



# **Çoklu-Uygulamalı Ortamlar için Kullanıcı Modelleme Sistemi**

**Program Kodu: 1001**

**Proje No: 112E111**

Proje Yürütücüsü:  
**Prof. Dr. Nihan Kesim Çiçekli**

Bursiyer(ler):  
Hilal Tarakçı  
Ali Karakaya  
Yunus Emre Işıklar  
Emrah Şamdan  
Arda Taşcı

Ekim 2014  
ANKARA



## ÖNSÖZ

Bu projede çoklu uygulamalı ortamlar için bir kullanıcı modelleme sistemi geliştirilmiştir. Önerilen sistemin ayırdedici özellikleri, kullanıcının sosyal ağlarda bıraktığı izleri toplayarak kullanıcı modelini oluşturması, bu modelin farklı türde uygulamalar üzerinde kullanılabilir olması, modelin saklanması için hiper-çizge yapısının kullanılması ve değişik uygulamalar için modelin gerekli olan kısmını, uygulamanın desteklediği formatta sunabilmesidir. Geliştirilen sistemin en önemli başka bir özelliği de, kişiselleştirilmiş içerik arama ve önerme uygulamaları için talep üzerine kişiselleştirme bilgisi sağlayabiliyor olmasıdır. Proje kapsamında değişik araştırma alanlarında çalışılmıştır. Öncelikle kullanıcı modelleme sistemleri araştırılmış, farklı kaynaklardan elde edilen kısmi kullanıcı profillerinin birleştirilmesi yöntemleri araştırılmıştır. Bu amaca en uygun olarak hiperçizge veri modelinin kullanılmasına karar verilmiştir. Geliştirilen veri modelinin kullanımı için değişik sosyal ağ kaynaklarından verilere erişme ve bunları analiz etme yöntemleri araştırılmıştır. En popüler olan Facebook, Twitter ve LinkedIn kaynaklarına öncelik verilmiş, bu ağlardan veri edinim yolları ve bu verilerin analiz yöntemleri geliştirilmiştir. Elde edilen ham verilerin içeriklerinin anlamsal olarak zenginleştirilmesi için ontolojiler, bilgi-tabanlı sistemler araştırılmış, var olan freebase ve dbpedia sistemlerinden faydalanılmıştır. Ayrıca kökleştirme ve varlık ismi tanıma algoritmaları geliştirilmiş, böylece daha anlamlı içerikler elde edilmiştir. Öneri sistemleri kullanıcı modelinin uygulama alanı olarak seçilmiş, öneri sistemlerinin detaylı araştırması yapıldıktan sonra kullanıcı modelini kullanan değişik öneri sistemleri geliştirilmiştir. Bunlardan biri de akıllı TV'ler için program öneri sistemidir. Çalışmalarımızın sonuçları beş konferans bildirisi olarak yayınlanmıştır. Proje kapsamında toplam iki yüksek lisans tezi tamamlanmıştır. İki yüksek lisans tezi ve bir doktora tezi de tamamlanma aşamasındadır. Bu rapor, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından desteklenen 112E111 numaralı, "Çoklu-Uygulamalı Ortamlar İçin Kullanıcı Modelleme Sistemi" başlıklı projenin iki yıllık proje süresinde yapılan çalışmaları özetleyen proje sonuç raporudur.



## ÖZET

Bu projede, kişinin kısmi profillerini birleştirmeyi kolaylaştıran ve bütün, çoklu-alanlı bir kullanıcı profili sağlayan bir kullanıcı modelleme sistemi geliştirilmiştir. Birleştirme işlemi, zenginleştirilmiş kullanıcı profillerine yol açan bir anlamsal zenginleştirme işlemi içermektedir. Oluşturulan kullanıcı modeli genel ve alan-tabanlı anlamsal kullanıcı profilleri çıkartabilme ve önerme gibi çeşitli bağlantılı veri sorgularına makul zamanda cevap verebilme yeteneklerine sahiptir. Bu projede, Facebook, Twitter ve LinkedIn'i de içeren sosyal hesaplardan kısmi profilleri içeri aktarabilen ve bu kısmi profilleri birleştirerek kullanıcının bütünsel ve anlamsal bir profilini oluşturabilen bir kullanıcı modelleme sistemi geliştirilmiştir. Ayrıca sistem; kitap, film, spor ve müzik alanları için çeşitli öneri modülleri içermektedir. Alandan-bağımsız genel bir öneri modülü ve belirtilen bir öge ile ilgilenmesi muhtemel kullanıcıların keşfine ve sistemdeki kullanıcı ile benzer ilgi alanlarına sahip kullanıcıların keşfine izin veren bir hesaplama modülü de mevcuttur. Bu modüllerle, sunulan kullanıcı modelini kullanarak bağlantılı veri problemlerine nasıl çözüm sunulabileceği gösterilmektedir. Modelin uygulanabilirliği sosyal ağlar ve akıllı TV kullanım alanlarında farklı öneri sistemleri geliştirilerek örneklenmiştir.



## **ABSTRACT**

In this project, a user modeling system which facilitates aggregating partial profiles of the individual and obtain a complete, multi-domain user model is developed. The aggregation involves a semantic enhancement procedure which results in enriched user profiles. The constructed user model is capable of extracting general and domain-based semantic user profiles and answering several connected data queries such as recommendation, in reasonable time. In this project, we implemented the user profiling system which enables importing partial profiles from social accounts including Facebook, Twitter and LinkedIn and aggregates these partial profiles and constructs a holistic semantic profile of the individual. Moreover, the system includes several recommendation modules for book, movie, sports and music domains. There is also a domain-independent general recommendation module and a computation module which enables discovery of users who might be interested in a specified item and users who have similar interests with the current user. With these modules, we illustrate how to present solutions to connected data problems by using the proposed user model. The use of the model has been illustrated by different recommendation systems in social network and smart TV usage data.

## İçindekiler

ÖZET .....	iii
ABSTRACT .....	iv
İçindekiler .....	v
Şekillerin Listesi .....	vi
Tabloların Listesi .....	vii
1. GİRİŞ .....	1
2. LİTERATÜR ÖZETİ .....	3
3. HİPER-ÇİZGE TABANLI KULLANICI MODELİNİN GELİŞTİRİLMESİ .....	6
3.1 Veri modeli .....	6
3.2 Değişik kaynaklardan oluşturulan kullanıcı modellerinin birleştirilmesi .....	9
4. KULLANICI MODELİNİN ÇOKLU-UYGULAMALI ORTAMLARDA KULLANIMI .....	13
4.1 Sorgular .....	14
4.2 Değerlendirmeler .....	19
5. KİŞİSELLEŞTİRİLMİŞ İÇERİK ARAMA VE ÖNERME UYGULAMALARI .....	21
5.1 Sosyal ağlardaki izlerden kullanıcı modeli oluşturulması .....	21
5.2 İsimlendirilmiş varlık tanıma ile anlam zenginleştirme .....	28
5.3 Akıllı TV'ler için öneri sistemi geliştirilmesi .....	34
5.3.1 Bağlam duyarlı işbirlikçi öneri sistemi .....	36
5.3.2 İçerik tabanlı öneri sistemi .....	40
6. SONUÇLAR .....	43
Kaynaklar .....	46

## Şekillerin Listesi

Şekil 1. Hiper-çizge kullanıcı modeli .....	7
Şekil 2. Kullanıcı Profili .....	8
Şekil 3. Örnek kullanıcı modeli.....	9
Şekil 4. İlgili alanı teriminin belirsizliğinin giderilmesi.....	10
Şekil 5. İlgili alanının bağlı olduğu alanların belirlenmesi .....	11
Şekil 6. İlgili alanının anlamsal olarak zenginleştirilmesi .....	11
Şekil 7. Metaschema örneği.....	12
Şekil 8. Zenginleştirilmiş ilgili alanının kullanıcı profiline bağlanması .....	13
Şekil 9. Hiper-çizge ve özellik-çizge yapılarının özdeşliği.....	14
Şekil 10. Anlamsal kullanıcı profili sorgusu .....	15
Şekil 11. Funguide Anlamsal Kullanıcı Profili Arayüzü .....	16
Şekil 12. Kitap öneri sorgusu .....	16
Şekil 13. Funguide kitap öneri modülü arayüzü .....	17
Şekil 14. Kullanıcının belli bir ilgili alanı öğesine ilgisinin hesaplanması sorgusu .....	17
Şekil 15. Belli bir ilgili alanı öğesi ile ilgili kullanıcıların keşfi sorgusu.....	18
Şekil 16. Benzer kullanıcıların keşfi sorgusu .....	18
Şekil 17. Funguide benzer kullanıcılar arayüzü .....	19
Şekil 18. Veritabanından bir kesit .....	22
Şekil 19. Birinci anket çalışması .....	26
Şekil 20. İkinci anket çalışması .....	27
Şekil 21. Varlık ismi tanıma kaynaklarının sınıflandırılmış gösterimi .....	29
Şekil 22. Yer ve organizasyon isimlerinin örüntü tabanlı kaynaklar kullanarak etiketlenmesi işlemi.....	30
Şekil 23. Kişi isimlerinin sözlük tabanlı kaynaklar kullanarak etiketlenmesi işlemi .....	31
Şekil 24. TV içeriği zenginleştirme uygulaması ekran görüntüsü.....	32
Şekil 25. Örnek program bilgisi .....	35
Şekil 26. Genel sistem akışı.....	36
Şekil 27. Örnek çizge .....	39
Şekil 28. Neo4j'de bir kullanıcı profili .....	41
Şekil 29. Çizgede oyunculara göre gezinim .....	42
Şekil 30. Çizgede yönetmen üzerinden gezinim .....	42



## Tabloların Listesi

Tablo 1. Tamı Tamına Eşleşme Yöntemine Göre Değerlendirme Sonuçları .....	33
Tablo 2. Parçalı Eşleşme Yöntemine Göre Değerlendirme Sonuçları .....	33
Tablo 3. Sistemin Zenginleştirme Oranı Performansları .....	34

## 1. GİRİŞ

Geçtiğimiz on yılda sosyal ağ sitelerinin popülaritesi büyük ölçüde artmıştır. Bir bireyin sosyal ağdaki davranışlarını inceleyerek kullanıcı profilinin çıkarılabilmesi mümkündür. (Gauch vd., 2007). Kullanıcının sosyal web sitelerindeki faaliyetleri, kişinin profili hakkında önemli bilgiler ortaya koymaktadır. Sosyal ağlar kullanım amaçları bakımından farklılık gösterir. Örneğin kullanıcı Facebook ağını sosyal etkileşim ve eğlence için kullanırken, LinkedIn ağını profesyonel profili için kullanabilir. Bu nedenle, sosyal ağlar üzerinde birbirinden bağımsız bir şekilde elde edilen kısmi kullanıcı profilleri, kullanıcının farklı özelliklerini yansıtmaktadır. Farklı bilgi kaynaklarından elde edilen kısmi profillerin sorunsuz bir şekilde birleştirilerek kişinin tam bir profilinin elde edilmesi sorunu literatürde henüz çözülmemiş bir problemdir. Bu projede; tam, anlamsal açıdan zenginleştirilmiş ve kullanıcıyı çoklu-alanda tanımlayan bir kullanıcı modeli oluşturmak amacıyla, kişinin kısmi profilleri, sunulan hiper-çizge kullanıcı modeli ile entegre edilmektedir. Ayrıca, bu hiper-çizge kullanıcı modelinin farklı öneri sistemi problemlerinde çözüm olarak kullanılması örneklenmektedir.

Sunulan hiper-çizge kullanıcı profili yapısı ile entegrasyon metodolojisi birbiriyle bağlantılıdır. Entegrasyon metodolojisi, önceden tanımlanmış kullanıcı modeli veri yapısına bağımlıdır ve bu veri yapısı entegrasyonun ana amaçları göz önünde bulundurularak tanımlanır. Eğer amaç, değişik sistemler arasında birlikte çalışabilir bir kullanıcı modeli oluşturmak ise, genellikle kullanıcı profili standart (Orlandi vd., 2012) veya kullanıcı tarafından tanımlanmış (Wischenbart vd., 2012; Ghosh ve Dekhil, 2008) bir ontoloji ile ifade edilir. Bu projede, ana amaçlarımızdan biri, öneri sistemleri gibi bağlantılı-veri problemlerine zahmetsiz çözümler üretecek bir çerçeve oluşturmaktır. Bağlantılı-veri problemleri için etkili bir çözüm stratejisi, veri yapısına bir giriş noktası oluşturmak ve komşu düğüm ve kenarları belirlenmiş bir algoritmaya göre gezmektir (Robinson vd., 2013). Düğümler genellikle öğeleri ve kullanıcıları temsil ederken, bir kullanıcı ve bir öğeyi bağlayan kenar, kullanıcının öğeye ilgisini ifade eder. Kenarlara ilişkin kuvvetini gösteren ağırlık değerleri de atanabilir. Çizgeler sadece ikili ilişkileri ifade etmeye yetenekli olduğundan, kullanıcı modeli alanındaki çoklu ilişkiler için farklı yaklaşımlar öne sürülmüştür. Kullanıcı modelini iki parçalı (Tiroshi vd., 2013) ve üç parçalı (Chen vd., 2012) tanımlayan bazı çalışmalar vardır. Genel olarak, eğer düğüm tipi sayısı  $n$  önceden biliniyorsa ve kullanıcı modelindeki ilişkiler ikili ise,  $n$ -parçalı çizge kullanıcı profilini ifade edebilir. Ancak, çoklu ilişkiler sözkonusu ise, kullanıcı modelini ifade etmek için hiper-çizge kullanmak daha uygundur (Li ve Li, 2013; Kramar vd., 2013; Bu vd., 2010). Bu





nedenle bu projede kullanıcı profillerinin modellenmesi konusunda hiper-çizge kullanımı fikri önerilmektedir.

Bu projede temel olarak (i) hiper-çizge veri modeli kullanarak kullanıcı profilinin ifade edilmesi, (ii) anlamsal bir entegrasyon metodolojisi ve (iii) bağlantılı-veri problemlerine kolay çözümler sunan bir öneri sistemi hedeflenmiştir. Proje çalışmaları sonunda bu temel amaca ulaşılmış, genel amaçlı, alan ve uygulamadan bağımsız, bütünlük ve hiper-çizge veri modeli kullanan bir kullanıcı modeli geliştirilmiştir. Bu kullanıcı modelinin elde edilmesi için anlamsal içerik zenginleştirme algoritmaları ve kısmi kullanıcı profillerinin entegrasyonu ile bütünlük kullanıcı modeli geliştirme algoritmaları üretilmiştir. Önerilen modelin bağlantılı-veri problemlerinin çözümlerini kolaylaştırdığını göstermek için hiper-çizge üzerinde farklı gezinim algoritmaları sunulmuştur. Bütün bu algoritmaları bir öneri sistemi kapsamında örnekleyen Funguide isimli bir uygulama geliştirilmiş ve <http://144.122.71.72:8080/signin> adresinde internet üzerinde kullanıma açılmıştır.

Bütünlük kullanıcı profilleri çoklu-uygulamalı ortamlarda farklı şekillerde kullanılabilir. Farklı uygulamalar bütünlük kullanıcı profili sağlayıcısından kendi amacına uygun olan bilgileri çekebilir. Örneğin belli bir alan (ör. Futbol) ile ilgilenen kullanıcılara erişmek için sağlayıcıya talep gönderebilir. Ya da belli bir kullanıcının farklı alanlardaki kısmi profillerini öğrenmek isteyebilir. Bu tip talepler bu raporda da anlatılan gezinim algoritmaları ile hiper-çizgeden bir kesit olarak kolayca karşılanabilir. Bütünlük kullanıcı profilleri kullanıcılar için de farklı yararlar sağlayabilir. Örneğin profili bilinen bir kullanıcıya farklı uygulamalar, farklı içerikler önerilebilir ya da kişiselleştirilmiş içerik arama uygulamaları yaratılabilir. Bu iki yaklaşımın (yani hem farklı uygulamaların hem de kullanıcıların yararına modelin kullanılabilmesi) yapılabilirliğini göstermek de projenin amaçları arasındadır.

Bu projede anlamsal olarak zenginleştirilmiş, bütünlük kullanıcı profillerinin kullanımlarını örneklemek amacıyla çeşitli alanlarda araştırmalar yapılmış ve farklı uygulamalar geliştirilmiştir. Projede ilk önce modelin kullanıcıların yararına kullanılabilmesini örneklemek amacıyla sosyal ağlardan veri çekilmesi ve bu verilerin analizi konusunda çalışmalar yapılmış, özellikle Facebook ve Twitter gibi popüler sosyal ağlar incelenmiştir. Baştaki amacımız bu sosyal ağlardan elde edilen kullanıcı verilerini bütünlük bir şekilde entegre ederek gerçek verilerden bütünlük kullanıcı modelleri üretebilmektir. Ancak yaptığımız araştırmalar sonucunda bu ağların verilerini paylaşma konusunda çok hassas olduklarını ve çok kısıtlı veriler sunduklarını farkettilik. Ayrıca her iki sosyal ağda birden hesabı bulunan kullanıcıların tespitinin, hesapları farklıysa eşleştirilmelerinin imkansız olduğunu gördük. Bu nedenle bu amacımızı kısmi olarak gerçekleştirmek zorunda kaldık. Kısıtlı sunulan facebook bilgilerinin kullanarak anlamsal zenginleştirme ve çizge-tabanlı kullanıcı modeli çıkarma çalışmaları yaptık ve bunu bir öneri sistemi olarak örneklendirdik.

Modelin farklı uygulamalar için yararını gösterebilmek amacıyla Akıllı TV'lerde öneri sistemi uygulamasını geliştirmeyi hedefledik. Sosyal ağlardan elde edilen kullanıcı profilinin belli alan ya da alanlardaki kesiti kullanılarak, o kullanıcıya TV programı önermek ya da yayıncılar için belli programlara potansiyel seyirciler bulabilmek geliştirdiğimiz algoritmalarla mümkün olabilmekteydi. Ancak burada da gerçek verilerle çalışmak mümkün olmadı. Arçelik'ten elde ettiğimiz Akıllı TV kullanım verisindeki kullanıcıların facebook ya da twitter bilgilerine ulaşmak mümkün değildi. Aynı şekilde bizim facebooktan elde ettiğimiz kullanıcıların TV seyretme bilgileri elimizde yoktu. Önerdiğimiz sistemin başarısını gerçek verilerle değerlendirmek mümkün olamadığı için sadece TV kullanım verisi olan kullanıcılar için anlamsal olarak zenginleştirilmiş, çizge tabanlı kullanıcı profilleri oluşturup öneri sistemi geliştirdik.

Örnek uygulamalarımız için elde edilen gerçek verilerle geliştirilen kullanıcı modelleri projede önerdiğimiz bütünleşik kullanıcı modelinin alt kümesidir. Bu uygulamalarda hep ikili ilişkiler olduğu için hiper-çizge modeli normal çizge modeline dönüşmüştür. Bu uygulamalar kapsamında yapılan yüksek lisans tezlerinde uygulamaların amacına uygun olarak orijinal gezinim ve öneri algoritmaları geliştirilmiştir.

Yapılan bütün bu çalışmalar, bu sonuç raporunda yenilikçi yönlerine ağırlık verilerek anlatılmıştır. 2. Kısımda Literatür özeti verilmiştir. Bu kısmın içeriği bütünleşik kullanıcı modeli geliştirme konusundaki çalışmaların özetiyle kısıtlanmıştır. Projenin diğer araştırma konularındaki literatür özetleri daha önceki gelişme raporlarında verilmişti. 3. Kısımda hiper-çizge tabanlı kullanıcı modeli anlatılmakta ve değişik kaynaklardan gelen kısmi kullanıcı profillerinin entegre eden algoritmalar sunulmaktadır. 4. Kısımda kullanıcı modelinin gerçekleştirimi ve çoklu-uygulamalı alanlarda kullanımı anlatılmaktadır. 5. Kısım projede geliştirilen farklı uygulamalara ayrılmıştır. Önce sosyal ağlardan toplanan verilerin anlamsal olarak zenginleştirilerek kullanıcı profili oluşturulması ve bunun bir öneri sisteminde kullanılması anlatılmaktadır. Daha sonra anlamsal zenginleştirme kapsamında yapılan isimlendirilmiş varlık tanıma çalışmaları sunulmakta ve bu çalışmanın sonucu olarak ortaya çıkan TV içerik zenginleştirme uygulaması anlatılmaktadır. Daha sonra da Akıllı TV'ler için geliştirdiğimiz içerik tabanlı ve işbirlikçi öneri sistemleri özetlenmiştir. 7. Kısımda ise elde edilen kazanımlar, sonuçlar ve gelecekteki araştırma konuları sunulmuştur.

## 2. LİTERATÜR ÖZETİ

Bu bölümde literatürde kullanıcı profillerinin entegrasyonu konusundaki çalışmalar özetlenmekte ve projede önerilen kullanıcı modelinin diğer çalışmalardan farklılıkları vurgulanmaktadır. Projede çalışılan diğer araştırma konuları olan öneri sistemleri, isimlendirilmiş varlık tanıma, anlamsal zenginleştirme ve sosyal ağ üzerinden kullanıcı modeli çıkarma alanlarının literatür özetleri geçen dönemlerdeki gelişme raporlarında bulunmaktadır. Bu raporda bu geniş literatür özeti kapsam dışında tutulmuştur.

Kısmi kullanıcı profillerinin entegrasyonu varlıkları eşleştirme, tekrarlayan veya çakışan verilerin düzeltilmesi, kısmi profillerinin heterojenliği gibi çeşitli konuları içermektedir (Orlandi vd., 2012). Ayrıca, entegre edilmiş profilin kullanım amacı, entegrasyon stratejisini de etkiler. Literatürde, farklı entegrasyon yaklaşımları vardır.

(Abel vd., 2013) çalışmasında, form tabanlı ve etiket tabanlı profiller ayrı ayrı yönetilmektedir. Form tabanlı profiller öznitelik-değer çiftlerinden oluşan bir liste iken, etiket tabanlı profiller ağırlıklandırılmış etiket kümelerinden oluşmaktadır. Form tabanlı profiller için entegrasyon stratejisi, öznitelik-değer listelerini birleştirmektir. Heterojen öznitelik sözlükleri meselesi, profilleri birleştirilmiş bir öznitelik-değer uzayına eşleştiren bir hizalama fonksiyonu kullanarak çözülmektedir. Ancak, bu hizalama fonksiyonu, nihai kullanıcı profilinde çift kayıtlara sebep olabilmektedir. Üstelik, entegre edilen profiller çelişen veri içeriyorsa, çelişen veriler de nihai profile yer almaktadır. Etiket tabanlı profillerin entegrasyonu, kısmi etiket tabanlı profillerin ağırlıklandırılmış toplamını alarak elde edilmektedir. Çalışmada, form tabanlı ve etiket tabanlı profiller birbirine entegre edilmemektedir. Bu projede ise, profiller arasında böyle bir ayırım yapılmaz ve alınan kısmi kullanıcı profillerin ağırlıklandırılmış toplamları alınarak tek bir kullanıcı profili oluşturulmaktadır. Heterojen sözlük problemi Freebase kullanılarak çözülmektedir.

(Orlandi vd., 2012) çalışmasında amaç, birçok sistemde ortak çalışabilir, kaynaktan bağımsız, çok alanlı kullanıcı profili elde etmektir. Bu sebeple, entegre edilen kullanıcı profili popüler standart ontolojiler kullanılarak ifade edilmektedir. Entegrasyon sırasında, tekrar eden öğeler sorunu, öğeler için global ağırlık hesaplayarak çözülmüştür. Bunun için, öğenin profile katılımına ait, öğenin geldiği bilgi kaynağı ve zaman gibi çeşitli metaveri saklanmıştır. İlgili alanı öğesine ait metaverilerin varlığı, kısmi profillerin entegrasyonu esnasında öğenin ağırlığının yeniden hesaplanmasına imkan vermektedir. Bu projede, öğeye ait geldiği bilgi kaynağı, kısmi profil tarihi ve anlamsal öğe için arama anahtar kelimesi gibi detaylar metaveri olarak saklanmaktadır. Bu metaveri, öğe ve ilgili kullanıcıya dair her yeni detay gelişinde güncellenmektedir.

(Wischenbart vd., 2012) çalışmasında entegrasyon, sosyal ağ verisinden yarı-otomatik olarak şema çıkarılması ve çıkarılan şemaların mevcut entegrasyon araçları ile entegre

edilmesi ile başarılmaktadır. (Plumbaum vd., 2011) çalışmasında, kısmi kullanıcı profillerini yarı otomatik olarak entegre edebilmek amacıyla bir entegrasyon ontolojisi sunulmaktadır.

Bu projede profil entegrasyonun amacı iki yönlüdür: (i) öneri sistemleri gibi bağlantılı veri problemlerini çizge gezme algoritmalarına indirgeyen bir hiper-çizge kullanıcı modeli sağlamak ve (ii) sunulan anlamsal iyileştirmeler ile önerilerin doğruluğunu arttırmak.

Literatürde anlamsal zenginleştirme, kavramı harici bir sözlüğe bağlamak, bağlanmadığında ikincil bir harici sözlük kullanmak, kavramın eşanlamlılarını da sisteme eklemek, önceden tanımlanmış bir gezme algoritması ile harici sözlükten kavramla alakalı diğer kavramları getirmek, arkadaşların veya benzer kullanıcıların profillerini kullanmak gibi yaklaşımlarla ele alınmaktadır (Abdel-Hafez ve Xu, 2013). Bu projede, anlamsal zenginleştirme için harici sözlüğün üzerinde bir ara ontoloji, sistemin ihtiyaçlarına göre kalibre edilerek kullanılmaktadır.

Çoğu kullanıcı modelleme ve öneri problemi özünde bağlantılı veri problemidir. Bağlantılı veri problemleri, problem ile alakalı alt-çizgeyi gezmek için uygun gezme algoritmaları üreterek çözülür. Bağlantılı veri problemleri ile uğraşırken, çizge veritabanlarının, ilişkisel ve NoSQL veritabanlarından daha hızlı olduğu iddia edilmektedir (Robinson vd., 2013). Çizge veritabanlarında, bir gezme sorgusunun performansı, sorgunun gezeceği alt çizgenin boyuna bağlıdır. Başka bir deyişle, tüm çizgenin boyutu, sorgu performansını etkilememektedir. Bu sebeple, projede hiper-çizge gerçekleştirimi için çizge veritabanı kullanılmaktadır.

(Tiroshi, 2012; Cena vd., 2013) çalışmalarında, çizge tabanlı kullanıcı modelleri sunulmaktadır. (Tiroshi, 2012) çalışmasında, kavramlar, harici ontolojiyi inceleyerek birbirleriyle bağlanmaktadır. Bu projede, entegrasyon sırasında, anlamsal açıdan birbirleriyle alakalı kavramlar ilişkilendirilmekte ve modeldeki önceden tanımlanmış farklı düğüm ve kenar tiplerinin varlığı özel gezinti algoritmaları yazılmasını kolaylaştırmaktadır.

Projede önerilen bütünleşik kullanıcı modelinin ilk taslak aşamasından başlayarak, son gelişmiş haline kadar olan basamakları farklı bildirilerle yayınlanmıştır (Tarakçı ve Çiçekli, 2012a, 2012b, 2014a, 2014b). (Tarakçı ve Çiçekli 2014a) KEOD 2014 konferansında “**En iyi bildiri**” ödülü almıştır. Halen bütün kullanıcı modelini ve çoklu-uygulamalı alanlarda kullanımını anlatan bir dergi makalesi hazırlıkları devam etmektedir. Sistemin gerçek verilerle değerlendirilmesi tamamlandığında makale yayın için teslim edilecektir.

### 3. HİPER-ÇİZGE TABANLI KULLANICI MODELİNİN GELİŞTİRİLMESİ

Bu projede, kişinin kısmi profillerini birleştirmeyi kolaylaştıran ve bütünleşik, çoklu-alanlı bir kullanıcı profili sağlayan bir kullanıcı modelleme sistemi geliştirilmiştir. Bu kısımda hiper-çizge tabanlı veri modeli anlatılmaktadır. Ayrıca, değişik kaynaklardan elde edilen kısmi profillerin entegrasyonu için geliştirilen yöntem sunulmaktadır.

#### 3.1 Veri modeli

Hiper-çizge (Ghoshal vd., 2009), çizgenin genelleştirilmesi ile elde edilir. Sıradan bir çizgede, kenarlar iki köşeyi birbirine bağlarken, hiper-çizgede istenen sayıda köşe, hiper-kenarlar ile birbirine bağlanabilir. Bu sebeple, sıradan çizge ikili ilişkileri verimli bir şekilde modelleyebilirken, ikiden fazla öge içeren n-li ilişkilerde hiper-çizge daha verimli bir çözüm olmaktadır. Ayrıca, iç içe geçebilen yapısı, hiper-çizgenin karmaşık kavramları ifade edebilme yeteneğini artırmaktadır. Kullanılacak hiper-çizge yapısının belirlenmesi, proje için öncelikli hedeflerden biri kabul edilmiştir. Literatürde, kullanıcı modeli için ikili çizge (Tiroshi vd., 2013) ve üçlü çizge (Bu vd., 2010) kullanan çalışmalar vardır. Genel olarak, eğer projedeki düğüm tiplerinin sayısı n önceden biliniyorsa ve kullanıcı modelindeki ilişkiler ikili ise, profil n'li çizge yapısı ile modellenabilir. Ancak daha yüksek dereceli ilişkiler söz konusu ise, kullanıcı modelini ifade etmek için hiper-çizge kullanmak daha uygundur (Li ve Li, 2013; Kramar vd., 2013; Bu vd., 2010)

Projede kullanılacak hiper-çizge yapısını belirlemek amacıyla, ilk dönemde Automed projesindeki (Theodoratos, 2012) hiper-çizge veri modeli incelenmiş, bu proje ile arasındaki benzerlik ve farklılıklar ortaya koyulmuştur. Projede sunulan hiper-çizge modelinin anlamsal zenginleştirme işlemini de desteklemesi gerekmektedir. Bu sebeple ilk dönemde anlamsal zenginleştirme amacıyla kullanılacak bilgi tabanları ve API'ler incelenmiştir. DBPedia (DBPedia, 2014) ve Freebase (Freebase, 2014) bilgi tabanlarının özellikleri karşılaştırılmış, bu karşılaştırmanın bulguları sonucu entegrasyon sırasında kullanılacak harici ontoloji olarak Freebase seçilmiştir. Freebase, varlıkları, çizge yapısında birbiri ile ilişkilendirir ve çeşitli API'ler ile sorgulama imkanı sunar.

Hiper-çizge yapısının ilk taslak formülasyonu projenin ikinci döneminde sunulmuş, üçüncü dönemdeki raporda bir hiper-çizge kullanıcı modeli örneği verilmiştir. Projenin son döneminde hiper-çizge kullanıcı modeline nihai hali verilmiş, proje gerçekleştirimi tamamlanarak konuyla ilgili iki çalışma yayınlanmıştır (Tarakçı ve Çiçekli, 2014a, 2014b) Bu

çalışmalarda Şekil 1' deki hiper-çizge sunulmuştur. Sunulan hiper-çizge kullanıcı modelinin genel özellikleri:

- Kullanıcılar ve ilgi alanlarını temsil eden öğeler farklı düğüm tipleri ile ifade edilmektedir.
- Ön tanımlı alanlar için alan düğümleri mevcuttur. Örneğin müzik, spor, kitap ve film alanları için ayrı alan düğümleri vardır. Böylece alan-tabanlı sorgular doğal olarak desteklenmiş olmaktadır.
- Sisteme girilen her öğe düğümü, bir veya daha fazla alana belli bir kenar tipi ile bağlanmaktadır. Bir öğenin bağlı olabileceği maksimum alan sayısı bir eşik değer ile kalibre edilmektedir.
- Sistemdeki öğeler arasında anlamsal ilişkiler, kullanıcı ve öğeler arasındaki ilgi ilişkileri, öğeler ile bağlı oldukları alanlar arasındaki ilişkiler, kullanıcılar arasındaki arkadaşlık ilişkileri modelde ayrı kenar tipleri ile tutulmaktadır. Bu ayırım, öneriyi de içeren bağlantılı veri sorgulamalarının sistem tarafından doğal olarak desteklenmesini sağlamaktadır.

Notation	Description	Type
$u$	a user	Node
$U$	Set of users	Hyperedge
$i$	an item	A Node
$I$	Set of items	Hyperedge
$D_{[d]}$	Domain starter node for each domain $d$	Node
$E_{bind}$	Metadata for user-item (interest) relation	Hyperedge
$E_{inner}$	The semantic relation between items	Hyperedge
$E_{domain}$	The domain bind between domain starter node and items	Hyperedge
$E_{friend}$	Friendship between users	Hyperedge
$H_u$	General (long term) user profile	A sub hypergraph

Şekil 1. Hiper-çizge kullanıcı modeli

Sunulan hiper-çizge kullanıcı modelinde kullanıcı, ilgi alanı öğeleri ve alanlar farklı düğüm tipleri ile ifade edilmiştir. Bir ilgi alanı öğesi en az bir, en çok belirtilen alan eşik değeri kadar alan ile ilişki olabilir. Projede kullanılan öntanımlı alanlar, Freebase commons paketindeki alanlardır. Sunulan modelde, kenar tipleri de düğümler gibi çeşitlilik göstermektedir. Bir kullanıcı ile öğeyi bağlayan kenar, kullanıcının öğeye olan ilgisini ifade etmekte, iki öğe düğümünü bağlayan kenar, bu öğeler arasındaki anlamsal alakayı saklamakta ve bir öğeyi, alan düğümlerine bağlayan kenar, ögenin bağlı olduğu alanları tutmakta, kullanıcı düğümlerini bağlayan kenarlar da bu kullanıcılar arasındaki arkadaşlık ilişkisini modellemektedir.

Sistemde bir kullanıcının profili, kullanıcının arkadaşları, bildirdiği profil ve zenginleştirilmiş profillerinin birleşiminden oluşan alt-çizgedir (Şekil 2). Kullanıcının arkadaşları, arkadaşlık kenar tipi ile kullanıcıya bağlı olan diğer kullanıcılardan oluşur. Kullanıcının bildirdiği profil, kullanıcının bağlı olduğu ilgi alanı öğeleri ve bu öğelerin alan düğümlerinden oluşur. Kullanıcının zenginleştirilmiş profili, kullanıcının bildirdiği profiledeki ilgi alanı öğeleri ile ilişkili olan diğer ilgi alanı öğeleri ve alanlarını da içeren alt-çizgedir.

$$\begin{aligned}
 H_u(u; min; max) &= U_{friends}(u) \\
 &\cup U_{explicit\ profile}(u) \\
 &\cup U_{enhanced\ profile}(u; min, max) \quad (1)
 \end{aligned}$$

$$\begin{aligned}
 U_{friends}(u) &= u \xrightarrow{follows} (u_f) \\
 &\cup (u_f) \xrightarrow{follows} u \quad (2)
 \end{aligned}$$

$$\begin{aligned}
 U_{explicit\ profile}(u) &= u \xrightarrow{interestedIn} (i) \xrightarrow{isInDomain} (d) \quad (3)
 \end{aligned}$$

$$\begin{aligned}
 U_{enhanced\ profile}(u; min; max) &= \\
 &u \xrightarrow{*min..max} (i) \xrightarrow{isInDomain} (d) \quad (4)
 \end{aligned}$$

Şekil 2. Kullanıcı Profili

Şekil 3'de örnek bir kullanıcı modeli alt-çizgesi verilmiştir. Kırmızı düğümler kullanıcıları, yeşil düğümler ilgi alanı öğelerini ve mor düğümler alanları göstermektedir. Örnekte IngridBergman ve GraceKelly kullanıcıları arasındaki *follows* ilişkisi, kullanıcıların birbiriyle arkadaş olduğunu göstermektedir. Şekilde, ilgi alanı öğeleri Freebase ID'leri ile isimlendirilmiştir. Her iki kullanıcı da /en/alfred\_hitchcock ilgi alanına *InterestedIn* ilişkisi ile



bağlıdır. Başka bir deyişle, /en/alfred\_hitchcock ilgi alanı, her iki kullanıcının bildirilmiş profilinde yer almaktadır. Bu düğüm, çeşitli ilgi alanlarına bağlıdır. Bu ilgi alanları, /en/alfred\_hitchcock ile anlamsal açıdan alakalı öğeleri göstermektedir. Örneğin, *ContributedTo* ile bağlı olan /en/saboteur, gerçekten de Alfred Hitchcock'un bir filmidir. /en/saboteur düğümü, FILM ve MEDIA alanlarına *IsInDomain* ilişkisi ile bağlıdır.

Şekil 3. Örnek kullanıcı modeli

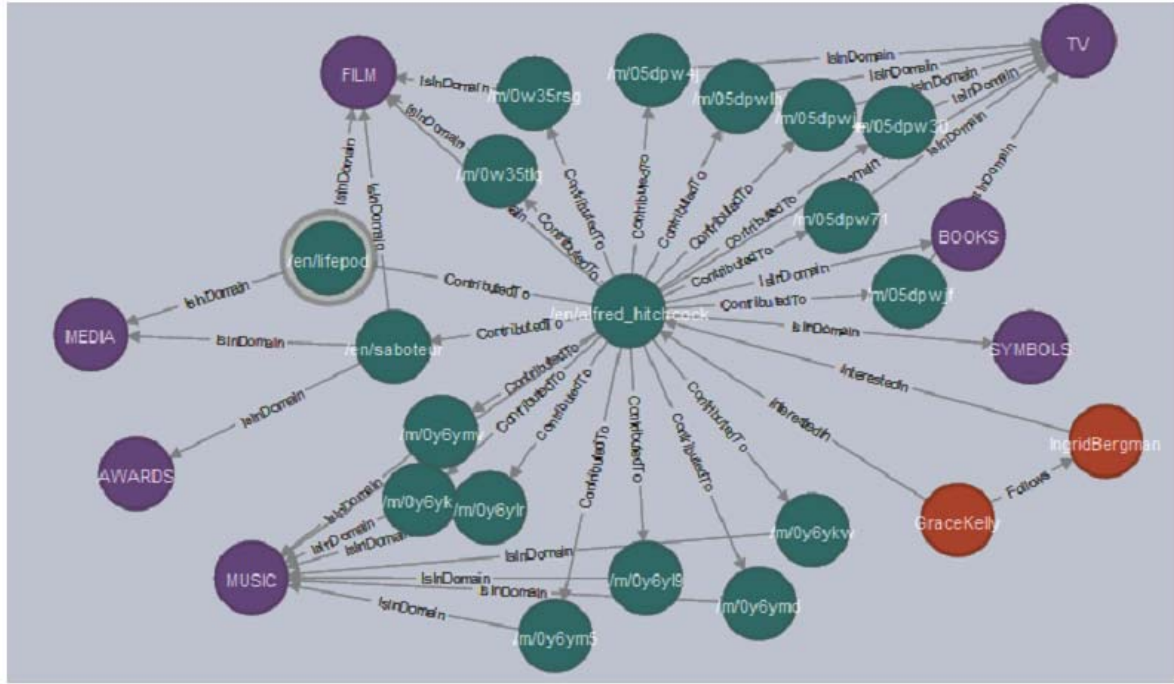


Figure 3: A Sample User Model

Bu kısımda anlatılan veri modeli ve kısmi profillerin entegrasyonu ayrı bir makale olarak yayınlanmıştır (Tarakçı ve Çiçekli 2014b).

### 3.2 Değişik kaynaklardan oluşturulan kullanıcı modellerinin birleştirilmesi

Şekil 1'de sunulan hiper-çizge yapısı, kısmi profillerin birleştirilmesini doğal olarak desteklemektedir. Kısmi profiller birleştirilirken, kısmi profili oluşturan her bir terim için sırasıyla ilgi alanı teriminin belirsizliği giderilmekte (Şekil 4), ilgi alanının bağlı olduğu alanlar belirlenmekte (Şekil 5), ilgi alanı anlamsal olarak zenginleştirilmekte (Şekil 6) ve zenginleştirilmiş ilgi alanı kullanıcı profiline bağlanmaktadır (Şekil 8).



---

**Algorithm 1** Disambiguation

---

```
1: procedure DISAMBIGUATE
2:   mqlQuery = makeJsonQuery(keyword)
3:   retJSONArray = executeFreebaseQuery(mqlQuery)
4:   freebaseData ← first of retJSONArray
5:   if freebaseData == null then
6:     keyword ← keyword.replace(ğ, g)
7:     keyword ← keyword.replace(ş, s)
8:     ...
9:     freebaseData = disambiguate(keyword)
10:  if freebaseData == null then
11:    keyword ← keyword.replace("Fans of", "")
12:    keyword ← keyword.replace("Fan Club", "")
13:    ...
14:    freebaseData = disambiguate(keyword)
15:    if freebaseData == null then
16:      freebaseData = disSplit("&", keyword)
17:      freebaseData = disSplit("and", keyword)
18:      ...
19:  return freebaseData
```

---

**Şekil 4. İlgili alanı teriminin belirsizliğinin giderilmesi**

İlgili alanı teriminin belirsizliği giderilirken, öncelikle terim, harici ontolojide (Freebase) aranmaktadır. Eğer eşleşme bulunamaz ise, terimde yer alan Türkçe karakterler, en yakın İngilizce karakterlerle değiştirilerek işlem yinelenmektedir. Yine eşleşme yakalanamazsa, "Fans Of", "Fanları", "Fan Club", "Fun Klübü" gibi ifadeler çıkarılarak terim tekrar aranmaktadır. Bu şekilde de bir eşleşme sağlanamazsa, "ve", "and", "&" gibi, iki terimin birleştiren ifadelerin varlığı kontrol edilip, gerekirse terim bölünerek parçalar ontolojide aranmaktadır.

---

**Algorithm 2** Decide Domains

---

```
1: procedure DECIDEDOMAINS
2:   mqlQuery = makeJsonQuery(freebaseID)
3:   retJSON = executeFreebaseQuery(mqlQuery)
4:   typeArray ← type property of(retJSON)
5:   foreach type in typeArray:
6:     domainType = convert2DomainType(type)
7:     // add to domain map,
8:     // increment frequency if already exists
9:     domainMap.Add(domainType)
10:    domainMap = pruneDomMap(domainMap)
11:    return domainMap
```

---

**Şekil 5.** İlgili alanının bağlı olduğu alanların belirlenmesi

Terim için Freebase'de eşleşme sağlandıktan sonra, ilgili alanı ögesinin ait olduğu alanlar belirlenmelidir. Freebase'deki tip bilgisi, ögenin ait olduğu alan hakkında fikir vermektedir. Örneğin Alfred Hitchcock için tip bilgisi:

```
"/film/writer",
"/film/producer",
"/common/topic",
"/people/person",
"/people/deceased_person",
"/film/actor",
"/film/director",
...
```

şeklinde. Tip bilgisindeki ilk kelime, ögenin ait olduğu alanı söylemektedir. Sistemde, ögenin tip bilgisinde aynı alana ait kaç tip olduğuna bakılarak alan bilgisi ağırlıklandırılmakta ve tanımlanan alan eşik değerine göre budanarak, öge, ait olduğu alanlara bağlanmaktadır. Ayrıca, alt alan bilgisi de metaveri olarak sistemde tutulmaktadır.

---

**Algorithm 3** Enhance

---

```
1: procedure ENHANCE
2:   mqlQuery = makeJsonQuery(freebaseID)
3:   retJSONArray = executeMetaschemaQuery(mqlQuery)
4:   retJSONArray ← limited(retJSONArray)
5:   metaschemaList ← parsed JSON(retJSONArray)
6:   return metaschemaList
```

---

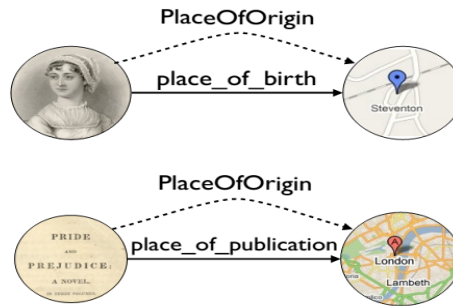
**Şekil 6.** İlgili alanının anlamsal olarak zenginleştirilmesi

İlgili alanının anlamsal olarak zenginleştirilmesi sırasında, Freebase'in üzerinde yer alan Metaschema Ontolojisi kalibre edilerek kullanılmaktadır. Metaschema ontolojisi, 46 kavramdan oluşur ve Freebase ontolojisi üzerinde genel sorgulama yapabilmeye imkan

verir. Şekil 7'de PlaceOfOrigin, JaneAusten ve Stevenson düğümleri arasında doğum yeri, PrideAndPrejudice ve London düğümleri arasında yayınlanma yeri olarak iki farklı anlam için kullanılmıştır. Projede, 46 metaschema kavramı incelenmiş ve kullanıcı modelleme alanında en çok kullanılacak 9 tanesi seçilmiştir. Buna göre, projede kullanılan metaschema kavramları şunlardır:

- BroaderThan/NarrowerThan
- ContributedTo/HasContributor
- PractitionerOf/HasPractitioner
- HasSubject/SubjectOf
- SuperclassOf/ SubclassOf
- Created/CreatedBy
- HasGenre/GenreOf
- HasName/NameOf
- HasChild/ HasParent

#### Generalized relationships



Şekil 7. Metaschema örneği

---

**Algorithm 4 Aggregation**

---

```
1: procedure AGGREGATION
2: foreach keyWord in keyWordList:
3:   freebaseData = disambiguate(keyword)
4:   freebaseID ← freebaseData.freebaseID
5:   if freebaseID in Hypergraph then
6:     if freebaseID already connected to User then
7:       increment frequency
8:     connect freebaseID to the user
9:   decideDomains (freebaseID )
10:  enhance (freebaseID )
```

---

**Şekil 8. Zenginleştirilmiş ilgi alanının kullanıcı profiline bağlanması**

Projede, kısmi profilde yer alan her terim için Freebase'deki anlamsal karşılığı bulunmakta, bağlı olduğu alanlar keşfedilmekte ve en alakalı kavramlar da beraberinde getirilerek çizgeye eklenmekte, böylece terim kullanıcı profiline entegre edilmiş olmaktadır.

#### 4. KULLANICI MODELİNİN ÇOKLU-UYGULAMALI ORTAMLARDA KULLANIMI

Projenin odak noktası sosyal ağlardan toplanan kısmi profillerin hiper-çizge yapısında tek ve bütün bir kullanıcı modeli olarak birleştirilmesi ve bu modelin bağlantılı veri problemlerinin çözümünü çok kolaylaştırdığının gösterilmesidir. Bu amaçla FunGuide adındaki ürün tasarlanmıştır. FunGuide <http://144.122.71.72:8080/signin> adresinde yayında olup, Facebook, Twitter ve LinkedIn den kullanıcının profilini çekerek birleştirmekte, kullanıcıya anlamsal profilini göstermektedir. Ayrıca, kullanıcının manuel olarak da ilgi alanı girmesine izin verilmiştir. Böylece sistem 4 farklı bilgi kaynağından alınan kısmi profilleri entegre etmiş olmaktadır.

Hiper-çizge yapısını tutmak için veritabanı olarak Neo4j (Neo4j, 2014) seçilmiştir. Neo4j, özellik-çizge yapısı ile kayıt yapmakta ve sorgulamaktadır. Hiper-çizge yapısı ile özellik-çizge yapısı aynı ifade kabiliyetine sahip, birbirine dönüşebilen özdeş yapılardır. Özellik-çizge, kabaca, çizgelerin düğüm ve bağlarına anahtar-değer formatında özelliklerin eklenebilmesine izin veren çizge olarak tanımlanabilir. (Tarakçı ve Çiçekli, 2014a) 'da hiper-çizge ve özellik-çizgenin birbirine dönüşebilmeleri örneklenmiştir (Şekil 9).



Şekil 10'da anlamsal kullanıcı profili belirlenirken kullanılan sorgu verilmiştir. Buna göre, temel olarak, kullanıcı düğümü hiper-çizge üzerinde bulunur ve kullanıcı düğümüne bağlı olan ilgi alanı öğeleri ve bu öğelere bağlı olan alan düğümlerinden oluşan alt-çizge kullanıcının anlamsal profilidir.

$$P_{general}(u;max) = u \xrightarrow{*0..max} (i) \xrightarrow{IsInDomain} (d)$$

```
@Query("START user=node({0}) "
+ "match user-[bind]->(item:Item)-[domainBind]->(domain:Domain) "
+ "return "
+ "item.id as itemId, "
+ "item.name as itemName, "
+ "bind.frequency as frequency, "
+ "bind.keywordList as keywordList, "
+ "collect(domainBind.domainDetailList) as domainDetailList, "
+ "collect(domain.name) as domainName "
+ "order by frequency desc, item.name asc "
+ "limit 100")
List<MR_UserProfile> getProfile( UserAccount user );
```

#### Şekil 10. Anlamsal kullanıcı profili sorgusu

Şekil 11'de FunGuide anlamsal kullanıcı profili arayüzü görülmektedir. Ekranın solunda kullanıcının anlamsal profil öğeleri yer almaktadır. Profil öğesinin yanında sırasıyla ait olduğu alan veya alanlar, alt alan bilgileri ve ağırlık değeri yer almaktadır. Aynı öğe kullanıcının farklı sosyal ağlarından veya farklı terimlerle bildirilerek gelmişse, ağırlık değeri artar. Ekranın sağında kullanıcının ilgili olduğu alanlar görülmektedir. Alanın yanında yer alan sayı, kullanıcının o alana ait bağlı olduğu öğe sayısını verir. Bu kısmın kullanıcıyı alan bazında modellediğini söyleyebiliriz. Örneğin örnek kullanıcı için kitaplardan ve bilgisayarlardan hoşlandığını söylemektedir.

Your semantic profile: 59

Add item to profile..

**Information retrieval** [BOOKS] [book subject] (Frequency: 2)

**Java Platform, Enterprise Edition** [COMPUTERS] [software,software genre] (Frequency: 2)

**Object-oriented programming** [EDUCATION] [field of study] (Frequency: 2)

**Service-oriented architecture** [COMPUTERS] [internet protocol] (Frequency: 2)

**Social networking service** [ORGANIZATION, FILM] [contact category,organization sector, film subject] (Frequency: 2)

**Agatha Christie** [BOOKS, TV] [author,book subject, tv writer] (Frequency: 1)

**Alfred Hitchcock** [FILM, BOOKS, ORGANIZATION, TV] [director,writer,producer,actor,film story contributor,film art director, person or entity appearing in film,film subject, author,book subject, organization founder, tv program creator,tv writer,tv director,tv actor,tv producer,tv personality] (Frequency: 1)  
**Angela's Ashes** [AWARDS, MEDIA, FILM] [award nominated work,ranked item,award winning work, adaptation,netflix title, film] (Frequency: 1)

**Animated cartoon** [TV] [tv genre] (Frequency: 1)

**Ant** [BIOLOGY, BOOKS] [organism classification,animal, book subject] (Frequency: 1)

**Apache Subversion** [COMPUTERS] [software] (Frequency: 1)

**Business Process Execution Language** [COMPUTERS] [programming language,internet protocol] (Frequency: 1)

**Business analysis** [BOOKS] [book subject] (Frequency: 1)

Your domains: 13

BOOKS(28)  
 COMPUTERS(17)  
 TV(9)  
 FILM(6)  
 ORGANIZATION(6)  
 INTERNET(6)  
 AWARDS(4)  
 MUSIC(4)  
 BUSINESS(3)  
 MEDIA(3)  
 CONFERENCES\_AND\_CONVENTIONS(2)  
 EDUCATION(2)  
 BIOLOGY(1)

## Şekil 11. FunGuide Anlamsal Kullanıcı Profili Arayüzü

Projede alan-bağımlı veya alandan bağımsız öneri yapabilme yetenekleri mevcuttur. Kitap, spor, film ve müzik alanlarında alan-bağımlı öneri modülleri yazılmıştır. Şekil 12'de kitap öneri sorgusu verilmiştir. Buna göre, kullanıcı düğümü bulunur, kullanıcının bağlı olduğu ilgi alanı öğelerine bağlı öğelerin ait olduğu alanlar arasında BOOKS olanlar sonuç kümesine alınır. Şekil 13'te FunGuide arayüzü görülmektedir.

```
@Query("START user=node({0}) "
+ "match user-[bind]->(temp:Item)->(item:Item)-[domainBind]->(domain:Domain) "
+ "where not user --> item and (domain.id = \"BOOKS\") "
+ "return "
+ "item.id as itemId, "
+ "item.name as itemName, "
+ "temp.size, "
+ "sum(bind.frequency) as frequency, "
+ "collect(domainBind.domainDetailList) as domainDetailList, "
+ "collect(domain.name) as domainName, "
+ "collect(bind.frequency) as frequencyList, "
+ "collect(bind.dateList) as dateList, "
+ "collect(temp.name) as tempName "
+ "order by temp.size, frequency desc, item.name asc "
+ "limit 100")
List<MR_BookRecommendation> getBookRecommendations( UserAccount user );
```

## Şekil 12. Kitap öneri sorgusu

Your book recommendations: 27

**BOOKS: 7**

**AntiPatterns** [book, written work]

**Hitchcock** [written work, book]

**Java Persistence** [written work, book]

**MySQL Crash Course** [written work, book]

**Software Estimation** [written work, book] a book about history of Hollywood directors including Hitchcock

**The Semantic Web** [written work, book]

**Who the devil made it** [book, written work] author of "Search Computing: Challenges and Directions"

**AUTHORS :4**

**Marco Brambilla** [author, author, author] an ontology architect and semantic engineer

**David Decraene** [author] works for WWW Consortium

**Ivan Herman** [author] a researcher on NLP, semantic web, big knowledge and linked data.

**Jie Bao** [author]

**LITERATURE SUBJECTS :11**

**Computer Science** [book subject, periodical subject, bc

**Facebook** [book subject]

Şekil 13. Funguide kitap öneri modülü arayüzü

Şekil 14'te kullanıcının bir ilgi alanı ögesine ilgisinin hesaplanması sorgusu verilmiştir. Buna göre, kullanıcı ve ilgi alanı ögesi arasındaki en kısa patika bulunur.

```
@Query("START item=node({0}), user=node({1}) "
+ "match p=shortestPath(user-[*]->item) "
+ "return "
+ "length(p) as length, "
+ "[ n in nodes(p) | n.name ] as nodes")
List<MR_InterestForItem> getInterestForItem( Item item, UserAccount user );
```

Şekil 14. Kullanıcının belli bir ilgi alanı ögesine ilgisinin hesaplanması sorgusu



Şekil 15'te belli bir ilgi alanı ögesi ile ilgilenen kullanıcıların keşfi için sorgu verilmiştir. İlgi alanı ögesi düğümü hiper-çizgede bulunur ve bu düğümüne direk veya dolaylı olarak bağlı tüm kullanıcı tipi düğümler alınır. Bağlı olduğu patika daha kısa olan kullanıcı, öge ile daha ilgilidir.

```
@Query("START item=node({0}) "
+ "match p=((user:UserAccount)-[*]->item) "
+ "return "
+ "[ n in nodes(p) | n.name ] as allWays, "
+ "length(p) as length, "
+ "user.name as userName, "
+ "user.login as userLogin, "
+ "item.name as itemName "
+ "order by length asc "
+ "limit 15")
List<MR_TopUsersForItem> getTopUsersForItem( Item item );
```

Şekil 15. Belli bir ilgi alanı ögesi ile ilgili kullanıcıların keşfi sorgusu

Direk veya dolaylı olarak benzer ilgi alanlarına ilgi duyan kullanıcıları benzer kullanıcılar olarak adlandırabiliriz. Şekil 16'da benzer kullanıcıların keşfi için kullanılan sorgu sunulmuştur. Buna göre kullanıcı düğümü hiper-çizgede bulunur ve kullanıcının direk veya dolaylı olarak bağlı olduğu ilgi alanı ögelerine direk veya dolaylı olarak bağlı diğer kullanıcılar bulunur. Daha fazla sayıda ilgi alanı ögesine, daha kısa patikalarla bağlı kullanıcılar benzerdir.

```
@Query("START user=node({0}) "
+ "match p=(user-[*]->(item:Item)<-[*]->(simUser:UserAccount)) "
+ "where not user.login = simUser.login "
+ "return "
+ "length(p) as length, "
+ "count(item) as simCount, "
+ "collect(item.name) as itemName, "
+ "simUser.name as userName, "
+ "simUser.login as userLogin "
+ "order by length asc, simCount desc "
+ "limit 25")
List<MR_SimilarUsers> getSimilarUsers( UserAccount user );
```

Şekil 16. Benzer kullanıcıların keşfi sorgusu

Şekil 17'de sol üst köşede kullanıcının Big Data ile ilgili olduğu, ayrıca Awadshesh kullanıcısının da Big Data ile ilgili olduğu görülmektedir. Benzer şekilde, sol alt köşede kullanıcının ve Zuhail kullanıcısının İşler Güçler ile ilgili olduğu görülmektedir. Sağ tarafta Awadshesh ve Zuhail, bu ilgi alanlarından dolayı, kullanıcı ile benzer kullanıcılar olarak getirilmiştir.

The item to compute..

Your interest for Big data :

path length: 1 nodes: [null, Big data]

Top 15 users for Big data

**Awadhesh** (path length: 1) ([null, Big data])

**hilal** (path length: 1) ([null, Big data])

---

Your interest for İşler Güçler :

path length: 1 nodes: [null, İşler Güçler]

Top 15 users for İşler Güçler

**zuhal** (path length: 1) ([null, İşler Güçler])

**hilal** (path length: 1) ([null, İşler Güçler])

Similar Users:

**zuhal** (path length: 2, similar items: [İşler Güçler])

**Awadhesh** (path length: 2, similar items: [Big data])

Şekil 17. Funguide benzer kullanıcılar arayüzü

## 4.2 Değerlendirmeler

Bu proje üç farklı şekilde değerlendirilmiştir: otomatize edilmiş soğuk başlangıç değerlendirmesi, öneri sistemi için otomatize edilmiş normal değerlendirme ve gerçek kişi değerlendirmesi.

### Değerlendirme Metriği:

Değerlendirme sürecinde kullanılan metrik, hits-to-total oranı olarak isimlendirdiğimiz kavramdır. Kısmi profil entegre edilirken, halihazırda kullanıcı profiline direk veya dolaylı olarak bağlı bir kavram geldi ise, bu, sistemin bir şekilde bu kavramı kullanıcıya zaten önereceği anlamına gelir ve hit kabul edilir. Orandaki total, kısmi profildeki terim sayısı toplamıdır. Yüksek hits-to-total oranı, sistemin kullanıcının gelecekteki profilini başarılı bir şekilde tahmin ettiğini gösterir.

### Veri Kümesi:

Veri kümesi, 204 kullanıcının Facebook hesaplarındaki kısmi profiller toplanarak hazırlanmıştır. Kullanıcıların profilleri 3,4,7 gün arasında değişen zaman aralıklarıyla her kullanıcı için 12 kısmi profil olacak şekilde toplanmıştır.



Veri kümesi eğitime ve test etme kümelerine ayrılmıştır. Her kısmi profilin %80'i eğitime kümesinde yer almış, geri kalan %20'si test kümesinde bırakılmıştır.

#### *Baz Alınan Sistem:*

Baz alınan sistemde, ontoloji kullanımı ve anlamsal zenginleştirme çıkarılmıştır.

#### *Konfigürasyon:*

Alan eşik değeri 4 alınmıştır. Bunun anlamı, her ilgi alanı ögesi en az bir, en çok 4 alana bağlı olabilir.

Anlamsal zenginleştirme eşik değeri 5 alınmıştır. Her ilgi alanı ögesi için en çok 5 metaschema özelliği ile anlamsal genişletme yapılabilir.

Bu değerleri arttırmak oluşan modelin anlamsal özelliğini artırır. Ancak değerlendirmenin çalışma zamanını da arttıracığından bu eşik değerler uygun görülmüştür.

#### *Soğuk Başlangıç Değerlendirmesi:*

Veri kümesindeki 20 kullanıcı soğuk başlangıç testi için kullanılmıştır. Soğuk başlangıç testinde, her kullanıcı veri kümesinden çıkarılmakta, geri kalan kullanıcılar ile hiper-çizge doldurulmakta ve en son test kullanıcısı sisteme girilerek başarıyı ölçülmektedir. 20 kullanıcı için başarı oranı **0.52** olarak ölçülmüştür. Baz alınan testteki başarı oranı **0.03**'tür. Bu durum göstermektedir ki, sistem soğuk başlangıç durumunda da kullanımın gelecekteki ilgi alanlarını tahmin edebilmektedir.

#### *Normal Değerlendirme:*

Değerlendirme için tüm veri kümesi kullanılmıştır. Öncelikle sistem eğitime kümesi ile doldurulmuş ve oluşan hiper-çizge analiz edilmiştir. Buna göre hiper-çizgede 22746 düğüm yer almaktadır ve bunların 204'ü kullanıcı tipi, 22466'sı ilgi alanı ögesi tipi düğümleridir. En popüler alanlar music, awards, people, business, media ve film iken, en popüler olmayan alanlar library, skiing, zoos and aquariums, bicycles ve physics alanlarıdır.

Baz alınan sistemin değerlendirilmesinde performans düşüklüğünden dolayı 40 kişi kullanılmıştır.

Hits-to-total oranı **0.61** olarak hesaplanmıştır. Baz alınan sistemde elde edilen oran **0.25**'tir. Bu durum göstermektedir ki kullanıcı profillerini anlamsal olarak zenginleştirmek, kullanıcının gelecekteki ilgi alanlarını başarılı bir şekilde tahmin edebilmeyi sağlamıştır.

### *Gerçek Kişi Değerlendirmesi:*

Gerçek kişi değerlendirme için sistem <http://144.122.71.72:8080/signin> adresinde kullanıma açılmıştır. Kullanıcılar facebook, twitter ve linkedIn hesapları ile sisteme girmekte, bu üç sosyal ağdaki hareketlerinden elde edilen veriler hiperçizge yapısında entegre edilmektedir. Elde edilen bu bütünleşik kullanıcı profiline göre sistem kullanıcıya farklı alanlarda öneriler sunmaktadır. Kullanıcılardan önerilerin geçerliliğini 1..5 arasında puanla değerlendirmeleri istenmektedir. Sistem halen bilgi toplamaya devam etmektedir. Yeterli sayıda anket sonucu döndüğünde sistemin genel başarı değerlendirme yapılabilecektir.

Bu kısımda anlatılan çoklu-uygulamalar alanında veri modelinin kullanımı ayrı bir bildiri olarak sunulmuştur (Tarakçı ve Çiçekli 2014b) ve bu bildiri KEOD 2014 konferansında “En iyi bildiri” ödülü almıştır.

## **5. KİŞİSELLEŞTİRİLMİŞ İÇERİK ARAMA VE ÖNERME UYGULAMALARI**

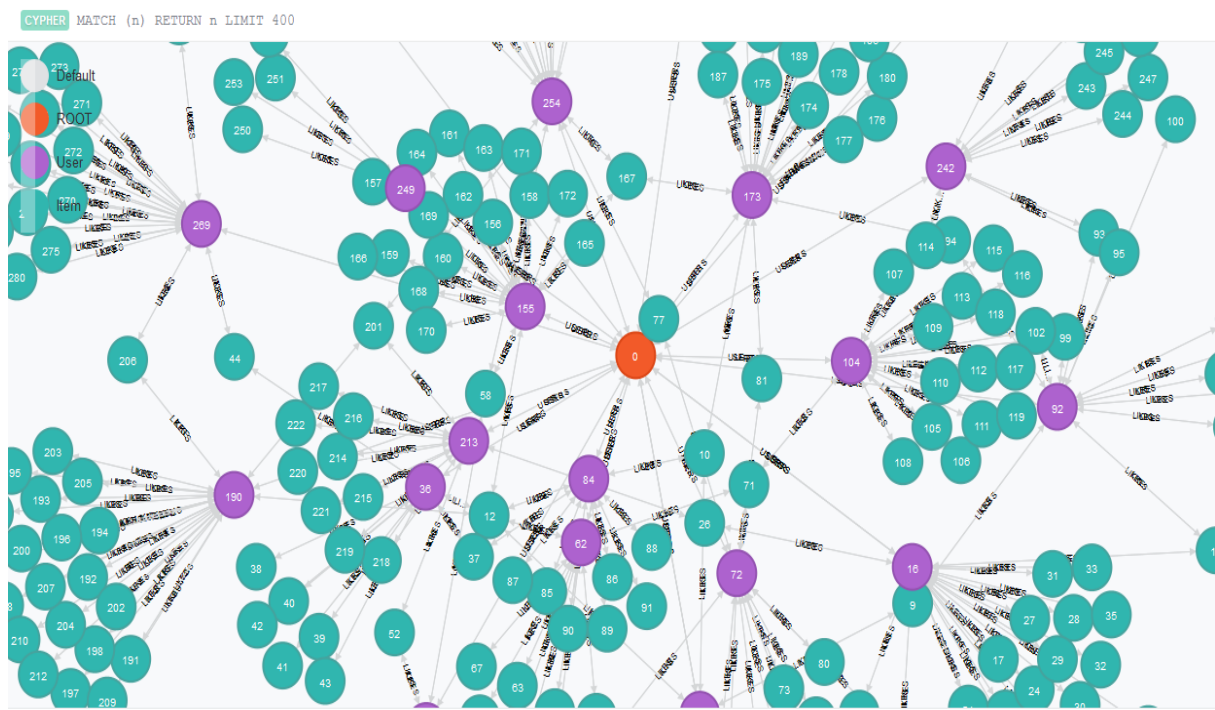
Raporun giriş bölümünde belirtildiği gibi hem kullanıcı profilinin farklı kaynaklardan alınan verilerle zenginleştirilmesi hem de kullanıcı profilinin değişik uygulamalarda kişiselleştirme amaçlı kullanımının sağlanması amacıyla örnek uygulamalar geliştirilmiştir. Bu amaçla proje kapsamında sosyal ağlardan elde edilen verilerin anlamsal olarak zenginleştirildikten sonra kullanıcı profili yaratmada ve yeni içerik önerisi yapmadaki katkısını gösterme için facebook verilerini kullanan bir öneri sistemi geliştirilmiştir. Ayrıca Akıllı TV’lerde kullanılmak üzere kullanıcıların izledikleri program içeriklerinden elde edilen profillerin anlamsal zenginleştirilmesini sağlayarak farklı öneri sistemi uygulamaları geliştirilmiştir. Anlamsal zenginleştirme çalışmaları kapsamında Türkçe için de çalışmalar yapılmış, ve Türkçe isimlendirilmiş varlıkları tanımak için algoritmalar geliştirilmiştir. Bu bölüm bütün bu yapılan çalışmaları anlatmaktadır.

### **5.1 Sosyal ağlardaki izlerden kullanıcı modeli oluşturulması**

Facebook, Twitter, LinkedIn gibi sosyal ağlar kullanıcıları hakkında farklı boyutlarda bilgi sunmaktadır. Kullanıcıların bu ağlarda bıraktıkları izlerden kullanıcı profili çıkarma işlemi son yıllarda araştırma konusu olmuştur. Bu projede bir yüksek lisans öğrencisi bu konuda literatür araştırması yapmış sosyal ağlardan Twitter ve Facebook için kullanıcı bilgilerinin çekilmesi için çalışmalar yapmıştır. Elde edilen verilerle yapılan deneysel çalışmalar neticesinde:

- Verilerin çok çeşitli olmasından dolayı gürültünün önlenmesi için bazı verilerin kullanılmaması,
- Verilerin birbirleri ile ilişkilerinin tanımlanarak veri kümesinin kontrollü daraltılarak anlamlandırılması,
- Bazı önemli verilerin tek başına kullanılmayacağı için çeşitli yollarla zenginleştirilerek kullanılmasına karar verilmiştir.

Sosyal ağlardan elde edilen veri Neo4j adlı çizge tabanlı veritabanında tutulmuştur. Elde edilen kullanıcı verisinin ham hallerinin (kullanıcı profillerinden direkt elde edilen veri) yanısıra Freebase kütüphanesi yardımıyla zenginleştirilmiş halleri de ilişki tipi ile beraber çizge tabanlı sisteme entegre edilmiştir. Veritabanının kısmi bir görüntüsü Şekil 18’de gösterilmiştir.



Şekil 18. Veritabanından bir kesit

Çizge tabanlı kullanıcı modeli elde edildikten sonra bu modeli temel alan bir öneri sistemi geliştirme çalışmaları yapılmıştır. İki farklı yöntem denenmiş, sonunda geliştirilen öneri sisteminin başarısı iki farklı anket çalışması yapılarak değerlendirilmiştir.

Öneri sisteminde kullanılacak kullanıcı modellerinin oluşturulan çizgeden elde edilebilmesi için iki yöntem belirlenmiştir. Bunlardan ilki vektör tabanlı kullanıcı modeli ve çizge tabanlı kullanıcı modelidir. Vektör tabanlı kullanıcı modelinde her bir kullanıcıya ait bilgiler ve bu bilgilerin model için ne kadar önem arz ettiğini gösteren bir değer vektörde tutulmaktadır.

Her bir kullanıcıyı merkez alarak, çizge üzerinde gezinerek kullanıcı verisinin doğrudan profilden mi geldiği, zenginleştirme sonucu mu geldiği bilgisine göre vektör inşa edilmektedir. Doğrudan gelen veriler için 1, birinci seviye zenginleştirme için 2, ikinci seviye için 3 ile işaretlenmiştir. Örneğin bir A kullanıcısının doğrudan profilden gelen sevdiği filmler, kitaplar, diziler 1 ile vektöre eklenirken; sevdiği filmlerin yönetmenleri, oyuncularını, kitapların yazarları 2 ile vektöre eklenmektedir. Bu yönetmenlerin diğer filmleri, oyuncuların oynadığı diğer filmler veya yazarların diğer kitapları ise 3 ile vektöre eklenmektedir. Bu derinlik seviyesini 2'de kesmemizin nedeni, derinlik arttıkça kullanıcı ile bu türetilmiş verilerin ilişkilerinin zayıflaması ve öneminin azalmasıdır. Kullanıcının sevdiği bir filmin yönetmeninin yönettiği başka bir filmde oynayan bir oyuncunun modelde derinlik seviyesi 3 olarak yer almasının çok etkisi olmayacağı varsayılmıştır.

Çizgeden elde edilen vektör tabanlı kullanıcı modeli ile tüm kullanıcılar bu şekilde vektörlere dönüştürülmüştür. Daha sonra bu vektörler arasında, kosinüs benzerliği ile benzerlikleri bulunmuş ve bu değerler normalize edilerek (0-1 aralığına sıkıştırılarak) bir matrise doldurulmuştur. Bu matrisin satır ve sütunlarında kullanıcı isimleri yer alırken, satır sütun kesişimlerinde bu iki kullanıcının bu yöntem ile hesaplanmış benzerlikleri yer almaktadır. Bu matriste her bir  $i=j$  değeri 1 olan (bir kullanıcı kendisine %100 benzerdir) asal köşegenine göre simetrik olan bir matristir.

Bu aşamada kümeleme yöntemi (clustering) ile kullanıcılar bir araya getirilir. Bu yöntemde oluşturulan baz matris üzerinde gezinerek en büyük benzerliğe sahip olan kullanıcılar bir araya getirilir. Bu iki kullanıcının vektörleri bir araya getirilerek yeni bir kullanıcıymış gibi diğer kullanıcılar ile benzerliği tekrar hesaplanarak matrise eklenir. Yeni elde edilen matriste yeni en yüksek değer bulunur ve bir araya getirme ve matrisi güncelleme işlemi bu şekilde devam eder. Her bir kullanıcı bir gruba eklendiğinde bu işlem sona erer. Bu işlemin nihayetinde, birbirine en çok benzeyen kullanıcılar bir araya getirilmiş olur. Günlük hayattan örnek vermek gerekirse, herkesin birbirine yabancı olduğu bir ortamda, ilk yakınlaşma en çok ortak zevklere sahip insanlar arasında olur. Bu iki kişi artık bir arkadaş grubudur ve grup halinde hareket ederler. Diğer kişiler, bu yeni grupla ortak zevklere sahiplerse bu gruba dahil olurlar değilse ortak zevkleri fazla olanlar kendi gruplarını kurarlar. Bu şekilde her bir kullanıcı bir gruba dahil olduğunda işlem tamamlanmıştır. Ortak zevkleri en çok olan insanlar bir araya gelmiştir. Bu ortak zevkleri olan gruplardan, bir üyenin beğendiğini diğerinin de beğeneceği varsayımı ile tavsiye sistemi geliştirilebilir.

Bu yöntemde en önemli sorun, hiçbir gruba dahil olamayan, kimseyle ortak zevki olmayan kullanıcılar olabilir. Sosyal ağdan toplanan verilerde böyle bir durumun oluşması neredeyse

imkansızdır. Kısıtlı verisi olan kullanıcılar için tavsiye sistemi verimli çalışmamaktadır. Verisi yeterli kullanıcılarda ise zenginleştirme işlemi sonrasında gelen veriler mutlaka eşleşmektedir. Bu kullanıcılarda filmler için tarz, kitaplar ve müzikler için tür, diziler ve spor takımları için kategori bilgileri gibi çok genel kavramlar mutlaka eşleşmektedir.

Modelleme ve tavsiye oluşturma aşamasında kullanılan diğer yöntem çizge tabanlı modelin direk olarak kullanılmasıdır. Standart ilişkisel veritabanları yerine kullanılan çizge tabanlı veritabanı sayesinde bu yöntem kullanılabilmiştir. Bu yöntemde, kullanıcılar arasında benzerlikler bulunurken, vektör tabanlı modellemede kullandığımız kosinüs benzerliği yerine ortalama uzaklık (average path length) hesaplama yöntemi kullanılmıştır. Bu yöntem ile her bir kullanıcıdan diğerine giden tüm yolların uzaklığı (her bir düğüm arası uzaklık 1 birim kabul edilmiştir.) hesaplanarak, bu yolların ortalamaları alınmıştır.

Kullanıcıların vektör tabanlı modellemede olduğu gibi, bu uzaklıklara göre bir araya getirilmesi ile kümeler oluşturulmuştur. Oluşturulan kümeler sanal kullanıcı olarak mevcut düğümlere yeni kenarlar eklenerek çizgeye dahil edilmektedir. Bu yöntem ile mevcut kullanıcılar korunmaktadır. Vektör tabanlı modellemede, her bir kullanıcı tek bir gruba ait olabilirken, çizge tabanlı modellemede farklı gruplarda yer alabilmektedir. Vektör tabanlı modellemede kullanıcı bir gruba dahil edildiğinde o kullanıcı için artık işlem yapılamamakta iken çizge tabanlı yaklaşımda, o kullanıcı çizge içerisinde halen yer almakta ve diğer kullanıcılar ve sanal kümelerle uzaklığına göre hesaplamalara dahil edilmektedir.

Bu yöntemde en önemli sorun düğümler arasında döngüler oluşması durumunda ya sonsuz sayıda çözüm çıkmakta ya da düğümler arası uzaklık olması gerekenden fazla çıkmaktadır. Bu sorunu çözmek amacıyla, döngü tespit algoritmalarından Floyd algoritması<sup>1</sup> kullanılmıştır. Kendi çizgemize göre özelleştirdiğimiz bu algoritma ile döngü oluşturan düğümlerden en kısa olanı işleme dahil edilmiştir.

Çizge tabanlı yöntem ile vektör tabanlı yöntemde kullandığımız 2 derinlik seviyesinin getirdiği kısıtlama aşılarak bir modelleme yapılması sağlanmıştır. Bu yöntemin dezavantajı ise çizge tabanlı işlemlerin performans noktasında sistemi oldukça zorlamasıdır.

Bu iki yöntem ile gelen tavsiyeler bir küme haline getirilerek kullanıcılarımıza anket olarak sunulmuştur. Anket sistemi olarak Limesurvey uygulama çatısı özelleştirilerek kullanılmıştır. İnternette online anketler veya hazır sistemler yerine bu şekilde özelleştirilmiş bir sistem kullanmamızın en önemli sebebi, anket cevaplarımızın kişilere özel olarak hazırlanması zorunluluğudur. Tavsiye edeceğimiz unsur her bir kullanıcı için ayrı olarak üretilmektedir.

Genel anket sistemlerinde ise, statik bir anket farklı kullanıcılara uygulanmakta ve sonuçlar incelenmektedir.

Özelleştirilmiş Limesurvey sisteminde iki tip anket hazırlanmıştır. Bu anketlerden ilkinde, tek cevaplı çoktan seçmeli sorular sorulmuştur. 4 şıklı çoktan seçmeli sorularda, bir adet doğru cevaba karşılık, aynı kategoriden 3 tane rastgele seçilmiş tavsiye kümesinde olmayan unsur da şıklarda yer almıştır. Bu soruların temelinde yer alan bazı tavsiye mantıkları şöyledir:

Film ile ilgili tavsiyelerde:

- Yer aldığı kümede kendisine çok benzeyen kullanıcıların sevdiği filmler,
- Sevdiği filmlerin yönetmenlerinden en fazla tekrar edenler,
- Bu sevdiği yönetmenlerin yönettiği diğer filmler,
- Sevdiği filmlerin oyuncularından en fazla tekrar edenler,
- Bu sevdiği oyuncuların oynadığı diğer filmler,
- Sevdiği filmlerin kategorilerinden en fazla tekrar edenler,

Müzik ile ilgili tavsiyelerde:

- Yer aldığı kümede kendisine çok benzeyen kullanıcıların sevdiği müzisyenler,
- Sevdiği müzisyenlerin kategorilerinden en fazla tekrar edenler,
- Bu müzik kategorilerinde müzik yapan diğer müzisyenler,

Kitap ile ilgili tavsiyelerde:

- Yer aldığı kümede kendisine çok benzeyen kullanıcıların sevdiği kitaplar
- Sevdiği kitapların kategorilerinden en fazla tekrar edenler,
- Sevdiği kitapların yazarlarından en fazla tekrar edenler,
- Bu sevdiği yazarların yazdığı diğer kitaplar,

Televizyon ile ilgili tavsiyelerde:

- Yer aldığı kümede kendisine en çok benzeyen kullanıcının sevdiği televizyon programları,
- Sevdiği programların türlerinden en çok tekrar edenler,
- Sevdiği programlarda oynayan oyunculardan en çok tekrar edenler,
- Bu sevdiği oyuncuların yer aldığı diğer televizyon programları,

Spor ile ilgili tavsiyelerde:

- Yer aldığı kümede kendisine en çok benzeyen kullanıcının sevdiği spor türleri veya spor takımları,

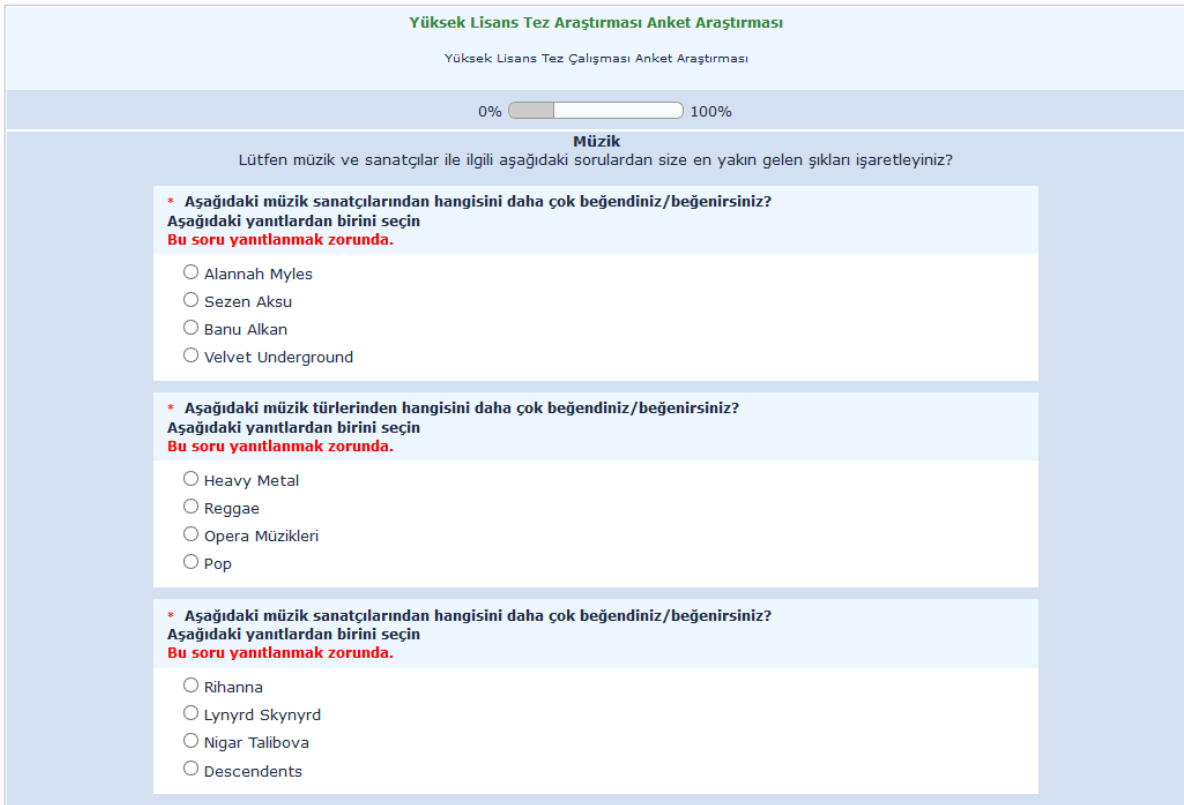
---

<sup>1</sup> en.wikipedia.org/wiki/Cycle\_detection



- Sevdiği spor takımlarının yer aldığı spor türleri,
- Sevdiği spor takımlarının bir oyuncusu,
- Sevdiği spor türlerinden ünlü bir oyuncu.

Bu tavsiye yöntemleri kişilerin profillerinin zenginleştirilmesi ile gelen içerik tabanlı tavsiyeler (content-based recommendation) ile kümeleme yöntemi yardımıyla benzer kullanıcıların bir araya getirilmesine dayalı işbirliği tabanlı tavsiyelerden (collaborative-based recommendation) gelen sonuçların bir araya getirilmesine dayanmaktadır. Bu tip ankete ait örnek bir ekran görüntüsü Şekil 19 deki gibidir.



**Yüksek Lisans Tez Araştırması Anket Araştırması**

Yüksek Lisans Tez Çalışması Anket Araştırması

0% 100%

**Müzik**

Lütfen müzik ve sanatçılar ile ilgili aşağıdaki sorulardan size en yakın gelen şıkları işaretleyiniz?

\* Aşağıdaki müzik sanatçılarından hangisini daha çok beğendiniz/beğenirsiniz?  
Aşağıdaki yanıtlardan birini seçin  
**Bu soru yanıtlanmak zorunda.**

Alannah Myles  
 Sezen Aksu  
 Banu Alkan  
 Velvet Underground

\* Aşağıdaki müzik türlerinden hangisini daha çok beğendiniz/beğenirsiniz?  
Aşağıdaki yanıtlardan birini seçin  
**Bu soru yanıtlanmak zorunda.**

Heavy Metal  
 Reggae  
 Opera Müzikleri  
 Pop

\* Aşağıdaki müzik sanatçılarından hangisini daha çok beğendiniz/beğenirsiniz?  
Aşağıdaki yanıtlardan birini seçin  
**Bu soru yanıtlanmak zorunda.**

Rihanna  
 Lynyrd Skynyrd  
 Nigar Talibova  
 Descendents

Şekil 19. Birinci anket çalışması

Diğer anket tipinde ise kullanıcılara, kategorilere ayrılmış tüm tavsiyeler bir küme olarak sunulmuş ve kendilerine yakın hissettiklerini işaretlemeleri istenmiştir. Her iki modelden gelen sonuçlar bir araya getirilerek iki modelden de gelen tavsiye unsurlarına öncelik verilecek şekilde, kullanıcılara çoklu seçim imkanı sağlanmıştır. Bu tip ankete ait örnek bir ekran görüntüsü Şekil 20 de verilmiştir.

**Yüksek Lisans Tez Çalışması**  
Yüksek Lisans Tez Çalışması Anket Araştırması

0%  100%

**Film**  
Lütfen beğendiğiniz/beğenebileceğiniz filmlerin ve oyuncuların yanındaki kutucukları işaretleyiniz?

\* Aşağıdaki filmlerden hangilerini beğenirsiniz?  
Uyanların tümünü seçin

- The Butterfly Effect
- The Karate Kid
- Lord of the Rings
- Spider Man
- Avatar
- Sherlock Holmes
- Captain America
- The Last Samurai
- gladiator
- A Beautiful Mind

\* Aşağıdaki film kategorilerinden hangilerini beğenirsiniz?  
Uyanların tümünü seçin

- Aksiyon
- Macera
- Bilim-Kurgu
- Gerilim
- Komedi
- Kara Mizah
- Psikolojik Gerilim
- Western
- Video Oyunu
- Biyografi
- Müzikal
- Savaş
- Uzay
- Melodram
- Kısa film
- Felaket

**Şekil 20. İkinci anket çalışması**

Hazırlanan bu iki anket tipinin bazı zayıflıkları ve güçlü yönleri bulunmaktadır. Tek cevaplı anket sisteminin olumsuz yönlerinden ilki, rastgele seçilen 3 diğer şık bizim tavsiyelerimizden daha yakın gelebilir. Örneğin profilinde yer almayan ama yeni izleyip beğendiği bir film bu 3 şık arasında kaldığında kullanıcı bu şıkkı seçebilmektedir. Ayrıca tek doğru cevap ürettiğimiz için, kullanıcının o tek doğru cevap hakkında bilgisi veya fikri yoksa bu şıkkı pas geçebilmektedir. Çoklu seçim anket yönteminde, ilk sorulardaki tüm cevaplar sistemimizin ürettiği tavsiyelerden oluşmaktadır. Anket kullanıcıları, bu soruda tüm şıkları işaretlemekten çekinebilmektedir. Sonuçları incelediğimizde, bu soru tipinde ilk birkaç şık mutlaka işaretlenirken, alt kısımda kalan cevaplar daha az işaretlenmektedir. Bu anketler sonunda elde ettiğimiz sonuçlarımız şu şekildedir.

Sistem 250'den fazla kullanıcı için test edilmiş ve performans sorunu gözlenmemiştir. Son kullanıcılara erişilerek tavsiye sistemimizin başarısının ölçülmesi gerekliliğinden dolayı 10 kullanıcı için bahsedilen iki anket uygulanmıştır. Tek cevap şıklı çoktan seçmeli anket çalışmamızda aşağıdaki sonuçlar elde edilmiştir. Sonuçların detaylarına indiğimizde, film yönetmenlerine bağlı tavsiyeler ile kitap kategorisine bağlı tavsiyelerin başarılarımızı düşürdüğü

gözlenmiştir. Anket yaptığımız kullanıcılara bu konu sorulduğunda, kullanıcıların filmlerin yönetmenlerine çok dikkat etmedikleri ve profillerindeki bazı kitapları okumadan beğendiklerini söylemişlerdir.

U1	U2	U3	U4	U5	U6	U7	U8	U9	U10
9/13	8/9	12/16	16/19	9/13	5/7	17/19	13/16	10/13	13/17
0.69	0.89	0.75	0.84	0.69	0.71	0.89	0.81	0.77	0.76

Max	0.89
Min	0.69
Avg	0.78

Çoklu seçim anket yönteminde elde ettiğimiz sonuçlar aşağıdaki gibidir. Bu anket tipinde ilk soruda o kategoriye ait tavsiyeler yer alırken diğer soruda profil zenginleştirilmesi sonucu elde edilen verilerin doğruluğu kontrol edilmiştir. Film kategorileri, dizi ve film türleri, müzik ve spor çeşitleri ikinci soruda yer almıştır.

U1	U2	U3	U4	U5	U6	U7	U8	U9	U10
Tahmin	Tahmin	Tahmin	Tahmin	Tahmin	Tahmin	Tahmin	Tahmin	Tahmin	Tahmin
0.67	0.65	0.67	0.61	0.50	0.63	0.67	0.80	0.67	0.78
Kategori	Kategori	Kategori	Kategori	Kategori	Kategori	Kategori	Kategori	Kategori	Kategori
0.86	0.90	0.83	0.75	0.82	0.88	0.65	0.67	0.81	0.80

	Avg	Min	Max
Tahmin	0.665	0.50	0.80
Kategori	0.797	0.65	0.90

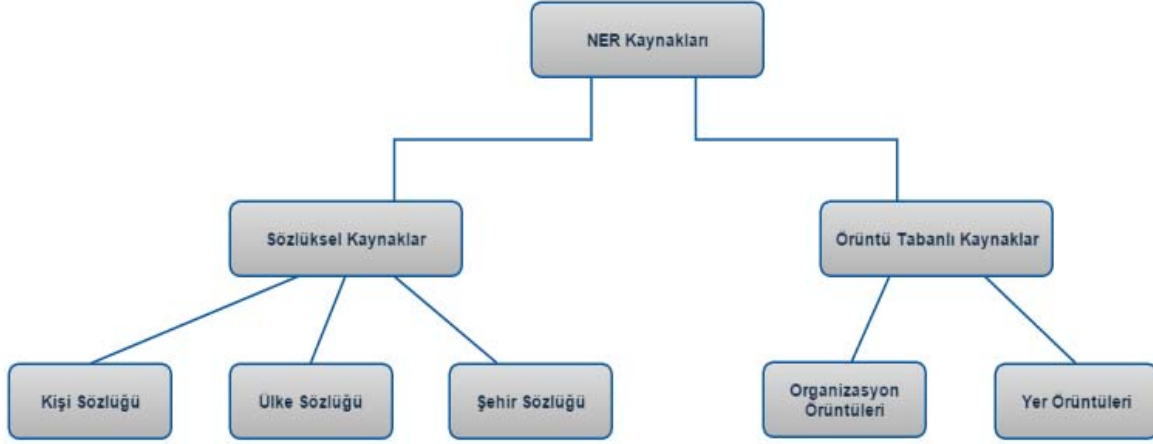
Özetle, sistemimiz çok fazla kullanıcı için çalışabilmektedir. Kısıtlı bir küme ile yapılan anket çalışması neticesinde, tek cevap şıklı çoktan seçmeli ankette ortalama %78 oranında başarı elde edilirken, çoklu seçim ankette %67'lik başarı oranı yakalanmıştır. Bahsedilen gerekçelerden dolayı filmler için yönetmen bilgisine dayalı tavsiyeler ve kitap tavsiyeleri sistemden çıkarıldığında daha yüksek başarı elde edilmesi mümkündür. Bu sistem bir yüksek lisans tezi kapsamında yapılmıştır. Öğrenci Ocak 2015'te jüriye girip mezun olmayı planlamaktadır.

## 5.2 İsimlendirilmiş varlık tanıma ile anlam zenginleştirme

Bu proje kapsamında öneri sistemlerinin performansını artırmak ve oluşturulan kullanıcı profillerini zenginleştirmek için varlık ismi tanıma yöntemlerinden faydalanılmıştır. Bu amaçla

gerçekleştirilen çalışmaları; kural tabanlı varlık ismi tanıma, bilgi toplama, anlam ayrımı, prototip uygulaması ve deneysel çalışmalar olarak sıralayabiliriz.

Kural tabanlı varlık ismi tanıma işlemi, sözlük ve örüntü tabanlı iki farklı tür kaynağın beraber kullanılmasıyla geliştirilmiştir. Şekil 21’de bu kaynakların sınıflandırılması gösterilmiştir.



**Şekil 21. Varlık ismi tanıma kaynaklarının sınıflandırılmış gösterimi**

Sözlük tabanlı kaynak; kişi, ülke ve şehir sözlüğünden oluşmaktadır. Bu sözlüklerin oluşması için çoğunluğu Türkçe olan yaklaşık olarak 12000 özel isim farklı kaynaklardan toplanmıştır. Örüntü tabanlı kaynak ise organizasyon ve yer isimlerini bulmaya yarayan örüntülerden oluşmaktadır. Bu örüntüler daha önceki raporda örnekleri verilen kuralların geliştirilmesiyle kapsamlı bir şekilde oluşturulmuştur. Örüntülerin metinlere uygulanması sırasında aşağıdaki algoritma izlenmiştir:

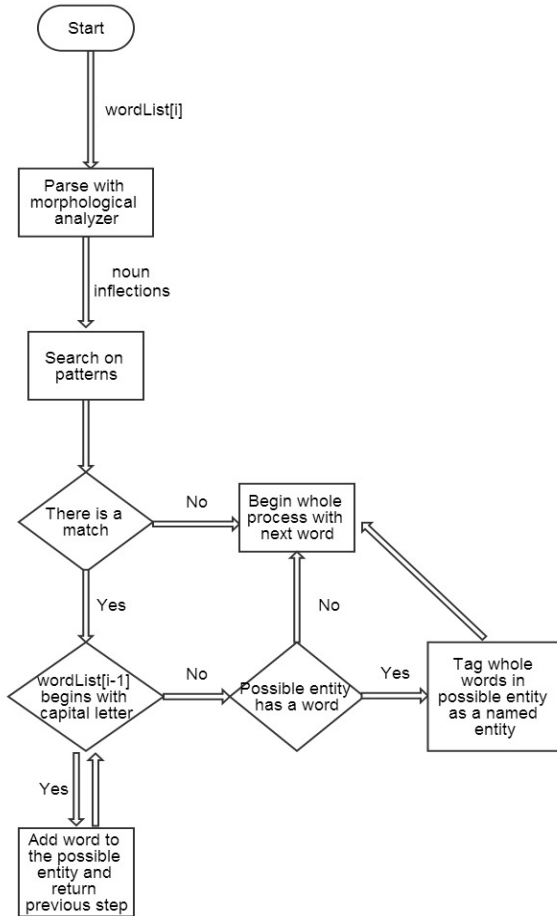
1. Bir metindeki cümlelerin sınırlarını Zemberek kütüphanesi yardımıyla belirle.
2. Her cümlenin kelimelerinden oluşan bir vektör oluştur.
3. Vektördeki her kelime için kısaltma, yer, organizasyon ve kişi ismi bulmak için geliştirilen örüntü ve sözlük tabanlı kuralları çalıştır.
4. Bir veya birden fazla kelime varlık ismi olarak etiketlenirse, bu kelime guruplarını etiketlenmiş isimler listesine ekle.

Bu algoritma içerisinde önemli bir yere sahip olan, her kelime tipi için çalıştırılan, örüntü ve sözlük tabanlı kurallardan oluşan varlık bulucular (entity finder) tanımlanmıştır. Bunlar kısaltma, yer, organizasyon ve kişi varlık bulucular olarak sıralanabilir.

**Kısaltma Varlık Bulucu:** Bu varlık bulucu basitçe her kelimeyi işleyerek, kelimenin bütün harflerinin büyük harflerden oluşup oluşmadığını kontrol eder. Eğer harflerin hepsi büyükse bu kelime bir kısaltma olarak etiketlenir. Harfler arasında ‘.’ işaretinin olması bu durumu bozmaz ve o kelimeler de kısaltma olarak etiketlenir.

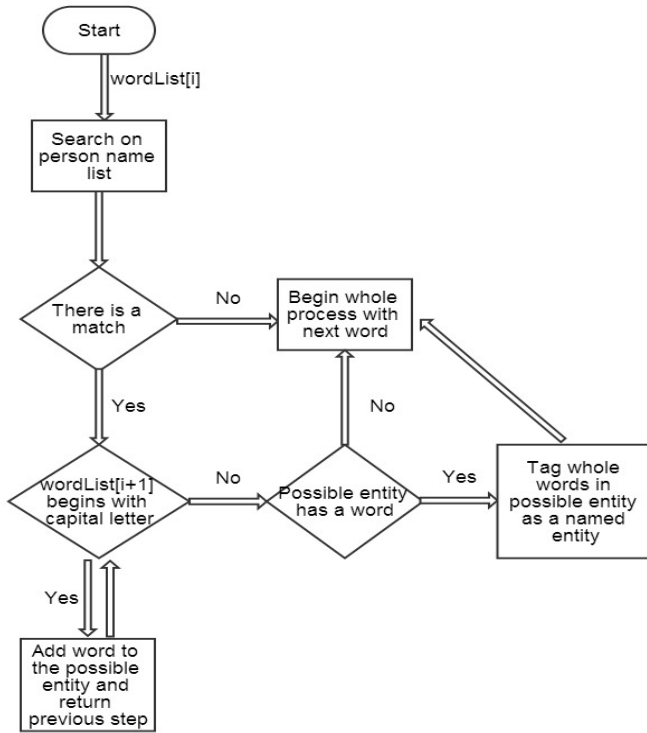
**Yer Varlık Bulucu:** Yer varlık bulucu ilk olarak yer sözlüğündeki kelimeleri verilen cümle içerisinde bütün olarak arar. Eğer bir eşleşme gerçekleşirse bu kelime ya da kelime gurubu yer varlık ismi olarak etiketlenir. Örneğin, verilen cümle ‘Çek Cumhuriyeti’ kelime gurubunu içeriyorsa; bu kelime gurubu yer sözlüğünde bulunduğundan yer ismi olarak etiketlenir. Sözlüksel kaynak kullanıldıktan sonra yer ismi kurallarından oluşan örüntü tabanlı kaynak kullanılır. Bunun için öncelikle her kelimenin kökü ve ekleri Zemberek yardımıyla ayrıştırılır. İsim olan kök ve türevleri örüntüler üzerinde aranır. Eğer bir eşleşme gerçekleşirse bu kelime muhtemel bir yer ismi olarak etiketlenir ve bu kelime öncesindeki kelimeler büyük harfle başlamayan bir kelime bulunana kadar tek tek analiz edilir. Örneğin “Akşam saatlerinde İstiklal Caddesi’ni dolduran göstericiler sloganlar atarak ilerledi” cümlesinde *İstiklal Caddesi* yer ismi olarak etiketlenir. Bu örnekte cadde kelimesi örüntüler içerisinde bulunduğundan bu kelime öncesindeki İstiklal kelimesi ile birlikte etiketlenir. Bu yaklaşımın grafiği Şekil 22’de görülmektedir.

**Organizasyon Varlık Bulucu:** Bu varlık bulucu da yer varlık bulucusu ile aynı işlemleri herhangi bir organizasyon sözlüğü kullanmadan gerçekleştirir. Şekil 22’de gösterilen örüntü tabanlı algoritma akışı bunun için de geçerlidir.



Şekil 22. Yer ve organizasyon isimlerinin örüntü tabanlı kaynaklar kullanarak etiketlenmesi işlemi

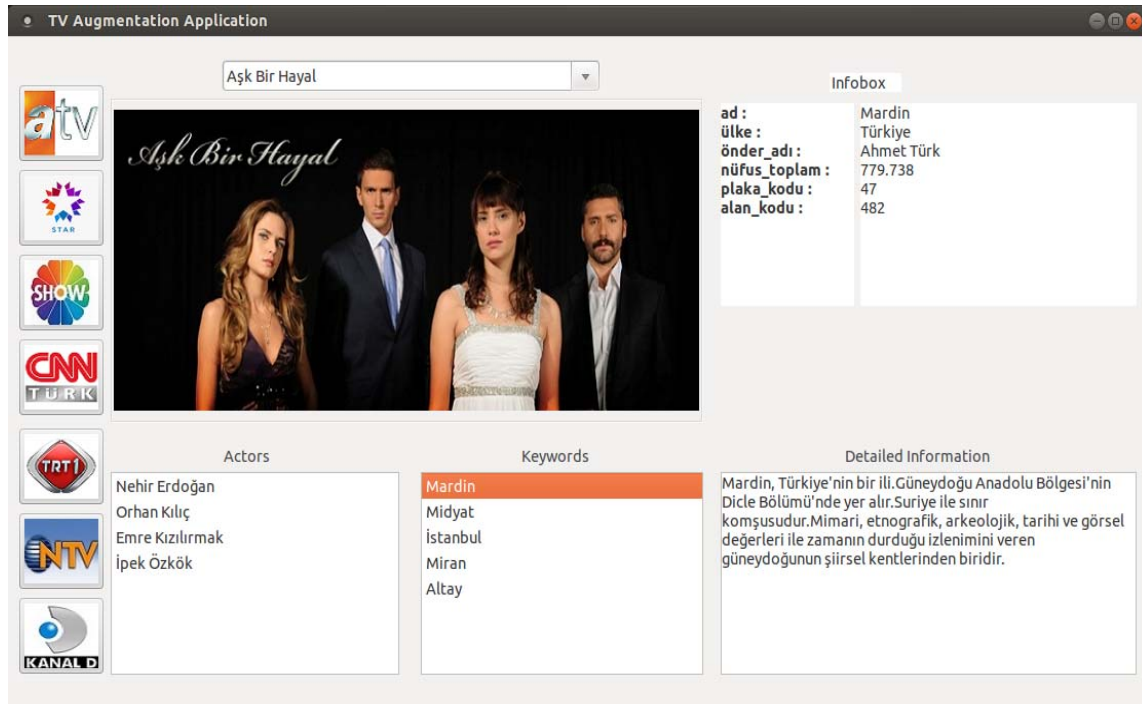
**Kişi Varlık Bulucu:** Kişi isimleri sadece sözlüksel kaynak kullanılarak bulunur. Bu varlık bulucu, sözlükle birlikte örüntü tabanlı bir kaynak kullanmak yerine isimlerden sonra gelen soy isimleri bulmak için de basit bir kuralı dikkate alır. Kelimeler sözlük listesinde arandıktan sonra eğer bir eşleşme gerçekleşirse, bu kelime kişi ismi olarak etiketlenir ve daha sonraki kelimeler küçük harfle başlayan bir kelime bulunana kadar incelenir. Bulunan bir kişi isminden sonra gelen büyük harfli kelimeler de kişi ismine eklenerek bütün isim etiketlenmiş olur. Örneğin “Akşam saatlerinde bir açıklama yapan Ersin Mustafa Özbükey, çalışmalarını tamamladığını söyledi” cümlesindeki *Ersin Mustafa Özbükey* kişi ismi olarak etiketlenir. Bu yaklaşımın grafiği Şekil 23’de görülmektedir.



**Şekil 23. Kişi isimlerinin sözlük tabanlı kaynaklar kullanarak etiketlenmesi işlemi**

Bulunan varlık isimlerini, kullanıcı profillerinde daha sağlıklı bir şekilde kullanmak için anlam ayrımı işlemi de gerçekleştirilmiştir. Bunun için her varlık isminin tanımı Vikipedi aracılığıyla internet üzerinden toplanmıştır. Eğer bir kelimenin birden fazla anlamı varsa; bizim bulduğumuz varlık isminin hangi anlamda kullanıldığını anlamak için, ismin geçtiği metin ve bütün Vikipedi anlamları arasında benzerlik hesaplaması yapılır. Bu hesaplama tf-idf vektör benzerliği yöntemi kullanılarak gerçekleştirilir. Varlık isminin bulunduğu metine en çok benzeyen Vikipedi tanımı o ismin gerçek anlamı olarak ele alınır. Ayrıca ilgili kelimenin Vikipedi sayfasında bulunan bilgi kutusu da ayrıştırılarak bu kelimenin önemli bilgileri çıkarılır. Daha sonra bu bilgiler kelimenin türüne göre özet haline getirilir. Örneğin kişi isimleri için doğum yeri, tarihi gibi bilgiler tutulurken; yer isimleri için nüfus ve coğrafi bilgiler özet bilgiye dâhil edilir.

Geliştirmesi tamamlanan bu varlık ismi tanıma ve anlam ayrımı modülünün akıllı TV'lerde kullanılabilecek bir de prototip uygulaması yapılmıştır. Bu TV içeriğini zenginleştirme uygulamasında; toplam 6 televizyon kanalı ve 32 programın EPG metinleri kullanılarak, bu metinlerde geçen önemli anahtar kelimeler çıkarılmış, tanımları ve özet bilgileri kullanıcıya TV programı ile birlikte sunulmuştur. Bu EPG metinleri, Radikal haber sitesinin TV rehberi sayfasından html ayrıştırma yöntemi ile elde edilmiştir. Yaklaşık 3 aylık bir TV program rehberi bilgisi veri tabanına kaydedilmiş ve kullanıma hazır hale getirilmiştir. Şekil 24'de bu prototip uygulamanın bir ekran görüntüsü görülmektedir.



Şekil 24. TV içeriği zenginleştirme uygulaması ekran görüntüsü

Bu ekran görüntüsünde sol tarafta görülen resimler kanalların ikonlarıdır ve her biri bir tuştur. Bir ikona basıldığında ilgili kanalın program listesi temsili resim üzerindeki listeli giriş kutusu içerisinde gösterilir. Bu listeden bir program seçildiğinde, o programın temsili resmi ekranda görülür ve programa ait çıkarılan önemli kelimeler ve oyuncularını bu temsili resim altında listelenir. Seçilen bir anahtar kelimenin veya oyuncunun detay bilgisi ve özet bilgisi de ekranın sağ tarafında görülür. Bu uygulamada TV kullanıcısına, bir programı izlerken o programla ilgili detay ve ek bilgiye gerçek zamanlı ulaşma imkânı sağlanmıştır.

Geliştirilen varlık ismi tanıma algoritmasının başarısını ölçmek için (Tatar, S., Cicekli, I. 2011) çalışmasında kullandığı, Türkçe haber metinlerinden oluşan ve varlık isimleri etiketlenmiş bir derleme kullanılmıştır. Bu deneysel çalışma sonrasında tamı tamına eşleşme

(exact matching) uygulandığında genel performans %79.81 olarak elde edilmiştir. Bu sonuç Türkçe benzer sistemlerle karşılaştırıldığında başarılı bir sonuç olarak değerlendirilebilir. Bu yöntemle göre elde edilen bütün sonuçlar Tablo 1’de gösterilmiştir.

**Tablo 1. Tamı Tamına Eşleşme Yöntemine Göre Değerlendirme Sonuçları**

<i>Varlık İsmi</i>	<i>Precision (%)</i>	<i>Recall (%)</i>	<i>F-score (%)</i>
<i>Kişi</i>	82.19	86.33	84.21
<i>Yer</i>	74.62	86.50	80.12
<i>Organizasyon</i>	76.38	76.00	76.19
<i>Genel</i>	76.76	83.12	79.81

Varlık isimlerinin örtüşen sınırlarına göre başarısı ölçüldüğünde ise bu sonuç %85’e kadar çıkmıştır. Bu sonuçlar; oluşturulan sözlüklerin ve geliştirilen kuralların gayet kapsayıcı olduğunu göstermektedir. Bu yöntemle göre elde edilen sonuçlar Tablo 2’de gösterilmiştir.

**Tablo 2. Parçalı Eşleşme Yöntemine Göre Değerlendirme Sonuçları**

<i>Varlık İsmi</i> <i>Kategorisi</i>	<i>Precision (%)</i>	<i>Recall (%)</i>	<i>F-score (%)</i>
<i>Kişi</i>	82.43	86.52	84.42
<i>Yer</i>	77.54	90.12	83.35
<i>Organizasyon</i>	86.20	85.75	85.97
<i>Genel</i>	81.57	88.52	84.90

Bu değerlendirmelerin dışında, varlık ismi tanıma yönteminin kullanıldığı TV içerik zenginleştirme sisteminin de değerlendirmesi yapılmıştır. Bu değerlendirmenin yapılması için aslında kullanıcı anketi tanımlamak ve anket sonuçlarına göre kullanıcı memnuniyetini ölçmek gerekmektedir. Fakat bunun için geliştirilen prototip uygulamasının akıllı TV’ler üzerinde çalıştırılarak kullanıcıların kullanımına açılması gerekmektedir. Uygulama akıllı TV’ler üzerinde henüz çalıştırılmadığı için bu değerlendirme Radikal program rehberinden toplanan programlar aracılığıyla gerçekleştirilmiştir. Televizyonun en çok izlendiği zamanlarda (prime time) yayınlanan yaklaşık 1700 programın EPG metinleri üzerinde varlık ismi tanıma algoritmamız çalıştırılmıştır. Bulunan varlık isimlerinin Vikipedi’den tanımları da çekilmiştir. Bu değerlendirmede, bulunan varlık isimleri ve detay bilgisi çekilen varlık isimleri oranlanarak içerik zenginleştirme ölçümü (Augmentation Rate) yapılmaya çalışılmıştır. Bu ölçüm şu şekilde tanımlanabilir:

**Zenginleştirilmiş Varlık İsmi = Vikipedi’de sayfası bulunan varlık ismi**



$$\text{Zenginleştirme Oranı} = \frac{\text{Zenginleştirilmiş Varlık İsmi Sayısı}}{\text{Sistem Tarafından bulunan Varlık İsmi Sayısı}}$$

Belirlenen 1700 program üzerinde yapılan bu değerlendirme sonuçları Tablo 3’de gösterilmiştir.

**Tablo 3. Sistemin Zenginleştirme Oranı Performansları**

	<i>Kişi İsimleri</i>	<i>Yer İsimleri</i>	<i>Organizasyon</i>	<i>Bütün İsimler</i>
	<i>(%)</i>	<i>(%)</i>	<i>İsimleri(%)</i>	<i>(%)</i>
<i>Zenginleştirme Oranı</i>	34.56	83.93	61.85	49.52

Tablo 3’de görüldüğü gibi zenginleştirme oranı en az kişi isimleri için gerçekleşmiştir. Bunun sebebi EPG metinlerinde geçen birçok yaygın kişi isminin Vikipedi sayfasında bulunmuyor olmasındandır. Örneğin ‘Ahmet’, ‘Zeynep’, ‘Bülent’ gibi isimler kişi olarak etiketlenmesine rağmen, Vikipedi sitesi için özel bir anlam içermediğinden zenginleştirilememektedirler. Yer isimleri için bunun tam tersini söyleyebiliriz. Birçok yer ismi Vikipedi sayfasında bulunduğu için etiketlenen varlık isimlerinin detayları bu siteden çekilebilmekte ve böylece zenginleştirme işlemi yapılabilmektedir. Bu yüzden yer isimleri için bu oran %84'lere kadar çıkmıştır. Organizasyon isimlerinde ise yapılan incelemelerde, birçok Türkçe organizasyonun Vikipedi internet sitesinde bulunmadığı ve bu yüzden bu oranları düşürdüğü tespit edilmiştir.

Bu sistemin detayları bir yüksek lisans tezi olarak sunulmuştur. Öğrenci Eylül 2014'te tezini tamamlamış, jüriye sunmuş ve mezun olmuştur.

### **5.3 Akıllı TV'ler için öneri sistemi geliştirilmesi**

Proje çalışmaları kapsamında Akıllı TV'ler (Smart TV, Connected TV) için bağlam duyarlı içerik tabanlı ve işbirlikçi tavsiyeler üretebilen öneri sistemleri geliştirilmiştir. Akıllı TV'lerden alınan verilerle kullanıcı profilleri ve program profilleri oluşturulmuş, kullanıcı-program ilişkileri hesaplanarak çizge tabanlı kullanıcı modeli gerçekleştirilmiştir.

Geliştirilen sistem Arçelik tarafından sağlanan 3 aylık (Ekim 2013-Ocak 2014) kanal kullanım verisi ile test edilmiştir. Bu veri 5,466 kullanıcıya ait 3,865,821 satır kanal kullanım verisi içermektedir. Bu veri içerisinde sadece kanal kullanımının başladığı ve bittiği zaman ve kanal adı bulunmaktadır. Kanal kullanım verisi izlenen programın içeriği hakkında herhangi bir

bilgi içermediği için aynı dönemi kapsayan Radikal TV rehberi verileri ayrıca çekilmiştir. Örnek program bilgisi Şekil 25’de görülebilir.



**Anasının Oğlu**  
Ne halimiz varsa gülümlü

**Anasının Oğlu**  
Anasının Oğlu hangi kanalda? : KANAL D  
Anasının Oğlu hangi gün? : 10.Ağustos.2014  
Anasının Oğlu saat kaçta? : 08:45 - 10:30 (105 dakika)  
Tür : Dizi (Komedi)  
Yönetmen : Sermiyan Midyat  
Oyuncular : Sermiyan Midyat, Melisa Sözen, Devrim Yakut, Köksal Engür, Durul Bazan, Gülhan Tekin, Çiğdem Tunç  
Anasının Oğlu özeti : Güneydoğulu bir aşiret ağası olan Cano Derikfan'ın yolunun İstanbul'a düşmesiyle komedi başlar...

Derikfan Aşireti'nin ağası Cano, kendisine yüz vermeyen sosyetik komşu Lal'e sınırlıklam aşık olmuştur ve artık resmen ona açılmaya karar vermiştir. Bu habere en çok seven ise bir torun sevdasıyla yanıp tutuşan annesi Sıdar'dır. Elbette ne Lal'in ne de ailesinin bu durumdan henüz haberi yoktur. Ve asıl sorun kendisini göstermemiştir henüz. Muhasebecisinin kendisinden habersiz çevirdiği dolaplar yüzünden Cano'nun tüm mallarına el konmuştur. Derikfan Aşireti'nin koca ağası birden bire beş parasız kalır. Ne var ki, sürprizler bitmez. Cano'nun kardeşi Fırat'ın sevgilisi Ece'yle birlikte kurduğu klinik de bu hengamede elden çıkmıştır. Artık Derikfanlar'ın hayatında yeni bir sayfa açılmıştır. Ancak en büyük şok daha kapıyı çalmamıştır. Ailenin reisi, 30 yıl önce Sıdar, Cano ve Fırat'ı terk eden Abdurrahim geri döner...

Şekil 25. Örnek program bilgisi

Program bilgisini kanal kullanım verisi ile birleştirebilmek için bu veri üzerinde bazı işlemler yapılmıştır. Bu işlemler şöyle listelenebilir:

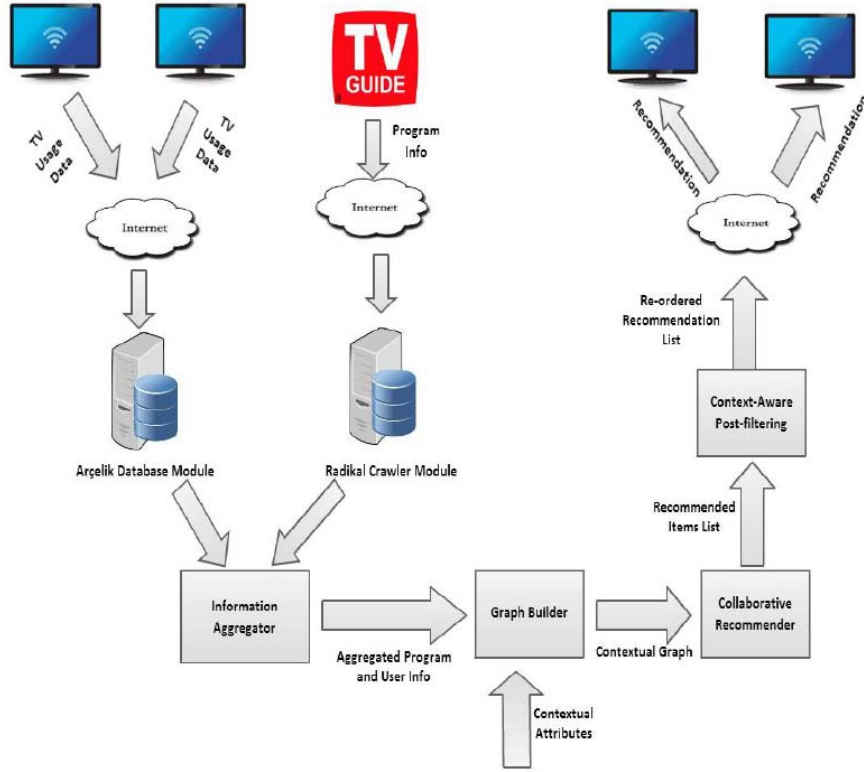
- Program bilgisindeki programın başlangıç ve bitiş zamanı Arçelik'in kanal kullanım verisi ile uyumlu hale getirilmiştir.
- Programın yayınladığı saat kullanılarak, programın dahil olduğu gün zamanı hesaplanmıştır (gece, sabah erken, kahvaltı, gündüz, akşam gibi). Uzun süreli programlar için birden fazla gün zamanı mümkün olabilmektedir.
- Pprogramların birden fazla türe sahip olmasına izin verilmiştir.
- Bir program için belirtilen bütün oyuncuların adları ve yönetmen bilgisi çıkarılmıştır.
- Program özeti ve uzun açıklamasındaki tüm kelimelerin kökü elde edilmiş ve bu kelime köklerinden fiil olanlar Türkçe'deki fiil köklerinin yarattığı belirsizlikten kaçınmak amacıyla elenmiştir. Elde edilen kelime kökleri "terim" olarak adlandırılmış ve program için oluşturulan terim dizisinde saklanmıştır. Kelime köklerini çıkarmak için, Zemberek isimli açık kaynaklı doğal dil işleme kütüphanesi kullanılmıştır.

Arçelik kanal kullanım verisi ile birleştirmek amacıyla 3 aylık (Ekim 2013-Ocak 2014) program verisi mümkün olan 39 kanal için Radikal TV Rehberi web sayfasından çekilmiştir. Bu işlem sonucunda; 45107 farklı program, 42 farklı tür, 4338 oyuncu, 1092 yönetmen, 4697 terim, elde edilmiştir.

Proje kapsamında akıllı TV'ler için geliştirilen her iki öneri sistemi de aynı verileri kullanmaktadır. Aşağıda önce bağlam duyarlı işbirlikçi öneri sistemi sonra da içerik tabanlı öneri sistemi anlatılmaktadır.

### 5.3.1 Bağlam duyarlı işbirlikçi öneri sistemi

Proje kapsamında geliştirilen çizge tabanlı işbirlikçi ve bağlam duyarlı öneri sisteminin genel yapısı Şekil 26'de görülebilir. Çizge tabanlı işbirlikçi öneri algoritması bağlam değişkenleri kullanılarak yapılan ön filtreleme ve sonradan filtreleme ile güçlendirilmiştir.



Şekil 26. Genel sistem akışı

Bu sistemin üç temel modülü vardır:

1. Bağlam duyarlı çizge oluşturma modülü
2. İşbirlikçi çizge tabanlı öneri modülü
3. Bağlam duyarlı yeniden değerlendirme modülü

#### Bağlam duyarlı çizge oluşturma modülü

Proje kapsamında bağlam bilgileri kullanılarak oluşturulmuş çizge üzerinde rastgele yürüyüşler yapan yaklaşım temel alınarak öneri sistemi geliştirilmiştir. Çizge üzerinde yürüyüşe başlamadan önce çizgedeki tüm düğümler arasındaki geçiş olasılığı hesaplanmalıdır. Bunun için düğüm tipleri arasındaki geçiş olasılıkları ayrı matrislerde hesaplanıp, daha sonra bu küçük matrislerden büyük bir matris oluşturulmaktadır. Eğer düğüm tipleri arasında belirli bir benzerlik ilişkisi yoksa bu durumda ya yeni bir benzerlik

ilişkisi çıkarılmalı ya da bu düğüm tipleri arasında hiçbir benzerlik olmadığı varsayılarak matriste ilgili hücreye sıfır değeri girilmelidir.

Radikal TV Rehberi sayfasından çekilen program bilgisi verisi, program türü, program gün zamanı, oyuncu ve yönetmen gibi bağlamsal değişkenler içermektedir. Bunlardan program türü ve program gün zamanı ön filtreleme için kullanılmıştır. Filtreleme kapsamında, belirlenen bağlam değişkeni dışında kalan programlar ihtimal dışında bırakılarak, çizgenin belirlenen bağlam değişkeni özelliğine sahip programlarla oluşturulması sağlanmıştır. Örneğin; program bilgisi verisinde eşleşmiş toplam 41357 program varken gün zaman değişkenlerinden “PRIME\_TIME” seçildiğinde yaklaşık 6500 program kalmaktadır. Kullanıcının bu zaman diliminde yer alan programlar arasından bir program izlemek istediği varsayılırsa, bu bağlam değişkeni özelliğine sahip olmayan programlar elenmiş olur. Böylece rastgele yürüyüş algoritması daha az karışık bir çizgede çalıştırılacağı için öneri sisteminin performansının artması beklenir.

Bağlam değişkinlerine göre filtreleme yapıldıktan sonra kalan veri kullanılarak 3 farklı düğüm tipine sahip çizge oluşturulmuştur. Çalışma kapsamında kullanılan düğüm tipleri: kullanıcı, program ve terim olarak belirlenmiştir. Çizgede yer alan tüm düğüm tipleri arasında bir geçiş olasılığı matrisi oluşturulmak amacıyla, farklı düğüm tipleri için farklı benzerlik fonksiyonları kullanarak düğüm tipleri arasındaki geçiş olasılığı hesaplanmıştır.

Kullanıcı ve program arasında kullanıcının programı izleme süresi ve programın toplam uzunluğu kullanılarak “rating” isimli bir benzerlik fonksiyonu tanımlanmıştır. Bu tanımlamada kullanıcının programı izleme süresi programın toplam uzunluğuna bölünerek, kullanıcının programa verdiği önem 0 ile 1 arasında bir değer ile temsil edilmiştir. Böylece kısa programların az izlenmesi nedeniyle önemini kaybetmesinin önüne geçilmesi amaçlanmıştır. Kullanıcılar kümesi  $U$  ile, programlar kümesi  $P$  ile gösterilirse, bütün kullanıcılar ve programlar arasındaki rating ilişkileri  $|U| \times |P|$  büyüklüğündeki  $UP = (u_{ij})$  matrisinde saklanır. Burada  $u_{ij}$  i numaralı kullanıcı ve j numaralı program arasındaki rating değeridir.

Program ve terim arasındaki ilişki TF-IDF ile belirlenmiştir. Programlar kümesi  $P$  ile, terimler kümesi  $T$  ile gösterilirse, bütün programlar ve terimler arasındaki TF-IDF değerleri  $|P| \times |T|$  büyüklüğündeki  $PT = (p_{ij})$  matrisinde saklanır. Burada  $p_{ij}$  i numaralı program ve j numaralı terim arasındaki TF-IDF değeridir.

Kullanıcı ve terim arasında direkt bir ilişki olmadığından, kullanıcı-terim ilişkisi kullanıcı-program ve program-terim ilişkileri kullanılarak hesaplanmıştır. Bir kullanıcının bir programa verdiği rating ve bir terimin aynı programla arasındaki TF-IDF bilgisi çarpılarak kullanıcı ve

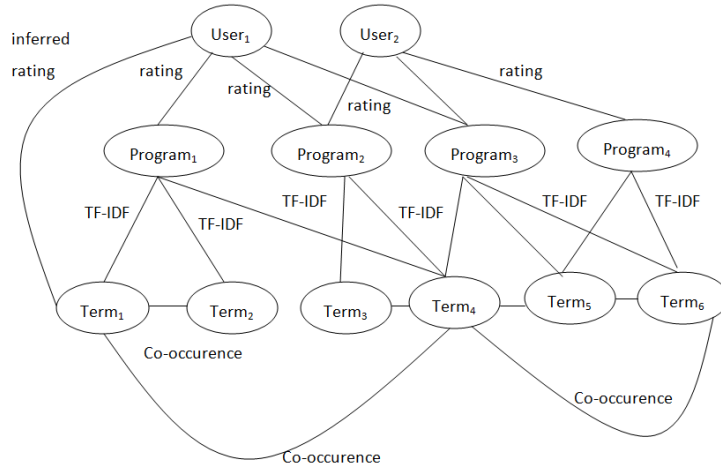
terim arasındaki bir benzerlik hesaplanabilir. Kullanıcı ve terim arasında bu yoldan farklı yollar da bulunuyorsa bu yollardaki benzerlik değerleri de hesaplanır. Kullanıcı ve terim arasındaki bütün yollardaki benzerlikler toplanarak kullanıcı-terim arasındaki toplam benzerlik hesaplanmış olur. Kullanıcılar kümesi U ile, terimler kümesi T ile gösterilirse, bütün kullanıcılar ve terimler arasındaki hesaplanan benzerlik değerleri  $|U| \times |T|$  büyüklüğündeki  $UT = (u_{ij})$  matrisinde saklanır. Burada  $u_{ij}$  i numaralı kullanıcı ve j numaralı terim arasındaki hesaplanan benzerlik değeridir.

Terimler arasında direkt bir ilişki olmamasına rağmen, terimlerin beraber geçtiği programlar kullanılarak terimler arasında co-occurrence (eşdizimlilik) isimli bir benzerlik tanımlanmıştır. Bu benzerlik hesaplanırken iki terimin birlikte yer aldıkları program sayısı, çizge de kullanılacak toplam program sayısına bölünmüştür. Terimler kümesi T ile gösterilirse, terimler arasındaki eşdizimlilik değerleri  $TT = (t_{ij})$  matrisinde saklanır.

Program-program ve kullanıcı-kullanıcı ilişkileri için daha önceden tanımlanmış bir benzerlik fonksiyonu olmadığından bu ilişkiler tanımlanırken birim matrisler kullanılmıştır.

Bütün düğüm tipleri için benzerlik fonksiyonları tanımlandıktan sonra, çizge oluşturulabilir. Örnek çizge Şekil 27'de görülebilir. Düğüm tipleri arasındaki benzerlikleri tutan küçük matrisler kullanılarak bütün bir çizge için kullanılacak geçiş olasılığı matrisi oluşturulur. Bunun için küçük matrisler büyük matrisin satır ve sütunları gibi yerleştirilir. Bu sayede,  $(|U| + |P| + |T|) \times (|U| + |P| + |T|)$  boyutlarında büyük geçiş olasılığı matrisi X aşağıdaki gibi oluşturulmuş olur.

$$X = \begin{bmatrix} U \times U & U \times P & U \times T \\ P \times U & P \times P & P \times T \\ T \times U & T \times P & T \times T \end{bmatrix}$$



Şekil 27. Örnek çizge

### İşbirlikçi çizge tabanlı öneri modülü

Rastgele yürüyüşe başlamadan önce büyük geçiş olasılığı matrisi ile aynı uzunluğa sahip yalnızca rastlantısal yürüyüşe başlanılacak düğümün değerinin 1, diğer değerlerin 0 olarak ayarlandığı bir başlangıç durum vektörü  $s_0$  oluşturulur. Daha sonra bu vektör seri olarak büyük geçiş olasılığı matrisi ile çarpıldığında  $n$  adım sonra ulaşılabilen düğümler için geçiş olasılıkları hesaplanmış olacaktır. Rastgele yürüyüş normal koşullarda geçiş olasılıkları belirli bir eşiğin altında değişim gösterene kadar devam eder ancak çok büyük verinin modellendiği çizgelerde yürüyüşün sınırlı adımla yapılabilmektedir. Bu çalışma kapsamında da sınırlı uzunlukta rastgele yürüyüşler yapılmıştır. Performans kısıtlarından ötürü en fazla 6 adım rastgele yürüyüş yapılabildiğinden, bu çalışmadan maksimum rastgele yürüyüş uzunluğu 6 adım olarak belirlenmiştir. Bu bağlamda, geliştirilen algoritmanın başarısı 1 ve 6 adım arasındaki rastgele yürüyüşler ile test edilmiştir.

Hedef kullanıcıdan sınırlı rastgele yürüyüş tamamlandığında, hedef kullanıcının diğer kullanıcı düğümlerine olan geçiş olasılıkları büyükten küçüğe göre sıralanarak en benzer  $k$  kullanıcı benzer kullanıcı seti olarak çıkarılmıştır. Bu işlem yapılırken, kaç tane benzer kullanıcı çıkarılacağını belirlemek için göz kararı kuralı (rule of thumb) kullanılarak testler yapılmıştır. Daha sonra bulunan yine rastgele yürüyüş kullanılarak bu benzer kullanıcılar için öneri listeleri oluşturulur. Bu öneri listelerinin birleştirilmesiyle kullanıcıya öneriler içerisindeki en yüksek skora sahip 10,20 ve 50 program önerilmektedir.

### Bağlam Duyarlı Yeniden Değerlendirme Modülü

Bu modül işbirlikçi çizge tabanlı öneri modülü tarafından üretilen öneri listesinin bağlam değişkenlerine göre yeniden düzenlenmesini yapmaktadır. Bu modül kapsamında oyuncu ve

yönetmen değerleri bağlam değişkeni olarak kullanılmıştır. Ancak, oyuncu ve yönetmen kavramları sadece belirli program tipleri için geçerli olduğundan bu modülün başarısı sadece dizi, sinema gibi belirli program tipleri için ayırt edici bir şekilde değişiklik göstermiştir.

Bu modülde, oyuncu ve yönetmen gibi bağlam değişkenlerini sonradan filtreleme için kullanabilmek için öncelikle oyuncu ve yönetmen gibi değişkenler için kullanıcıların beğenisi ölçülmelidir. Bu işlem için oyuncu ve yönetmenlerin yer aldığı programların kullanıcılar tarafından hangi “rating” skoru ile beğenildiği incelenir. Bir oyuncunun ya da bir yönetmenin rol aldığı filmlere kullanıcı tarafından verilmiş olan toplam “rating”, kullanıcının ilgili yönetmen ya da oyuncu için verdiği “rating” olarak hesaplanır. Modül kapsamında öneri listesinin yeniden sıralanması için, öneri modülü tarafından üretilen programların skoru içerdikleri bağlam değişkenleri (oyuncu ve yönetmenler) için hesaplanan “rating”lerle çarpılır. Daha sonra oluşan yeni skorlar yeniden sıralanarak öneri listesi oluşturulur. Bu sayede daha önceden öneri listesi içinde yer almamış programlar önerileceği için öneri sisteminin başarısına olumlu katkı sağlanır.

### 5.3.2 İçerik tabanlı öneri sistemi

Kullanıcılara içerik tabanlı program önerileri yapabilmek için, kullanıcı düğümleri bulunduktan sonra aktivasyon dağılımı (spreading activation) methodu kullanılmıştır. Aktivasyon dağılımı bir düğümden yola çıkarak gidilebilecek düğümlere aktivasyon değerleri atayarak çizge gezinimi yapar. Atanacak aktivasyon değerleri belirli bir değer altına düştüğü zaman gezinim durur. Projede geliştirilen içerik tabanlı öneri sisteminde algoritma şu şekilde uyarlanmıştır:

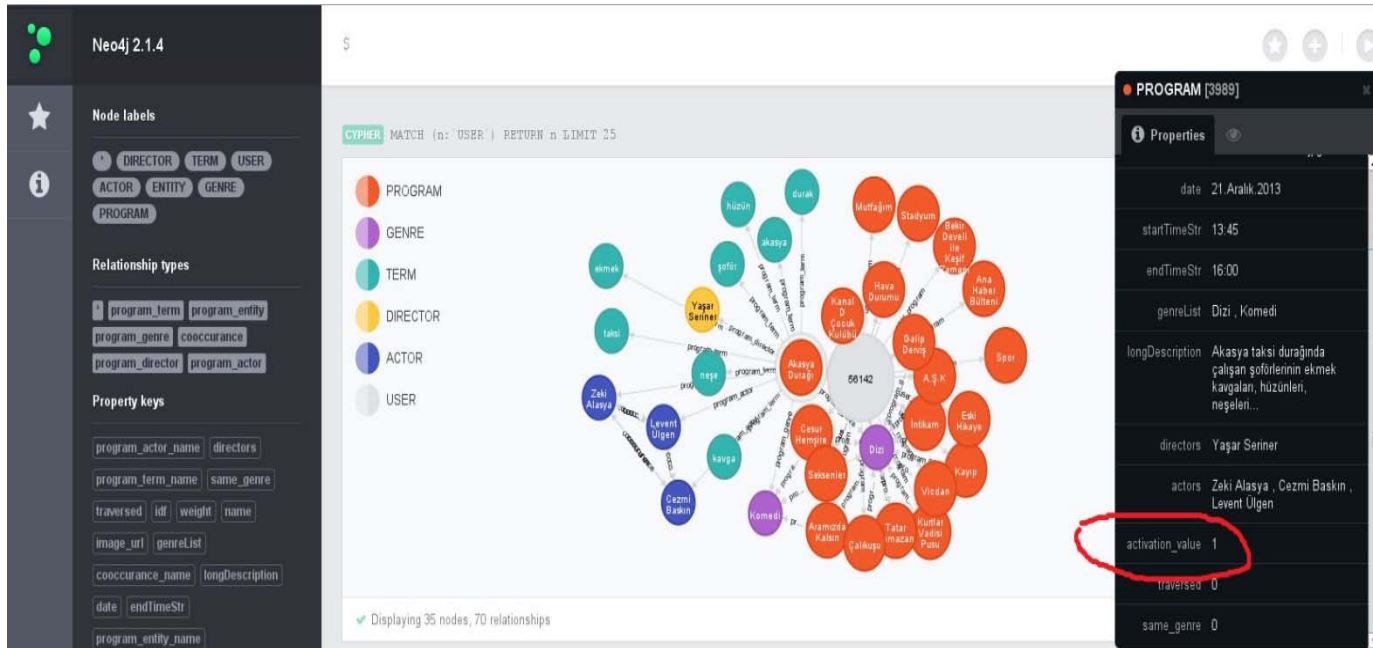
- Program önerisi sunulacak kullanıcı düğümü bulunur
- Bu düğümden yola çıkarak program düğümlerine geçilir.
- Program düğümlerinin her biri için,
  - Aktivasyon değeri 1.0 olarak atanır,
  - Program ile aynı türe sahip olan program düğümleri işaretlenir
  - Programın aktivasyon değeri  $\text{Kullanıcı} - \text{Program Bağ ağırlığı} \times 1.0 \times 0,6$  (seviye değer kaybı) olarak atanır.
  - Aktivasyon değeri atandıktan sonra bu programdan çıkılarak gidilebilecek varlık, terim, yönetmen, oyuncu düğümleri gezilir ve bunların da aktivasyon değerleri kendi aktivasyon değerleri + programın aktivasyon değeri  $\times \text{tf-idf} \times 0,6$  (seviye değer kaybı) olarak hesaplanır. Bu sayede birden fazla programdan gidilebilecek aynı düğümlerin değeri diğerlerinden daha yüksek olur.
  - Bu düğümlerden de aynı şekilde gidilebilecek program düğümlerinin aktivasyon değerleri hesaplanır.
  - Aktivasyon değeri 0,2 nin altına düştüğü zaman gezinim durdurulur.
- Kullanıcı düğümü ile bağı olan her program düğümü için aynı işlem yapıldıktan sonra, aktivasyon değeri hesaplanmış program düğümleri çıkartılır ve aktivasyon değerlerine göre sırandırılmış olarak kullanıcıya sunulur.



Ayrıca bu algoritmaya ek olarak, bir programı ifade eden düğüm tipleri de ifade etme yeteneklerine göre derecelendirilmiştir. Bir kullanıcının izlediği programın oyuncu, yönetmen, isimlendirilmiş varlık gibi özelliklerinin ham terim bilgilerine kıyasla daha anlamlı sonuçlar verdiği görülmüştür. Bu derecelendirme şu şekilde yapılmıştır;

- Program düğümünden oyuncu düğümüne geçerken seviye kaybı 0,6
- Program düğümünden yönetmen düğümüne geçerken seviye kaybı 0,5
- Program düğümünden varlık düğümüne geçerken seviye kaybı 0,6
- Program düğümünden terim düğümüne geçerken seviye kaybı 0,2

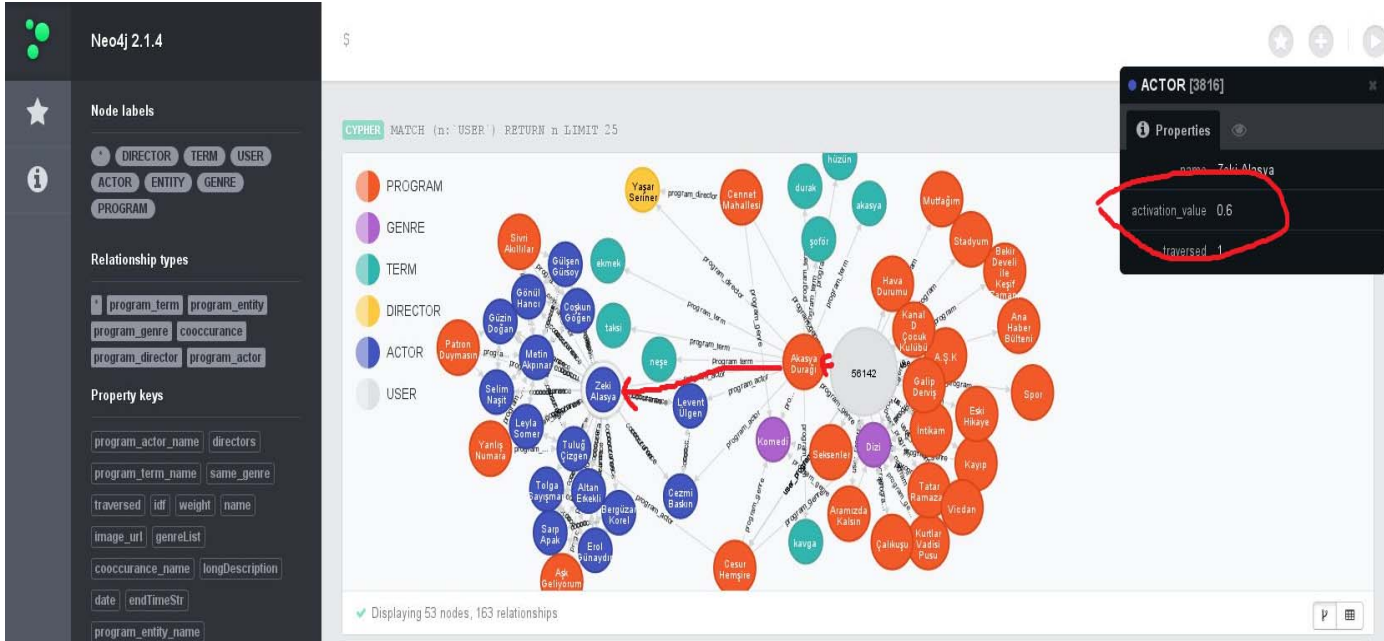
Şekil 28, Neo4j yapısında aktivasyon değerleri hesaplandıktan sonraki bir kullanıcı profilini göstermektedir. Bu profilde kullanıcı “Akasya Durağı” izlemiş fakat “Cennet Mahallesi” izlememiştir. Bir kullanıcı düğümünden başlanır. Bu kullanıcı düğümünden yola çıkarak gidilebilecek program düğümleri bulunur ve bu program düğümlerinin aktivasyon değeri 1.0 olarak atanır. “Akasya Durağı” program düğümünün aktivasyon değeri 1.0 olarak atanmıştır. Daha sonra bu düğüm üzerinden aktivasyon yayılımı yapılarak diğer düğümlere geçilmiştir.



Şekil 28. Neo4j’de bir kullanıcı profili

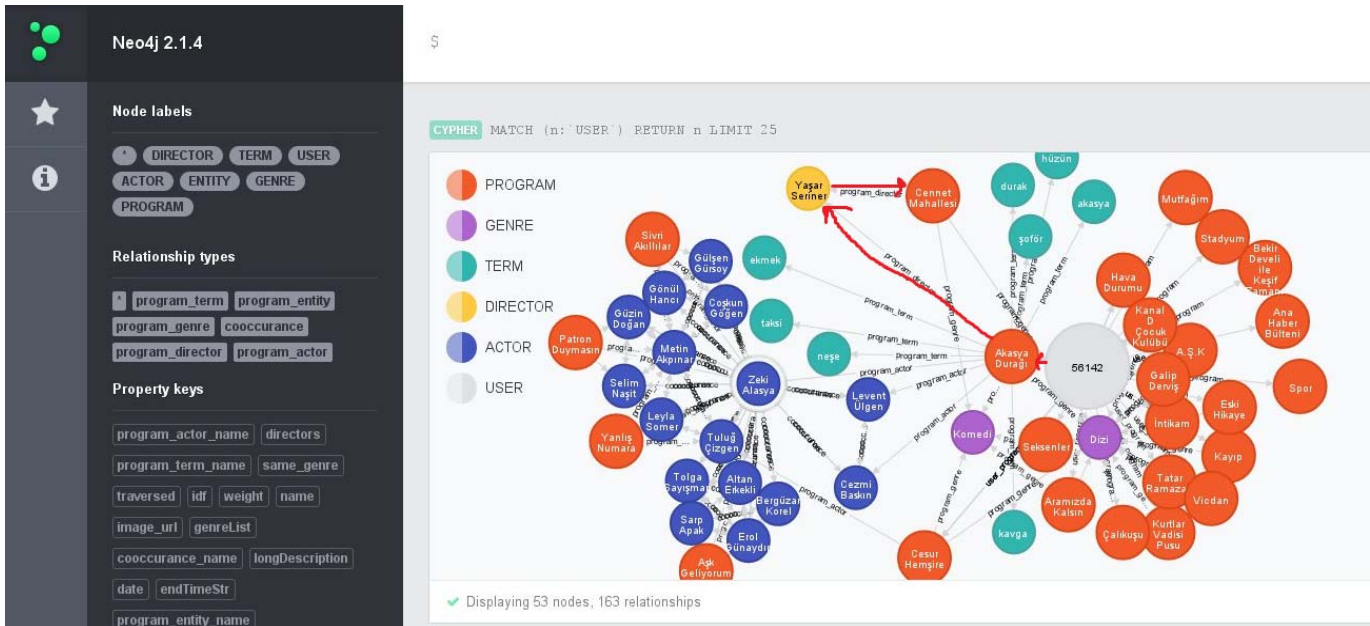
Daha sonra bu program düğümünden yola çıkarak gidilen oyuncu, yönetmen, varlık, terim düğümlerinin aktivasyon değerleri hesaplanmıştır.





Şekil 29. Çizgede oyunculara göre gezinim

Bağlı bulunan düğümler gezildikten sonra program düğümlerine geçilerek program düğümlerinin de aktivasyon değerleri hesaplanır. Örnek olarak, Şekil 30'de görüldüğü gibi yönetmen düğümü üzerinden "Cennet Mahallesi" düğümüne geçilmiştir aktivasyon değeri  $1.0 \times 0.6 \times 0.6 = 0.36$  olarak hesaplanmıştır.



Şekil 30. Çizgede yönetmen üzerinden gezinim

Bu şekilde yinelemeli olarak gezilebilecek düğümlerden aktivasyon değeri 0.2 nin altına düşen düğümlerden daha fazla ileri gidilmez ve gezinme durdurulur.



Bu örnekte kullanıcıya sadece “Akasya Durağı” düğümünden gidilerek aşağıdaki öneriler sunulmuştur :

**Program Name** : Gurbette Aşk Bir Yastıkta  
**Path** : {"value"=>0.36, "path"=>["Cezmi Baskın : ACTOR"]}

**Program Name** : Patron Duymasın  
**Path** : {"value"=>0.36, "path"=>["Zeki Alasya : ACTOR"]}

**Program Name** : Yanlış Numara  
**Path** : {"value"=>0.36, "path"=>["Zeki Alasya : ACTOR"]}

**Program Name** : Sivri Akıllılar  
**Path** : {"value"=>0.36, "path"=>["Zeki Alasya : ACTOR"]}

**Program Name** : Aşk Geliyorum Demez  
**Path** : {"value"=>0.36, "path"=>["Zeki Alasya : ACTOR"]}

**Program Name** : Cennet Mahallesi  
**Path** : {"value"=>0.25, "path"=>["Yaşar Seriner : DIRECTOR"]}

Yukarıdaki örnekte “Akasya Durağı” programı için kullanıcıya önerilen programlar ve bu programların değerleri (aktivasyon değeri) ve nasıl ulaşıldığı gösterilmiştir.

Bu kısımda anlatılan bağlam duyarlı işbirlikçi öneri sistemi bir yüksek lisans tezi olarak tamamlanmış, öğrenci Eylül 2014’te tezini sunmuş ve mezun olmuştur (Şamdan E., 2014). Bu tezi anlatan ve RecSysTV 2014 çalıştayında sunulan bir bildiri de yayınlanmıştır (Şamdan, E., vd. 2014). İçerik tabanlı öneri sisteminin de gerçekleştirilmesi tamamlanmış, değerlendirme sonuçlarının kesinlik kazanması beklenmektedir. Bu kısım da ayrı bir yüksek lisans tezi kapsamında çalışılmıştır. Öğrenci Ocak 2015’te mezun olmayı planlamaktadır.

## 6. SONUÇLAR

Projede, kısmi kullanıcı profillerini birleştirmeyi kolaylaştıran bir hiper-çizge kullanıcı modeli yapısı sunulmuştur. Ayrıca bağlantılı-veri problemlerinin çözümünü kolaylaştırmak hedef alınmış ve sunulan FunGuide uygulaması ile bu durum örneklenmiştir. FunGuide, kullanıcının izni dahilinde Facebook, LinkedIn ve Twitter ağlarından veri çekebilir ve bu veri ile entegre edilmiş anlamsal kullanıcı profilini sunar. Ayrıca, kullanıcıya alandan-bağımsız veya alan tabanlı (kitap, film, müzik ve spor) önerilerde bulunabilir, kullanıcının bir öğeye ilgi duyup duymadığını hesaplayabilir, bir öğeye ilgi duyan kullanıcıları listeleyebilir ve benzer kullanıcıları keşfedebilir. Önümüzdeki iki ay içinde FunGuide’ın gerçek kullanıcılarla

değerlendirmesi tamamlanacak ve sistemin başarımının kesin değerlendirilmesi tamamlanacaktır.

Projede önerilen bütünleşik kullanıcı modelinin ilk taslak aşamasından başlayarak, son gelişmiş haline kadar olan basamakları farklı bildirilerle yayınlanmıştır (Tarakçı ve Çiçekli, 2012a, 2012b, 2014a, 2014b). (Tarakçı ve Çiçekli 2014b) KEOD 2014 konferansında “En iyi bildiri” ödülü almıştır. Halen bütün kullanıcı modelini ve çoklu-uygulamalı alanlarda kullanımını anlatan bir dergi makalesi hazırlıkları devam etmektedir. Sistemin gerçek verilerle değerlendirilmesi tamamlandığında makale dergi yayını için teslim edilecektir.

Proje başvurusunda öngördüğümüz nihai amacımız gerçek kullanıcıların Facebook, Twitter, LinkedIn gibi sosyal ağlardaki ve mobil telefon, akıllı TV gibi kişisel cihazlardaki izlerini entegre ederek bütünleşik kullanıcı profili sunabilen bir sistemdi. Entegre edilen bütünleşik kullanıcı profilinin uygun altkümelerini hem bu uygulamalara hem de daha başka uygulamalara (ör. Arama motoru, ürün tavsiye sistemleri, web tv uygulamaları gibi) gönderebilmek, ya da farklı uygulamalara istedikleri profile kullanıcı sağlayabilmektir. Bu amacımıza Facebook, AkıllıTV öneri sistemleri ve FunGuide ile parça parça ulaştık. FunGuide uygulaması ile Facebook, Twitter ve LinkedIn’den verileri çekilebilen kullanıcılar için bütünleşik kullanıcı modelleri oluşturduk ve onlara farklı alanlarda öneriler sunan bir sistem geliştirdik. Başka deyişle nihai amacımızı FunGuide vasıtasıyla simule ederek kullanıcı modelimizin istenilen şekilde kullanılabilirliğini ispatladık. Eksik olan şey geliştirdiğimiz sosyal ağ öneri sistemi ve TV programı öneri sistemlerini bütünleşik kullanıcı modeli ile entegre bir şekilde çalıştıramamızdır. Bunun nedeni de çalıştığımız bütün ortamlarda izlerini çıkartabileceğimiz ortak kullanıcı verisini bulamamızın imkansızlığı olmuştur. Böyle bir veri olmadığı için facebooktan toplanan kullanıcılar için TV programı önerebilsek de bunun doğru değerlendirmesini yapamadık. Aynı şekilde akıllı tv kullanıcılarının facebook bilgilerine erişim imkansız olduğu için onların sosyal ağ profilleri ile TV kullanım profillerini entegre edemedik.

Ancak 4. Kısımda anlatılan teorik yaklaşım böyle veri kümesi olsaydı nihani amacımızın ulaşılabilirliğini örneklerle göstermektedir. Nitekim FunGuide uygulamasını kullanan kullanıcılar Facebook, twitter ve linkedIn ağlarının her üçünde de hesabı olan kullanıcıları kapsamaktadır. Bu ortak kullanıcılar için bütünleşik model kurulabilmekte ve onlara alana bağımlı, alandan bağımsız ya da alanlar arası (cross-domain) öneriler yapılabilmektedir.

Proje çıktılarının gelecekteki çalışmalarla daha geliştirilmesi planlanmaktadır. Kullanıcı profili zamana ve bağlama göre değişmektedir. Bu amaçla sistemin kısa ve uzun vadeli kullanıcı



profillerini ayırtedebilmesi, profili buna göre dinamik olarak düzenleyebilmesi ve bağlamı da hiper çizgeye entegre edebilmesi için model geliştirme çalışmaları yapılacaktır. Ayrıca, bu projede edindiğimiz tecrübe ile anlamsal içerik arama, kişiselleştirilmiş arama motorları ve reklam önerme konularında yeni projeler başlatmayı planlıyoruz.

Bu proje kapsamında 5 öğrenci burs aldı. 1 doktora tezi ve 4 yüksek lisans tezi çalışmaları yürütüldü. Yüksek lisans tezlerinden iki tanesi tamamlandı (Şamdan, E. 2014) (Işıklar, Y.E. 2014), 2 tanesi Ocak 2015'te bitirilme aşamasındadır. Proje çalışmalarını anlatan 5 uluslararası konferans yayını yapıldı (Tarakçı ve Çiçekli, 2012a, 2012b, 2014a, 2014b) (Şamdan E. vd., 2014). Proje kapsamında iki yurtdışı seyahat yapıldı. ACL 2014 konferansı için ABD, Baltimore'a, KEOD 2014 konferansı için Roma'ya gidildi.

## Kaynaklar

- 1 Bu, J., Tan, S., Chen, C., Wang, C., Wu, H., Zhang, L., He, X. 2010. "Music Recommendation by Unified Hypergraph: Combining Social Media Information and Music Content", In Proceedings of the International Conference on Multimedia, MM '10, 391-400.
- 2 Chen, B., Wang, J., Huang, Q., Mei, T. 2012. "Personalized Video Recommendation Through Tripartite Graph Propagation", Proceedings of ACM Multimedia , 1133-1136.
- 3 DBpedia Social Website, <http://dbpedia.org/About> Son erişim tarihi: 15 Ekim 2014.
- 4 Facebook Social Website, <https://www.facebook.com/> Son erişim tarihi: 15 Ekim 2014.
- 5 Facebook Knowledge Base Website, <http://www.freebase.com/> Son erişim tarihi: 15 Ekim 2014.
- 6 Ghoshal, G., Zlatic, V., Caldarelli, G., Newman, M. E. J. 2009. "Random Hypergraphs and Their Applications", Phys.Rev.E, 79, 6.
- 7 Goetschel, R.Jr. 1995. "Introduction to Fuzzy Hypergraphs and Hebbian Structures", Fuzzy Sets and Systems, 7, 113-130.
- 8 Işıklar, Y.E., 2014, Kural Tabanlı Varlık İsmi Tanıma Yöntemi Kullanarak Televizyon İçeriğini Zenginleştirme Sistemi, Yüksek lisans tezi, ODTÜ, Eylül 2014.
- 9 Kramar, T., Barla, M., Bielikova, M. 2013. "Personalizing Search Using Socially Enhanced Interest Model Built from the Stream of User's Activity", J. Web Eng., 12(1-2), 65-92.
- 10 Li, L., Li, T. 2013. "News Recommendation via Hypergraph Learning: Encapsulation of User Behavior and News Content". Proceedings of the sixth ACM international conference on Web search and data mining, 305- 314.
- 11 Neo4j Graph Database Website, <http://www.neo4j.org/> Son erişim tarihi: 15 Ekim 2014.
- 12 Şamdan, E., Taşçı, A., Çiçekli, N.K. 2014. A Graph-based Collaborative and Context-aware Recommendation System for TV Programs, RecSysTV 2014, Foster City, Silicon Valley, USA, 6th-10th October 2014.
- 13 Şamdan, E. 2014. "A Graph-based Collaborative and Context-aware Recommendation System for TV Programs", Yüksek lisans tezi, ODTÜ, Eylül 2014.

- 14 Tarakçı, H., Cicekli, N.K. 2012. "UCASFUM: A Ubiquitous Context-Aware Semantic Fuzzy User Modeling System", Intl. Conf. on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KEOD 2012), SciTePress, 278-283.
- 15 Tarakçı, H., Cicekli, N.K. 2012. "Ubiquitous Fuzzy User Modeling for Multi-Application Environments by Mining Socially Enhanced Online Traces", Proc. of the 20th Intl. Conference on User Modeling, Adaptation, and Personalization (UMAP 2012), Springer-Verlag, 387-390.
- 16 Theodoratos, D. 2002. "Semantic Integration and Querying of Heterogeneous Data Sources Using a Hypergraph Data Model", Proceedings of the 19th British National Conference on Databases: Advances in Databases, 166-182.
- 17 Tiroshi, A., Berkovsky, S., Kaafar, M. A., Chen, T., Kuflik, T. 2013. "Cross Social Networks Interests Predictions Based on Graph Features", In Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13, 319-322.
- 18 Tarakçı, H., Cicekli, N.K. 2014. "Using Hypergraph-Based User Profile in a Recommendation System", International Conference on Knowledge Engineering and Ontology Development (KEOD 2014).
- 19 Tarakçı, H., Cicekli, N.K. 2014 "A Formal Framework for Hypergraph-Based User Profiles", International Symposium on Computer and Information Sciences (ISCIS 2014)
- 20 Tatar, S., Cicekli, I. 2011 "Automatic rule learning exploiting morphological features for named entity recognition in Turkish" Journal of Information Science 37.2 137-151.

**TÜBİTAK**  
**PROJE ÖZET BİLGİ FORMU**

Proje Yürütücüsü:	Prof. Dr. FEHİME NİHAN ÇİÇEKLİ
Proje No:	112E111
Proje Başlığı:	Çoklu-Uygulamalı Ortamlar İçin Kullanıcı Modelleme Sistemi
Proje Türü:	Araştırma
Proje Süresi:	24
Araştırmacılar:	
Danışmanlar:	
Projenin Yürütüldüğü Kuruluş ve Adresi:	ORTA DOĞU TEKNİK Ü. MÜHENDİSLİK F. BİLGİSAYAR MÜHENDİSLİĞİ B.
Projenin Başlangıç ve Bitiş Tarihleri:	01/10/2012 - 01/11/2014
Onaylanan Bütçe:	142280.0
Harcanan Bütçe:	74577.19
Öz:	<p>Bu projede, kişinin kısmi profillerini birleştirmeyi kolaylaştıran ve bütün, çoklu-alanlı bir kullanıcı profili sağlayan bir kullanıcı modelleme sistemi geliştirilmiştir. Birleştirme işlemi, zenginleştirilmiş kullanıcı profillerine yol açan bir anlamsal zenginleştirme işlemi içermektedir. Oluşturulan kullanıcı modeli genel ve alan-tabanlı anlamsal kullanıcı profilleri çıkartabilme ve önerme gibi çeşitli bağlantılı veri sorgularına makul zamanda cevap verebilme yeteneklerine sahiptir. Bu projede, Facebook, Twitter ve LinkedIn'i de içeren sosyal hesaplardan kısmi profilleri içeri aktarabilen ve bu kısmi profilleri birleştirerek kullanıcının bütünsel ve anlamsal bir profilini oluşturabilen bir kullanıcı modelleme sistemi geliştirilmiştir. Ayrıca sistem; kitap, film, spor ve müzik alanları için çeşitli öneri modülleri içermektedir. Alandan-bağımsız genel bir öneri modülü ve belirtilen bir öge ile ilgilenmesi muhtemel kullanıcıların keşfine ve sistemdeki kullanıcı ile benzer ilgi alanlarına sahip kullanıcıların keşfine izin veren bir hesaplama modülü de mevcuttur. Bu modüllerle, sunulan kullanıcı modelini kullanarak bağlantılı veri problemlerine nasıl çözüm sunulabileceği gösterilmektedir. Modelin uygulanabilirliği sosyal ağlar ve akıllı TV kullanım alanlarında farklı öneri sistemleri geliştirilerek örneklenmiştir.</p>
Anahtar Kelimeler:	Kullanıcı modeli, öneri sistemleri, anlamsal zenginleştirme, sosyal ağlar, bilgi gösterimi
Fikri Ürün Bildirim Formu Sunuldu Mu?:	Hayır