



## Araştırma Makalesi / Research Article

## Sinema verilerinin Neo4j çizge veritabanı ile modellenmesi ve analizi

## Modeling and analysis of cinema data using Neo4j graph database

<sup>1</sup>Gizem Tangalay Dalgın, <sup>2</sup>Resul Daş<sup>1</sup>Fırat Üniversitesi, Teknoloji Fakültesi, Yazılım Mühendisliği Bölümü  
[gizemtangalay@gmail.com](mailto:gizemtangalay@gmail.com) ORCID: 0009-0003-0285-5360<sup>2</sup>Fırat Üniversitesi, Teknoloji Fakültesi, Yazılım Mühendisliği Bölümü  
[rdas@firat.edu.tr](mailto:rdas@firat.edu.tr) ORCID: 0000-0002-6113-4649

## MAKALE BİLGİLERİ

## Makale Geçmişi:

Geliş 18.08.2023  
Revizyon 20.11.2023  
Kabul 30.11.2023  
Online 29.03.2024

## Anahtar Kelimeler:

Neo4j, Sinema Verileri, Çizge  
Modelleme, Çizge Görselleştirme.

## ÖZ

Günümüzde kurum veya kuruluşlar için veri, önemli bir varlık haline gelmiştir. İnternet ortamındaki hızlı gelişmeler ile sinema verileri de hızla büyümekte ve bu veriler arasındaki ilişki giderek daha karmaşık hale gelmektedir. Çizge veri tabanı mimarisi, elde bulunan verilerin model tasarımında varlıklar arasındaki ilişkiyi vurgulayan, verileri modellemek için etkili bir araçtır. Geleneksel veri tabanlarına göre daha karmaşık verileri depolamakta ve daha hızlı bir şekilde sunmaktadır. Bu çalışmada; sinema verileri ve karmaşıklaşan ilişkiler incelenerek neo4j çizge veri tabanı ile modellenmektedir. Bu sayede veriler arası ilişkiler daha kolay bir şekilde görüntülenmekte, sorgulanabilmekte ve daha kolay bir öğrenme imkanı sunmaktadır. Çalışma ile Neo4j çizge veri tabanı karmaşık ve aralarında birden fazla ilişki bulunan verileri modellemede oldukça başarılı sonuçlar sunduğu görülmüştür. Sinema verilerini depolamak ve yönetmek için Neo4j veri tabanını kullanmak, bir sinema web sitesi için kullanıcının gereksinimlerini ve ilgi alanlarını karşılamasını kolaylaştırdığı görülmektedir. Sinema adı, yılı, yönetmeni, oyuncularını, sinema filminin hasılatı, seyirci sayısı ve sinemanın türü gibi değişkenler dikkate alınarak çizge modellemesi ve analizi yapılarak başarılı sonuçlar elde edilmiştir.

## ARTICLE INFO

## Article history:

Received 18.08.2023  
Received in revised 20.11.2023  
Accepted 30.11.2023  
Available online 29.03.2024

## Keywords:

Neo4j, Cinema Data, Graph  
Modeling, Graph Visualization

Doi: 10.24012/dumf.1343167

\* Sorumlu Yazar

## ABSTRACT

Nowadays, data has become a crucial asset for institutions and organizations. With the rapid developments in the online environment, film data is also rapidly growing, and the relationships among these data are becoming increasingly complex. Graph database architecture is an effective tool for modeling data by emphasizing the relationships between entities. It can store more complex data and present it more quickly compared to traditional databases. In this study, cinema data and the growing complexities in relationships are examined, and they are modeled using the Neo4j graph database. This allows for easier visualization and querying of relationships between data, providing an easier learning opportunity. The study shows that the Neo4j graph database provides successful results in modeling complex data with multiple relationships. Using the Neo4j database to store and manage cinema data facilitates meeting the requirements and interests of users for a cinema website. By considering variables such as the movie's name, year, director, cast, box office revenue, audience size, and genre, successful results were achieved through graph modeling and analysis.

## Giriş

İnternetteki gelişmeler, fiziksel cihazlar ve sanal bileşenler sayesinde dünya sınırları ortadan kalkmaktadır. İnsanlar giderek daha fazla bilgiye daha hızlı ve kolay bir şekilde ulaşabilir hale gelmektedir. Bilginin az olduğu, ulaşamadığı bir çağdan aşırı bilgi çağına girilmektedir. Zengin içeriğe sahip bu ortamda pek çok karmaşık ve değişken bilgi ile karşılaşmaktadır ve bu kadar karmaşıklaşan yapı içerisinde faydalı içeriğe hızlı bir şekilde ulaşmamız giderek zorlaşmaktadır. Bu zorlukları hafifletmek veya ortadan kaldırmak için verilerin işlenmesi, sınıflandırılması, ortaya anlamlı verilerin çıkarılması, verilerin görselleştirilmesi oldukça faydalıdır. Veriler arasında kurulacak anlamsal ağ sağlanan bilgileri daha sağlıklı tanımlar ve bilgiyi zenginleştirir.

Veri tabanları verileri kategorize bir şekilde tutarak, kolay bir şekilde ulaşılmasını ve yönetilmesini sağlamaktadır. İlişkisel veri tabanlarının ardından Çizge veri tabanları ortaya çıkmıştır. Son yıllarda, ürün ortamı için Neo4j, Infinite, Graph, DEX, InfoGrid, HyperGraphDB, Trinity vb. gibi bir dizi yüksek performanslı Çizge veri tabanı bulunmaktadır [1]. Bu veri tabanlarının arasında Neo4j; Java temelli, açık kaynak kodlu bir veri tabanı yazılımıdır. Elde bulunan veriler tablolarla değil düğümler ve düğümler arasındaki ilişkilerden oluşan ağ yapısında tutulur. Düğümler varlıkları temsil eder. Şekil 1’de Neo4j platformu örnek çizge model görünümü yer almaktadır. Günümüzde verilerin boyutu giderek artmaktadır ve verilerin arasındaki ilişki giderek karmaşıklaşmaktadır. Genellikle verinin kendisine değil, anlamlandırmak için verilerin arasındaki ilişkiye odaklanmak gerekir ve geleneksel olan veri tabanları bu isteği karşılayamamaktadır. Neo4j çizge veri tabanı bu konuda işimizi kolaylaştırmaktadır. Bir Çizge veri tabanındaki öğeler, olası herhangi bir şekilde birbirine bağlanabilir. İnsanlar günlük hayatında temel ihtiyaçlarının yanı sıra eğlenceye de ihtiyaç duyarlar. Bu eğlencelerden biri de film seyretmektir. Filmler türü ve izleyenlerinin yaşı fark etmeksizin genellikle her insanın dikkatini çekmektedir. Son yıllarda çok fazla sinema filmi ve televizyon dizisi çekilmektedir, ancak bu film ve dizilerin çok azı piyasada başarılı olmaktadır. Bu nedenle, filmlerin ve televizyon dizilerinin başarı faktörlerinin analizi, yapımcılar ve yatırımcılar için çok önemlidir. Büyük verilerin bulunduğu bu çağda, veri kaynakları çok değerlidir ve yapımcılar, yatırımcılar dâhil olmak üzere tüm tarafların dikkatini çekmiştir. Bunu sinema filmleri alanında kullanmak, sinema film endüstrisinin yeniliği ve gelişimi için önemli bir destek sağlayabilir. Manevi kültürün bir yansıması olan sinema filmleri, gelir elde etmek için önemli araçlardan biridir; Ekonomik kazancı arttırabilmek için yapımcı, yatırım planını yaparken yatırım riskini ve yatırımın gelirini tam olarak hesaplayacak ve seyircinin ihtiyaçları ile birlikte potansiyel gelir hesaplanacaktır. Yapımcı, bu büyük verileri analiz ederek, kendi aralarında sıralayarak filmlerin yapım ekibi, sektördeki pazar ortamı gibi etkileyen temel faktörleri detaylı bir şekilde değerlendirebilir, yatırımcılara kapsamlı ve bilimsel veriler sunabilir. Film verileri içeren bir web sitesinde yalnızca film videoları depolanmaz. Yer alan filme ait yönetmenler, yazarlar, oyuncular vb. bilgilerin de depolanması gerekir. Çok sayıda ilişki bulunduğu geleneksel veri tabanları istenilen performansı gösteremez ve bu yer alan tabloların güncellenmesi de zorlaşmaktadır.

Örneğin, iki filmin yönetmenlerinin başka filmlerle ortak olup olmadığı bulunmak istenildiğinde, iki filmin yönetmenlerinin yönetmenliğini üstlendiği diğer filmleri bulmak ve daha sonra bu filmler arasında bir ortaklık olup olmadığını incelenmesi gerekir. Bu nedenle film verilerinin araştırılması, incelenmesi ve işlenmesinde ilişkisel olmayan veri tabanlarını kullanmak daha uygun bir seçimdir. Çizgeler tarih boyunca çözülmemiş birçok problemin çözümü ve analizi için modeller oluşturmuştur. Bu modellerin çok çeşitli tür ve formatları vardır. Çizge görselleştirme problemlerin anlaşılması ve takip edilmesi açısından kolaylık sağladığından büyük önem taşımaktadır. Günümüzde Çizgeler o kadar popüler hale geldi ki bilim dünyasında hemen her alanda Çizgelerle modellenen bir algoritmayla karşılaşmak mümkündür. Bilgisayar bilimi, sosyal bilimler, dil bilimi, mühendislik, matematik, tıp gibi her alanda Çizgelerle karşılaşmaktadır [2]. Neo4j çok büyük sayıda karmaşık, dinamik, yapılandırılması düşük verileri işlemek ve yönetmek için uygun olan bir veri tabanıdır. Veri tabanını oluşturmak, veri tabanında sorgulamalar yapmak, verilerde güncellemeler vb. işlemler yapmak için uygundur. Sosyal ağlar ve bilgilerin görselleştirilmesi vb. alanlarda veri depolama sorunlarını çözmek için oldukça etkili bir yöntemdir.

Bu makale çalışmasında çizge modelleme incelenmiş ve sinema verilerinin Neo4j ile çizge modelleme uygulaması gerçekleştirilmiştir. Birinci bölümde konu ile ilgili açıklamalar yapılmış ve izlenecek yollar anlatılmıştır. İkinci bölümde çizge veri tabanları ile ilgili bilgiler verilmiştir. Üçüncü bölümde Neo4j veri tabanı tanıtılmıştır. Dördüncü bölümde asıl çalışmamız olan Neo4j ile sinema verilerinin çizge modellemesi ve analizi anlatılmıştır. Son olarak beşinci bölümde elde edilen sonuçlar yazılmıştır ve çalışma sonuçlandırılmıştır.

## Literatür taraması

Çizge teorisi, uygulamaları ve görselleştirilmesi üzerine son yıllarda oldukça yoğun bir ilgi artmıştır. Yazılım dünyasında çizge uygulamalarının etkisi hem geliştirici hem de kullanıcılar açısından büyük anlama kolaylığı sağlamaktadır. Bu nedenle hem akademik çalışmalarda hem yazılım firmalarının uygulamalarında aktif olarak kullanılmaktadır. Bu bölümde literatürden bazı önemli çalışmalara yer verilmektedir.

Zhu Z. ve diğerleri çalışmalarında karmaşık fabrikasyon ürünlerin imalatını planlamak için Neo4j tabanlı yeni bir yaklaşım önermişlerdir. Dünyadaki üretim konfigürasyonunu göz önünde bulunduran bu çalışma, iş önceliği kısıtlamaları, zaman kısıtlamaları ve stok kısıtlamaları ile esnek atölye çizgeleme problemini ele almaktadır. Bu kısıtlamaları aynı anda dikkate alabilmek için Çizge veri tabanı, anlamsal ağ ve bilgi yakalamayı içeren Neo4j tabanlı yaklaşım önerilmiştir. Çeşitlendirilmiş bağlı çizgeleme verilerini yönetmek ve incelenen çizgeleme problemini Neo4j Çizge veri tabanını kullanarak esnek Çizgelerle modellemek uygun ve verimlidir [3]. Chen A. ve diğerleri çalışmalarında Synthea aracı tarafından oluşturulan 150.000 sentetik hastayla elektronik tıbbi kayıt (EMR) veri tabanını simüle etmişlerdir. Örnek bir hedef hastalık olarak akciğer kanseri için Neo4j Çizge veri

tabanında hasta sağlık faktörü grafiği oluşturulmuştur. Çizge, faktör düğümleri ile hasta düğümleri arasında 990.000'den fazla ilişki içermektedir ve Birleşik Tıbbi Dil Sistemi (UMLS) akciğer kanseri alt grafiğine bağlanmıştır. Böyle bir entegre biyomedikal Çizge, hasta sağlık faktörlerini ve biyomedikal bilgiyi aynı Çizgeye görüntüleyebildi ve karşılaştırabildi. En üst sıralarda yer alan faktörler, bu Çizge yönteminin geçerliliğini gösteren literatürdeki raporlarla büyük ölçüde doğrulanmıştır [4]. Hu H. ve diğerleri çalışmalarında büyük veri karakteristiğine dayalı elektrik güç sistemlerinin dinamik yıldırımdan korunma yöntemi önermişlerdir. Bu yöntem temel olarak iki bölümden oluşmaktadır: Birinci bölüm, güç sisteminin büyük verilerini ve güç sisteminin dinamik düzenlenmesini analiz etmek için kullanılan Neo4j çerçeve modeli ve farklı çerçeve modellerini karşılaştırmak ve analiz etmek için kullanılan Python yazılımı; İkinci bölüm, dinamik yıldırım ve geleneksel koruma yöntemleri arasındaki karşılaştırma. Sonuçlar, Neo4j geçiş hızının Hadoop ve Spark'tan sırasıyla %87,5 ve %89,1 daha hızlı olduğunu, kümeleme etkisinin Hadoop ve Spark'tan sırasıyla %12,5 ve %17,8 daha yüksek olduğunu gösteriyor. Sonuç olarak Neo4j çerçeve modeli, güç sistemindeki büyük verinin özelliklerine daha uygundur [5]. Shi Y. ve diğerleri çalışmalarında işle ilgili suçlar için oluşturulmuş bir bilgi grafiği üzerinde çalışmışlardır. Suç vakalarının artan sayısı, savcılık organlarının çalışmalarına büyük zorluklar getirdi. Aynı zamanda, geleneksel veri tabanları büyük ölçekli, yapılandırılmamış ve karmaşık dinamik verilerle uğraşırken giderek daha dezavantajlı hale geldi. Çıkarılan bilgileri depolamak ve işle ilgili suç vakaları için bir bilgi grafiği oluşturmak üzere Neo4j Çizge veri tabanını kullanarak, eldeki verilerin görselleştirilmesini ve derinlemesine madenciliğini kolaylaştırdılar. Savcılık organlarının çalışmaları için suç vakalarını analiz ederek yeni bir yol sundular [6]. Kuhn M. ve diğerleri çalışmalarında statik ve dinamik veri alt modellerini, bağlantılı tek bir izlenebilirlik grafiğinde birleştiren Çizge tabanlı bir izlenebilirlik modeli geliştirmişlerdir. Akıllı üretimde, artan miktarda heterojen veri girişinin bütünsel bir izlenebilirlik modeline entegre edilmesi gerekir. Temel zorluklardan biri, gerçek zamanlı izleme amaçları için dinamik ve olaya dayalı sensör verilerinin, çoğunlukla geriye dönük izleme amaçları için gerekli olan statik verilerle tutarlı bir şekilde eşleşmesidir. Bir otomotiv kullanım senaryosu için yeni izlenebilirlik modeli uygulanmıştır. Neo4j Çizge veri tabanına ve Apache Kafka yazılımına dayalı hibrit bir mimari kullanılarak test edilmiştir [7]. Lu H. ve diğerleri çalışmalarında film verilerine odaklanarak Neo4j tabanlı analiz yapılmıştır. İlk olarak Neo4j ve Cypher Query Language tanıtmışlardır. Daha sonra film verilerindeki yönetmenler, oyuncular vb. gibi anahtar nesnelere arasındaki ilişkileri analiz etmişlerdir. Neo4j veri tabanı karmaşık ve çok bağlantılı verilerle başa çıkmada iyidir, film verilerini depolamak ve yönetmek için Neo4j veri tabanını kullanmak film verilerinin analizini kolaylaştırmaktadır [8]. Shuai Q. Ve diğerleri çalışmalarında film kültürü alanı için, kullanıcıların film kültürüyle ilgili sorunları hızlı ve doğru bir şekilde sorgulamasına yardımcı olabilecek akıllı bir Çince soru ve yanıtlama sistemi oluşturuyor. Öncelikle film kültürünün bilgi grafiği oluşturulmuş ve veri depolamak için Neo4j Çizge veri tabanı kullanılmıştır. Daha sonra, kullanıcının sorunlarını sınıflandırmak için saf Bayes modeli kullanılır. Son olarak,

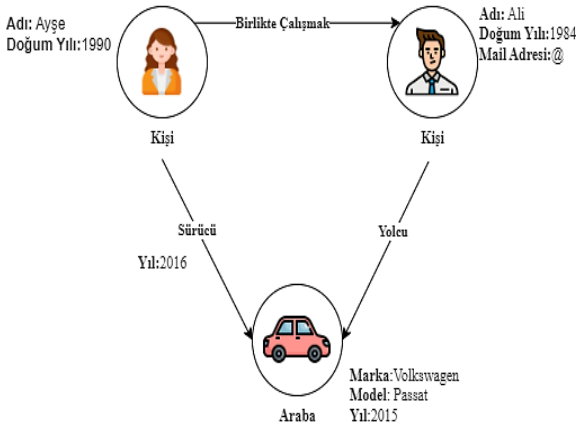
kullanıcının niyetine ve anahtar kelimelerine göre sorular bilgi grafiği sorgu ifadelerine dönüştürülür ve veri tabanı sorgulamasından sonra cevaplar döndürülür [9]. Yi N. ve diğerleri çalışmalarında Çizge veri tabanına dayalı film öneri sistemi tasarımı ve uygulaması yapmışlardır. Bu çalışmada uygulanan film tavsiye sistemi, geleneksel kullanıcı tabanlı işbirlikçi filtreleme algoritmasına dayanmaktadır ve kullanıcı projesi puanlama matrisi önceden doldurulmuştur. Aynı zamanda, bu sistemin veri tabanı teknolojisi, karmaşık ilişkilerle başa çıkmada iyi olan Çizge veri tabanını kullanır. Veri görselleştirmede, kullanıcı deneyimini iyileştirmek için bir filmin öneri derecesi, düğümün boyutu ve kenarın kalınlığı ile ifade edilir [10]. Goyanı M. ve diğerleri çalışmalarında film öneri sistemine ilişkin inceleme yapmışlardır. Bu çalışmada film önerisi için son teknoloji İçerik Tabanlı Filtreleme, İşbirlikçi Filtreleme, Hibrit Yaklaşım ve Derin Öğrenme Temelli Yöntemleri inceledik. Farklı benzerlik ölçülerini de incelemiştirlerdir. Arkadaş tavsiyesi veren Facebook, iş tavsiyesi LinkedIn, müzik tavsiyesi Pandora, Netflix film tavsiyesi, Amazon ürün tavsiyesi gibi birçok firma tavsiye sistemini karlarını artırmak ve müşterilerine fayda sağlamak için kullanmaktadır. Bu makale esas olarak film tavsiyesi için farklı tekniklerin ve yöntemlerinin kısa bir incelemesine odaklanmaktadır [11]. Raghavendra C. Ve diğerleri çalışmalarında film öneri sistemleri için benzerliğe dayalı işbirlikçi filtreleme modeli uygulaması yapmışlardır. Bu çalışmada, kullanıcı tabanlı ve öge tabanlı filtreler kullanarak filmler için öneri sistemi geliştirmiş ve MovieLens veri kümesi dikkate alınarak öneri görevlerinde uygulanan çeşitli benzerlik ölçütlerinin performansı karşılaştırılmıştır [12].

## Çizge veri tabanı

Geleneksel veri tabanlarının uygulama alanlarındaki gereksinimleri karşılama konusundaki sınırlamaları, Graf ya da çizge veri tabanı adı verilen teknolojilerin ortaya çıkmasına yol açmıştır. Çok büyük ölçüde ve aralarındaki ilişkilerin karmaşık olduğu verilerin gün geçtikçe artması Çizge veri tabanlarının kullanımını şart haline getirmektedir. Graflar ile verilerin incelenmesi ve analizi birçok farklı disiplin ve alt disiplinde uygulanmaktadır. Böylece iletişim ağlarında optimum yolların seçimi, nesnelere arasındaki ilişkilerin tanımlanması, sistemi oluşturan bileşenlerin temsili gibi çok sayıda farklı problemin aşılmasına ve/veya çözülmesine olanak sağlar. Bilgisayar bilimlerinden fizik, kimya ve biyolojiye, hatta ekonomiden tarihe kadar geniş bir kullanım alanına sahiptir [13]. En genel ifade ile çizge modelleme varlıkların temsillerinin ve bu varlıkların arasındaki ilişkilerin düğüm ve çizgi kullanarak gösterilmesidir.

Çizge, kenarlardan ve tepe noktalarından oluşan bir veri yapısıdır [14]. Çizge veri tabanı teknolojisi, bir veri modelinin tasarımında varlıklar arasındaki ilişkiye odaklanma önemli bir etken olduğundan, verileri modellemek için etkili bir araçtır [15]. Nesnelere ve aralarındaki ilişkileri modellemek, hemen hemen her şeyin karşılık gelen bir Çizgeye temsil edilebileceği anlamına gelmektedir. Çizge veri tabanı, yoğun ve birbirleriyle ilişkili veri kümelerinin verimli bir şekilde işlenmesi için optimize edilmiştir [14]. Bu tasarım, tahmine dayalı modellerin oluşturulmasına ve korelasyonların ve modellerin saptanmasına izin verir [15]. Tüm düğümlerin ilişkilerle birbirine bağlandığı bu son derece

dinamik veri modeli, köşeler arasındaki kenarlar boyunca hızlı geçişlere izin verir. Geçişlerin yerleştirilmiş olması ve alakasız veri kümelerini hesaba katmak zorunda olmaması özel bir avantajdır [16]. Şekil 2’de insanlar ve nesnelar arasında bulunan ilişkileri temsil eden örnek bir graf model yer almaktadır. Graf, düğümler ve bu düğümler arasındaki ilişkilerden oluşan bir nesnedir. Her düğümün kendine ait özellikleri vardır ve yine özellikleri bulunan ilişkiler tarafından organize olup, düzenlenirler. Graflarda ilişkilerin yönünü belirten oklar bulunur. Şekil 3’te bir grafın ayrı ayrı bileşenleri yer almaktadır ve bu bileşenlerin aralarında nasıl bir ilişki bulunduğunu gösteren bir örnek bulunmaktadır.



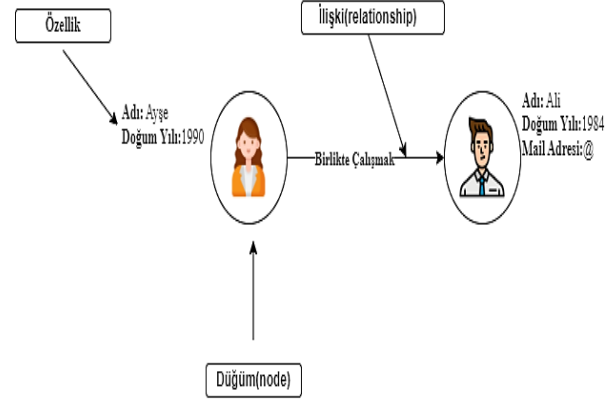
Şekil 2. Örnek graf model görünümü

Yazılımcıların, kodlama aşamasına geçmeden önce veri modelini sabitlemek yerine veri modelinin artarak değiştiğini düşünmelerini sağlayacak bir araca sahip olmaları gerekmektedir. Graf veri tabanları, mevcut bulunan sorgularda bir değişikliğe gitmeden yeni ilişkilerin, düğümlerin ve özelliklerin eklenebilmesine izin vermektedir. Yeni verilerin daha kolay eklenebilmesi daha karmaşık sorgulamanın yapılabilmesine de imkân tanır. Temel veri modelinde herhangi bir değişiklik olması halinde geçişler daha kolaydır. Graf veri tabanlarında şema bulunmadığından, şemalar veri tabanı uygulaması tarafından belirlenir ve bu sayede daha hatasız olarak doğrulanırlar.

### Neo4j çizge veritabanı

Neo4j, Nosql ve ilişkisel veri tabanlarının en güçlü yanlarını bünyesinde bulunduran çok büyük sayıda veri ve veri ilişkilerinden oluşan, graph veri tabanıdır. Neo4j veri tabanı, matematikte yer alan graph teorisinden esinlenilmiş ağ odaklı bir veri tabanıdır. İlişkisel ve nosql veri tabanlarının güçlü yönlerini birleştirmek hedeflenmiştir. Verileri tablolar yerine ağlarda yapılandırılmış durumda depolamaktadır. Neo4j, ACID (Atomicity, Consistency, Isolation, Durability) prensiplerini barındıran bir veri tabanıdır. Yapılan işlemlerden herhangi bir işlemin başarısız gerçekleşmesi durumu ile karşılaştığında veri tabanının durumunu değiştirmeden bırakarak atomikliği sağlamaktadır. Veri tabanında yapılan herhangi bir değişiklik verilere zarar vermez veya yok etmez böylece tutarlılığı sağlamaktadır. Gerçekleştirilen bir işlem sonucu değişen veriler, işlenen

Graf veri tabanları, veri bağlantısı veya topoloji hakkındaki bilgilerin önemli olduğu alanlarda çalışırken ciddi derecede ön plana çıkmaktadır. Bu tür uygulamalarda veri ile verinin kendisi arasındaki ilişkiler genellikle aynı düzeydedir [17]. Graf veri tabanları sosyal graflarda, öneri sistemlerinde, bioinformatik vb. alanlarda kullanılabilir. Yeni nesnelar ile kullanıcının daha önce ilgilendiğini belirttiği nesnelar arasındaki ilişkilere bakılarak kullanıcıya ürünler önerilebilir. Esneklik ve hızlilik, iş gereksinimlerinin devamlı olarak değiştiği günümüzde en önemli unsurlardandır



Şekil 3. Örnek graf bileşenleri

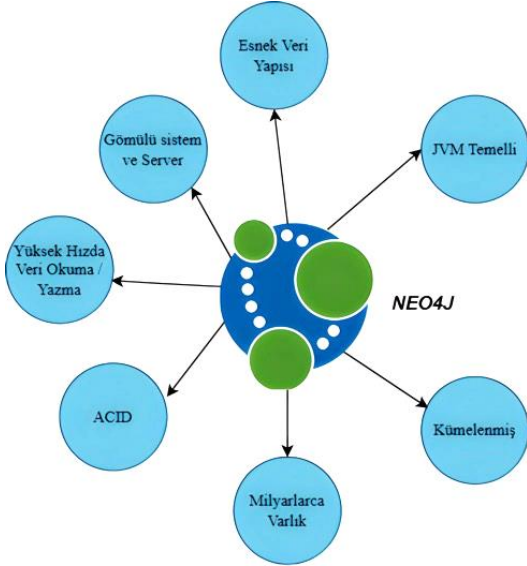
kadar diğer işlemlerden yalıtılır ve böylece izole edilmesi sağlanır. Neo4j kalıcı bir Çizge veri tabanı olması nedeniyle, taahhüt edilen bir işlemin çıktılarını her koşulda alınabilir, bu sayede dayanıklı olması sağlanmaktadır.

Her ilişkinin bir ilişki türü olmalıdır. İlişkilerde, özellikler bir anahtar/değer çifti olabilir. Hem düğümler hem de ilişkiler özelliklere sahip olabilir, bu özellikler ilişkileri ölçekbilir. Bir düğüm için sıfır veya daha fazla etiket atanabilir. Bu etiketler rolleri, kategorileri veya türleri temsil etmek için kullanılan etiketlerdir. Düğümlerin indekslerini ve kısıtlamalarını tanımlamak için kullanılır. Neo4j veri tabanı, Çizge veri tabanının daha hızlı ve verimli bir şekilde sorgulanmasına, güncellenmesine ve silinmesine olanak tanıyan bir sorgulama dili olarak cypher'ı kullanır [18]. Cypher sorgulama dili bir Çizgeten hangi bilgilerin alınması gerektiğine odaklanır. Kullanım olarak basit bir dil olsa da çok güçlü bir sorgulama dildir. Karmaşık veri tabanı sorguları kolayca Cypher sorgulama dili ile yazılabilir. Şekil 4’te Neo4j’nin genel özellikleri yer almaktadır. Şekil 5’te ise Neo4j bileşenleri ve bu bileşenler arasındaki ilişkiler gösterilmiştir. Neo4j çizge veri tabanının önemli özellikleri aşağıda sıralanmıştır.

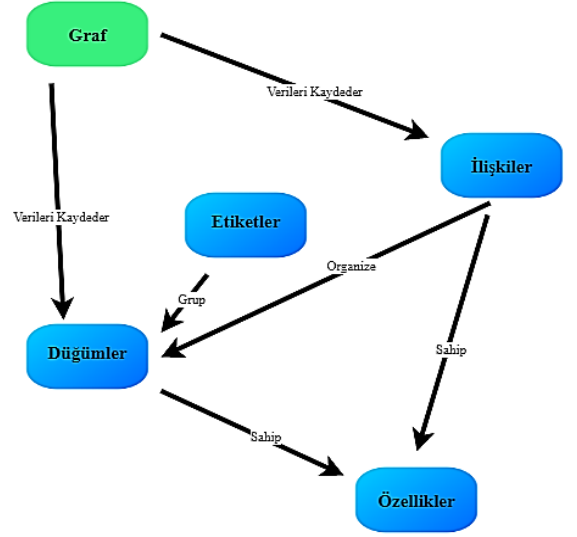
- Kolay ve esnek bir veri yapısına sahip olması nedeniyle uygulamanın ihtiyaçlarına göre kolayca şekillendirilebilir bir yapıdadır.
- Gerçek verilere dayanan sonuçlar sunmaktadır.
- Şemalar içermez, bağlantılı verilerin düğümler ile basit temsilini sunar.

- Çizgeleri yerel olarak depolayıp işlediği için yüksek performans gösterir.
- Muadil veri tabanlarına göre daha hızlı sonuçlar üretir ve düğümler arasında daha hızlı bir geçiş/gezime imkanı sunar.
- ASCII-art sözdizimi kullanılarak Çizge olarak görselleştirmek için Cypher sorgulama dili kullanılır. Cypher sorgulama dilinin komutlarının anlaşılması ve okunması kolaydır.

- Neo4j en büyük ve aktif bir Çizge topluluğuna sahiptir.
- Veri bütünlüğü korunurken yüksek hızda okuma ve yazma performansı sunmaktadır.
- Eğitim ekosistemi, eğitim materyali ve yazılmış kitaplar sayesinde kolay öğrenilebilir.
- Düşük bellek kullanımı ile büyük veri boyutlarını yüksek hızda işe aktarır.



Şekil 4. Neo4j veritabanı genel özellikler



Şekil 5. Neo4j veritabanı bileşenleri ve ilişkiler

## Metodoloji ve uygulama

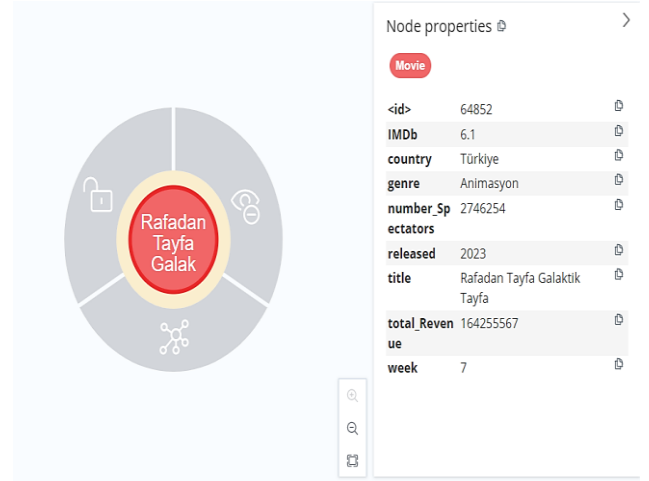
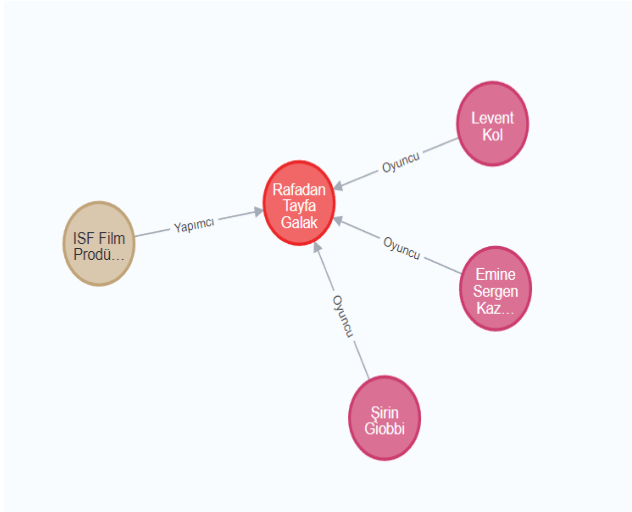
Neo4j'de veri modeli düğüm, düğümler arası ilişkiler ve özelliklerden oluşmaktadır. Düğümler nesnelere ve düğümler farklı ilişkilerle birbirine bağlıdır. Çalışmamızda film bir düğümdür. Filmin türü, vizyon tarihi vb. bilgiler ise filmin özellikleridir. Filmler, yönetmenler ve oyuncular depolanacak olan farklı varlıklardır. Veri tabanında saklanması gereken yalnızca farklı varlıklar değildir, bu varlıklar arasındaki ilişkilerin de veri tabanında depolanması gerekmektedir.

Neo4j'de veri tabanı oluşturmak kolaydır. Düğümleri, ilişkileri ve ilişkiler arası özellikleri veri tabanına eklemek yeterlidir. Aşağıdaki örnekte film verilerinin yer aldığı örnek bir Neo4j veri tabanının nasıl oluşturulacağı yer almaktadır. Şekil 6'da örnek kod bloğu, Şekil 7'de kod bloğunun çıktısı yer almaktadır. Şekil 8'de örnek filme ait özellikler yer almaktadır.

```
neo4j$
1 CREATE (RafadanTayfaGalaktikTayfa:Movie {title:'Rafadan Tayfa Galaktik Tayfa', released:2023, genre:'Animasyon',
2 number_Spectators:'2746254', total_Revenue:
3 164255567, IMDb:6.1, week:7, country:'Türkiye'})
4 CREATE (ISFFilmProdüksiyon:Yapımcı {name:'ISF Film Prodüksiyon'})
5 CREATE (ŞirinGiobi:Oyuncu {name:'Şirin Giobbi'})
6 CREATE (LeventKol:Oyuncu {name:'Levent Kol'})
7 CREATE (EmineSergenKazbek:Oyuncu {name:'Emine Sergen Kazbek'})
8 CREATE
9 (ŞirinGiobi)-[:Oyuncu]→(RafadanTayfaGalaktikTayfa),
10 (LeventKol)-[:Oyuncu]→(RafadanTayfaGalaktikTayfa),
11 (EmineSergenKazbek)-[:Oyuncu]→(RafadanTayfaGalaktikTayfa),
12 (ISFFilmProdüksiyon)-[:Yapımcı]→(RafadanTayfaGalaktikTayfa)
```

Şekil 6. Uygulamanın Neo4j örnek kesit kod bloğu

Şekil 7. Neo4j örnek kod bloğu çıktısı



Şekil 8. Örnek film düğüm özellikleri

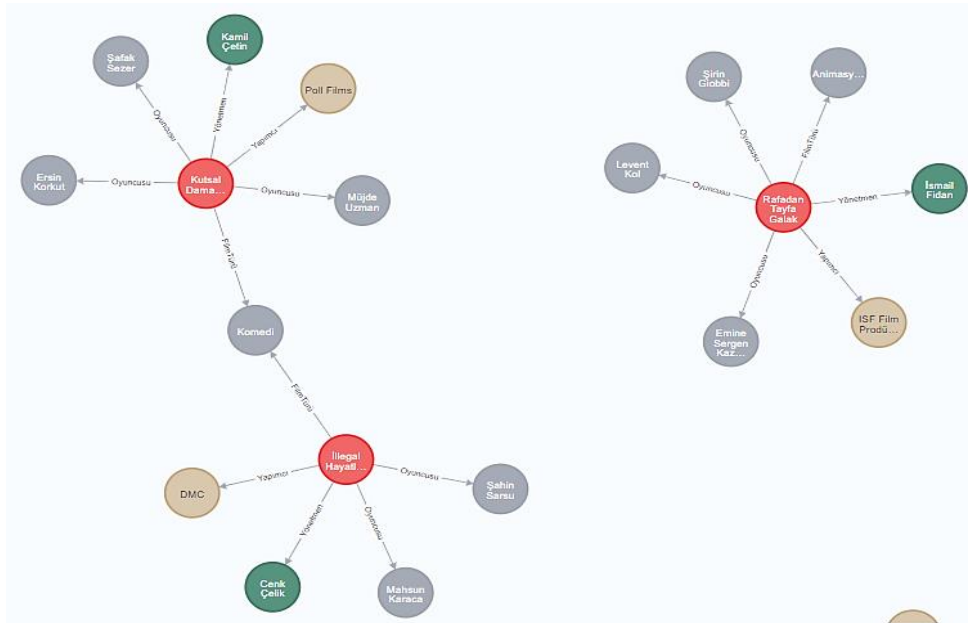
Neo4j çizge veri tabanında yer alan etiketler, düğüm koleksiyonlarını tanımlamak için kullanılmaktadır. Etiketler sayesinde kısıtlamalar, sorgulamalar yapılmaktadır. Bu örnekte, 'Rafadan Tayfa Galaktik Tayfa' adında bir film yaratılır. Özelliklerine bakılarak filmin gösterimde 7 hafta kaldığı, 2023 yılında vizyona girdiği vs. bilgiler görülmektedir.

```
(ŞirinGiobbi)-[:Oyuncu]->(RafadanTayfaGalaktikTayfa)
```

Yukarıda yer alan kod Şirin Giobbi'nin Rafadan Tayfa Galaktik Tayfa'da oyuncu olduğunu tanımlamaktadır.

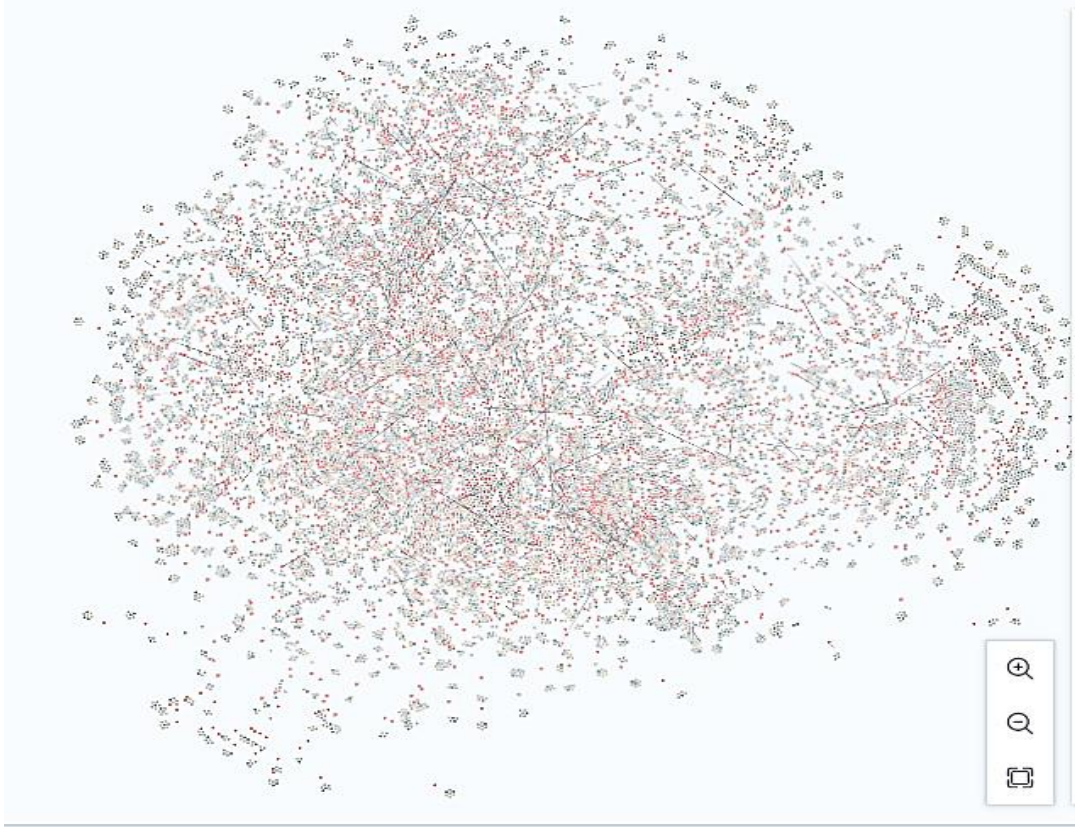
## Veri seti

Bu çalışma da 3440 film ve 25787 ilişkidenden oluşan bir veri tabanı oluşturulmuştur. Şekil 9'da sorgu limiti 25 verilerek veri tabanından küçük bir bölüm gösterilmektedir. Kırmızı düğümler filmleri, yeşil düğümler yönetmenleri, gri düğümler oyuncularını ve kahverengi düğümler yapımcıları



temsil etmektedir. Şekil 10'da çalışmada yer alan tüm düğüm ve ilişkiler yer almaktadır. Film verilerini analiz etmek için ana yöntem, veri tabanını dolaşarak sorgulamaktır. Neo4j, çok sayıda bağlantı içeren geleneksel RDBMS sorgularındaki performans düşüşü sorununu çözmeye odaklanır. Veri tabanı, düğümleri ve kenarları aynı hızda hareket ettirecektir [19]. Cypher aslen Neo Technology tarafından Çizge veri tabanı Neo4j için yaratılmıştır. Neo4j, çok karmaşık Çizge geçiş işlemlerini destekler, Cypher sorgulama dili, veri tabanından veri sorgulamayı oldukça kolaylaştırır. Cypher çeşitli yan tümceler içerir. En yaygın olanları şunlardır: MATCH ve WHERE. Bu işlevler SQL'dekinden biraz farklıdır. MATCH, öncelikle ilişkilere dayalı olarak aranan örüntünün yapısını açıklamak için kullanılır. WHERE kalıplara ek kısıtlamalar eklemek için kullanılır. Cypher ayrıca veri yazmak, güncellemek ve silmek için maddeler içerir. CREATE ve DELETE, düğümleri ve ilişkileri oluşturmak ve silmek için kullanılır [20].

Şekil 9. Film veri tabanı limitli sorgu sonucu

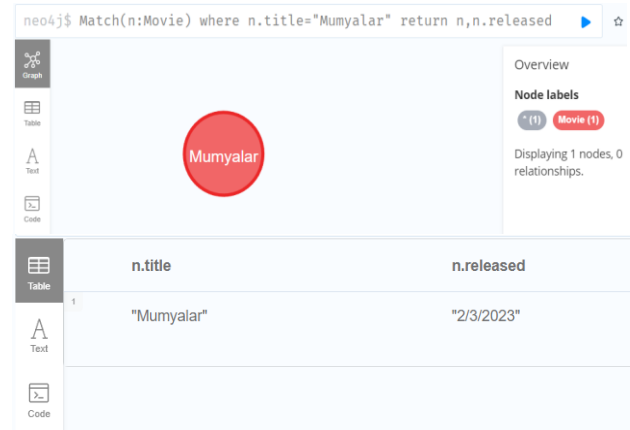


Şekil 10. Film veri tabanı limitsiz sorgu sonucu

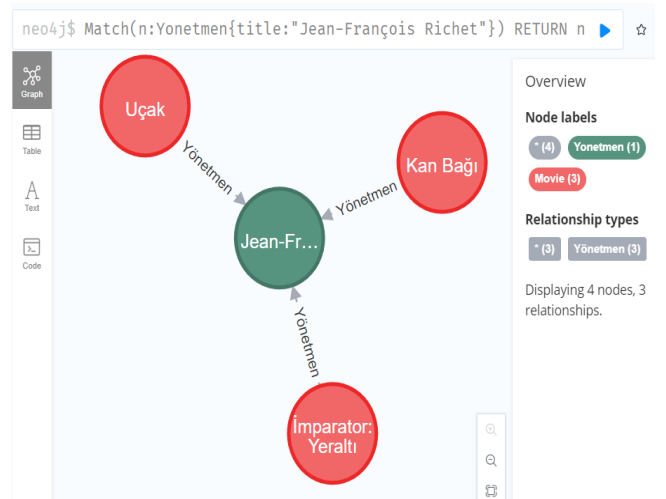
Filmin sadece adı bilindiğinde ve diğer özellikleri sorgulanmak istendiğinde, düğümün özelliklerinin sorgulanması gerekir. Bir düğümün özellikleri sorgulanmak istendiğinde kullanılacak Cypher sorgusu Şekil 11’de yer almaktadır. Şekil 12’de ise sorgunun sonucu bulunmaktadır.

```
1 Match (n:Movie)
2 Where n.title="Mumyalar"
3 Return n, n.released
```

Şekil 11. Örnek Cypher sorgusu ve çıktısı



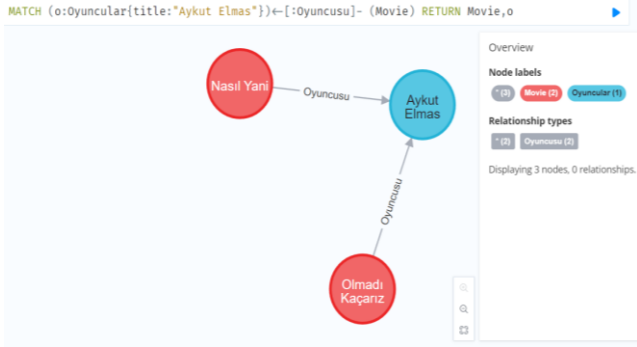
Şekil 12. Örnek Cypher sorgu sonucu



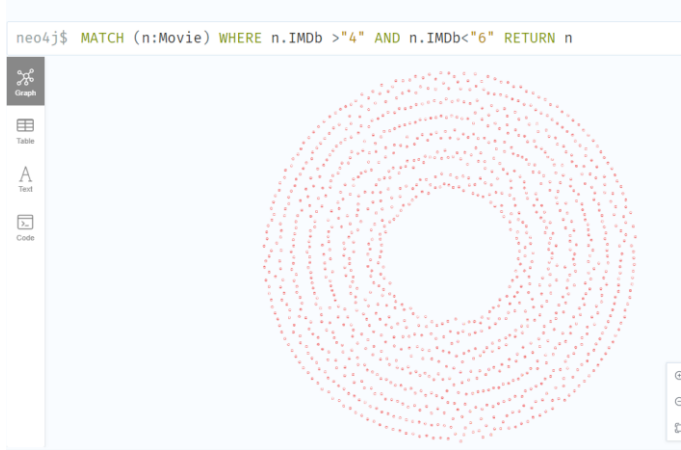
Şekil 13. Örnek yönetmenin yönettiği filmler sorgu sonucu

Şekil 13’te örnek bir yönetmen adı girilerek veri tabanında yönettiği filmleri tek bir sorgu ile bulabiliriz. Örnek koddaki “Jean-François Richet” adlı yönetmenin ülkemizde son 10 yılda gösterime giren 3 filmde yönetmenlik yaptığı görüntülenmektedir. Şekil 14’te ise örnek bir oyuncu adı girilerek veri tabanında oynadığı 2 film görüntülenmektedir. Şekil 15’te IMDb puanı 4 ve 6 arasında olan filmleri görüntülemekteyiz. Veri tabanımızda IMDb puanı 4 ve 6 arasında olan 1175 filmin yer aldığını görüntülemekteyiz. Limit değerini sorguda 10 girerek IMDb puanı 4 ve 6 arasında

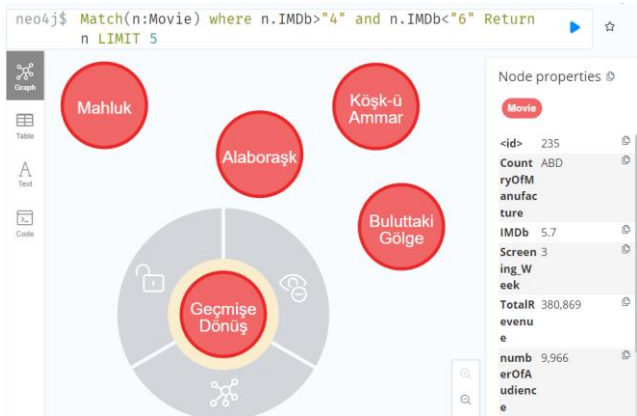
olan ilk 10 filmi görüntülenebilir. Bu filmlerden birini seçerek sorgu sonucu doğrulanmaktadır. Şekil 16'da limit değeri 5 verilerek "Geçmişe Dönüş" isimli filmin IMDB puanının "5.7" ve 4-6 aralığında olduğu görüntülenmektedir. Şekil 17'de 8 haftadan fazla gösterimde kalan filmler sorgulanmaktadır. Sorgu sonucunda veritabanında 8 haftadan fazla gösterimde olan film sayısının 162 olduğu görülmektedir. Şekil 18'de örnek olarak "Taşyıcı Son Hız" filmine bakıldığında 9 (dokuz) hafta gösterimde kaldığı görülmektedir.



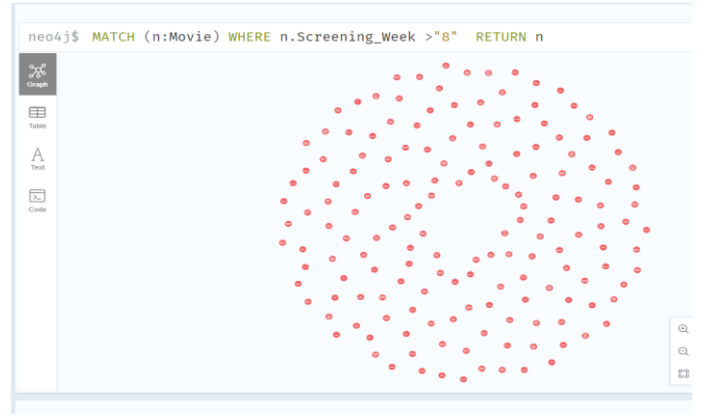
Şekil 14. Örnek oyuncu ve oynadığı filmler sorgu sonucu



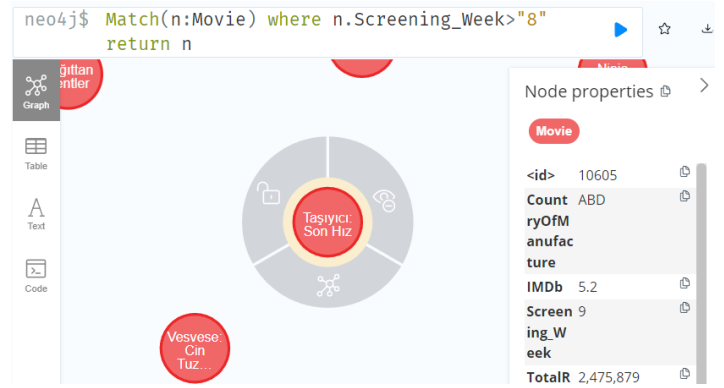
Şekil 15. IMDB puan sorgu sonucu



Şekil 16. IMDB puan limitli sorgu sonucu

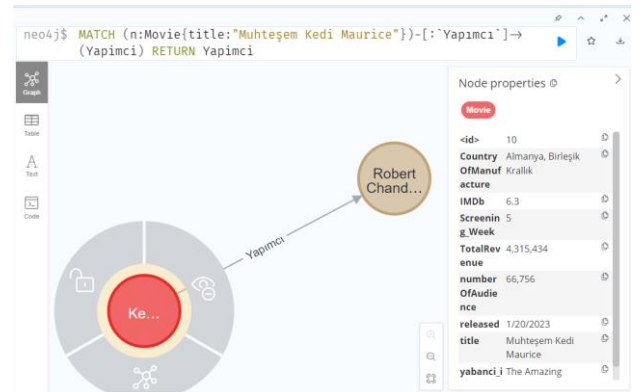


Şekil 17. Gösterim haftası sorgu sonucu



Şekil 18. Gösterim haftası detaylı sorgu sonucu

Bir film adı girilerek o filmin sadece oyuncusu veya sadece yapımcısı vb. özellikleri bulunabilmektedir. Şekil 19'da örnek bir film adı girilerek yapımcısı sorgulanmıştır. Veritabanında bulunan filmler türlerine göre sınıflandırılabilir. Şekil 20'de örnek olarak film türü dram olan filmler sorgulanmıştır ve sorgu sonucu olarak türü dram olan 1276 adet film bulunmuştur.



Şekil 19. Örnek Cypher sorgu sonucu

Çalışmada kullandığımız veri tabanımızın veri profilini Neo4j ile görüntüleyebilmekteyiz. Veritabanında 16265 düğüm ve 25787 ilişki bulunmaktadır.



```
neo4j$ MATCH (n:Movie)-[:`FilmTürü`]-> (f:Film_Turu{title:"Dram"}) RETURN n.title,f.title
```

n.title	f.title
"Kabile"	"Dram"
"Hile Yolu"	"Dram"
"Sıradışı Anne"	"Dram"
"Birdman veya (Cahilliğin Umulmayan Erdemi)"	"Dram"
"Gece Hayvanları"	"Dram"
"Para Tuzağı"	"Dram"

Started streaming 1276 records after 3 ms and completed after 8 ms.

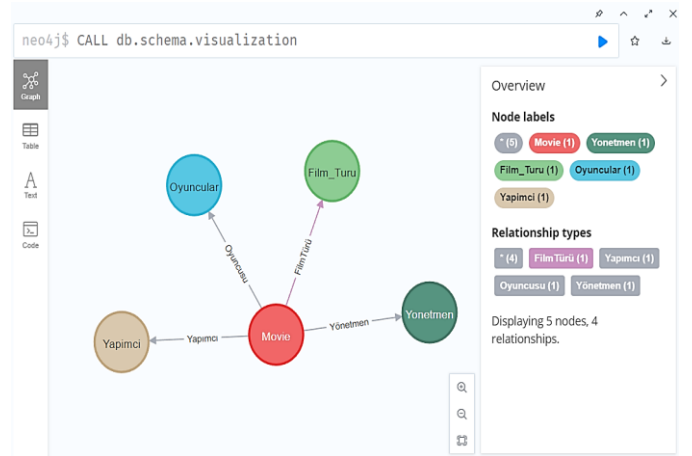
Şekil 20. Örnek film türü sorgu sonucu

```
neo4j$ MATCH (n) RETURN count(n)
```

count(n)
16265

Şekil 21. Düğüm sayısı sorgu sonucu

Şekil 21 ve Şekil 22’de sorgu sonuçları görüntülenmektedir. Şekil 23’te veriler arasındaki şema sorgulanmıştır ve sorgu sonucunda oluşan şema yer almaktadır. Çalışmamızda 5 adet düğüm vardır ve Şekil 24’te sorgulanmış düğüm etiket listesi yer almaktadır.



Şekil 23. Veri şeması sorgu sonucu

```
neo4j$ MATCH ()->() RETURN count(*)
```

count(*)
25787

Şekil 22. İlişki sayısı sorgu sonucu

```
neo4j$ CALL db.labels()
```

label
"Movie"
"Yapımcı"
"Yönetmen"
"Film_Turu"
"Oyuncular"

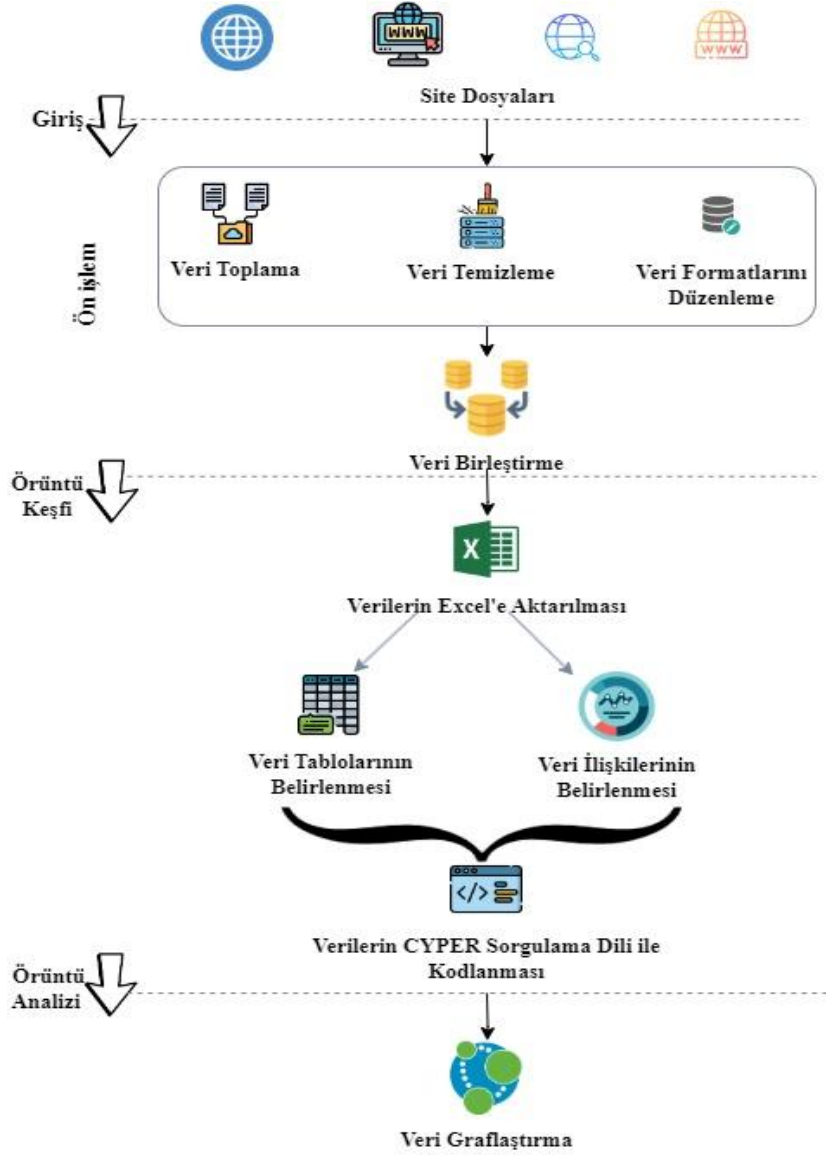
Started streaming 5 records after 2 ms and completed after 2 ms.

Şekil 24. Düğüm etiket listesi sorgu sonucu

Çalışmada izlenen uygulama adımları giriş, örüntü keşfi ve örüntü analizi olmak üzere 3 adımda incelenerek bu aşamalar Şekil 25’te yer almaktadır. Giriş bölümünde verilerin nasıl ve nereden bulunduğu gösterilmektedir. Ön işlem aşamasında,

web sitelerinde bulunan karmaşık düzensiz verilerin belirli bir düzen içerisinde kullanıma hazır hale getirmek için arındırılması, temizlenmesi ve dönüştürülmesi aşamaları gösterilmiştir. Verilere uygulanan bu işlemler sonrasında elde edilen kullanılacak olan veriler birleştirilmiştir. Birleştirilen

veriler Excel dosyasına aktarılmıştır. Bu verilerin tabloları ve veriler arasındaki ilişkiler belirlenmiştir. Veriler CYPHER sorgulama dili kodlanarak Neo4j veri tabanına aktarılmıştır.



Şekil 25. Neo4j çizge veritabanı ile sinema verilerinin çizge modellenmesinin uygulama adımları

### Veri toplanması

Büyük ölçekli film verilerini elde etmek zorlu bir süreçtir. Bulunan verilerin içerisinde ilgili kısımları elde etmek daha zor bir süreçtir. Elde edilen bu verilerin anlaşılabilirliği ayrı bir zorluktur. Öncelikle verilerin toplanmasında hangi kaynaklardan hangi türlü verilerin toplanacağı belirlenmelidir. İlk zorluk hangi verilerin toplanacağına karar vermektir.

### Veri temizlenmesi

Karmaşık ve büyük ölçekli film verilerinin toplanmasından sonraki adım, verilerin temizlenmesi ve kullanılabilir hale getirilmesidir. Farklı formatlarda ve büyük boyutlarda çeşitli parametreler bu film verilerinin temizlenmesi oldukça zor bir

süreçtir. Elde bulunan bu veriler Neo4j veri tabanına aktarılmadan önce temizlenmeli ve hazırlanmalıdır. Geçersiz, eksik ve tutarsız verileri düzenlemek için birçok yaklaşım bulunmaktadır. Bu sorunlu kayıtlardaki eksik değerleri bulmak veya verileri normalleştirmek faydalı bir yaklaşımdır.

### Veri organizasyonu

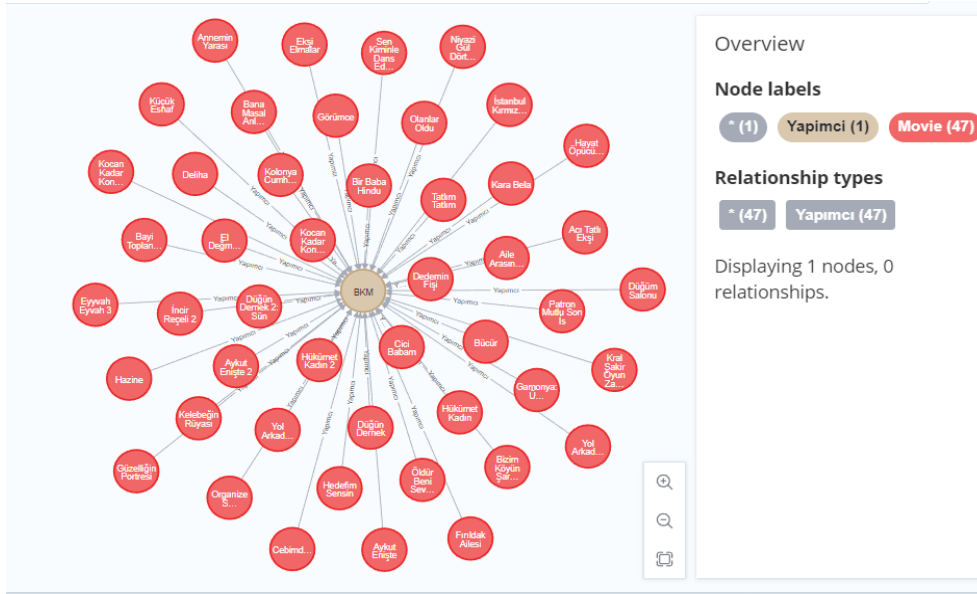
Film verilerini düğüm ve kenar şeklinde tanımlanır. Bu tanımlama veri anlaşılabilirliğini mümkün kılacaktır. Film verileri, oyuncular, yönetmenler, yapımcılar ve bunların arasındaki ilişkilerden oluşur. Film adı, yönetmen, aktörler vb. özellikler düğüm olarak kullanılır ve toplam hasılat, seyirci sayısı, gösterim yılı vb. özellikler varlıkların nitelikleri olarak kullanılır.

## Çizge görselleştirme

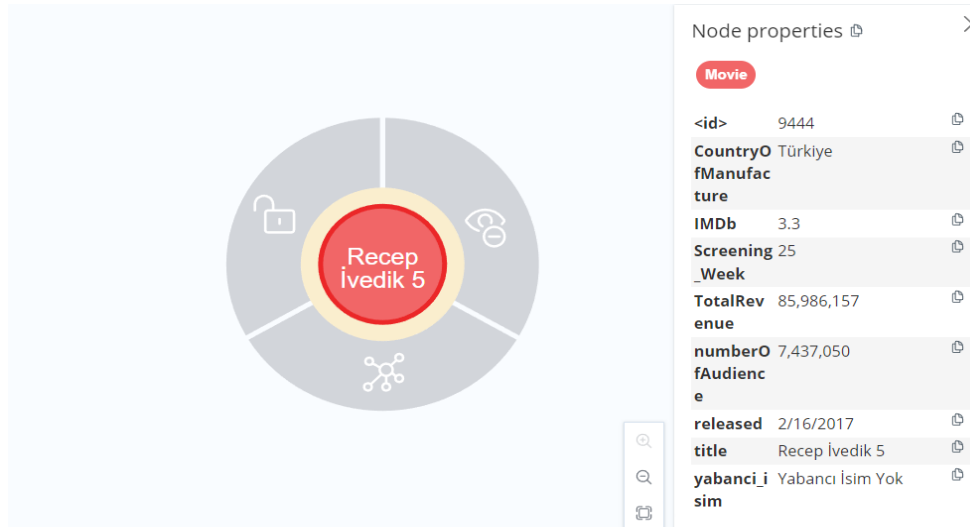
Çizge görselleştirme aşamasında önemli görülen belirli özellikleri vurgulamak için çeşitli teknikler kullanılabilir. Grafiğin anlaşılmasını geliştirmek için renklendirme, soldurma ve etiketleme gibi teknikler kullanılabilir [21]. Veri toplama, temizleme ve verilerin organizasyonu adımlarından sonra film bilgileri, oyuncular, yapımcılar ve diğer faktörleri içeren bilgi grafiği ve veri görselleştirme işlemi yapıldı. Veri görselleştirme işlemi için Neo4j veri tabanı kullanıldı. Kırmızı düğümler filmleri, yeşil düğümler yönetmenleri, gri düğümler oyuncuları ve kahverengi düğümler yapımcıları temsil etmektedir. Film verileri toplamak için 01.01.2013 - 13.03.2023 tarihleri arasında Türkiye’de gösterime girmiş filmler araştırıldı, elde edilen sonuçlar incelendi. İnceleme sonrası çalışmada kullanılacak film verilerine karar verildi ve gereksiz veriler silme, veriler arasındaki gereksiz boşluklar

kaldırma vs. işlemlerden geçirilerek veriler temizlendi. Veri tabanının düğüm ve aralarında hangi ilişkiler olacağı belirlenerek veri organizasyonu gerçekleştirildi.

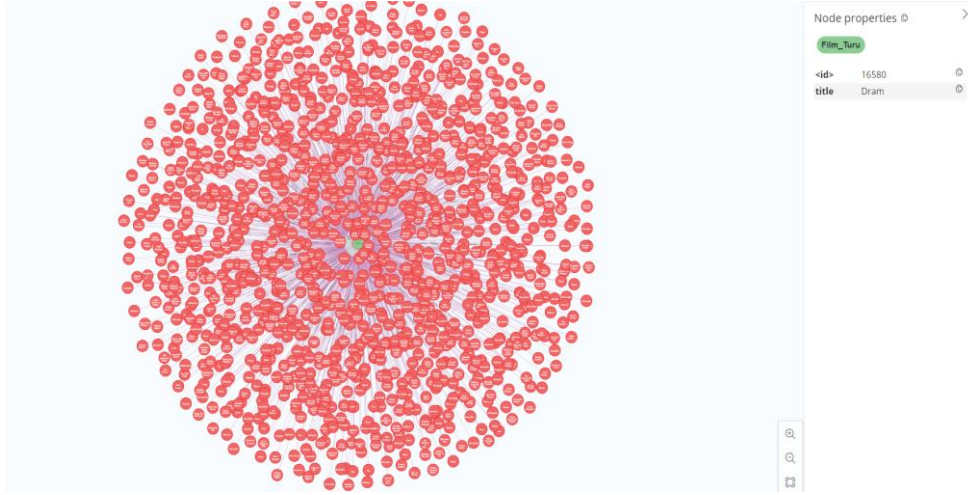
Çalışmamızda literatürden farklı olarak veri tabanında 01.01.2013 - 13.03.2023 tarihleri arasında Türkiye’de gösterime girmiş sinema filmleri yer almaktadır. Film analizleri sonucunda ele alınan yıllar arasında en çok film yapan yapımcı 47 film ile “BKM” olmuştur ve sorgu sonucu Şekil 27’de yer almaktadır. En fazla filmin gösterime girdiği yıl 429 film ile 2018 yılı olmuştur. Çalışmada kullanılan tarihler arasında en fazla seyircisi olan film 7437050 kişi ile “Recep İvedik 5” olmuştur, sorgu sonucu Şekil 27’de yer almaktadır. En fazla gösterime giren film türü 1276 film ile “Dram” türü olmuştur ve sorgu sonucu Şekil 28’de yer almaktadır.



Şekil 26. En çok film yapan yapımcı



Şekil 27. En fazla seyircisi olan film



Şekil 28. Gösterime giren en fazla film türü

## Sonuç

Günümüzde her alanda olduğu gibi film sektöründe de veriler katlanarak büyümektedir, bu verilerin arasındaki ilişkiler de giderek daha karmaşık bir hale gelmektedir. Genellikle sadece verinin kendisine odaklanmak değil, verilerin arasında yer alan ilişkilere de odaklanmak gerekmektedir. Geleneksel veri tabanları bu talepleri karşılayamamaktadır. Sinema film verilerini depolamak ve bu verileri yönetmek için Neo4j veri tabanını kullanmak, bir film web sitesi için kullanıcıların gereksinimlerini karşılamak ve ilgi alanlarını bulmak gibi işlemleri kolaylaştırır. Verilerin yönetimi söz konusu olduğunda, Neo4j veri tabanının NoSQL veya MYSQL veri tabanlarına güçlü bir alternatif olacağı sonucuna varabiliriz. Daha önce de belirtildiği gibi, verilerin yönetimi ve alımı, istenilen çıktıyı elde etmek için çok hızlı ve görsel olarak çok hassastır. Toplam hasılat, seyirci sayısı gibi özelliklere bakılarak yeni çıkarılacak sinema filmleri hakkında çıkarımda bulunulabilir, bu veriler yol göstermede kullanılabilir. Son yıllarda sinema sektörü için çok ciddi bütçeler ayrılarak filmler yapılmaktadır ve dolayısıyla film yapımcıları veya film sponsorları bu graf modellemelerini kullanarak daha net verilere ulaşabilir ve elde edilen veriler ışığında daha kolay yoldan seyircilerin hangi yıllarda hangi tür filmlere talep gösterdiği, hangi yönetmenin filmlerinin ya da hangi oyuncuların rol aldığı filmlerin daha çok izlendiği bilgilerine ulaşabilir. Film sektörüne yapılacak yatırımlarda nasıl bir yol izlenilmesi gerektiği konusunda yol gösterici olabilir.

## Kaynakça

- [1] S. Sencer and K. Eren, "Graph Database for Agent Based Emergency Response Model", Proceedings of the 2014 international conference on advances in big data analytics, July 21-24, 2014.
- [2] M.Sülü, R. Daş, "Graph visualization of cyber threat intelligence data for analysis of cyber attacks", Balkan Journal of Electrical and Computer Engineering (BAJECE), (2022),10(3), 300-306.
- [3] Z. Zhu X.Zhou, K. Shao, A novel approach based on Neo4j for multi-constrained flexible job shop scheduling problem. Computers & Industrial Engineering, 2019, 130: 671-686.
- [4] A. Chen, A novel graph methodology for analyzing disease risk factor distribution using synthetic patient data. Healthcare Analytics, 2022, 2: 100084.
- [5] H. Hu, M. Fang, Y. Zhang, L. Jing and F. Hu, "Dynamic lightning protection method of electric power systems based on the large data characteristics", Int. J. Elect. Power Energy Syst., vol. 128, no. 1, pp. 1-14, 2021.
- [6] Y. Shi, et al. A knowledge graph constructed for job-related crimes. Procedia Computer Science, 2022, 199: 540-547.
- [7] M. Kuhn, E. T. Kaminski and J. Franke, "Track and Trace: Integrating static and dynamic data in a hybrid graph based traceability model", Procedia CIRP, vol. 112, pp. 250-255, 2022.
- [8] H. Lu, Z. Hong and M. Shi, "Analysis of film data based on Neo4j", 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), pp. 675-677, 2017.
- [9] Q. Shuai and C. Zhang, "Question Answering system based on Knowledge Graph of Film Culture", 2020 International Conference on Culture-oriented Science & Technology (ICCST), pp. 150-153, 2020.
- [10] N. Yi, C. Li, X. Feng and M. Shi, "Design and implementation of movie recommender system based on graph database", 2017 14th Web Information Systems and Applications Conference (WISA), pp. 132-135, 2017.
- [11] M. Goyani and N. Chaurasiya, "A Review of Movie Recommendation System", ELCVIA: electronic letters on computer vision and image analysis, vol. 19, no. 3, pp. 18-37, 2020.
- [12] C. K. Raghavendra and K. C. Srikantaiah, "Similarity Based Collaborative Filtering Model for Movie Recommendation Systems", 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1143-1147, 2021.
- [13] R. Das and M. Soylu, "A key review on graph data science: The power of graphs in scientific studies", Chemometrics and Intelligent Laboratory Systems, c. 240, sy 104896, Haz. 2023
- [14] M. Rodriguez, A. Marko and P. Neubauer, "Constructions from dots and lines", Bulletin of the American Society for

- Information Science and Technology, vol. 36, no. 6, pp. 35-41, 2010.
- [15] A. Silvescu, D. Caragea and A. Atramentov, "Graph Database", Artificial Intelligence Research Laboratory Department of Computer Science Iowa State University, 2012,
- [16] M. A. Rodriguez and P. Neubauer. The graph traversal pattern. Graph Data Management: Techniques and Applications, 2011.
- [17] C. Berge, The theory of graphs. Courier Corporation, 2001.
- [18] G. M. D'silva, S. Thakare and V. A. Bharadi, "Real-time processing of IoT events using a Software as a Service (SaaS) architecture with graph database. In: Computing Communication Control and automation (ICCUBEA)", 2016 International Conference on, pp. 1-6, 2016.
- [19] Y. Wang, "A Comparative Study of Graph Database NEO4J and Relational Database", Modern Electronic Technology, vol. 35, no. 20, pp. 77-79, 2012.
- [20] L. Wang, "Construction of Document Resources Association Network Based on graph database technology", Digital Library Forum, pp. 59-65, 2014.
- [21] M. Soylu, A. Soylu, ve R. Das, "A new approach to recognizing the use of attitude markers by authors of academic journal articles", Expert Systems with Applications, c. 230, sy 120538, Kas. 2023