoritmaları; en çok veri entegrasi, ontolojilerin birleştirilmesi ve veritabanı üzerindeki verilerin açıklaması ya da yorumlanması gibi işlemler için kullanılmaktadır (Ivanova ve diğ. 2017; Lambrix ve diğ. 2016; Zablith ve diğ. 2015).

Ontolojiler üst seviye, orta seviye, alan ontolojileri olarak üç grupta toplanabilir. Üst seviye ontolojiler, farklı kaynaklardan elde edilen heterojen verinin birleştirilmesidir. Daha genel kavramları ve ilişkileri kapsar.

- SUMO (Suggested Upper Merged Ontology): Fiziksel nesneler, süreçler, ilişkiler gibi çok geniş ve genel kavramları ve ilişkileri barındırır (Niles ve Pease 2001).

- DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering): Varlıklar, olaylar, nitelikler gibi genel kategorileri içerir (Gangemi ve diğ. 2001).

Orta seviye ontolojiler belirli bir alan içindeki kavram ve ilişkileri kapsar. Üst seviye ontolojilerdeki soyut kavramların somut gösterilmiş halidir. Orta seviye ontolojilere aşağıdaki örnekler verilebilir.

- SWEET Ontoloji (Semantic Web for Earth and Environmental Terminology): Çevre bilimleri ve Dünya sistemleri hakkında bilgi sağlayan bir ontolojidir. Çevresel araştırmalar ve uygulamalar için kullanılır. İklim, atmosfer, okyanuslar, kara yüzeyleri ve çevresel süreçler hakkında kavramlar bulunur (Raskin ve Pam 2005).
- Biomedical Ontolojisi (OBO Faoundry): Biyomedikal araştırmalar için bir dizi orta seviye ontoloji sağlar. Bu ontolojiler, biyolojik ve tıbbi kavramları kapsar. Genler, hastalıklar, biyolojik süreçler ve fenotipler gibi kategorileri içerir (Smith ve diğ. 2007).

Alan ontolojisi, belirli alana ait özel kavramları ve kavramların ilişkilerini gösterir. Aynı kavramlar birçok alanda varken, kavramların gösterimi alan içeriklerinin farklılıklarından dolayı değişmektedir. Alan ontolojisine Gene ontolojisi ve FOAF ontolojisi örnek verilebilir.

- Gene Ontolojisi (GO): Genetik araştırmalar için, genlerin ve onların ürünlerinin fonksiyonlarını tanımlar (Ashburner ve diğ. 2000).

- FOAF (Friend of a Friend): Sosyal ağlardaki bireyler ve onların ilişkileri hakkında bilgileri bulundur (Brickley ve Miller 2000).

Alan ontolojisi, üst ve orta seviye ontolojilerine kavramların eklenmesiyle genişleyebilir.

Orta seviye ontolojiler, alan ontolojileri içinde belirlenen basit alt seviyede alana özel kavramlar ile üst seviye ontolojilerinde tanımlı soyut kavramlar arasında bir bağlantı görevi görmektedir. Herhangi bir seviye arasında ontoloji eşlemesi gerçekleştirilebilir.

3.7.2 Birleştirme

Dağıtık sorgular yazılırken, yazılan sorgular verilerin kaynaklarına göre alt parçalara bölünebilir. Bu alt sorgularda dağıtık sorgu motoru tarafından ilgili kaynaklara yönlendirilir. Ancak alt sorgular zaman zaman aynı kaynaklara yönlendirilebilirler ve dağıtık sorgulama motorları yönlendirme sonuçlarını tek bir sonuç olarak göstermek adına birleştirebilir.

Alt sorgu sonuçlarının hızlı bir şekilde birleştirilebilmesi için birleştirme algoritmasının seçimi önemlidir. Birleştirme işlemi için özel yöntemler tasarlanabileceği gibi, bu amaç için sıklıkla kullanılan yöntemlerde bulunmaktadır. Bu yöntemler arasında nested loop join, hash join ve bind join de vardır (Saleem 2023).

3.7.2.1 Nested Loop Join

İç içe yazılan sorgularda öncelikli olarak en iç sıradaki çalıştırılır ve bir sonraki sorgu ile birleştirilir (Rakhmawati ve diğ. 2013). Yani birleştirme işlemi sırasında tüm veriler için ayrı ayrı filtreleme mekanizması işletilir ve en son olarak elde edilen tüm veriler birleştirilir. Kısaca sorgulanan tüm veri kaynaklarına aynı anda erişim sağlandığı ve büyük çaplı veriler eş zamanlı işlendiği için performans bakımından uygun değildir.

3.7.2.2 Hash Join

Bu birleştirme algoritması, ilk olarak Görlitz ve Staab (2011) tarafından önerilmiştir. Bu algoritmanın temel mantığı yazılan SPARQL sorguları alt sorgu

parçalarına ayrıştırıldıktan sonra elde edilen bu alt parçalar paralel olarak çalıştırılmaktadır. Elde edilen sonuçlar ise, sorguyu çalıştıran sistemin ya da kişinin yerel sisteminde birleştirilmektedir. Eğer sistem üzerinde sorgulanan ve dönen veri yüksek boyutlu değilse, sorgu hızının iyi olduğu görülmüştür. Ancak sistem üzerinde sorgulanan ve dönen veri boyutu yüksek ise; ağ üzerindeki iletim maliyetlerinden dolayı sorgu hızının düştüğü belirtilmiştir (Rakhmawati ve diğ. 2013).

3.7.2.3 Bind Join

Haas ve diğ. tarafından (1997) tanıtılan Bind Join algoritması, alt sorguları sıra ile çalıştırmaktadır. Alt sorgulara ayrıştırılan SPARQL sorgunun ilk çalıştırılan parçası, ikinci sırada çalıştırılan sorguya filtre olarak eklenir. Dolayısıyla doğrudan veri seti kısıtlanarak birleştirme ve ağ iletim maliyeti azaltılır. Ancak paralel olarak sorguların çalıştırılması mümkün değildir. Bunun sonucunda çalışma süresinin uzadığı fark edilmiştir (Rakhmawati ve diğ. 2013).

4. MİMARİ

SPARQL ontolojiler üzerinde sorgulama yapabilmek için geliştirilmiş bir RDF sorgu dilidir. Sorgu ifade etmek için kullanılan bir protokoldür. SPARQL standardı W3C tarafından tasarlanıp onaylanmıştır. Veriler RDF olarak depolanabilir. Özne-Yüklem-Nesne üçlülerini kullanarak sorgulama yapar. SPARQL gerekli ve isteğe bağlı grafik modellerini sorgulama yeteneğine sahiptir. Kullanıcıların bir veritabanının nasıl organize olduğundan ziyade ne bilmek istediklerine odaklanmaları için yardımcı olur.

4.1 SPARQL

SPARQL dağıtık veri kaynaklarını sorgulayabilmek için de kullanılan bir dildir. Bu sebeple öncelikli olarak SPARQL kullanarak sorgulamaların nasıl yapılabileceğine dair örnekler anlatılmıştır.

```
PREFIX dbpedia: <http://dbpedia.org/resource/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?person ?name ?birthDate ?place
WHERE {
    ?person a foaf:Person .
    ?person foaf:name ?name .
    ?person dbo:birthPlace dbpedia:Turkey .
    ?person dbo:birthPlace ?place .
    ?person dbo:birthDate ?birthDate .
}
ORDER BY ?birthDate
LIMIT 100
```

Şekil 4.1: Örnek Sparql Sorgusu (Dbpedia 2024).

SPARQL sorgu dili temelde aşağıdaki sorgu sözcüklerini barındırır.

- SPARQL, RDF verileri için tanımlanmış SQL benzeri bir sorgu dilidir.
- PREFIX ön ek bildirimidir. Burada belirtilen bir kaynak URL (Uniform Resource Locator)'sidir. Fakat bu kaynak uzun olduğu ve diğer kaynaklarla benzer ilk kısımlar olabileceğinden ilk kısımdan tanımlama yapılır. Şekil 4.1'de örnek bir Sparql sorgusu görülmektedir.
- SELECT içinde nelerin dönmesini istediğimizi (? ile başlayan her bir yapı bir değişkeni ifade eder) belirtir.
- Sorgu sonucunda kişi, isim, doğum tarihi, doğum yeri ekranda görüntülenecektir.
- WHERE kısmındaki her iki. (nokta) arasında bir üçlü bulunur. Bu üçlü tanımları değişken (? ile başlar) içerebilir. Üçlü tanımlar bir web kaynağını (URI, unified resource identifier) içerebilir. Veya eğer içeriği string olarak ifade edilebilen bir türden veri var ise bunu da içerebilir. Bu türden tanıma örnek olarak "20080101"^^xsd:date verilebilir. Burada string olarak verilen kısımdan sonra, hangi türden bir ifade yapıldığı ^^ işaretinden sonra gelen bir kaynak ile ifade

edilir. Sorgu çalıştırıldığında çizge üzerinde bu tanımlamalara uyan üçlüler seçiliyor. Bu üçlülerden SELECT içinde bulunan değişkenler seçilip döndürülüyor.

- Şekil 4.1'teki sorguda WHERE ifadesiyle Dbpedia'dan alınan veriler içinden Türkiye'de doğmuş olan kişiler sorgulanmıştır.
- DISTINCT ifadesi sorgu sonucunda tekrar eden sorgu sonuçlarının bir kez yazılması için kullanılır. Şekil 4.1'deki sorguyu şekil 4.2'deki gibi güncellersek sorgu sonucunda tekrar eden veri bir kez yazılacak.

```
SELECT DISTINCT ?person
```

Şekil 4.2: Distinct İfadesinin Kullanımı.

- FILTER ihtiyacınıza bağlı olarak sorguya filtre ekleyebilirsiniz.

Şekil 4.3'teki sorguya FILTER eklenirse ismi “Evliya Çelebi” olanları listeler.

```
FILTER(?name in ("Evliya Çelebi"@en))
```

Şekil 4.3: Filter İfadesinin Kullanımı.

- ORDER BY: Sorgu sonucu istediğiniz değişken üzerinden sıralar. ?birthDate verileri doğum tarihlerine göre sıralar.
- ASC: Veriyi küçükten büyüğe sıralama yapmayı sağlar.
- DESC: Veriyi büyükten küçüğe sıralama yapmaya yarar.
- OFFSET: Belirli sayıda sonucun atlanmasına ve sonucun bir sonraki değerden başlayarak gösterilmesine olanak sağlar.
- OPTIONAL: Belirtilen özellik varsa sonucu getirir, yoksa boş döndürür.
- LIMIT: Belirli sayıda sonuç elde etmek için kullanılır.
- CONSTRUCT: Sorgu sonuçlarına göre yeni bir RDF grafiği oluşturur.
- DESCRIBE: Veri kaynağının istatistiklerini sorgular.

4.2 Dağıtık Sorgu

Dağıtık sorgular büyük veri bulutlarının veya dağıtık veri depolarının birbiriyle eş zamanlı olarak sorgulanıp hızlı sonuç almak amacıyla tasarlanmış yöntemlerdir. Bir veri kümesindeki bir varlığın, başka bir veri kümesindeki varlığa referans vererek iki farklı veri kümesi arasında bağlantı kurarak tek bir sorguda sorgulamaya imkân tanır.

Bağlı verinin etkin bir şekilde sorgulanabilmesi için dağıtık sorgulama motorları kullanılmıştır. Bu motorlar kullanılarak tek bir nokta üzerinden bağlı veri bulutu üzerinde bulunan RDF verisi sorgulanabilir.

4.2.1 Dağıtık Sorgulama Motorlarının Mimari Yapısı

İstemci Arabirimi (Client Interface), Dağıtık Sorgu Motoru (Distributed Query Engine), Dağıtık Veri Tabanı: (Distributed Engine) ve Veri Parçalama (Data Sharding) dağıtık bir sistemde olması gereken faktörlerdir (Tanenbaum ve Steen 2007, s.4).

4.2.2 Dağıtık Sorgulama Motoru Seçerken Dikkate Almanız Gereken Faktörler

- Performans: Sorgu motorunun hızı ve büyük veri miktarı karşısındaki uyumu önemlidir, performansı olarak değerlendirilir.
- Ölçeklenebilirlik: Zamanla artan veri miktarına karşı sorgu motorunun kendini uyarlaması olarak ifade edilebilir.
- Yüksek kullanılabilirlik: Veri kaybının engellenmesi ve veriler arasında kesintilerin olmaması önemlidir.
- Açıklık
- Kaynakların kolay erişilebilir olması

4.3 Dağıtık Sorgulama Motorları

Bu bölümde dağıtık sorgulama motorlarından örneklere yer verilmiştir.

4.3.1 DAW (Duplicated Aware)

Son yıllarda artan veri ağı, çeşitli ortamlardan gelen verinin bir ortamda saklanması ihtiyacını doğurmuştur. Ancak bu verilerin tekrar eden verilerden oluştuğu gözlenmiştir (Saleem ve diğ. 2013). DAW sorgulama motoruyla, tekrar eden verinin sorgu hızına etkisini incelemiştir. DAW min-wise kombinasyonu ve kompak veri özetleri mantığına dayanır. Minimum düzeyde bağımsız permütasyonların ve farklı kaynakların sonuçları arasındaki örtüşmeyi tahmin etmek için üçlü seçicilik bilgisi kullanır. Bu bilgi, veri kaynaklarını sıralamak için kullanılır.

DAW;

- Sorgu yürütme sürelerini iyileştirir.
- Geri çağırmanın değişken düzeyde olduğunu gösterir.
- Sabit bir sayıyı geri çağırmayı en üst düzeye çıkarmak için uygundur.

4.3.2 FEDX

Schwarte ve arkadaşları, heterojen veri kaynakları üzerindeki SPARQL sorgularının verimli bir şekilde işlenmesini sağlayan ve aynı zamanda önerilen çerçevenin bir dizi gerçek dünya sorgusu üzerinde uygulanabilirliğini ve verimliliğini gösteren FedX adında bir çerçeve önerdi(Schwarte ve diğ. 2011).

Sesame üzerinde geliştirilmiş SPARQL 1.0 üzerinde çalışan bir dağıtık sorgulama motorudur. VOID (zhao ve diğ. 2009) istatistik bilgilerini ve ASK sorgularını kullanarak veri kaynağı eleme işlemini gerçekleştirir. ASK sorgularının maliyetinin azaltmak için sonuçlarını ön bellekte tutar ve bu sonuçları gelecekte

kullanabilir. FEDX aracında alt sorguların sonuçları elde edildikten sonra maliyet hesaplaması yapılarak, maliyete göre birleşme sırası belirlenmektedir.

4.3.3 DARQ

Quilitz ve Leser DARQ adında bir birleşik sorgu motoru önermektedir. (Quilitz ve Leser 2008) DARQ sorgu işleme süresini ve bant genişliği kullanımını azaltmak için hizmet açıklamalarını, sorgu yeniden yazma mekanizmalarını ve maliyete dayalı bir optimizasyon yaklaşımını birleştirir.

4.3.4 Splendid

Splendid (Görlitz ve Staab 2011) sesame SPARQL sorgu dili üzerine geliştirilmiş bir dağıtık sorgulama aracıdır. SPARQL 1.0 üzerine dağıtık sorgulama yapılmasını desteklemektedir. SPARQL sorgularıyla veri kaynağı seçimi için VOID açıklamalarını, ASK sorgularını kullanır. Splendid'in ana bileşenleri sorgu ayrıştırıcısı, sorgu optimizasyonu ve sorgu yürütücüsüdür. Veri kaynağı seçimi ve birleşme alanında aşağıdaki iyileştirmeleri önermiştir.

Veri kaynağı seçimi ön bellekte bulunmayan üçlüler için ASK sorgusu oluşturarak veri kaynağı seçme işlemi yapmaktadır.

Bkz. bölüm 3.7.2 deki üç farklı birleşme yöntemi yaygın olarak kullanılmaktadır. Splendid dağıtık sorgulama motoru küçük veri kaynakları için “Hash Join” büyük veri kaynakları için “Bind Join” kullanmaktadır. (Görlitz ve Staab 2011)

4.3.5 Ontario

Veri gölü, büyük miktarda ham veriyi orijinal biçimde tutan büyük boyutlu bir depolama havuzudur. Veri gölü(Data Lake) derinlemesine analiz isteyenler için idealdir (Endris 2019).

Ontario, büyük ölçekli heterojen veri için uyarlanmış birleşik sorgu işleme yaklaşımıdır. Mevcut SPARQL yaklaşımı karmaşıklık getirir, ontario bu karmaşıklığı ortadan kaldırmayı amaçlar.

Bir SPARQL sorgusu verildiğinde, Ontario semantik veri gölündeki RDF şablonlarıyla eşleşen bir dizi yıldız şekilli alt sorgu oluşturur. Farklı şekildeki yıldız sorguları ayırt edebilir ve uygunluğuna karar verebilir. Ontario ayrıca yürütme süresini azaltabilen ve yanıt eksiksizliğini artırabilen gür ağaç planlarını da tanımlar; Ontario, yıldız şeklindeki gruplar düzeyinde tanımlanan bir dizi buluşsal yöntem ve bunların yürütüleceği veri motorları tarafından yönlendirilir.

4.3.6 Metafactory

İki OULAD ve üç MUD SPARQL uç noktasına bağlanan bir FEDX federasyonu ile bağlanmıştır. Birden çok verilerin entegrasyonu için grafik veri modelleri kullanır. Farklı veri modellerinde birleştirici bir veri modeli olarak grafikler kullanılır. Bilgi temsil katmanını kullanarak anlamsal entegrasyon sağlar. Grafikleri sorgulamak için SPARQL kullanır (Schwarte ve diğ. 2011).

4.3.7 ColChain

Bakış açısını Semantik web'in verimli kullanılmamasının, verilerin güncel olmaması ve bazen de erişim problemlerinin yaşanması sebebine dayandıran ColChain mimarisi bu problemleri çözmeyi hedeflemektedir (Aebeloe ve diğ. 2021). Hedeflenen mimaride verilerin her zaman erişilebilir olması, veri sağlayıcıların yükünü azaltan ve kullanıcıların verileri güncelleme yeteneğine sahip olması hedeflenmiştir. Mimaride verileri güncelleme imkânıyla birlikte güncellenmiş verileri kökenlerine kadar izleme, güncellenmemiş veriye sorgu yapma imkanı sağlaması da hedeflenmiştir.

4.3.8 Comunica

TPF arayüzleri için optimize edilmiş federasyon algoritmasıdır. Sorguyu üçlü desenlere bölmek için Bind join algoritmasını kullanır. Heterojen arayüzleri ve SPARQL uç noktalarını önemsiz olarak farzeder. Yapılan çalışmalar Comunica algoritmasının karmaşık sorgular ve birçok kaynak için beklenenden daha iyi sonuçlar alındığı görülmektedir (Taelman ve diğ. 2018).

4.4 Dağıtık Sorgu Yapmanın Adımları

Verinin Dağıtılması ve Depolanması: İlk olarak büyük veri kümelerini dağıtık sistemlere dağıtarak veya dağıtık veri tabanlarına yükleyerek kullanılacak ontolojileri hazırlamakla ilgili olan kısımdır.

Sorgu Gönderme ve Dağıtık İşleme: Sorgular, genellikle bir istemci tarafından veri depolama sistemi üzerinde gönderilir. Dağıtık sorgulama motoru bu sorguyu alır genellikle birden fazla düğümde eş zamanlı olarak işlemeye başlar. Bu düğümler verinin depolandığı ve işlendiği birimlerdir.

Paralel İşleme ve Veri Akışı: Dağıtık sorgu motoru, aldığı sorguyu veri parçalarına bölerek paralel olarak işler. Her bir parça farklı düğümlerde veya veri bölümlerinde işlenir. İşlenen sonuçlar daha sonra birleştirilir ve istemciye gönderilir.

Sonuçların Toplanması ve İşlenmesi: İşlenen sonuçlar genellikle dağıtık sorgu motoru tarafından toplanır ve birleştirilir. Daha sonra istemciye geri döndürülmeden gereksiz veriler temizlenir veya istenmeyen veriler filtrelendir.

Sonuçların İstemciye Dönmesi: Son olarak işlenen ve birleştirilen sonuçlar istemciye döndürülür. İstemci bu sonuçları kullanıcıya sunar veya başka işlemler için kullanır.

4.5 Dağıtık Sorgulama Motorlarının Hızı

Dağıtık sorgulama motorlarının hızını aşağıdaki faktörler etkiler.

- Sorgu motorunun veri kümesi seçimi
- Alt sorguların mümkün olduğunca ilişkili veri kümesi seçimini etkin yapması
- Sorgular için etkin işletim planı oluşturulmalı
- Alt sorgulardan dönen cevapların en az sayıda HTTP isteği oluşturacak şekilde birleştirilmesi
- Önbellekleme yöntemi ile sorgu işletim maliyetinin ortadan kaldırılması

5. TEST

5.1 Data Set

Knowledge4Covid19 çerçevesi Covid19 hastalığı tedavi süreçleri ve beraberinde bulunan hastalıkla uygulanan tedavi adımlarını inceler uygulanan ilaçlar arasındaki etkileşimleri incelemek amacıyla ilgili bilgi kaynaklarını entegre ederek birleşik sonuçlar almamızı sağlayan bir çerçevedir.

Knowledge4Covid-19 çerçevesi bir veri ekosistemi olarak tasarlanmıştır. Bu sistemde meta veriler tanımlanır. Bu verileri DBpedia, Wikidata, Drugbank'daki kaynaklara bağlar.

Drugbank (Drugbank 2024) veri kaynağından aşağıdaki veriler alınmıştır.

Drug Interactions,

- Adverse Effects
 - ✓ Adverse Effect Drugs
 - ✓ Adverse Effect Condition
 - ✓ Adverse Effect Incidences
 - ✓ Adverse Effect Attributes
- Drug Indications
- Drug Allergies

5.2 Dağıtık Sorgu Örnekleri

5.2.1 Sorgu 1

Şekil 5.1'deki sorgu belirli bir ilaç (DB01234) ile ilgili COVID-19 ve komorbidite (bir kişide aynı anda birden fazla hastalığın bulunması) tedavileri hakkında bilgi sağlayarak, ilaçların vücuttan atılımı, ticari adları ve ilgili tedavi bilgilerini ortaya çıkarmayı amaçlamaktadır.

```
PREFIX db: <http://dbpedia.org/>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX k4covid: <http://research.tib.eu/covid-19/vocab/>

PREFIX k4covid: <http://research.tib.eu/covid-19/entity/>

PREFIX dbo: <http://dbpedia.org/ontology/>

PREFIX dbp: <http://dbpedia.org/property/>

SELECT DISTINCT ?dbDrugBank ?treatment ?sameAsCovidDrug ?comorbidityDrug
?excretion ?tradeName

WHERE {
  SERVICE <https://labs.tib.eu/sdm/covid19kg/sparql> {
    ?treatment k4covid:hasCovidDrug ?covidDrug.
    ?treatment k4covid:hasComorbidityDrug ?comorbidityDrug.
    ?comorbidityDrug k4covid:hasCUIAnnotation ?CUIComorbidityDrug.
    ?CUIComorbidityDrug owl:sameAs ?sameAsComorbidityDrug .
    ?covidDrug k4covid:hasCUIAnnotation ?CUICovidDrug.
    ?CUICovidDrug owl:sameAs ?sameAsCovidDrug .
  }
  SERVICE <https://dbpedia.org/sparql> {
    ?sameAsCovidDrug dbp:excretion ?excretion.
    ?sameAsCovidDrug dbo:drugbank ?dbDrugBank .
    ?sameAsCovidDrug dbp:tradenname ?tradeName.
  }
  FILTER (?dbDrugBank = "DB01234")
} ORDER BY ?dbDrugBank
```

Şekil 0.5: Belirli İlacın Ticari Adını ve Atılım Yolunu Getiren Sorgu.

PREFIX db: <<http://dbpedia.org/>>: Bu PREFIX ön eki DBpedia’da genel kaynaklara referans vermek için kullanılır.

PREFIX owl: <<http://www.w3.org/2002/07/owl#>>:Bu ön ek Web Ontology Language(OWL) bulunan terimlere referans vermek için kullanılır.

PREFIX k4covid: <<http://research.tib.eu/covid-19/vocab/>> :Bu ön ek Covid hastalığı ile ilgili özel olarak tanımlanmış kelime dağarcığını ifade eder.

PREFIX k4covid: <<http://research.tib.eu/covid-19/entity/>>:Bu ön ek Covid hastalığı ile ilgili tanımlanmış veri kümesini ifade eder.

PREFIX dbo: <<http://dbpedia.org/ontology/>>:Bu ön ek DBpedia ontolojinde bulunan sınıflara referans vermek için kullanılır.

PREFIX dbp: <<http://dbpedia.org/property/>>:Bu ön ek DBpedia’da özelliklere referans vermek için kullanılır.

SELECT DISTINCT ?dbDrugBank ?treatment ?sameAsCovidDrug
?comorbidityDrug ?excretion ?tradeName

Şekil 5.1’deki ifade drugbank’dan alınan ilaç ID’si, uygulanan tedavi kodları, eşlik eden hastalık için kullanılan ilaçlar, dbDrugBank’daki ID’si belirtilen ilacın atılım yolu ve ticari adı görüntülenmektedir.

WHERE : Veri kümesinden hangi verilerin çekileceği bu kısımda oluşturulur.

SERVICE <<https://labs.tib.eu/sdm/covid19kg/sparql>>: İfadesi sorgu esnasında başka bir SPARQL uç noktasına sorgu yapılmasını sağlar.Burada Covid-19 hastalığı ile ilgili verilerin olduğu bir RDF formatında uç noktaların bulunduğu bilgi grafiğidir.Dağıtık sorgu yapılabilmesini sağlayan ifadedir.

?treatment k4covid:hasCovidDrug ?covidDrug. : İfadesi tedavilerde kullanılan covid ilaçlarını belirtir.

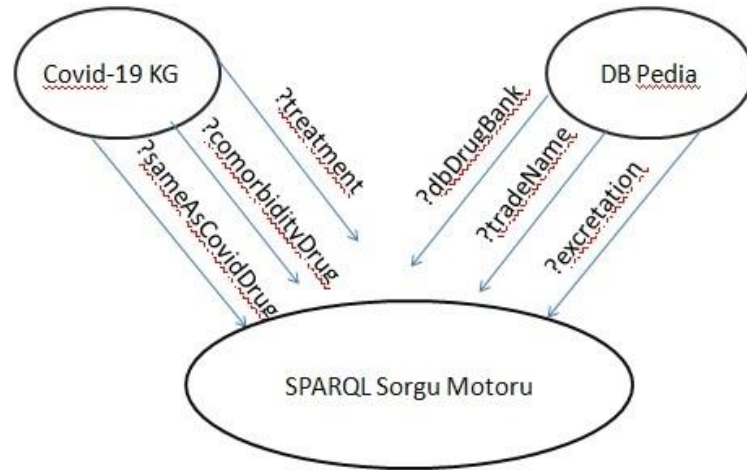
?treatment k4covid:hasComorbidityDrug ?comorbidityDrug. İfadesi uygulanan farklı tedavilerde Covid ile beraber bulunan hastalık için kullanılan ilaçları ifade eder.

?comorbidityDrug k4covid:hasCUIAnnotation ?CUIComorbidityDrug.

?CUIComorbidityDrug owl:sameAs ?sameAsComorbidityDrug ifadeleri eşlik eden hastalığın arasında CUIAnnotation ilişkisi olduğunu ifade eder. Bir başka ifade olan “sameAs” ifadesi ?CUIComorbidityDrug ve ?sameAsComorbidityDrug ifadelerinin aynı şeyleri ifade ettiğini belirtir.

DBpedia’den alınan ilaç bilgisi FILTER (?dbDrugBank = “DB01234”) atılım hızı ve ticari adı ve beraberinde bulunan hastalık için kullanılan ilaçları bu tedavilerin numaraları dağıttık sorgu ile tek bir sorgu ifadesi ile çekilmesi amaçlanmıştır. Burada ?sameAsCovidDrug ifadesi iki farklı veri kaynağı arasında ilişki kurulmasını sağlayan ifadedir. Bu ifade dağıttık sorgu yapılmasına olanak sağlar.

Sorgu 1 için, hangi kaynaktan hangi ilgili verilerin getirildiğini gösteren veri ilişkileri diyagramı Şekil 5.2’de verilmiştir.



Şekil 5.6: Sorgu 1 İçin Veri İlişkileri Diyagramı.

Sorgu sonucunda dbDrugBank= “DB01234” ilacının hangi tedavilerde kullanıldığını, atılım yolu, ticari adı gibi bilgilere kolaylıkla ulaşabiliyoruz. Bu sayede iki hastalığın

bir arada tedavi edilmesi gereken durumlarda ilaçlarla ilgili bilgilere erişmekte kolaylık sağlanır.

Sorgu sonucunda 246 sonuç 1.949 saniyede elde edilmiştir.

dbDrugBank	treatment	sameAsCovidDrug	comorbidityDrug	excretion	tradeName
1 DB01234	k4covid:Treatment_18	db.resource:Dexamethasone	k4covid:DB00440	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
2 DB01234	k4covid:Treatment_18	db.resource:Dexamethasone	k4covid:DB00584	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
3 DB01234	k4covid:Treatment_19	db.resource:Dexamethasone	k4covid:DB00584	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
4 DB01234	k4covid:Treatment_24	db.resource:Dexamethasone	k4covid:DB00584	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
5 DB01234	k4covid:Treatment_26	db.resource:Dexamethasone	k4covid:DB00584	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
6 DB01234	k4covid:Treatment_18	db.resource:Dexamethasone	k4covid:DB00338	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
7 DB01234	k4covid:Treatment_19	db.resource:Dexamethasone	k4covid:DB00338	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
8 DB01234	k4covid:Treatment_20	db.resource:Dexamethasone	k4covid:DB00338	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
9 DB01234	k4covid:Treatment_21	db.resource:Dexamethasone	k4covid:DB00338	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
10 DB01234	k4covid:Treatment_23	db.resource:Dexamethasone	k4covid:DB00338	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
11 DB01234	k4covid:Treatment_24	db.resource:Dexamethasone	k4covid:DB00338	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
12 DB01234	k4covid:Treatment_25	db.resource:Dexamethasone	k4covid:DB00338	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
13 DB01234	k4covid:Treatment_26	db.resource:Dexamethasone	k4covid:DB00338	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
14 DB01234	k4covid:Treatment_18	db.resource:Dexamethasone	k4covid:DB01001	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
15 DB01234	k4covid:Treatment_18	db.resource:Dexamethasone	k4covid:DB00437	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
16 DB01234	k4covid:Treatment_19	db.resource:Dexamethasone	k4covid:DB00437	db.resource:Urine	Decadron, Ozurdex, Dexycu, others
17 DB01234	k4covid:Treatment_20	db.resource:Dexamethasone	k4covid:DB00437	db.resource:Urine	Decadron, Ozurdex, Dexycu, others

Şekil 0.7: Sorgu 1 Sonucunda Elde Edilen Veriler.

5.2.2 Sorgu 2

Şekil 5.4’te verilen sorgu obezite hastalığında kullanılan ilaçlar, bu ilaçların atılım yolu, ilaçların uygulanma yolları ve ilaçların prospektüs bilgisini farklı iki ontoloji kaynağından çekerek tek bir sorguda listelenmesini sağlamaktadır.

PREFIX wd: <http://www.wikidata.org/entity/>: Bu ön ek PREFIX anahtar sözcüğüyle tanımlanmış isim alanı “wd” isimlendirilmiştir. Bu sayede kod yazarken kolaylık sağlanır. Bu adres wikidata’da bulunan veri kümesine referans vermek için kullanılır.

PREFIX wdt: <http://www.wikidata.org/prop/direct/>: Bu ön ek PREFIX anahtar sözcüğüyle wdt isim alanı ile isimlendirilmiştir. Bu sayede daha kısa ve okunabilir kod yazmamızı sağlar. Wikidata’daki doğrudan özelliklere referans vermek için kullanılır.

```

PREFIX wd: <http://www.wikidata.org/entity/>

PREFIX wdt: <http://www.wikidata.org/prop/direct/>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX dbp: <http://dbpedia.org/property/>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX k4covid: <http://research.tib.eu/covid-19/vocab/>

PREFIX k4covid: <http://research.tib.eu/covid-19/entity/>

SELECT DISTINCT ?comorbidity ?sameAsComorbidityDrug ?excretion ?routes
?prospectus WHERE {

    SERVICE <https://labs.tib.eu/sdm/covid19kg/sparql> {

        ?treatment k4covid:hasCovidDrug ?covidDrug.

        FILTER(?comorbidity=k4covid:Obesity)

        ?treatment k4covid:hasComorbidity ?comorbidity.

        ?treatment k4covid:hasComorbidityDrug ?comorbidityDrug.

        ?comorbidityDrug k4covid:hasCUIAnnotation ?CUIComorbidityDrug.

        ?CUIComorbidityDrug owl:sameAs ?sameAsComorbidityDrug .

        ?covidDrug k4covid:hasCUIAnnotation ?CUICovidDrug.

        ?CUICovidDrug owl:sameAs ?sameAsCovidDrug .

    }

    SERVICE <https://dbpedia.org/sparql> {

        ?sameAsComorbidityDrug dbp:excretion ?excretion.

        ?sameAsComorbidityDrug dbp:routesOfAdministration ?routes.

        ?sameAsComorbidityDrug rdfs:comment ?prospectus.

        FILTER(LANG(?prospectus) = "en").

    }

} ORDER BY ?comorbidity

```

Şekil 0.8: Obesite Hastalarında Covid-19 İle Birlikte Kullanılan İlaçların Bilgisini Getiren Sorgu.

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>:Bu ifade RDF(Resource Description Framework) dilinde kullanılan bir PREFIX

kısaltmasıdır. Bu sayede rdfs PREFIX'i <http://www.w3.org/2000/01/rdf-schema#>'sına atanmıştır.

PREFIX owl: [<http://www.w3.org/2002/07/owl#>](http://www.w3.org/2002/07/owl#): Bu ön ek Web Ontology Language(OWL) bulunan terimlere referans vermek için kullanılır.

PREFIX k4covid: [<http://research.tib.eu/covid-19/vocab/>](http://research.tib.eu/covid-19/vocab/) : Bu ön ek Covid hastalığı ile ilgili özel olarak tanımlanmış kelime dağarcığını ifade eder.

PREFIX k4covid: [<http://research.tib.eu/covid-19/entity/>](http://research.tib.eu/covid-19/entity/): Bu ön ek Covid hastalığı ile ilgili tanımlanmış veri kümesini ifade eder.

PREFIX dbp: [<http://dbpedia.org/property/>](http://dbpedia.org/property/): Bu ön ek DBpedia'da özelliklere referans vermek için kullanılır.

Şekil 5.4'teki sorguda hastalık adı, kullanılan ilaç, ilacın atılımı bilgisi, uygulanma yolu ve prospektüs bilgilerini sorgulayıp görüntülemektedir.

WHERE ifadesi sorguda çekilmek istenen verileri nasıl çekileceğini ifade etmektedir.

SERVICE ifadesi sorgu esnasında farklı bir RDF uç noktasına erişilmesini sağlar bu sayede farklı iki ontolojiden aynı anda veri çekilebilir. Covid-19 ile ilgili verilerin elde edilebilmesi için kullanılmıştır.

?treatment k4covid:hasCovidDrug ?covidDrug: Bu ifade ?treatment nesnesinin ?covidDrug ile hasCovidDrug ilişkisi içerisinde olduğu belirtir.

FILTER(?comorbidity=k4covid:Obesity): İfadesiyle Obezite hastalığı filtrelenmiştir.

?treatment nesnesi ?comorbidity özelliği ile hasComorbidity ilişkisi içerisinde, aynı şekilde ?comorbidityDrug özelliği ile hasComorbidityDrug ilişkisi içerisinde olduğunu belirtir.

Şekil 5.4'te verilen sorguda ?comorbidityDrug nesnesi ?CUIComorbidityDrug özelliği ile hasCUIAnnotation ilişkisi içerisinde.

?CUIComorbidityDrug sameAs ilişkisi ile ?sameAsComorbidityDrug ile aynı ifade olduğu belirtilmektedir.

?covidDrug nesnesi ?CUICovidDrug özelliği ile hasCUIAnnotation ilişkisi içerisinde.

?CUICovidDrug ve ?sameAsCovidDrug sameAs ilişkisi ile aynı ifadeler olduğu belirtilmiştir.

?covidDrug nesnesi ?CUICovidDrug özelliği ile hasCUIAnnotation ilişkisi içerisinde.

?CUICovidDrug ve ?sameAsCovidDrug sameAs ilişkisi ile iki kaynaktaki ifadelerin eşdeğer olduğunu belirtmektedir.

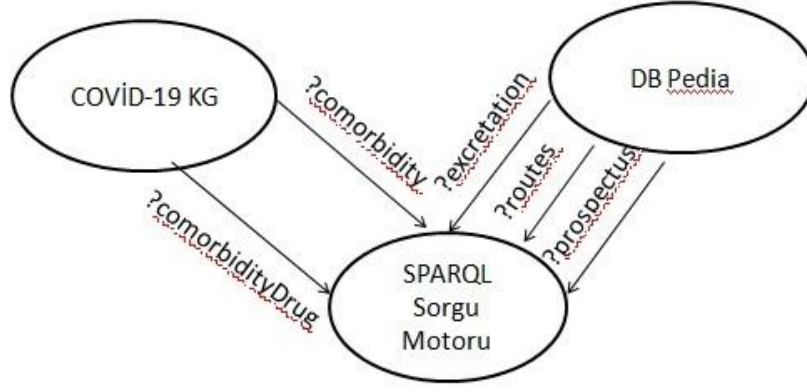
Sorgunun ikinci kısmında yeni bir SERVICE ifadesi görülmektedir. Aynı sorguda SERVICE ifadesi ile <<https://dbpedia.org/sparql>> uç noktasından veri çekilebilmesini sağlar.

?sameAsComorbidityDrug nesnesi dbp:excretion ilişkisi ?excretion özelliğini sorgulamızı sağlar. Bu ifade ilacın atılım hakkında bilgiyi ifade eder.

?sameAsComorbidityDrug nesnesi dbp:routesOfAdministration ilişkisi ile ?routes özelliğini sorgulamızı sağlar. Bu ifade ilacın uygulanma yolunu ifade eder.

?sameAsComorbidityDrug nesnesi rdfs:comment ilişkisi ile ?prospectus özelliğini sorgular ve ilacın prospektüs bilgilerini çeker. FILTER(LANG(?prospectus) = "en") ifadesi bu bilgileri İngilizce'ye çevirir.

Sorgu 2 için veri kaynaklarındaki ilişkileri gösteren diyagram Şekil 5.5'de verilmiştir.



Şekil 0.9: Sorgu 2 İçin Veri İlişkileri Diyagramı.

Sorgunun sonucunda elde edilen verilere ait ekran görüntüsü Şekil 5.6'da verilmiştir.

Table Response 6 results in 33.419 seconds Simple view Ellipse Filter query results Page size: 50

	comorbidity	sameAsComorbidityDrug	excretion	routes	prospectus
1	k4covid:Ob...	<http://dbpedia.org/resource/...	Kidney 13%, f...	<http://dbpe...	Simvastatin, sold under the brand name Zocor among others, is a statin, a type of lipid-lowering medication. It is...
2	k4covid:Ob...	<http://dbpedia.org/resource/...	<http://dbpedi...	<http://dbpe...	Amlodipine, sold under the brand name Norvasc among others, is a calcium channel blocker medication used to ...
3	k4covid:Ob...	<http://dbpedia.org/resource/...	Urine	<http://dbpe...	Paracetamol, also known as acetaminophen, is a medication used to treat fever and mild to moderate pain. Com...
4	k4covid:Ob...	<http://dbpedia.org/resource/...	Urine	<http://dbpe...	Paracetamol, also known as acetaminophen, is a medication used to treat fever and mild to moderate pain. Com...
5	k4covid:Ob...	<http://dbpedia.org/resource/...	Urine	<http://dbpe...	Paracetamol, also known as acetaminophen, is a medication used to treat fever and mild to moderate pain. Com...
6	k4covid:Ob...	<http://dbpedia.org/resource/...	<http://dbpedi...	<http://dbpe...	Atorvastatin is a statin medication used to prevent cardiovascular disease in those at high risk and to treat abnor...

Şekil 5.10: Sorgu 2 Sonucunda Elde Edilen Veriler.

Sorgu sonucunu incelediğinizde 6 kayıt görüntülenmiştir. Sorgu motorunda yaklaşık 33.419 saniyede sorgu sonucu elde edilmiştir.

5.2.3 Sorgu 3

Şekil 5.7'daki sorguda Kardiyopatili hastalarda Covid-19 tedavisiyle birlikte uygulanan tedaviler, kullanılan ilaçlar, bu ilaçların etken maddesi, etken maddelerinde hangi ilaçlarla etkileşime girdiği listelenmektedir.

```

PREFIX dbp: <http://dbpedia.org/property/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX k4covid: <http://research.tib.eu/covid-19/vocab/>
PREFIX k4covid: <http://research.tib.eu/covid-19/entity/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wikibase: <http://wikiba.se/ontology#>
PREFIX p: <http://www.wikidata.org/prop/>
PREFIX ps: <http://www.wikidata.org/prop/statement/>
PREFIX pq: <http://www.wikidata.org/prop/qualifier/>
SELECT DISTINCT ?treatment ?sameAsComorbidityDrug ?covidDrug ?idDrug ?activeIngredient
?activeIngredientLabel ?interactingDrug ?interactingDrugLabel
WHERE {
  SERVICE <https://labs.tib.eu/sdm/covid19kg/sparql> {
    ?treatment k4covid:hasCovidDrug ?covidDrug.
    ?treatment k4covid:hasComorbidity k4covid:Cardiopathy .
    ?treatment k4covid:hasComorbidityDrug ?comorbidityDrug.
    ?comorbidityDrug k4covid:hasCUIAnnotation ?CUIComorbidityDrug.
    ?CUIComorbidityDrug owl:sameAs ?sameAsComorbidityDrug .
    ?covidDrug k4covid:hasCUIAnnotation ?CUICovidDrug.
    ?CUICovidDrug owl:sameAs ?sameAsCovidDrug .
  }
  SERVICE <https://query.wikidata.org/sparql> {
    ?sameAsComorbidityDrug wdt:P592 ?idDrug .
    ?sameAsComorbidityDrug wdt:P769 ?interactingDrug.
    ?interactingDrug rdfs:label ?interactingDrugLabel.
    FILTER(LANG(?interactingDrugLabel) = "en").
    ?sameAsComorbidityDrug wdt:P3780 ?activeIngredient .
    ?activeIngredient rdfs:label ?activeIngredientLabel.
    FILTER(LANG(?activeIngredientLabel)="en").
  }
}
ORDER BY ?treatment

```

Şekil 5.11: Belirli Bir Hastalıkta Covid-19 ile Beraber Uygulanan Tedaviler ve Kullanılan İlaçların İçeriği ve Diğer İlaçlar ile Etkileşimini Getiren Sorgu.

Sorguyu incelediğimizde PREFIX’le ifade edilen ön ekler bize bu adresteki üçlülere kullanma imkânı sağlar.

SELECT ifadesi tedavi, eşlik eden hastalığın ilacı, covid için kullanılan ilaç, eşlik eden hastalığın ilacının wikidata’da temsil ettiği id’si, ilacın içindeki etken madde, etken maddenin literatürdeki adı, etken maddenin etkileşime girdiği ilaç ve etkileşime girdiği ilacın literatürdeki adını listelemektedir.

?treatment: Tedavi veya kullanılan ilaç kombinasyonunu temsil eder.

?covidDrug: COVID-19 tedavisinde kullanılan ilaçları temsil eder.

?comorbidityDrug: Kardiyopati (kalp hastalığı) gibi eşlik eden hastalık durumlarında kullanılan ilaçları ifade eder.

?CUIComorbidityDrug: Komorbidite ilaçlarının klinik tanımlayıcılarını (Concept Unique Identifier - CUI) temsil eder.

?sameAsComorbidityDrug: Komorbidite ilaçlarının farklı veri kaynaklarındaki eşdeğer tanımlayıcılarını ifade eder.

?CUICovidDrug: COVID-19 ilaçlarının klinik tanımlayıcılarını (CUI) temsil eder.

?sameAsCovidDrug: COVID-19 ilaçlarının farklı veri kaynaklarındaki eşdeğer tanımlayıcılarını ifade eder.

SERVICE <<https://labs.tib.eu/sdm/covid19kg/sparql>>: Bu ifade, TIB (Technische Informationsbibliothek) tarafından sağlanan COVID-19 bilgi grafiğine yönlendirilmiştir. Bu bilgi grafiği, COVID-19 ile ilgili verileri içermektedir.

?treatment k4covid:hasCovidDrug ?covidDrug.: Bu satır belirli bir ?treatment(tedavi)’nin covid-19 ilacını (?covidDrug) içerdiğini belirtir.

?treatment k4covid:hasComorbidity k4covid:Cardiopathy :Kardiyopatili (kalp rahatsızlığı) hastalarda uygulanan tedavileri ifade eder.

?treatment k4covid:hasComorbidityDrug ?comorbidityDrug :Tedavilerde (?treatment) eşlik eden bir hastalık varsa uygulanan ilacı(?comorbidityDrug) ifade eder.

?comorbidityDrug k4covid:hasCUIAnnotation ?CUIComorbidityDrug :Eşlik eden hastalığın ilacının klinik tanımlayıcıya(CUI) sahip olduğunu belirtir.

?CUIComorbidityDrug owl:sameAs ?sameAsComorbidityDrug :Bu ifade komorbidite ilac tanımlayıcısının diğer kaynaklardaki eşdeğer kaynakla aynı olduğunu ifade eder.Bu ifade iki farklı ontolojiden ortak özellik bularak veri çekmemizi sağlar.

?covidDrug k4covid:hasCUIAnnotation ?CUICovidDrug :Covid-19 ilaçlarının klinik tanımlayıcıya (CUI) sahip olduğunu belirtir.

?CUICovidDrug owl:sameAs ?sameAsCovidDrug :Covid-19 ilacının tanımlayıcısının diğer veri kaynaklarındaki tanımlayıcı ile aynı olduğunu ifade eder.

İkinci SERVICE ifadesi ile <https://query.wikidata.org/sparql> kaynağındaki verileri sorgulamamızı imkan sağlar.

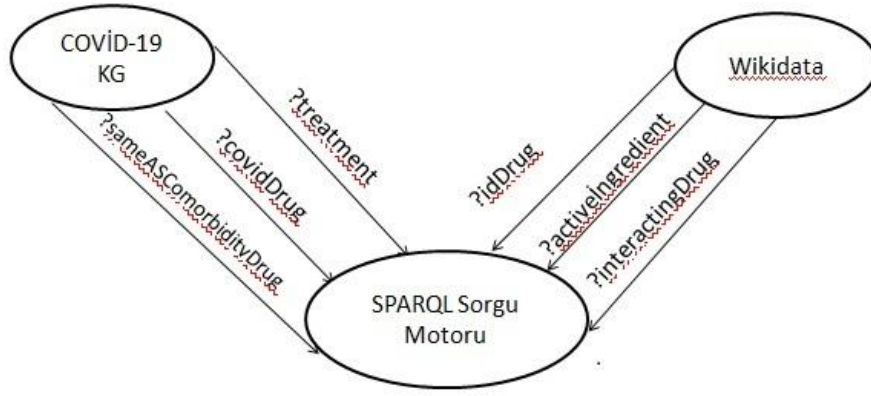
?sameAsComorbidityDrug wdt:P592 ?idDrug :Korbidite ilacının wikidata’da temsil ettiği id’sini ifade eder.

?sameAsComorbidityDrug wdt:P3780 ?activeIngredient :Korbidite ilacının etken maddesini ifade eder.

?sameAsComorbidityDrug wdt:P769 ?interactingDrug :Korbidite ilacının etken maddesine bağlı olarak etkilediği ilacı ifade eder.

?activeIngredient rdfs:label ?activeIngredientLabel :İlacın etken maddesinin etiketini(adını) sorgular.

Sorgu 3’e ait veri kaynaklarındaki verilere ait ilişkilerin diyagramı Şekil 5.8’de verilmiştir.



Şekil 0.12: Sorgu 3 İçin Veri İlişkileri Diyagramı.

Sorgu sonucunda 49271 satır veri elde edilmiştir ve sorgu süresi 7,823 saniyedir. Sorgu 3 sonucunda elde verilere ilişkin ekran görüntüsü Şekil 5.9’da verilmiştir.

Table		Response		49271 results in 7.823 seconds		Simple view		Ellipse		Filter query results		Page size: 50			
	treatment	sameAsComorbidityDrug	covidDrug	idDrug	activeIngredient	activeIngredientLabel	interactingDrug	interactingDrugLa							
1	k4covid:Treatment_0	wd:Q57055	k4covid:DB12...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
2	k4covid:Treatment_0	wd:Q57055	k4covid:DB04...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
3	k4covid:Treatment_...	wd:Q57055	k4covid:DB00...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
4	k4covid:Treatment_...	wd:Q57055	k4covid:DB01...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
5	k4covid:Treatment_...	wd:Q57055	k4covid:DB00...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
6	k4covid:Treatment_...	wd:Q57055	k4covid:DB00...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
7	k4covid:Treatment_...	wd:Q57055	k4covid:DB00...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
8	k4covid:Treatment_...	wd:Q57055	k4covid:DB01...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
9	k4covid:Treatment_...	wd:Q57055	k4covid:DB01...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
10	k4covid:Treatment_...	wd:Q57055	k4covid:DB04...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
11	k4covid:Treatment_...	wd:Q57055	k4covid:DB04...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
12	k4covid:Treatment_18	wd:Q57055	k4covid:DB01...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
13	k4covid:Treatment_18	wd:Q57055	k4covid:DB14...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
14	k4covid:Treatment_...	wd:Q57055	k4covid:DB12...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
15	k4covid:Treatment_...	wd:Q57055	k4covid:DB12...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							
16	k4covid:Treatment_...	wd:Q57055	k4covid:DB13...	CHEMBL112	wd:Q3245302	Tylenol	wd:Q177094	imatinib							

Şekil 0.13: Sorgu 3 Sonucunda Elde Edilen veriler.

Bu sorgu Covid-19 hastalığında kullanılan ilaçlar ve beraberinde bulunan hastalıklarda (örneğin, Kardiyopati) kullanılan ilaçlar hakkında bilgi sağlar. Bu bilgi ilaçların klinik araştırmalarında, ilaçların etkileşimlerin belirlenmesinde ve uygulanacak tedavilerin belirlenmesinde önemli rol oynar. Sorgu, ilaçların tanımlayıcılarını kullanarak farklı veri kaynaklarındaki eşdeğer tanımlayıcılarını bir araya getirir. Bu sayede ilaçlarla ilgili daha geniş ve kapsamlı bilgi edinmemizi sağlar.

5.2.4 Sorgu 4

Şekil 5.10’da verilen sorgu, COVID-19 tedavisinde kullanılan ilaçların yanı sıra obezite gibi komorbiditelere sahip hastalar için kullanılan diğer ilaçlarla ilgili çeşitli bilgileri sağlar. Sorgu sonucunda ilaçların DrugBank kimlikleri, ticari isimleri, eliminasyon yarı ömrü ve ilgili sayfa bağlantıları bilgileri elde edilir. Sorgu, COVID-19 bilgi grafiği ve DBpedia gibi iki farklı veri kaynağından bilgi çekerek bu bilgileri entegre eder.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX db: <http://dbpedia.org/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX k4covid: <http://research.tib.eu/covid-19/vocab/>
PREFIX k4covide: <http://research.tib.eu/covid-19/entity/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbp: <http://dbpedia.org/property/>
SELECT DISTINCT ?dbDrugBank ?tradeName ?comorbidity ?wikiPageWikiLink
?wikiPageWikiLinkLabel ?elimination
WHERE {
  SERVICE <https://labs.tib.eu/sdm/covid19kg/sparql> {
    ?treatment k4covid:hasCovidDrug ?covidDrug.
    ?treatment k4covid:hasComorbidityDrug ?comorbidityDrug.
    ?treatment k4covid:hasComorbidity ?comorbidity.
    ?treatment k4covid:hasComorbidity k4covide:Obesity .
    ?comorbidityDrug k4covid:hasCUIAnnotation ?CUIComorbidityDrug.
    ?CUIComorbidityDrug owl:sameAs ?sameAsComorbidityDrug .
    ?covidDrug k4covid:hasCUIAnnotation ?CUICovidDrug.
    ?CUICovidDrug owl:sameAs ?sameAsCovidDrug .
  }
  SERVICE <https://dbpedia.org/sparql> {
    ?sameAsCovidDrug dbo:drugbank ?dbDrugBank .
    ?sameAsCovidDrug dbp:tradenname ?tradeName.
    ?sameAsCovidDrug dbp:eliminationHalfLife ?elimination.
    ?sameAsCovidDrug dbo:wikiPageWikiLink ?wikiPageWikiLink.
    ?wikiPageWikiLink rdfs:label ?wikiPageWikiLinkLabel.
    FILTER(LANG(?wikiPageWikiLinkLabel)="en").
  }
} ORDER BY ?dbDrugBank
```

Şekil 0.140: Herhangi Bir İlacın Drugbank Bilgisi ve Ticari Bilgilerini Görüntüleyen Sorgu.

Sorgunun kodlarını incelediğimizde, PREFIX ön ekleri ile ifade edilmiş etiketlerin anlamları belirtilmiştir.

rdfs: RDF Schema Vocabulary (<http://www.w3.org/2000/01/rdf-schema#>)

db: DBpedia Namespace (<http://dbpedia.org/>)

owl: Web Ontology Language (<http://www.w3.org/2002/07/owl#>)

k4covid: TIB COVID-19 Vocab (<http://research.tib.eu/covid-19/vocab/>)

k4covid: TIB COVID-19 Entity (<http://research.tib.eu/covid-19/entity/>)

dbo: DBpedia Ontology (<http://dbpedia.org/ontology/>)

dbp: DBpedia Property (<http://dbpedia.org/property/>)

SERVICE kısmı sorgunun belirli bir SPARQL endpoint'inde (<<https://labs.tib.eu/sdm/covid19kg/sparql>>) çalıştırılacağını gösterir.

?treatment k4covid:hasCovidDrug ?covidDrug:Bu kısımda ?treatment'in (tedavi) ?covidDrug (covid-19)ilacına sahip olduğunu belirtir.

?treatment k4covid:hasComorbidityDrug :Bu ifade ile ?comorbidityDrug ?treatment'in(tedavi) ?comorbidityDrug (eşlikeden ilaç) sahip olduğunu belirtir.

?treatment k4covid:hasComorbidity ?comorbidity :Bu ifade ?treatment'in ?comorbidity özelliğine sahip olduğunu belirtir.

?treatment k4covid:hasComorbidity k4covid:Obesity :Bu kısımda ?treatment'in ?comorbidity özelliğinin Obesity olma durumunu belirtir.Yani Obesitesi olan Covid-19 hastalarını ifade eder.

?comorbidityDrug k4covid:hasCUIAnnotation ?CUIComorbidityDrug:Bu ifade Covid-19 ile beraber bulunan hastalıkların ilaçlarının CUI özelliğinin bulunduğunu ifade eder.

?CUIComorbidityDrug owl:sameAs ?sameAsComorbidityDrug :Bu ifade komorbidite ilaç tanımlayıcısının diğer kaynaklardaki eşdeğer kaynakla aynı olduğunu ifade eder.Bu ifade iki farklı ontolojiden ortak özellik bularak veri çekmemizi sağlar.

?covidDrug k4covid:hasCUIAnnotation ?CUICovidDrug:Bu ifade Covid ilaçlarının CUI koduna sahip olduğunu belirtir.

?CUICovidDrug owl:sameAs ?sameAsCovidDrug: Bu ifadede covid ilaç tanımlayıcısının diğer kaynaklardaki eşdeğer kaynakla aynı olduğunu ifade eder.

Sorgudaki ikinci SERVICE kısmı sorgunun belirli endpointte (<https://dbpedia.org/sparql>) çalıştırılacağını gösterir.

?sameAsCovidDrug dbo:drugbank ?dbDrugBank:Bu kısımda ilk SERVICE kısmındaki sameAsCovidDrug özelliğini alıp onun drugbanktaki herhangi bir ilaçla aynı olup olmadığını sorgular.

?sameAsCovidDrug dbp:tradeName ?tradeName: Bu kısımda covid ilaçlarının ticari adını(?tradeName) sorgular.

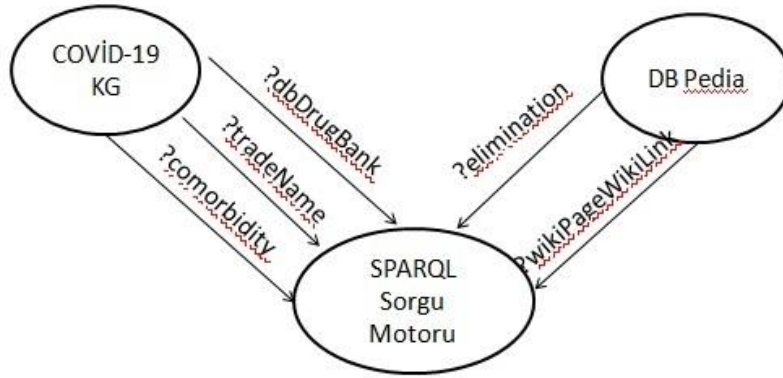
?sameAsCovidDrug dbp:eliminationHalfLife ?elimination: Bu kısımda covid ilaçlarının emilim hızını(?elimination) sorgular.

?sameAsCovidDrug dbo:wikiPageWikiLink ?wikiPageWikiLink: Covid ilaçlarının başka sayfalarla olan bağlantısını verir.

?wikiPageWikiLink rdfs:label ?wikiPageWikiLinkLabel:
?wikiPageWikiLinkLabel sayfanın alan adını sorgular.

FILTER(LANG(?wikiPageWikiLinkLabel)="en"):İngilizceye çevirir.

Sorgu 4 için verilerin kaynakları ile ilişkilerini gösteren diyagram Şekil 5.11’de verilmiştir.



Şekil 0.15: Sorgu 4 İçin Veri İlişkileri Diyagramı.

Sorgu sonucunda elde ettiğimiz veriler Şekil 5.12’ de görülmektedir. Sorgu 5.725 saniyede 4425 sonuç üretmektedir.

Table	Response	4425 results in 5.725 seconds	Simple view	Ellipse	Filter query results	Page size: 50
dbDrugBa...	tradeName	comorbidity	CUICovidDrug	wikiPageWikiLink	wikiPageWikiLinkLabel	elimination
1 DB00635	Deltasone, Liquid Pred, O...	k4covid:Obesity	k4covid:C0032...	db:Cancer	Cancer	7200.0
2 DB00635	Deltasone, Liquid Pred, O...	k4covid:Renal%	k4covid:C0032...	db:Cancer	Cancer	7200.0
3 DB00635	Deltasone, Liquid Pred, O...	k4covid:Asthma	k4covid:C0032...	db:Cancer	Cancer	7200.0
4 DB00635	Deltasone, Liquid Pred, O...	k4covid:Obesity	k4covid:C0032...	db:Cataracts	Cataracts	7200.0
5 DB00635	Deltasone, Liquid Pred, O...	k4covid:Renal%	k4covid:C0032...	db:Cataracts	Cataracts	7200.0
6 DB00635	Deltasone, Liquid Pred, O...	k4covid:Asthma	k4covid:C0032...	db:Cataracts	Cataracts	7200.0
7 DB00635	Deltasone, Liquid Pred, O...	k4covid:Obesity	k4covid:C0032...	db:Category:Antiasthmatic_drugs	Antiasthmatic drugs	7200.0
8 DB00635	Deltasone, Liquid Pred, O...	k4covid:Renal%	k4covid:C0032...	db:Category:Antiasthmatic_drugs	Antiasthmatic drugs	7200.0
9 DB00635	Deltasone, Liquid Pred, O...	k4covid:Asthma	k4covid:C0032...	db:Category:Antiasthmatic_drugs	Antiasthmatic drugs	7200.0
10 DB06273	Actemra, RoActemra	k4covid:Obesity	k4covid:C1609...	db:Benign_tumor	Benign tumor	-1209600.0
11 DB06273	Actemra, RoActemra	k4covid:Renal%	k4covid:C1609...	db:Benign_tumor	Benign tumor	-1209600.0
12 DB06273	Actemra, RoActemra	k4covid:Asthma	k4covid:C1609...	db:Benign_tumor	Benign tumor	-1209600.0
13 DB00635	Deltasone, Liquid Pred, O...	k4covid:Obesity	k4covid:C0032...	db:Prednisolone	Prednisolone	7200.0
14 DB00635	Deltasone, Liquid Pred, O...	k4covid:Renal%	k4covid:C0032...	db:Prednisolone	Prednisolone	7200.0
15 DB00635	Deltasone, Liquid Pred, O...	k4covid:Asthma	k4covid:C0032...	db:Prednisolone	Prednisolone	7200.0
16 DB00635	Deltasone, Liquid Pred, O...	k4covid:Obesity	k4covid:C0032...	db:Pregnancy	Pregnancy	7200.0
17 DB00635	Deltasone, Liquid Pred, O...	k4covid:Renal%	k4covid:C0032...	db:Pregnancy	Pregnancy	7200.0

Şekil 0.16: Sorgu 4 Sonucunda Elde Edilen Veriler.

5.2.5 Sorgu 5

Sorgu 5'te üç farklı veri kaynağından aynı anda veri çekilmiştir. SPARQL, RDF veri yapısını kullanan veri kaynaklarından dağıtık sorgulama yaparak veri çekmeye olanak sağlar. Bu sayede farklı veri kaynaklarından veri çekerek daha zengin ve kapsamlı verilere erişme imkânı sağlar.

Sorgu 5'in SAPARQL kodları incelendiğinde Covid-19 (<https://labs.tib.eu/sdm/covid19kg/sparql>) uç nokta (endpoint), wikidata kaynağına ait uç nokta (<https://query.wikidata.org/sparql>) ve dbpedia kaynağına ait uç nokta (<https://dbpedia.org/sparql>) kullanarak dağıtık sorgulama yapıldığı görülmektedir.

SELECT ifadesi dağıtık veri kaynaklarından covid hastalığı tedavisinde aynı anda farklı hastalığı (örneğin, kardiyopati) bulunan tedavileri, komorbidite ilaçları, komorbidite ilaçlarının kimlikleri (id), moleküler kütlesi, endeksiyonları, atılım yolu, metabolizması, gülümseme notasyonu ve ticari adını görüntülenir.

SERVICE anahtar sözcüğüyle COVID-19 ile ilgili bilgilerin depolandığı TIB tarafından sağlanan endpoint kullanılır.

?treatment değişkeni bir tedavi değişkenini ifade eder.

?covidDrug tedavi değişkeninin bir covid ilacını içerdiğini temsil eder.

k4covid:hasComorbidity, tedavinin bir komorbidite içerdiğini gösterir. Burada k4covid:Cardiopathy(kardiyopati) komorbiditesi belirtilmiştir.

k4covid:hasComorbidityDrug, tedavinin kkomorbidite için bir ilaç (?comorbidityDrug) içerdiğini gösterir.

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbp: <http://dbpedia.org/property/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX k4covid: <http://research.tib.eu/covid-19/vocab/>
PREFIX k4covide: <http://research.tib.eu/covid-19/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT DISTINCT ?treatment ?sameAsComorbidityDrug ?sameAsCovidDrug ?idDrug ?mass
?indication ?indicationLabel ?excretion ?metabolism ?dbSmiles ?tradeName
WHERE {
  SERVICE <https://labs.tib.eu/sdm/covid19kg/sparql> {
    ?treatment k4covid:hasCovidDrug ?covidDrug.
    ?treatment k4covid:hasComorbidity k4covide:Cardiopathy .
    ?treatment k4covid:hasComorbidityDrug ?comorbidityDrug.
    ?comorbidityDrug k4covid:hasCUIAnnotation ?CUIComorbidityDrug.
    ?CUIComorbidityDrug owl:sameAs ?sameAsComorbidityDrug .
    ?covidDrug k4covid:hasCUIAnnotation ?CUICovidDrug.
    ?CUICovidDrug owl:sameAs ?sameAsCovidDrug . }
  SERVICE <https://query.wikidata.org/sparql> {
    ?sameAsComorbidityDrug wdt:P592 ?idDrug .
    ?sameAsComorbidityDrug wdt:P2067 ?mass .
    ?sameAsComorbidityDrug wdt:P2175 ?indication.
    ?indication rdfs:label ?indicationLabel.
    FILTER(LANG(?indicationLabel)="en"). }
  SERVICE <https://dbpedia.org/sparql> {
    ?sameAsCovidDrug dbp:excretion ?excretion.
    ?sameAsCovidDrug dbp:metabolism ?metabolism.
    ?sameAsCovidDrug dbp:smiles ?dbSmiles .
    ?sameAsCovidDrug dbp:tradenname ?tradeName.
  }
}

```

Şekil 5.17: İlaçları ve Bu İlaçların Ticari İsimlerini Birlikte Getiren Sorgu.

k4covid:hasCUIAnnotation, komorbidite ilacının CUI anotasyonunu (?CUIComorbidityDrug) içerdiğini belirtir.

owl:sameAs, CUI anotasyonunun diğer veri kaynaklarındaki eşdeğerini (?sameAsCovidDrug) gösterir.

owl:sameAs, CUI anotasyonunun diğer veri kaynaklarındaki eşdeğerini (?sameAsComorbidityDrug) gösterir.

?CUICovidDrug owl:sameAs ?sameAsCovidDrug Covid ilaçlarının CUI anotasyonunu ile diğer veri kaynaklarındaki eşdeğerini temsil eder.

İkinci SERVICE yapısında, wikidata veritabanına erişmek için <https://query.wikidata.org/sparq> adresi kullanılır. SERVICE ifadesi, sorgu içerisinde farklı bir uç noktaya erişmemizi bu sayede dağıtık sorgu yazabilmemizi sağlar.

SERVICE ifadesi içerisinde komorbidite ilaçlarının kimlik numaraları (?idDrug) ifade eder.

Komorbidite ilaçlarının kütlelerini (?mass) ifade eder.

Komorbidite ilaçlarının indikasyonlarını (?indication) ifade eder.

Üçüncü SERVICE ifadesinde <https://dbpedia.org/sparq> uç noktasını kullanır. SERVICE ifadesi SPARQL sorgularını daha güçlü ve dinamik hale getirir. Çünkü bu sayede aynı anda birden fazla kaynaktan dağıtık sorgu yapmamıza imkan sağlanmış olur.

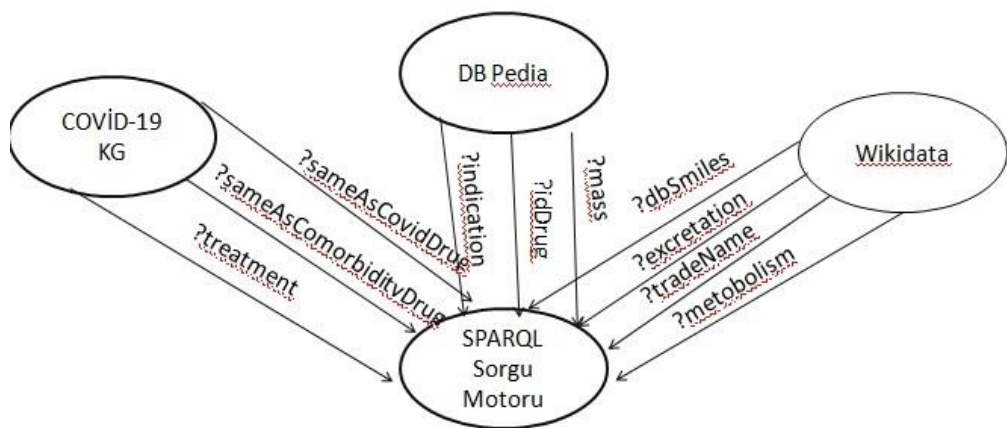
SERVICE ifadesinin içeriğini incelediğimizde tedavide kullanılan ilacın vücuttan nasıl atıldığını (?excretion) ifade eder.

?sameAsCovidDrug değişkenine ait metabolizmasını(?metabolism) ifade eder

?sameAsCovidDrug değişkenine ait gülümseme notasyonunu (?dbSmiles) ifade eder.

?sameAsCovidDrug değişkeninin ticari adını(?tradeName) ifade eder.

Sorgu 5 için verilerin kaynakları ile ilişkilerini gösteren diyagram Şekil 5.14'te verilmiştir.



Şekil 5.18: Sorgu 5 İçin Veri İlişkileri Diyagramı.

Sorgu sonucunda elde edilen veri şekil 5.16'da görülmektedir. Sorgu çalışma süre 4.948 saniye sürerken, 3180 veri elde edilmiştir.

Table Response 3180 results in 4.948 seconds

Simple view ☐ Ellipse ☒ Filter query results Page size: 50

	treatment	sameAsComorbidityDrug	sameAsCovidDrug	idDrug	mass	indication	indicationLabel	excretion	metabolism	dbSmiles	tradeName
1	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	358.178	<http://www.wikid...	ulcerative colitis	Mostly kidney, al...	<http://dbpedia.or...	C1c1cc2ncccNCCCCNCCO	Plaqueul, others
2	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	361.144	<http://www.wikid...	staphylococcal infection	Mostly kidney, al...	<http://dbpedia.or...	C1c1cc2ncccNCCCCNCCO	Plaqueul, others
3	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	358.178	<http://www.wikid...	myasthenia gravis	Mostly kidney, al...	<http://dbpedia.or...	C1c1cc2ncccNCCCCNCCO	Plaqueul, others
4	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	358.178	<http://www.wikid...	cancer	Mostly kidney, al...	<http://dbpedia.or...	C1c1cc2ncccNCCCCNCCO	Plaqueul, others
5	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	416.202	<http://www.wikid...	arterial hypertension	Mostly kidney, al...	<http://dbpedia.or...	C1c1cc2ncccNCCCCNCCO	Plaqueul, others
6	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	408.1452	<http://www.wikid...	arterial hypertension	Mostly kidney, al...	<http://dbpedia.or...	C1c1cc2ncccNCCCCNCCO	Plaqueul, others
7	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	263.188...	<http://www.wikid...	pain	Mostly kidney, al...	<http://dbpedia.or...	C1c1cc2ncccNCCCCNCCO	Plaqueul, others
8	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	206.131	<http://www.wikid...	pain	Mostly kidney, al...	<http://dbpedia.or...	C1c1cc2ncccNCCCCNCCO	Plaqueul, others
9	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	206.131	<http://www.wikid...	ankylosing spondylitis	Mostly kidney, al...	<http://dbpedia.or...	C1c1cc2ncccNCCCCNCCO	Plaqueul, others
10	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	358.178	<http://www.wikid...	enlaseopathy	Mostly kidney, al...	<http://dbpedia.or...	C1c1cc2ncccNCCCCNCCO	Plaqueul, others
11	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	358.178	<http://www.wikid...	ulcerative colitis	mostly fecal	<http://dbpedia.or...	CCc4nccc4	Norvir
12	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	358.178	<http://www.wikid...	ulcerative colitis	mostly fecal	<http://dbpedia.or...	CCc4nccc4	Norvir
13	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	361.144	<http://www.wikid...	staphylococcal infection	mostly fecal	<http://dbpedia.or...	CCc4nccc4	Norvir
14	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	361.144	<http://www.wikid...	staphylococcal infection	mostly fecal	<http://dbpedia.or...	CCc4nccc4	Norvir
15	k4covid.Tre...	<http://www.wikidata.org/entit...	<http://dbpedia.org/resource...	CHEMB...	358.178	<http://www.wikid...	myasthenia gravis	mostly fecal	<http://dbpedia.or...	CCc4nccc4	Norvir

Şekil 5.19: Sorgu 5 Sonucunda Elde Edilen Veriler.

6. SONUÇ VE ÖNERİLER

İnternet üzerindeki veri miktarı günden güne artmış ve bu veriler insanlar tarafından hızlı bir şekilde anlaşılıp, yorumlanabilmektedir. Ancak bu verileri anlayabilen, yorumlayabilen ve anlamlı sonuçlar çıkartabilen modellerin geliştirilmesi gereklidir. Bu amaç doğrultusunda anlamsal web teknolojileri kullanılmaya başlanmıştır. Verilerin kavramsallaştırılması ve anlamlı bilgiler elde edilmesi için kullanılan anlamsal web teknolojileri, birçok farklı veri kaynağını kullanabilmektedir. Basit sorgulamalar kısmında tek bir kaynaktan veri çekilmesi yeterli iken, zamanla tek kaynak ve basit sorgulamalar yetersiz kalmıştır. Bu tez kapsamında farklı veri kaynakları kullanarak dağıtık sorgulama mimarileri ve motorları araştırılmıştır.

Tezin kapsamında dağıtık veri kaynakları kullanılarak SPARQL dilinde birleşik bir sorgular geliştirilmiştir. Bu amaçla; Dbpedia, Wikidata, Drugbank, ...vb gibi farklı fiziksel alanlardaki veri kaynaklarının birbirleriyle sorunsuz bir şekilde haberleşmesi ve istenilen farklı parametrelere göre verilerin filtrelenip getirilmesi sağlanmıştır. Öncelikli olarak tüm bu farklı veri kaynaklarına erişim sağlanmasında ve sorguların çalıştırılmasını sağlayacak olan dağıtık sorgu motoru belirlenmiştir. Bunun için dağıtık sorgulama motorları incelenmiş olup, Sakor ve diğ. (2023) tarafından tanımlanan web ortamında sorgulama imkânı sağlayan dağıtık sorgulama motoru seçilmiştir. Diğer dağıtık sorgulama motorlarında, veri setinin sabit olması, verilerin boyutlarının az olması ya da sadece kendileri tarafından belirlenen sorguların yazılması gibi farklı problemlerle karşılaşmış ve bu sebeple ilgili sorgu motorları seçilmemiştir. Dağıtık sorgu motoru seçimi yapıldıktan sonra SPARQL sorguları oluşturulmuştur. Oluşturulan sorgularda, Dbpedia, Wikidata, Drugbank, gibi farklı veri kaynaklarından veriler kullanılmıştır.

Geliştirilen tez ile kullanıcıların kullandıkları verileri tek bir merkezde toplamadan yani verileri kendi sistemlerine kopyalamadan verilerin bulundukları kaynaklara erişim sağlayarak istenilen sorgulamaların yapılmasını sağlamaktadır. Bu sayede veri yönetimi, veri güvenliği gibi ek maliyet oluşturabilecek işlemlerden kaçınılmıştır. Ayrıca kendi sistemleri üzerine kopyalanan verilerin birleştirilmesi gibi

bir işlemde gerçekleştirilmemiştir. Gerçek zamanlı olarak veriler sorgulandığı için verilerin güncellenmesi ya da değişmesi gibi bir problem de ortaya çıkmamaktadır. Bu artıların yanı sıra, geliştirme aşamasında farklı problemler ile karşılaşılabilinmektedir. Bunların en başında karmaşık sorguların oluşturulması işlemi gelmektedir. Her ne kadar SPARQL dili standart bir söz dizimine sahip olsa da, seçilen dağıtık sorgu motoruna göre bazı kelimeler için kısıtlamalar getirilebilmektedir. Bu sebeple sorguların yazılması zaman alabilmektedir. Ayrıca, veriler farklı kaynaklardan alındığı için, verilerin yapısının ve içeriklerinin bilinmesi büyük önem taşımaktadır. Verilerin yapısı bilinmeden, hangi kaynakta hangi alanın ortak olduğu bulunamamakta bu da sorgulama yapıldığında hiçbir verinin gelmemesine sebep olmaktadır.

Tez kapsamında gelecekteki çalışmalar için, karşılaşılan zorluklara çözümler üretilebilir. Örneğin, iki farklı kaynaktaki verinin ortak yapısı doğal dil işleme, yapay zekâ, vb. gibi yöntemlerle tespit edilebilir. Diğer sorgulama motorlarında belirli kaynakların kullanılarak kısıtlı veriler üzerinde sorgulamalar yapılmasına olanak sağladığı görülmektedir. Bu kısıtlamaları ortadan kaldıracak nitelikte yöntemler geliştirilebilir. Ek olarak, dağıtık sorgulama motorlarında veri boyutlarından kaynaklı zaman problemleri ile karşılaşmıştır. Bu problemi çözmek içinde hızlı ve etkin yöntemler geliştirilebilir.

7. KAYNAKLAR

Aebeloe, C., Montoya, G., Hose, K., "Colchain: Collaborative linked data networks", (2021, April), *In Proceedings of the Web Conference 2021*, pp. 1385-1396, (2021).

Apaydin, A., Çelik, D., Elçi, A., "Semantic image retrieval model for sharing experiences in social networks", (2010, July), *In 2010 IEEE 34th annual computer software and applications conference workshops*, pp. 1-6, IEEE, (2010).

Arora, M., Sharma, R. L., "Artificial intelligence and big data: ontological and communicative perspectives in multi-sectoral scenarios of modern businesses", *foresight*, 25(1), 126-143, (2023).

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Sherlock, G., "Gene ontology: tool for the unification of biology", *Nature genetics*, 25(1), 25-29, (2000).

Akyokuş, S., "Anlamsal Webin Mimarisi[online]", (Temmuz 2024), <https://anlamsalweb.wordpress.com/2009/01/20/anlamsal-webin-mimarisi/>, (2009).

Barba-González, C., Caballero, I., Varela-Vaca, Á. J., Cruz-Lemus, J. A., Gómez-López, M. T. & Navas-Delgado, I., "BIGOWL4DQ: Ontology-driven approach for Big Data quality meta-modelling, selection and reasoning", *Information and Software Technology*, 167, 107378, (2024).

Bayraktar, H., Gülşen, B., Sengör, E. S. & Badem, H., "Akıllı şehirlerde veri", *VIII. UZAKTAN ALGILAMA-CBS SEMPOZYUMU* (UZAL-CBS 2022), Ankara, (17-19 Kasım 2022).

Beden, Ş., "Bir Semantik Web-Tabanlı Öğrenme Yönetim Sistemi Modeli", Yüksek Lisans Tezi, *İstanbul Üniversitesi Fen Bilimleri Enstitüsü*, Enformatik Ana Bilim Dalı, İstanbul, (2012).

Berners-lee, B. T., & Hendler, J., "The Semantic Web", 21. doi:10.1007/978-3-642-29923-0, (2001).

Breitmann, K.K., Casanova, M.A. and Truszkowski, W., *Semantic Web: Concepts, Technologies and Applications*, USA:Springer-Verlag London Limited, 327p, (2007).

Borowicc, S. L., Alves-Souza, S. N., “Heterogeneous Data Integration: A Literature Scope Review”, In *Proceedings of the 26th International Conference on Enterprise Information Systems (ICEIS 2024)*, pages 189-200, (2024).

Brickley, D., & Miller, L., "FOAF Vocabulary Specification." FOAF Project, (2000).

Chong, E. I., Das, S., Eadon, G., & Srinivasan, J., “An efficient SQL-based RDF querying scheme”, *Proceedings of the 31st international conference on Very large data bases*, VLDB Endowment, 1216-1227, (2005).

De Almeida, V. P., Campos, J. G., de Armas, E. M., da Silva, G. M. H., Neves, H. F., Corseuil, E. T. L., & Gonzalez, F. R., “INSIDE: Semantic Interoperability in Engineering Data Integration”, In *ICEIS (1)*, pp. 107-114, (2023).

DBpedia[online], (25,06,2024), <https://www.dbpedia.org/>, 2024.

DrugBank, Biomedical Knowledge Data Package[online], (26.06.2024), https://go.drugbank.com/data_packages/biomedical_knowledge, 2024.

Fisseha, F., “The Basics of Ontologies, Nordic Agricultural Ontology Service (AOS) Workshop”, Royal Veterinary and Agricultural University, Copenhagen, (2003).

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., & Schneider, L., "Sweetening ontologies with DOLCE." In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web (EKAW)*, pp. 166-181. Springer, Berlin, Heidelberg, (2002).

GeoNames, Geographical Database[online], (26.06.2024), <https://www.geonames.org/>, 2024.

Gruber, T. R., “A translation approach to portable ontology specifications”, *Knowledge acquisition*, 5(2), 199-220, (1993).

Görlitz, O. & Staab, S., “SPLENDID: SPARQL Endpoint Federation Exploiting VOID Descriptions”, *COLD*, 782, (2011).

Haas, L., Kossmann, D., Wimmers, E. & Yang, J., “Optimizing queries across diverse data sources”, (1997).

Huang, L., Hu, G. & Yang, X., “Review of ontolog mapping. In Consumer Electronics”, Communications and Networks(CECNet), 2012 2nd International Conference on.IEEE,p.537–540,(2012).

Ivanova, V., Bach, B., Pietriga, E. And Lambrix, P., “Alignment cubes: Towards interactive visual exploration and evaluation of multiple ontology alignments”, *In International Semantic Web Conference*, Springer, p.400–417, (2017).

Janev, V. “Semantic Intelligence in Big Data Applications”, (eds: Jain, S., Murugesan, S.), *Smart Connected World*, Springer, Cham. https://doi.org/10.1007/978-3-030-76387-9_4,(2021).

Kabakuş, A. T., & Doğdu, E., “Linked Data Üzerinden Doğal Dil Sorgularını Cevaplayan Sistem”, *In CEUR Workshop Proceedings*, CEUR-WS, (2013).

Keskin, Ö. & Sezer, E., “Kaynak Kod Sorgulamada Ontoloji Kullanımı”, *Ulusal Yazılım Mühendisliği Sempozyumu*, (2009).

Kiong, Y. C., Palaniappan, S. & Yahaya, N. A., “Health ontology system”, *In 2011 7th International Conference on Information Technology in Asia*, pp. 1-4, IEEE, (2011, July).

Koivunen, M. R., & Miller, E., “W3C semantic web activity”, *Semantic Web Kick-Off in Finland*, 2, 27-44, (2001).

Kolli, R., “Scalable matching of ontology graphs using partitioning”, Ph.D. thesis, *University of Georgia*, Athens, (2008).

Kulkarni, A., Ramanathan, C. & Venugopal, V. E., “Ontology mediated document retrieval for exploratory big data analytics”, *In 2023 IEEE 17th International Conference on Semantic Computing (ICSC)*, pp. 100-103, IEEE, (2023, February).

Lambrix, P., Dragisic, Z., Ivanova, V. and Anslow, C., “Visualization for ontology evolution”, *In 2nd International Workshop on Visualization and Interaction for Ontologies and Linked Data*, Kobe, Japan, Rheinisch-Westfaelische Technische Hochschule Aachen University, p.54–67, (17 October 2016).

Lu, Z., Wang, L. & Wen, J.R., “Direct Semantic Analysis for Social Image Classification”, *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 1258-1264, Québec, Canada, (2014).

Lymperis, D. & Goumopoulos, C., "Sedia: A platform for semantically enriched IOT data integration and development of Smart City Applications", *Future Internet*, 15(8), 276, (2023).

Niles, I., & Pease, A., "Towards a Standard Upper Ontology." In *Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS)*, pp. 2-9. ACM, (2001).

Mann, V., Viswanath, S., Vaidyaraman, S., Balakrishnan, J. & Venkatasubramanian, V., "SUSIE: Pharmaceutical CMC ontology-based information extraction for drug development using machine learning", *Computers & Chemical Engineering*, 179, 108446, (2023).

Ma, L. & Hartmann, T., "A proposed ontology to support the hardware design of building inspection robot systems", *Advanced Engineering Informatics*, 55, 101851, (2023).

Mosharraf, M. & Taghiyareh, F., "Federated search engine for open educational linked data", *Bull. IEEE Tech. Comm. Learn. Technol*, 18(6), 6-10, (2016).

Öztürk, Övünç., "Anlamsal Web için Bir Ontoloji Ortamı Tasarımı ve Gerçekleştirimi", Yüksek Lisans Tezi, *Ege Üniversitesi Fen Bilimleri Enstitüsü*, Bilgisayar Mühendisliği Ana Bilim Dalı, İzmir, (2004).

Pascal, H., Krötzsch, M., Parsia, B., Patel-Schneider, P. F. & Rudolph, S., "OWL 2 Web Ontology Language Primer[online]", (19 Ağustos 2016), <https://www.w3.org/TR/owl2-primer/>, (2012).

Paneque, M., del Mar Roldán-García, M. & García-Nieto, J., "e-LION: Data integration semantic model to enhance predictive analytics in e-Learning", *Expert Systems with Applications*, 213, 118892, (2023).

Quilitz, B., & Leser, U., "Querying distributed RDF data sources with SPARQL", In *The Semantic Web: Research and Applications: 5th European Semantic Web Conference, ESWC 2008, Tenerife, Canary Islands, Spain, June 1-5, Proceedings* 5 pp. 524-538, Springer Berlin Heidelberg, (2008).

Rakhmawati, N. A., Umbrich, J., Karnstedt, M., Hasnain, A. & Hausenblas, M., "Querying over Federated SPARQL Endpoints - A State of the Art Survey", CoRR, abs/1306.1, *Databases; Distributed, Parallel, and Cluster Computing*, (June 2013).

Raskin, R. G., & Pan, M. J., “Knowledge representation in the semantic web for Earth and environmental terminology (SWEET)”. *Computers & geosciences*, 31(9), 1119-1125, (2005)

Sakor, A., Jozashoori, S., Niazmand, E., Rivas, A., Bougiatiotis, K., Aisopos, F., ... & Vidal, M. E., “Knowledge4COVID-19: A semantic-based approach for constructing a COVID-19 related knowledge graph from various sources and analyzing treatments toxicities”, *Journal of Web Semantics*, 75, 100760, (2023).

Saleem, M., “Storage, indexing, query processing, and benchmarking in centralized and distributed RDF engines: a survey”, *Authorea Preprints*, (2023).

Saleem, M. & Ngonga Ngomo, A. C., “Hibiscus: Hypergraph-based source selection for sparql endpoint federation”, *In The Semantic Web: Trends and Challenges: 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings 11* (pp. 176-191). Springer International Publishing, (2014).

Saleem, M., Ngonga Ngomo, A. C., Xavier Parreira, J., Deus, H. F. & Hauswirth, M., “Daw: Duplicate-aware federated query processing over the web of data”, *In The Semantic Web–ISWC 2013: 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I 12* (pp. 574-590). Springer Berlin Heidelberg, (2013).

Schmidt, M., Görlitz, O., Haase, P., Ladwig, G., Schwarte, A. & Tran, T., “FedBench: A benchmark suite for federated semantic data query processing”, *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7031, (2011).

Schwarte, A., Haase, P., Hose, K., Schenkel, R., Schmidt, M., “Fedx: Optimization techniques for federated query processing on linked data”, *In: International Semantic Web Conference*, (2011).

Sen, P. S. & Mukherjee, N., “An ontology-based approach to designing a NoSQL database for semi-structured and unstructured health data”, *Cluster Computing*, 27(1), 959-976, (2024).

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., ... & Lewis, S., “The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration”, *Nature biotechnology*, 25(11), 1251-1255, (2007).

SPARQL, SPARQL Query Language for RDF[online], (15.01.2008), <http://www.w3.org/TR/rdf-sparql-query/>,(2008).

Sridhar, S. ve Sanagavarapu, S., “Twitter Social Networking Graph using Hypernym based Semantic Similarity Detection”, *International Conference on Smart Electronics and Communication (ICOSEC)*, 28- 35, Trichy, India, (2020).

Taelman, R., Van Herwegen, J., Vander Sande, M., ve Verborgh, R., “Comunica: a modular SPARQL query engine for the web”, *In The Semantic Web–ISWC 2018: 17th International Semantic Web Conference*, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part II 17 (pp. 239-255). Springer International Publishing, (2018).

Tanenbaum, A. ve Steen, M., “Distributed systems: principles and paradigms”, 2. Baskı. Amsterdam: Pearson Prentice Hall, Inc., ss 2-4-17-18-25, (2006).

The Gene Ontology Consortium, "The Gene Ontology Resource: 20 years and still GOing strong." *Nucleic Acids Research*, 47(D1), D330-D338, (2019).

Türkyılmaz, İ., "Semantik Web Teknolojileri", Akademik Bilişim 2008, Çanakkale Onsekiz Mart Üniversitesi, Çanakkale, (30 Ocak 2008).

Wikipedia[online], (26.06.2024), <https://wikipedia.org/>, 2024.

Yücebaş, S. C., Tiryaki, A. M. & Boğa, M., “E-Ticaret Sitelerinde Serbest Metin Aramaları için Anlamsal Çıkarılma Yapan bir Metin Analizcisi Yöntemi”, *20. Akademik Bilişim 2018 Konferansı Bildirileri*, (2018).

Zablith, F., Antoniou, G., d’Aquin, M., Flouris, G., Kondylakis, H., Motta, E., Plexousakis, D. and Sabou, M., “Ontology evolution: a process-centric survey”, *The knowledge engineering review*, 30(1):45–75, (2015).