

(Geliş Tarihi / Received Date: 23.02.2021, Kabul Tarihi/ Accepted Date: 08.04.2021)

Yeni Bir İçerik-Tabanlı Akademik Makale Tavsiye Sistemi Prototipi Geliştirilmesi

Veysel Karani ÖZ, Emine DENİZ, Sinem BOZKURT KESER*, Yusuf KARTAL, Savaş OKYAY

¹ Eskişehir Osmangazi Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, 26480, Eskişehir

Anahtar Kelimeler:
Akademik Makale Tavsiye Sistemi, Tavsiye Sistemleri, İçerik Tabanlı Filtreleme.

Özet: Dijital bilgi ve ziyaretçi sayısındaki hızlı artış, ilgi çekici öğelere erişimde zaman kaybına neden olmaktadır. Aşırı bilgi yükünü hafifletmek için, bilgiyi filtreleme ve önceliklendirme ile verimli bir şekilde sunmak gerekir. Bu durum tavsiye sistemlerinin ortaya çıkmasını ve önem kazanmasını sağlamıştır. Öneriler, kullanıcılara kişiselleştirilmiş ürün veya hizmet sunmak gibi amaçlarla oluşturulur. Tavsiye sistemleri kitap, müzik, film, ticari ürünler ve akademik makale gibi farklı sistem uygulamalarında karşımıza çıkmaktadır. Özellikle akademik alanda dijital bilimsel içeriğin genişlemesi, tavsiye sistemlerinin önemini vurgulamaktadır. Akademik makale tavsiye sistemlerinde, çeşitli yöntemler uygulanmaktadır. Bu çalışmada ise, başlık ve özet bilgileri ile girdi olarak alınan herhangi bir makaleye en benzer makaleleri listeleyen yeni bir içerik-tabanlı akademik tavsiye sistemi prototipi tanıtılmaktadır. Farklı veri işleme yöntemleri ile oluşturulan öneri listeleri kıyaslanarak prototipin başarımı değerlendirilmektedir.

The Development of a Novel Content-Based Academic Paper Recommendation System

Keywords:
Academic Paper Recommendation System, Recommender Systems, Content Based Filtering.

Abstract: The rapid increase in digital information and visitors' number causes a waste of time to access attractive items. To alleviate information overload, it is necessary to present information with filtering and prioritization efficiently. This situation has enabled the recommender systems to emerge and gain importance. Suggestions are created to provide users with personalized products or services. Recommendation systems appear in different system applications such as books, music, movies, commercial products, and academic articles. The expansion of digital scientific content emphasizes the importance of recommendation systems, especially in the academic field. Various methods are used in academic article recommendation systems. In this study, a new content-based academic recommendation system prototype is introduced that lists the articles most similar to any given input with title and summary information. The performance of the prototype is evaluated by comparing the suggestion lists created with different data processing methods.

*İlgili yazar: sbozkurt@ogu.edu.tr

1. GİRİŞ

İnternet teknolojisinin gelişmesi ve dijitalleşmenin artması ile birlikte, bilgi üretimi ve tüketimi de artmaktadır. Bu durum, kullanıcıların kendisi için faydalı olacak bilgiyi bulmasında zorlanmasına neden olmaktadır. Kullanıcıların aradıkları içeriğe hızlı ve doğru bir şekilde erişebilmeleri için tavsiye sistemleri geliştirilmektedir [1]. Tavsiye sistemleri güncel bir araştırma konusu olduğu için birçok pratik uygulamada kullanılmaktadır. Kitap, müzik, film, ticari ürünler ve akademik makale gibi farklı sistem uygulamaları en belirgin örneklerdir[2].

Araştırmacılar, son teknoloji ürün çalışmalarındaki eğilimlerin farkında olmak, yeni ürünler icra etmek gibi çeşitli amaçlarla akademik makaleler aramaktadır. Bununla birlikte, bilimsel yayınların fazlalığı nedeniyle ilgili akademik makalelerinin değerlendirilmeme olasılığı ve harcanan zaman artmaktadır. Bilim topluluğu son 17 yılda akademik makale tavsiye sistemlerini iyileştirmeye odaklanmıştır [3]. Bu dönemde, akademik makale tavsiye sistemleri oluşturmak için 80'den fazla yöntem öneren yaklaşık 220 araştırma makalesi ve patent yayınlanmıştır [3]. Araştırmacıların çalıştıkları alan ile ilgili akademik makalelerini bulmaya yönelik mevcut yaklaşımlar, içerik tabanlı, iş birliğine dayalı, bibliyografik bilgi tabanlı, graf tabanlı, global alaka düzeyi ve hibrit yaklaşımlar olarak sınıflandırılabilir [3, 4]. Bu yaklaşımlar arasında çoğunlukla içeriğe dayalı filtreleme yöntemi uygulandığı görülmektedir.

İşbirlikçi filtreleme, "benzer düşünen insanlar aynı öğeleri sever veya sevmez" fikrine dayanır. Aynı ilgi alanlarına sahip bir kullanıcı grubu oluşturulur ve akranların tercihlerine göre öneriler sunulur. Bu sistemler araştırma makalelerinin içeriğine dayanmaz; bunun yerine, kullanıcının ilgi alanlarına ve tercihlerine güvenirlir. Dolayısıyla, işbirlikçi filtreleme tabanlı sistemler içerikten bağımsız sistemlerdir. Kullanıcılara güvenmenin bir sonucu olarak, bu yaklaşım soğuk başlatma (*cold start*), seyreklik (*sparsity*), işlem süresi, ölçeklenebilirlik (*scalability*) ve kullanıcı müdahalesi gibi sorunlara sahiptir. Yeni bir kullanıcı hiç makale değerlendirmese veya az sayıda makale değerlendirirse, sistem benzer düşünen araştırmacıları bulamaz. Aynı durum, yeni bir makalenin hiç beğenilmediği veya çok az beğenildiği ve önerilemeyeceği yeni bir araştırma makalesinde de geçerlidir [5]. Bu sistemlerin karşılaştığı diğer bir genel sorun, araştırma makalelerinin sayısı çok yüksek ve kullanıcı sayısı az olduğu için "seyrekliktir". Bu durumda, benzer fikirlere sahip kullanıcılar bulmak zor bir iştir. Bu sistemler ayrıca, genellikle daha az ölçeklenebilir olduklarından ve daha fazla çevrimdışı veri işleme gerektirdiklerinden zaman maliyeti yüksek bir yaklaşıma sahiptirler.

Graf tabanlı yaklaşımlar, akademik çevrede var olan doğal bağlantıları kullanır [6]. Bu bağlantılara dayanarak, makalelerin alıntılarla nasıl ilişkilendirildiğini gösteren graf ağlar oluşturur. Graf; alıntılar, yazarlar, ortak yazarlar, konular ve yayın yılı gibi çeşitli ilişkilere dayanılarak oluşturulmaktadır. Graf

tabanlı yaklaşımlarda kullanıcıların geçmişte ilgilendikleri alanları gelecekte de ilgilenecekleri varsayılır. Çalışmalarda, yazarlar ve makaleler düğüm olarak kabul edilerek graf oluşturur. Makaleler arasındaki ilişki, kullanıcılar arasındaki ilişki ve kullanıcılar ile makaleler arasındaki ilişki kenar olarak kabul edilir. Ardından, kullanıcılar ve makaleler arasındaki ilişkiyi hesaplamak için graf üzerinde rastgele yürüyüş veya diğer algoritmalar kullanılır. Bu yaklaşım diğer yaklaşımlarla birleştirilerek tavsiye sisteminde daha iyi performans sağlamaktadır.

Global alaka düzeyi en basit haliyle, bir tavsiye sistemi herkese uyan bir yaklaşımı benimseyerek en yüksek global alaka düzeyine sahip öğeleri önermektedir [7]. Bu durumda, alaka düzeyi bir kullanıcıya özgü olarak hesaplanmaz. Bunun yerine, genel popülerlik gibi bazı global önlemler kullanılır. Örneğin, bir film kiralama sistemi, en sık kiralanan veya tüm kullanıcılar arasında en yüksek ortalama dereceye sahip filmleri önerebilir. Bu durumda, temel varsayım kullanıcıların diğer kullanıcıların çoğunun hoşlandığı şeylerden hoşlanacağı yönündedir.

Bibliyografik bilgiye dayalı yaklaşımlar, makaleler arasındaki alaka ve ilişkileri analiz etmek için her makalede sağlanan alıntılarını kullanır. Bazı araştırmacılar yalnızca alıntıda verilen küçük bilgilere güvenmektedirler. Bu araştırmacıların yaklaşımları alıntı ve referans listesine dayanmaktadır [8]. Bibliyografik temelli yaklaşımlarda, en iyi sonuçlar bibliyografik eşleştirme ve ortak alıntı ile üretilir [9]. Genel olarak, alıntı yapma ve alıntı yapılan makaleler arasındaki en güçlü ilişki (bu ilişki makalenin yazarı tarafından kurulduğundan) atf olarak kabul edilir. Bu yaklaşımlar, içeriğin kullanılmaması durumunda uygulanamaz.

İçerik tabanlı filtreleme, öğelerin içeriklerinin hedef kullanıcının geçmişte kullandığı öğelerin içeriklerine benzer olmasına dayanır [3]. İçerik tabanlı filtreleme, temel olarak kullanıcıların geçmiş tercihlerini dikkate alır ve kullanıcı profili olarak adlandırılan kullanıcıların ilgi alanı modelini oluşturmak için kişisel kütüphane oluşturur. İçerik tabanlı bir tavsiye sisteminde, anahtar kelimeler öğeyi tanımlamak için kullanılmaktadır. Bazı araştırmacılar başlıktan anahtar kelimeleri çıkarırken, diğerleri özetlerden anahtar terimleri çıkarmaktadır. Bir araştırma makalesinde yazar tarafından tanımlanan anahtar kelimeler de araştırmacılar tarafından analiz edilmektedir [11]. Araştırmacıların çoğu, anahtar terimleri makalenin tüm içeriğinden (*full-text*) çıkarmaya odaklanmaktadır. İçeriğe dayalı yaklaşımlar yaygın olarak kullanılmasına ve tüm içeriğin temel bilgilerini taşımasına rağmen, tüm içeriğe erişmek her zaman mümkün değildir. Bu nedenle, tüm içerik mevcut değilse makalenin başlık veya özetinin kullanılmasını önerilmektedir [11]. Bu çalışmada, yeni bir içerik tabanlı akademik makale tavsiye sistemi prototipi geliştirilmektedir. Kullanılan yöntem kapsamında;

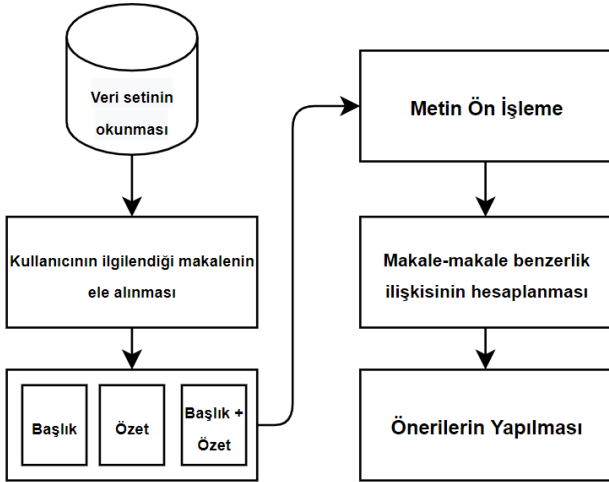
- Makale başlığı
- Özet
- Makale başlığı ve özet

girdi şeklinde ele alınarak prototipin fonksiyonelliği farklı çıktılar üzerinden doğrulanmaktadır. Lineer kernel yöntemi makaleler arasındaki benzerlikleri hesaplamada kullanılmaktadır. Hesaplanan benzerlik oranlarına göre makaleler sıralandıktan sonra, yüksek benzerliğe sahip ilk on makale kullanıcılara önerilmektedir.

Çalışmanın takip eden bölümünde materyal ve yöntem anlatılmaktadır. Üçüncü bölümde önerilen akademik makale tavsiye sistemi tanıtılmakta ve dördüncü bölümde deneylerden elde edilen sonuçlar verilmektedir. Son bölümde ise sonuçlar ve gelecek çalışmalar anlatılmaktadır.

2. MATERYAL VE METOT

Bu çalışmada, yeni bir içerik-tabanlı akademik makale tavsiye sistemi prototipi tanıtılmaktadır. Farklı alanlardan bilimsel çalışmalar içeren bir veri seti üzerinden kullanıcının seçmiş olduğu herhangi bir makalenin diğer makaleler ile olan ilişkisi elde edilmektedir. Makalelerin başlıkları, özetleri; ayrıca, başlıkları ve özetleri birlikte girdi olarak ele alınmaktadır. Önerilen sisteme entegre edilen yöntem Şekil 1’de gösterilmektedir.



Şekil 1. Önerilen akademik makale tavsiye sistemine ait akış diyagramı

Şekil 1 ile verilen akış diyagramında da görüldüğü üzere, önerilen sistemin ilk aşamasında içerik tabanlı filtreleme yönteminde girdi olarak kullanılmak üzere makalelerin özet ve başlık bilgileri çıkarılmaktadır. Sonraki aşamada, tavsiye sisteminin performansını iyileştirebilmek için özet ve başlıklara metin ön işleme yapılmaktadır. Metin ön işleme yapıldıktan sonraki aşama ise sırasıyla başlık, özet ve ayrıca başlık ve özetin birlikte ele alınması ile lineer kernel yöntemi kullanılarak makaleler arasındaki benzerlik değerlerinin elde edilmesidir. Lineer kernel, polinomsal kernel yönetiminin özel bir durumudur. Polinomsal kernel Denklem (1) ile hesaplanır:

$$k(x, y) = (\gamma x^T + c_0)^d \quad (1)$$

Lineer kernel, polinomsal kernel fonksiyonundaki γ ve d parametrelerinin 1, c_0 ise 0 olarak ele alınmasıyla elde edilir. Lineer kernel fonksiyonu, x ve y sütun vektörleri ise, bunların doğrusal çekirdeği Denklem (2) ile hesaplanır:

$$k(x, y) = x^T y \quad (2)$$

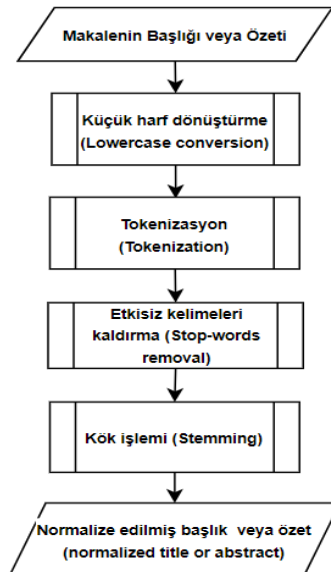
Son aşamada, hesaplanan benzerlik değerlerine göre makaleler sıralanarak kullanıcıya önerilmektedir.

2.1 Veri Setinden İçerik Çıkarılması

İçerik tabanlı yöntemlerde, makalelerin tüm metninden yararlanılır ve diğer yaklaşımlardan daha umut verici sonuçlar elde edilmektedir. Ancak çoğu içerik, abonelik gereksinimleri nedeniyle ücretsiz olarak kullanılamamaktadır. Bu nedenle, başlık ve özet makalelerin içeriği olarak değerlendirilmektedir [11]. Bu çalışmada, ARXIV veri setinden elde edilen makalelerin başlığı ve özeti kullanılmaktadır [10]. ARXIV veri setinden her bir alandan makale içerecek şekilde 6000 makale rastgele seçilerek çalışmada kullanılacak veri seti oluşturulmuştur. JSON formatında tutulan veri setinde 37.810 adet kelime grubu yer almaktadır.

2.2 Elde Edilen Metinlere Ön İşlem Uygulanması

Bir metin üzerindeki noktalama işaretleri, dilde sık kullanılan kelimeler, yeni satır için belirtilen ‘\n’, girintilemek (tab) için kullanılan ‘\t’ gibi işaretler metindeki benzerlik oranını etkilemektedir. Metinler arasındaki benzerlik oranını hesaplamadan önce veri setine bazı ön işlemler uygulanmaktadır. Ön işlem aşaması yalnızca hesaplama karmaşıklığını azaltmakla kalmayıp aynı zamanda tavsiye sisteminin performansını da iyileştirmektedir. Şekil 2 ile önerilen akademik makale tavsiye sisteminde kullanılan ön işlem aşamasına ait akış diyagramı verilmektedir.



Şekil 2. Metin ön işlemenin akış şeması

Ön işlemin ilk adımı, beyaz boşluk (*white-space*) ve noktalama işaretlerinin kaldırılmasıdır. Ön işlemin ilk adımıyla başlık ve özet, noktalama işaretleri ve beyaz boşluk denilen *tab*, *space*, *enter* gibi karakterlerden temizlenmektedir. Bundan sonra, yardımcı fiiller, edatlar, bağlaçlar ve ünlemler gibi anlamsız simgeleri filtrelemek için göz ardı edilebilecek kelime listeleri (*stop-words*) kullanılmaktadır. Aynı kelime köküne sahip ancak farklı biçimlerde olan bazı kelimeler de vardır, örn. "Oluştur": "oluşturuldu". Bu nedenle, bir kelimenin farklı dil bilgisel biçimlerini kök biçimine çevirmek için kökten türetme de gereklidir [12]. Yaygın olarak kullanılan ve literatürde geçerliliği benimsenmiş kök belirleme algoritması olan "*Porter Stemmer*", önerilen sistemde kullanılmaktadır [13]. İngilizce kelimelerin iyi bilinen soneklerini kaldırmak ve değiştirmek için tasarlanmış en yaygın kök belirleme algoritmalarından biridir.

2.3 Makale-Makale Benzerlik İlişkisinin Hesaplanması

Bu aşamada, başlık ve özet bilgisi ön işlemde geçirildikten sonra sırasıyla başlık, özet ve hem başlık hem özeti birlikte ele alınarak Terim Frekansı – Ters Doküman Frekansı (TF-IDF, *Term Frequency – Inverse Document Frequency*) oranları hesaplanır. TF-IDF bir doküman içerisinde geçen terimlerin tespit edildiği ve bu terimlerin geçtiği miktara göre çeşitli hesapların yapıldığı istatistiksel bir yöntemdir. TF-IDF ağırlığı iki terimden oluşur. İlk terim bir makaledeki bir sözcüğün görünme sayısının o makaledeki toplam sözcük sayısına bölünmesiyle elde edilen sayı olan normleştirilmiş TF, ikinci terim ise kitaplardaki (*corpus*) makale sayısının logaritmasının belirli terimin görüldüğü makale sayısına bölünmesiyle hesaplanan IDF terimidir.

$$Tf(t) = \frac{\text{Bir dokümanda } t \text{ teriminin görülme sıklığı}}{\text{Bir dokümandaki terimlerin toplam sayısı}} \quad (3)$$

IDF ise bir kelimenin önemini ölçer. TF hesaplanırken, tüm kelimeler eşit önemde düşünülür. Fakat, İngilizcedeki '*is*', '*of*', '*that*' vb. kelimeler yaygın olarak kullanıldığından belirleyiciliğinin daha düşük olması beklenir. Kısaca IDF, sık geçen bu tip kelimelerin etkisini düşürürken, nadir geçen kelimelerin önemini artırır.

$$Idf(t) = \log_2\left(\frac{\text{Dokümanların toplam sayısı}}{t \text{ terimini içeren dokümanların sayısı}}\right) \quad (4)$$

Geliştirilen prototipte, makaleler arasındaki benzerlik değerlerini hesaplamak için TF-IDF oranları ve lineer kernel yöntemi birlikte kullanılmaktadır.

2.4 En Benzer İlk On Makalenin Önerilmesi

Bu adımda, girdi olarak ele alınan makalenin başlığı, makalenin özeti veya makalenin özeti ve başlığı bilgilerine karşılık gelen rakamsal indeks hesaplanmaktadır. Daha sonra o indekse sahip makalenin

diğer makaleler ile olan benzerlik skorlarına göre büyükten küçüğe sıralama işlemi gerçekleştirilmektedir. Son olarak en benzer ilk on makalenin kullanıcıya önerilmesi gerçekleştirilmektedir.

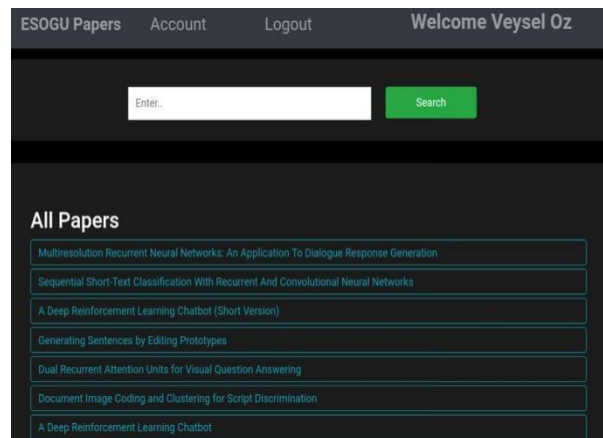
3. AKADEMİK MAKALE TAVSİYE SİSTEMİ PROTOTİP TASARIMI

Önerilen akademik makale tavsiye sistemi prototipi web tabanlı olarak geliştirilmiştir. Önerilen tavsiye motoru; makale-makale ilişkisini ele almaktadır. Web arayüzünden kaydolun kullanıcıların bilgilerini tutmak için ilişkisel veri tabanı olan Microsoft SQL Server kullanılmaktadır. Kullanıcılar için veri tabanında *id*, *username*, *name*, *email*, *phone*, *birthofdate* özellikleri tutulmaktadır. Platform kullanıcılar tarafından kullanılmaya başlandıktan sonra kullanıcı geçmişi, makale tıklanma sayısı, derecelendirme gibi özellikler izlenmektedir. Önerilen sistem tarafından kullanılan makale içerikleri, NoSQL veri tabanı olan MongoDB üzerinde tutularak geliştirilen uygulama tarafından hızlı, güvenilir ve kolay bir şekilde okunmaktadır. Esnek yapı, büyük verili ve çok sayıda aktif kullanıcı sistemlerde yüksek performans ve yönetim kolaylığı sunan MongoDB, önerilen akademik makale tavsiye sisteminde makalelerin depolanmasında kullanılmaktadır. Uzun vadede çok sayıda kullanıcı girişi ve hareketi olması durumunda kullanıcı bilgilerinin de MongoDB'ye aktarılması planlanmaktadır.

3.1 Akademik Makale Tavsiye Sistemi Grafikal Kullanıcı Arayüzleri

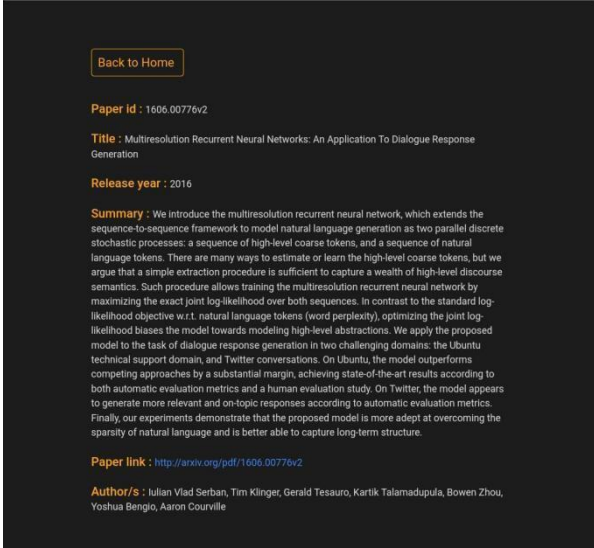
Tavsiye sisteminin uygulandığı web platformu Php dilinde yazılmaktadır; *front-end* tasarımında Html, CSS ve Bootstrap kullanılmaktadır.

Kullanıcı e-posta ve şifresi ile giriş yaptıktan sonra Şekil 5'te görüldüğü gibi ana sayfaya yönlendirilmektedir. Ana sayfa ekranı, her bir sayfada 100 tıklanabilir makale bulunacak şekilde alt sayfalardan oluşmaktadır.



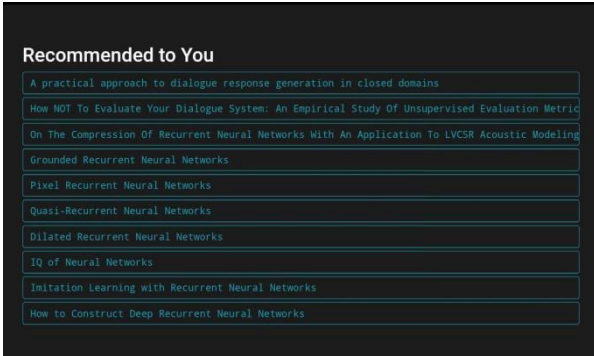
Şekil 3. Ana sayfa

Kullanıcı herhangi bir makaleye tıkladığında makale bilgisi sayfasına yönlendirilmektedir. Bu sayfa; makale ile ilgili bilgiler (makalenin başlığı, özeti, yayın yılı vb.) içermektedir.



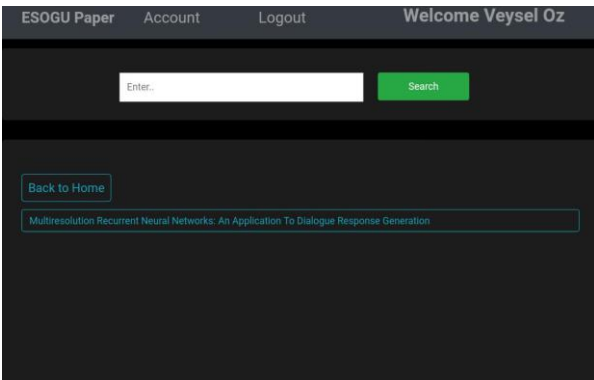
Şekil 4. Makale bilgi sayfası

Sayfanın altında bu makale ile ilgili Şekil 5'te görüldüğü gibi ilk on makale önerisi bulunmaktadır. Kullanıcılar, aradıkları alana uygun sistem tarafından önerilen makaleleri ziyaret edebilmektedirler.



Şekil 5. İlgili makale altındaki öneri listesi

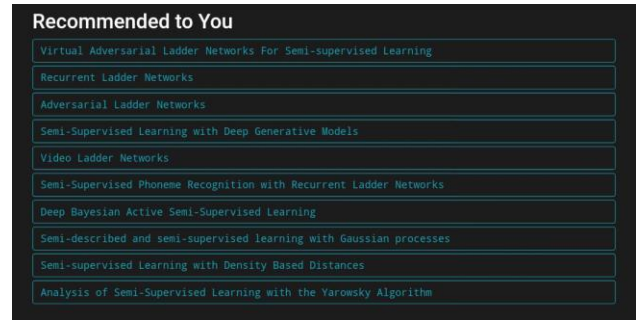
Şekil 6 arama sonuçlarının listelendiği ekranı göstermektedir. Kullanıcılar, arama kutucuğuna aradıkları makalenin tam başlığını girerek arama işlemini gerçekleştirebilmektedirler.



Şekil 6. Arama arayüzü

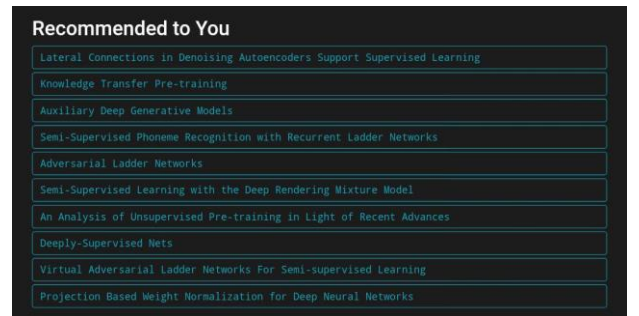
4. BULGULAR

Bu çalışmada Akademik Makale Tavsiye Sistemi için içerik tabanlı filtreleme yöntemi kullanılmaktadır. Literatürdeki içerik tabanlı filtreleme yöntemlerine kıyasla makalenin başlığı, makalenin özeti ve hem makale özeti ve hem makale başlığı birlikte girdi olarak kullanılmaktadır. Makale başlığını yazarın belirlemesinden dolayı, önerilerde hataya neden olabileceği gözlemlenmektedir. Bir makalenin içeriğini tanımlayabilecek en iyi kriterlerden birinin özet olabileceği sonucuna varılmaktadır. Bu yüzden makale başlığı ve özeti birlikte girdi olarak kullanılarak tavsiye sisteminin performansı iyileştirile bilinmektedir. Makale başlığına göre benzerlik skorlarına bakıldığında bir makalenin benzerliğinin yüksek olduğu ve en benzer on makale sıralamasında en üst sıralardayken, girdi olarak makale başlığı ve özeti birlikte kullanıldığında bu makalenin benzerlik skorunun ve en benzer on makale sıralamasının azaldığı görülmektedir. Şekil 7 ile verilen sonuçlar incelendiğinde aynı başlık bilgisi verildiğinde lineer kernel benzerlik ölçüm yöntemi ile ilk on makalenin önerildiği görülmektedir.



Şekil 7. "Semi-supervised Learning with Ladder Networks" başlıklı bilimsel çalışmanın bilgi sayfasında başlık bilgisine göre üretilen ilk on öneri

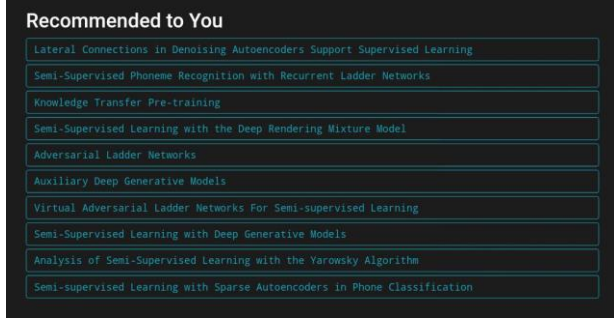
İlgili makale girildiğinde lineer kernel yöntemi için özet bilgisine göre önerilen ilk on makale Şekil 8'de görülmektedir.



Şekil 8. "Semi-supervised Learning with Ladder Networks" başlıklı bilimsel çalışmanın bilgi sayfasında özet bilgisine göre üretilen ilk on öneri

Şekil 7 ve Şekil 8 karşılaştırıldığında özet bilgisi bir metnin benzerlik oranı açısından daha doğru değerlendirilmektedir. Şekil 8'de verilen sonuçlarda bazı makalelerin sıralamasının değiştiği veya skorunun

değiştirdiği gözlemlenmektedir. “*Semi-Supervised Phoneme Recognition with Recurrent Ladder Networks*” makalesi sadece başlık kullanılarak sıralandığında altıncı sırada iken makalenin özeti dikkate alındığında ise bu makalenin dördüncü sıraya yerleştiği görülmektedir. Bu oranlar bir metnin sadece başlıklarına bakılarak benzer olduğunun söylenemeyeceğini göstermektedir. Aynı makale için özet ve başlık bilgisinin birlikte kullanılmasıyla önerilen ilk on makale ise Şekil 9’da görülmektedir.



Şekil 9. “*Semi-supervised Learning with Ladder Networks*” başlıklı bilimsel çalışmanın bilgi sayfasında özet ve başlık bilgisine göre üretilen ilk on öneri

Şekil 7 ve Şekil 8’e bakıldığında “*Semi-Supervised Phoneme Recognition with Recurrent Ladder Networks*” makalesi iki yöntemde de sıralamada alt sırada bulunmaktadır. Fakat Şekil 9 ele alındığında ikinci sıraya çıktığı görülmektedir. Makale başlığına göre en benzer ilk on makale listesinde bir makale alt sıralardayken, girdi olarak makale başlığı ve özet birlikte kullanıldığında bu makalenin daha üst sıralara yükseldiği görülmektedir.

5. TARTIŞMA VE SONUÇ

Bu çalışmada, yeni bir içerik tabanlı akademik makale tavsiye sistemi önerilmektedir. Web tabanlı bir uygulama olarak sunulan tavsiye sistemi, veri seti için lineer kernel benzerlik ölçümü kullanarak öneri sunmaktadır. Analiz sonuçlarında başlık bilgisi, özet bilgisi ve hem özet ve hem başlık bilgisi için veri seti içerisinde ilk on makalenin önerildiği ve bu makaleler için elde edilen sonuçlara göre makalenin başlığının tek başına doğru bir öneri sunamayacağı, makalenin özeti ve başlığı birleştirilerek kullanılmasının daha doğru öneriler sunacağı tespit edilmiştir.

Gelecekte yapılacak çalışmalar arasında, yazar, anahtar kelimeler, etiketler, yazarların h-indexleri ve makalenin referans kombinasyonları ile tavsiye eden sistem analiz edilerek kullanıcı memnuniyetinin de dikkate alındığı önerilerin gerçekleştirilmesi hedeflenmektedir. Bu tavsiye sistemi makale, dergi, konu ile ilgili varsa kod deposu, konuyla daha önce ilgilenen başka yazarlar ve konu ile ilgili herkese açık olan veri setlerini tavsiye edecek bir sistem olarak geliştirilebilir. Geliştirilecek olan tavsiye sistemine kullanıcı bilgilerinin sağlanabilmesi için sistem üzerindeki kullanıcı hareketlerinin tutulması sağlanacaktır.

Teşekkür

Bu çalışma Eskişehir Osmangazi Üniversitesi, Bilimsel Araştırma Projeleri Komisyonu’nun (ESOGÜ-BAP) 14.01.2021 tarih ve 61690618-622.03 sayılı kararı ile (proje numarası 202115003) desteklenmektedir. Katkılarından dolayı teşekkür ederiz.

KAYNAKÇA

- [1] Ruotsalo, Tuukka. (2010). Methods and applications for ontology-based recommender systems.
- [2] Lu, Jie & Wu, Dianshuang & Mao, Mingsong & Wang, Wei & Zhang, Guangquan. (2015). Recommender System Application Developments: A Survey. *Decision Support Systems*. 74. [10.1016/j.dss.2015.03.008](https://doi.org/10.1016/j.dss.2015.03.008).
- [3] Beel J, Gipp B, Langer S, Breiteringer C. Paper recommender systems: a literature survey. *International Journal on Digital Libraries* 2016; 17 (4): 305-338.
- [4] Ricci F, Rokach L, Shapira B. Introduction to Recommender Systems Handbook. *Recommender Systems Handbook*; 2011. p. 1-35
- [5] Schafer JB, Frankowski D, Herlocker J, Sen S. Collaborative filtering recommender systems. In: *The Adaptive Web*; Berlin, Germany; 2007. pp. 291-324.
- [6] X. Bai, M. Wang, I. Lee, Z. Yang, X. Kong and F. Xia, "Scientific Paper Recommendation: A Survey," in *IEEE Access*, vol. 7, pp. 9324-9339, 2019.
- [7] P. Lops, M. Gemmis, and G. Semeraro, “Content-based recommender systems: State of the art and trends,” *Recommender Systems Handbook*, pp. 73–105, 2011.
- [8] Liu C. The proximity of co-citation. *Scientometrics* 2012; 91 (2): 495-511.
- [9] Liu S, Chen C. The effects of co-citation proximity on co-citation analysis. In: *13th International Conference of the International Society for Scientometrics and Informetrics*; Durban, South Africa; 2011. pp. 474-484.
- [10] ARXIV. (2020). Retrieved from <https://www.kaggle.com/neelshah18/arxivdataset> (Erişim Tarihi: 09.04.2021)
- [11] Ahmad, Shahbaz & Afzal, Muhammad. (2017). Combining Co-citation and Metadata for Recommending More Related Papers. 218-222.
- [12] Otair, Mohammed. (2013). Comparative Analysis of Arabic Stemming Algorithms. *International Journal of Managing Information Technology*.
- [13] Porter, M.F.. (2006). An algorithm for suffix stripping. *Program*. 14. 130-137.