**Neo4j Ödev:** Movies Dataset ile Graph Keşfetme ve Cypher Sorguları

**Hazırlayan:** Mert OLÇAMAN

**Hafta:** 1

**patika.dev**

**NewMind AI Bootcamp**

**Bolum 1: Generative AI (GenAI) Hakkında Kısa Bilgiler**

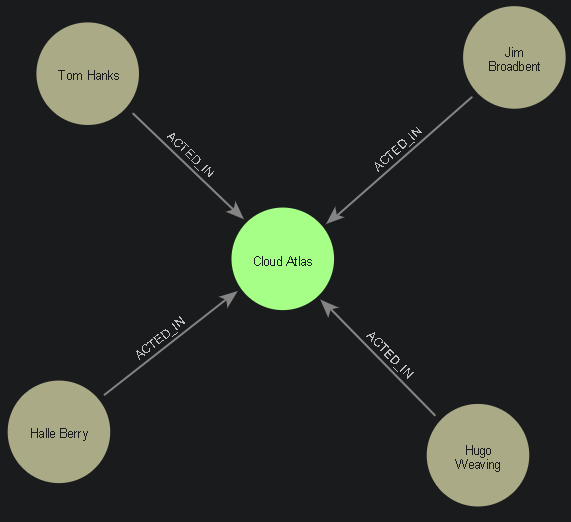
-**Generative AI Nedir?**  
Generative AI, üretken yapay zeka modeli anlamına gelir. Bu model çok fazla dataset ile eğitilir ve böylece verilerin arasındaki ilişkiyi anlayarak yeni çıktılar üretebilir.  
  
-**Hangi alanlarda kullanılır?**  
**Sanat ve Tasarım:** Görsel oluşturma alanında son zamanlarda çok etkili olabilmektedirler. Bu görseller pazarlama, reklam, oyun gibi sektörlerde kullanılabilir. Bazen gerçeğinden ayırmak oldukça zordur. Yine de bazı gerçek dışı oluşumlar farkedilerek bu ayrım sağlanabilir.  
**Text:** Bir çok alan için yazı içeriği üretmede kullanılabilir. Bu yazılar; blog, sosyal medya, iş başvurusu, senaryo ve yaşam koçluğu gibi hayatımızı şekillendirebilecek alanlarda kullanılabilir.  
**Yazılım:** Text kısmına benzer olarak istenilen bir görev için kod yazma yapabilir. Bu sayede program yazmak gerektiğinde, istenilen kod betimlenerek kodun çıktısı direkt olarak alınabilir ve programcıya oldukça hız kazandırır.  
  
-**GenAI ile grafik veritabanları birleştirildiğinde ne tür uygulamalar geliştirilebilir?**  
Grafik veritabanları, klasik veritabanlarına göre ilişkilerin doğrudan daha kolay görülebilmesini sağlar. Generative AI ise veriler arasındaki ilişkileri iyi bir şekilde öğrendiği için, başka bir veri kümesindeki ilişkileri de anlamakta zorluk çekmez. Her ne kadar %100 kusursuz çalışmasalar da bir çok görünmeyen noktayı ortaya çıkartabilir.  
**Sanayi:** Örneğin üretim ve montaj yapan ve birçok operatörün farklı vardiyalarda birçok proses ile birlikte çalıştığı bir fabrika veri seti düşünüldüğünde, üretimin hangi kısmında ne gibi bir sorun var ve neyden dolayı kaynaklı olduğu grafik veritabanı ve GenAI sayesinde kolayca anlaşılabilir. Böyle bir uygulama yapmak mühendislerin işini kolaylaştırır ve bu sorunu nasıl ortadan kaldırabilecekleri konusunda bile Generative AI yardımcı olabilir.  
**E-ticaret:** Bir e-ticaret firmasında, her gün binlerce sipariş verilir ve bunların içinde birçok sebepten kaynaklı sorunlar ortaya çıkabilir. Bunların satıcıdan mı, lojistik kısmından mı, yoksa kullanıcı hatasından mı kaynaklı olduğunu ortaya çıkarmak GenAI ve grafik veritabanları kombinasyonu ile mümkün kılınabilir.  
**Sağlık:** Bazen insanların hastalıklarının asıl sebebinin ortaya çıkarılması testlere bakıldığında mümkün olmayabilir. Bazı hastalar diğer hastalarla benzer özellik gösterirken, bir başkaları diğer gruptaki hastalarla benzer semptomlar ortaya çıkarabilir. Belki bir hastanın geçmişindeki akciğer sorunu, başka bir hastalığın ortaya çıkmasını kolaylaştırabilir ve burada asıl sebebin akciğerden kaynaklı olduğu ve gelecekte insanları bilinçlendirmenin önünü açabilir. Yine ilişkileri incelemek söz konusu olduğu için, Gen AI bu konuda grafik veritabanları ile çok iyi bir iş çıkarabilir.  
  
-**Riskleri veya etik tartışmaları var mıdır?**  
GenAI kullanılarak üretilen birçok görsel ve video içerikleri gerçeğinden ayırmakta zorluk çekilebilecek noktalara gelmiş bulunmaktadır. Bunlar kötü amaçlı insanlar tarafından kullanıldığında tehdit oluşturabilmektedir. Örneğin bir politikacı veya ünlü birisi için sahte bir görsel içerik oluşturularak tüm hayatının zedelenmesine ve hatta tehlikeye girmesine sebep olabilir. Aynı şekilde, kötü birisinin eline geçen ve ünlü güvenilir birisinin yüzünü kullanarak sahte görsel ve video içeriği oluşturularak insanların dolandırılması gibi tehlikeli bir sonuç olabilir.  
Bunun yanında içerik üreticileri sık sık GenAI’dan yardım almaktadırlar ve bazı bilgileri kontrol etmek mümkün olmayabilir. AI tarafından halüsinasyon sebebi ile üretilen sahte bir bilgi çok popüler bir platformda teyit edilmeden paylaşıldığında aynı veya başka bir Generative AI model tarafından örnek data olarak kullanıldığında sahte bilgi paylaşımı artabilir ve insanların doğru bilgiye erişmesi gittikçe zorlaşabilir.  
  
İnsanlar sık sık bu modelleri kullandıkları için bazı yeteneklerini kaybederek tembelleşmeye başlayabilirler. Örneğin önceden kod yazan programcılar bir sorun olduğunda kodu tek tek inceleyip sebebini anlamaya çalışırlardı, fakat artık bu modeller sayesinde sebebi bularak çözüme yönelik kod ürettirebiliyorlar ve yeni öğrenen programcılar bu özellikten dolayı daha az düşünmeye başlıyorlar. Bir fikir üretmek gerektiğinde çoğu zaman üretken yapay zekadan destek almak insanları tembelliğe iterek ileride daha az düşünen ve sorgulayan bir neslin oluşmasına sebep olabilir.

**Bölüm 2: Veri Yükleme ve Explore Arayüzü ile Çalışma**

**Sorgu1:** 

2010’dan sonra vizyona giren filmler ve bu filmlerde rol almış kişileri veren sorgu.

**Görsel1:**

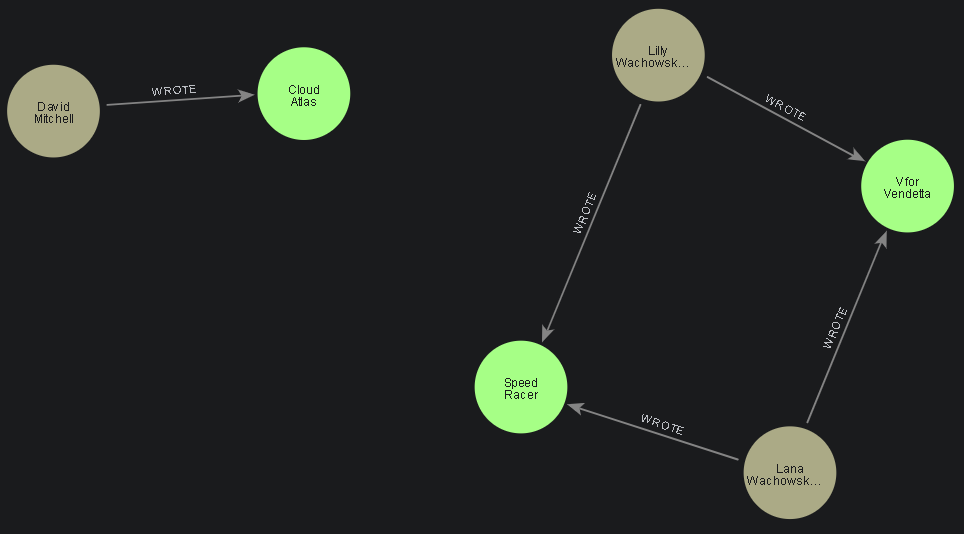


Yukarıdaki görselde, 2010 yılı ve sonrasında vizyona girmiş filmlerde rol almış oyuncular gözükmektedir. Görsele göre, sadece 1 tane film (Cloud Atlas) vardır ve bu oyuncular; Tom Hanks, Jim Broadbent, Halle Berry ve Hugo Weaving’tir.

**Sorgu2: **

1965 ve 1970 yılları arasında doğan kişiler ve yazdıkları filmleri veren sorgu.

**Görsel2:**

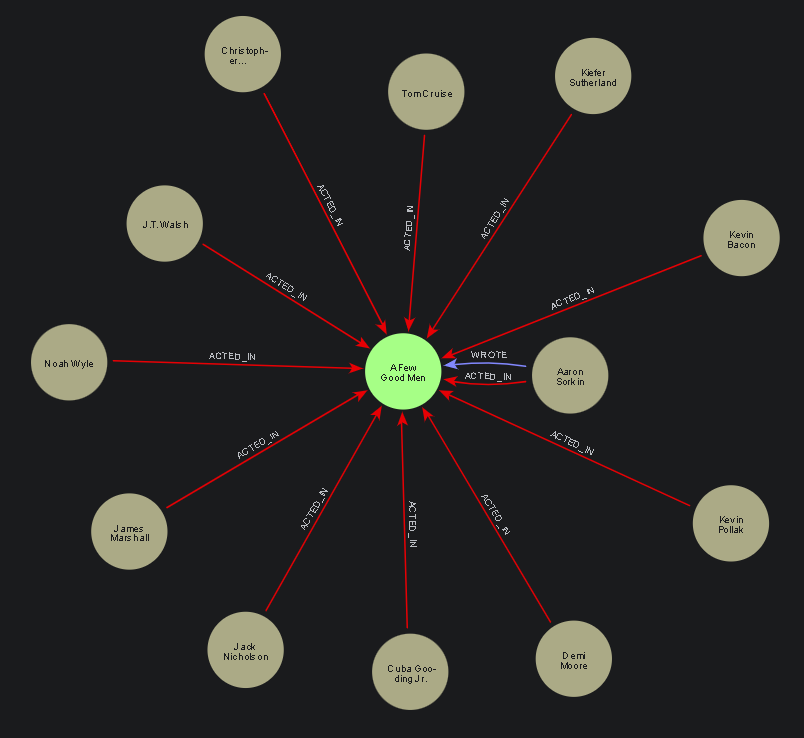


Yukarıdaki görselde, 1965 ve 1970 yılları arasında doğan kişilerin yazdıkları filmler gözükmektedir. Bu filmler sırasıyla V for Vendetta, Speed Racer ve Cloud Atlas filmleridir. Ayrıca Lana Wachowski ve Lilly Wachowski’nin aynı filmleri yazdıkları görülebilir. Soyadları aynı olduğundan, bu kişilerin akraba oldukları anlaşılabilmektedir.

**Sorgu3:** ****

Kendi yazdığı filmde oynayan kişinin, oynadığı film ve diğer oynayan kişileri veren sorgu.

**Görsel3:**

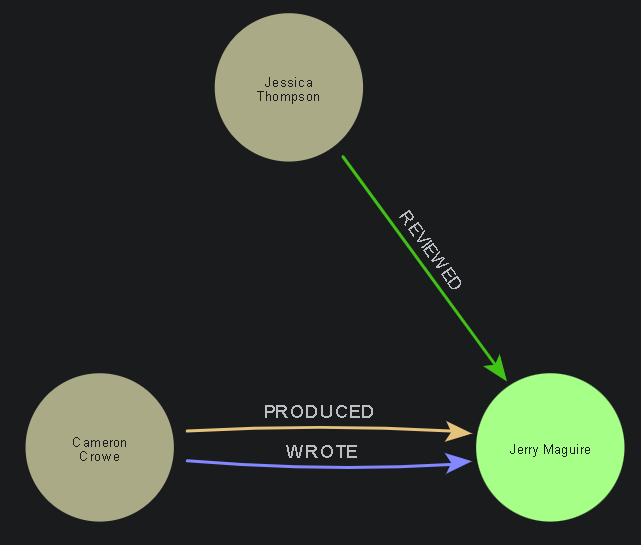


Aaron Sorkin’in, kendi yazdığı “A Few Good Men” isimli filmde diğer birçok oyuncu ile birlikte rol aldığı görülebilmektedir. Veri seti içerisinde bu kişi haricinde başka kimsenin kendi yazdığı filmde oynamadığı da bariz bir şekilde görülebilir.

**Sorgu4:** 

Yapımcısı ve senaristi aynı olan kişinin filmi ve bu filmi eleştiren kişiyi veren sorgu.

**Görsel4:**



Cameron Crowe’un, kendi yazdığı ve yapımcılığını üstlendiği film *Jerry Maguire*’dır. Bu film yalnızca Jessica Thompson tarafından yorumlanmıştır.

**Bölüm 3: Cypher Sorguları ile Veriyi Keşfetme**

**1. Veritabanındaki tüm film adlarını listeleyin.**

**Cypher Query:**

MATCH (m:Movie)

RETURN m.title as movie\_name

**Query Result:**

****

**Yorum:** Speed Racer, V for Vendetta, The Matrix ve serileri gibi birçok 38 tane film bulunmaktadır. Çıktının sadece bir kısmı alınmıştır.

**2. “Tom Hanks’in oynadığı filmleri bulun.**

**Cypher Query [WHERE]:**

MATCH (p:Person)-[r:ACTED\_IN] -> (m:Movie)

WHERE p.name = "Tom Hanks"

RETURN m.title as MovieName\_Tom\_Hanks

**Cypher Query [without WHERE]:**

MATCH (p:Person {name: "Tom Hanks"}) -[r:ACTED\_IN]->(m:Movie)

RETURN m.title as MovieName\_Tom\_Hanks

**Query Result:**

****

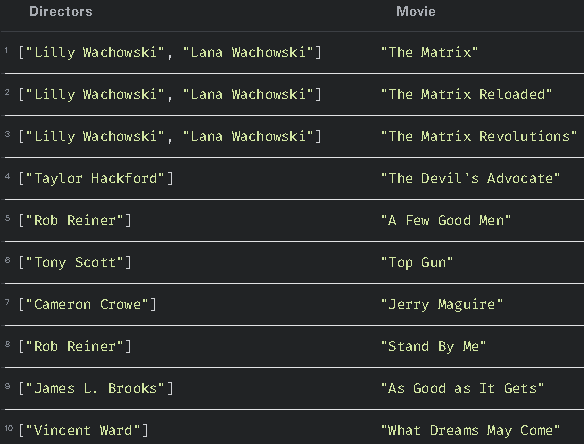
***query hakında açıklama*:** 2.query ilkinden daha hızlıdır. Çünkü ilkinde tüm filmler getirilir ve daha sonra oynayan kişilerin içinde Tom Hanks keyword’u aranır. Fakat 2.query’de Tom Hanks isimli oyuncunun oynadığı filmler getirilir.

**Yorum:** Tom Hanks’in rol aldığı toplam 12 film bulunmaktadır. (The Green Mile, The Da Vinci Code, Apollo 13 vb.)

**3. Her film için yönetmenlerini listeleyin.**

MATCH (p:Person)-[r:DIRECTED]->(m:Movie)

RETURN collect(p.name) as Directors, m.title as Movie



**Yorum:** Toplamda 38 film ve 28 kisi gözükmektedir. Graph sonucu verecek sorgu ile incelendiğinde, en çok ağ yapısına sahip olan son cok film yoneten kisilerin Wachowski kardesler oldugu gorulmektedir.

**4. Aynı filmde oynamış iki farklı oyuncu çiftini listeleyin (tekrarsız).**

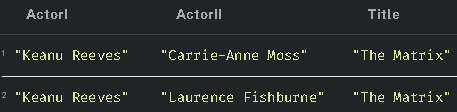
MATCH (a:Person)-[:ACTED\_IN]-> (m:Movie) <- [:ACTED\_IN]-(b:Person)

WHERE elementId(a) < elementId(b)

RETURN a.name as ActorI,b.name as ActorII, m.title as Title

LIMIT 2

**Query Result:**

****

***query hakinda aciklama*:** Duplicate degerleri onlemek icin ***“WHERE elementId(a) < elementId(b)”*** kosulu kullanilir. Normalde name kullanmak ta ayni sonucu verir, fakat ayni isim ve soyisme sahip 2 farkli oyuncu olmasi durumunda hatali sonuc elde edilebilir. Her ne kadar bu dataset icin boyle bir durum soz konusu olmasa da boyle bir kullanim daha uygundur.

**Yorum:** Query sonucuna gore Matrix filminde oynamis 2 farkli oyuncu ciftini gorebiliriz. Keanu Reeves her 2 cift icin de ortak isim olarak gozukmektedir.

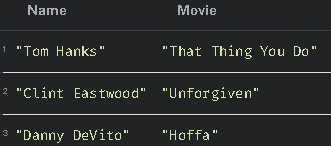
**5. Hem yönetip hem oynadığı bir film olan kişileri ve filmleri bulun.**

**Cypher Query:**

MATCH (a:Person)-[r1:DIRECTED]->(m:Movie)<-[r2:ACTED\_IN]-(a:Person)

RETURN a.name as Name,m.title as Movie

**Query Result:**



**Yorum:** Hem kendi filminde rol almis, hem de yonetmis kisiler, film isimleri ile birlikte sirasi ile tablodaki gibi gorulebilir.

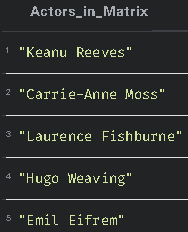
**6. "The Matrix" filminde oynamış oyuncuları listeleyin.**

**Cypher Query:**

MATCH (m:Movie {title:"The Matrix"})<-[:ACTED\_IN]-(p:Person)

RETURN p.name as Actors\_in\_Matrix

**Query Result:**



**Yorum:** The Matrix filminde rol almis oyuncular tablodaki gibi gorulmektedir.

**7. En çok filmde oynamış 5 kişiyi bulun.**

**Cypher Query:**

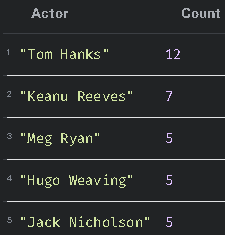
MATCH (p:Person)-[:ACTED\_IN]->(m:Movie)

RETURN p.name as Actor, COUNT(p) as Count

ORDER BY Count DESC

LIMIT 5

**Query Result:**



**8. “Tom Hanks” ile aynı filmde oynamış diğer oyuncuları listeleyin.**

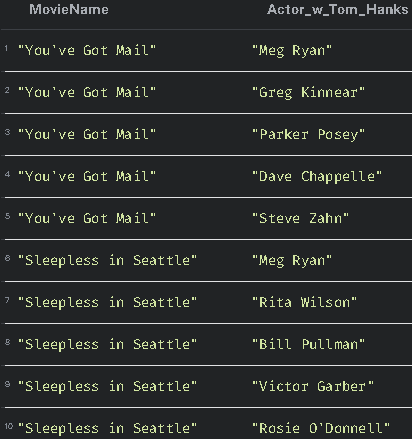
**Cypher Query:**

MATCH (a:Person {name:"Tom Hanks"})-[:ACTED\_IN]->(m:Movie)<-[:ACTED\_IN]-(b:Person)

RETURN m.title as MovieName, b.name as Actor

LIMIT 10

**Query Result:**



**9. Yönetmeni "Lana Wachowski" olan filmleri bulun.**

**Cypher Query:**

MATCH (a:Person {name:"Lana Wachowski"})-[:DIRECTED]->(m:Movie)

RETURN m.title as Movie

**Query Result:**



**10. 2000 yılından sonra yayınlanmış filmleri listeleyin.**

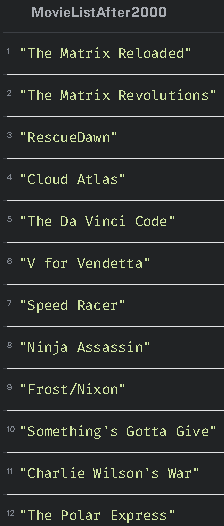
**Cypher Query:**

MATCH (m:Movie)

WHERE m.released> 2000

RETURN m.title as MovieListAfter2000

**Query Result:**

****