

Türk Kütüphaneciliği Dergisi Dijital Arşivindeki Bilimsel Makalelere Erişimin Artırılması: Türkçe Özlerden Anahtar Kelimeler Atanması*

Enhancing Access to Scientific Articles in the Turkish Librarianship Journal Digital Archive: Assigning Keywords from Turkish Abstracts

Kemal Yayla** 

Öz

Amaç: Bu çalışmanın amacı, Türk Kütüphaneciliği dergisinin web sitesinde yer alan ve anahtar kelime atanması yapılmamış hakemli makalelere Türkçe özlerini kullanarak anahtar kelime atamaları gerçekleştirmektir. Bu sayede anahtar kelime eksikliği nedeniyle erişimde güçlük çekilen web arşivindeki çalışmalara daha etkin bir şekilde erişim sağlanması hedeflenmektedir.

Yöntem: Çalışmada, 1995 ile 1999 yılları arasında anahtar kelime olmaksızın yayımlanmış olan 58 hakemli makale incelenmiştir. Anahtar kelime ataması için Türkçe özleri girdi olarak kullanılan YAKE algoritması kullanılmış, metinlerin işlenmesi için Zemberek doğal dil işleme aracı ve Python programlama dili tercih edilmiştir. Atanan anahtar kelimelerin anlamlılığı, Anlamlılık Kontrol Oranı (AKO) ve Ortalama Mutlak Hata (OMH) değerleri ile ölçülmüştür. Atanan anahtar kelimelerin bağlamsal geçerliliği üç uzman değerlendirici ile algoritmanın atadığı anahtar kelimeler arasında değerlendirici uyumunu ölçen ACI oranı ile ölçülmüştür.

Bulgular: Anahtar kelime ataması yapılan makalelerde en sık atama yapılan kelimelerin “kütüphane”, “bilgi”, “hizmet” ve “makale” olduğu tespit edilmiştir. Algoritmanın OMH değeri 0,099 olarak belirlenmiş, bu da algoritmanın yüksek doğrulukla anahtar kelime atadığını göstermektedir. Ancak, uzman değerlendiriciler ile algoritma tarafından atanan anahtar kelimeler arasında düşük seviyede uyum tespit edilmiştir.

Sonuç: Çalışma, Türkçe özlerden anahtar kelime çıkarımı yapmanın dijital belge erişimini artırmada önemli bir yöntem olduğunu göstermektedir. Algoritmaların daha yüksek performans göstermesi için, uzmanlar tarafından belirlenen veri setleri ile eğitilmeleri gerekmektedir. Ayrıca, yapılandırılmış öz kullanımı ve öz metinlerin uzunluğunun artırılması önerilmektedir.

Özgünlük: Bu çalışma bilimsel makalelerin Türkçe özlerden anahtar kelime çıkarımı yaparak dijital belge erişimini artırmaya yönelik özgün bir yaklaşım sunmaktadır. Kütüphanecilik ve Bilgi Bilim alanındaki dijitalleşmiş makaleleri web ortamından erişimin artırılmasına yönelik yapılan ilk çalışmalardan biridir.

Anahtar Sözcükler: Anahtar kelime atanması; YAKE; Türkçe öz; dijital arşiv.

*Bu çalışma “Arşivlenmiş Makalelere Erişimin İyileştirilmesi: Türk Kütüphaneciliği Dergisi Makalelerine Türkçe Özlerinden Anahtar Kelime Ataması” isimli Cumhuriyet’in 100. Yılında Türk Kütüphaneciliği Sempozyumunda sunulan bildiri özütünün gözden geçirilmiş ve genişletilmiş halidir.

This study is a revised and expanded version of the abstract of the paper presented at the Turkish Librarianship Symposium on the 100th Anniversary of the Republic titled “Improving Access to Archived Articles: Assigning Keywords from Turkish Abstracts to Turkish Librarianship Journal Articles”.

** İzmir Katip Çelebi Üniversitesi, Sosyal ve Beşeri Bilimler Fakültesi, Bilgi ve Belge Yönetimi Bölümü, İzmir, Türkiye. E-posta: kemal.yayla@ikcu.edu.tr

İzmir Katip Çelebi University, Faculty of Social and Human Sciences, Department of Information and Document Management, İzmir, Türkiye. E-mail: kemal.yayla@ikcu.edu.tr

Abstract

Purpose: This study aims to assign keywords to peer-reviewed articles on the Turkish Librarianship Journal's website that do not have keyword assignments by using their Turkish abstracts. The goal is to improve access to works in the web archive that are difficult to reach due to the lack of keywords.

Method: The study examined 58 peer-reviewed articles published without keywords between 1995 and 1999. The YAKE algorithm was used for keyword assignment, with Turkish abstracts as input, and the Zemberek natural language processing tool and Python programming language were utilized for text processing. The meaningfulness of the assigned keywords was measured using the Significance Control Ratio (SCR) and Mean Absolute Error (MAE) values. The contextual validity of the assigned keywords was assessed by measuring the inter-rater agreement (ACI ratio) between three expert evaluators and the algorithm-assigned keywords.

Findings: The most frequently assigned keywords in the articles were "library," "information," "service," and "article." The algorithm's MAE value was determined to be 0.099, indicating that the algorithm assigned keywords with high accuracy. However, low agreement was found between the keywords the experts assigned and those the algorithm assigned.

Implications: The study demonstrates that extracting keywords from Turkish abstracts is essential for enhancing access to digital documents. Algorithms need to be trained on datasets determined by experts to perform better. Additionally, using structured abstracts and increasing the length of abstract texts are recommended.

Originality: This study presents a novel approach to improving digital document access by extracting keywords from Turkish abstracts of scientific articles. It is one of the first studies aimed at increasing web access to digitized articles in the Librarianship and Information Science field.

Keywords: Keyword assignment; YAKE; Turkish abstract; digital archive.

Giriş

Akademisyenlerin çalışma rutinleri genellikle öğretim, öğrenci performansının değerlendirilmesi, idari işler ve toplumsal katkıyı içerir; ancak asıl odak noktaları, araştırma monografileri ve bilimsel dergi makalelerine yönelik araştırma ve yazım faaliyetleridir (Yılmaz ve Memişoğlu, 2019, ss. 543-544). Beşerî bilimler için araştırma monografileri akademik başarının zirvesi olarak görülse de sosyal bilimlerde yüksek etkili dergilerde makale yayımlamak araştırmacıların kariyer hedefleri için daha belirleyici bir etkiye sahiptir. Bu süreçte akademik dergiler araştırma bulgularının kamuoyuyla paylaşılması ve bilimsel uzmanlığın toplumla paylaşılması gibi önemli işlevleri yerine getirerek bir aracı rolü üstlenirler (Brienza, 2012; Moulaison ve Million, 2015). Dijital yayıncılık ve bilimsel araştırma pratiklerinin artan kullanımıyla birlikte akademisyenler, araştırmalarını yürütme, meslektaşlarıyla iş birliği yapma, bulgularını paylaşma, konferans bildirimlerini yayımlama ve veri setlerini paylaşabilme gibi çeşitli faaliyetler için gerekli araç ve kaynaklara daha fazla erişim sağlamıştır (Çapkın, 2014; Kayaoğlu, 2007).

İnternetin geniş kapsamlı, sürekli değişen ve çeşitlilik gösteren yapısı, dijital bilimsel yayınların kullanıcılar tarafından keşfedilmesini zorlaştırabilir. Kullanıcılar, arama yaparken kapsamı genişlettikçe indeksleme sorunları, büyük erişim kümeleri ve disiplinler arası anlam farklılıkları gibi zorluklarla karşılaşabilir. Bu durum, sayfa içeriğine yalnızca sınırlı sayıda kullanıcının ulaşabilmesine neden olabilir (Mohamed, 2006). Bir web sayfasını çevrimiçi yayımlamak veya arama motorlarına göndermek, o sayfanın arama motoru sonuçlarında başarılı bir şekilde bulunmasını garanti etmez (J. Zhang ve Dimitroff, 2005). Sınırlı teknik uzmanlığa ve kaynaklara sahip yayıncılar, arama motoru optimizasyonu gibi uygulamaları ihmal ederek daha fazla görünürlük fırsatını kaçırabilirler. Bununla birlikte internette yer alan bilgilerini etkili ve verimli bir şekilde otomatik olarak indekslemeye yardımcı olacak basit, uyumlu ve kullanışlı bir bilgi açıklama standardı olan doldurulmuş HTML etiketleri, yani üst veriler, belge erişimini ciddi biçimde artırmaktadır (Allen ve Weber, 2015). Özellikle dijital olarak oluşturulan materyaller için üst veriler standart bir biçimde web sitesine özgü olarak oluşturulmaktadır.

Değişen bilimsel yayıncılık dinamikleri, basılı olarak arşivlenen akademik dergilerin arşivleme süreçlerinde önemli değişikliklere yol açmıştır. Elektronik yayınlar için uygulanamayan geleneksel arşivleme yöntemlerinin yerini, web ortamı için geliştirilen dijital arşivleme süreçleri almıştır (Honey, 2005). Teknik altyapının gelişmesiyle birlikte sadece dijital üretilen yayınlar değil geçmişte fiziksel olarak saklanan kopyalarda sayısallaştırılarak dijital ortama aktarılmaya başlanmıştır. Bu yaklaşım fiziksel kopyaların uzun vadeli korunmasıyla ilgili problemlerin üstesinden gelmeyi önemli ölçüde kolaylaştırmıştır (Küçük ve Alır, 2003; Aldemir ve Oğuz, 2006; Bayter, 2009).

Dijitalleştirmeyle birlikte basılı araştırma makalelerine erişimin artması, Kütüphanecilik ve Bilgi bilim (KB) disiplininin geçmişten günümüze anlaşılması için önemli fırsatlar sunarken, kullanıcıların geçmişteki akademik yayıncılık anlayışına göre hazırlanmış bilimsel çalışmalarını güncel bilgi erişim sistemlerine uygun olarak sınıflandırma ve erişime sunma gibi yeni zorlukları da beraberinde getirmiştir (Gil-Leiva ve Alonso-Arroyo, 2007).

Özellikle web ortamında kolayca erişilebilen akademik dergi yayınları için anahtar kelime kullanarak arama yapmak, günümüzde en sık başvuru alan akademik bilgi arama yaklaşımlarından birisidir (Howcroft, 2007). Akademik yayıncılıkta dergiler, yazarlardan sunulan makalelerin ana hatlarını ve temel konularını özetleyen anahtar kelime listesi sağlamlarını sıklıkla talep eder. Anahtar sözcükler, özetleme, indeksleme ve daha hassas arama işlemleri gibi çeşitli amaçlara hizmet eder ve belge üst verilerinin kritik bir parçasını oluşturur. Belge koleksiyonlarında arama yapılırken dizin terimleri olarak işlev görür (Turney, 2002). Çevrimiçi olarak erişilebilen geniş belge havuzunun yalnızca küçük bir bölümü bu önemli özelliklere sahiptir (Chi vd., 2019). Bu durum, araştırmacıların ve kullanıcıların ilgili bilgilere erişimini zorlaştırmakta ve akademik çalışmaların etkili bir şekilde indekslenmesi ve sınıflandırılması sürecini karmaşık hale getirmektedir (Blank vd., 2016; Dişli ve Tonta, 2023).

İnternet ortamında saklanan dijital belgelerin etkin bir şekilde tanımlanabilmesi ve erişilebilirliklerinin kolaylaştırılması, detaylı bir sınıflandırma ve indeksleme sürecini zorunlu kılmaktadır. Bu süreç anahtar sözcüklerin etkin kullanımını da içermekte olup, belgelerin sınıflandırılmasını ve erişilebilirliğini iyileştirirken aynı zamanda akademik çalışmaların görünürlüğünü ve erişimini de artırma potansiyeline sahiptir (Gillis vd., 2019; Terra vd., 2021). Bu bilgilerden hareketle çalışmanın amacı, Türk Kütüphaneciliği (TK) dergisi web sitesinde yer alan ve anahtar kelime ataması yapılmamış hakemli makalelere, Türkçe özetlerini kullanarak anahtar kelime atamalarını gerçekleştirmektir. Bu sayede anahtar kelime eksikliği nedeniyle erişimde güçlük çekilen web arşivindeki çalışmalara daha etkin bir şekilde erişim sağlanması amaçlanmaktadır. Bu çalışmada elde edilen bulguların, web ortamında yer alan dijital belge koleksiyonlarını kullanmayı hedefleyen araştırmacılar ve araştırmalarının görünürlüğünü artırmak isteyen yayıncılar tarafından kullanılabilmesi düşünülmektedir.

Literatür Değerlendirmesi

Dijital ortamda yer alan belgelere erişim noktası oluşturma, belgenin içerdiği konuların belirlenmesi ve bunların kontrollü bir yapıda temsil edilmesi üzere iki aşamadan oluşmaktadır. Bu aşamalar kimi zaman uzman bilgisine (Grey ve Hurko, 2012) kimi zamanda belge metinlerini analiz eden algoritmalara (C. Zhang vd., 2008; Binici, 2019; W. Lu vd., 2021) göre gerçekleştirilmektedir. Belgelerin sınıflandırılması ve erişilebilirliğinin artırılması amacıyla en sık başvuru alan yöntemlerden birisi belgelerin içeriğine tanımlayan anahtar kelimelerin çıkarımıdır. Literatürde anahtar kelime çıkarımı için çeşitli yöntemler geliştirilmiştir. Siddiqi ve Sharan (2015) ise anahtar kelime çıkarım süreçlerini kural tabanlı dilbilimsel yaklaşımlar, istatistiksel yaklaşımlar, makine öğrenimi yaklaşımları ve araştırma alanına özgü yaklaşımlar olarak dört ana grup altında incelemiştir. Papagiannopoulou ve Tsoumakas (2020) anahtar kelime çıkarım yöntemlerini denetimli (supervised) ve denetimsiz (unsupervised) olmak üzere iki ana kategoride sınıflandırmışlardır. Yine aynı çalışmada denetimsiz yöntemler istatistiksel, graf (graph) temelli, gömülü özellikler ve dil modeli temelli olarak dört farklı alt başlık altında incelenmiştir. Son on yılda araştırmaların odağı denetimsiz yöntemlere, özellikle de ağ ve graf tabanlı anahtar kelime çıkarım metodlarına doğru kaymıştır (Beliga vd., 2016).

Dijital belgelerin 'makine tarafından okunabilir ve işlenebilir' hale getirilerek dijital olarak kullanılabilir formatlarda sunulması, yayın altyapısındaki etiketler aracılığıyla akademik bilgi sistemlerine erişim sağlayan dergiler için önemli bir süreçtir. Ancak basılı arşivlerini sayısallaştırma yoluyla çevrim içi erişime açan dergilerin sundukları bu koleksiyonları zenginleştirilmiş bir bilgi erişim perspektifinden uzak, sınırlı bir üst veri sunumu yaklaşımı benimsedikleri görülmektedir (Conway, 2015). Bilimsel bir yayın için üst veri örnekleri arasında belgenin başlığı, oluşturulma tarihi, yazarı, veri formatı, öz metni ve varsa anahtar kelimeler bulunur. Ancak orijinal belgede bulunmayan bazı özelliklerin algoritmik yöntemlerle eklenmesi mümkündür. Anahtar kelime çıkarım yöntemleri eksik olan üst verilerin tamamlanması için kullanılacak alternatif çözümlerden birisi olarak değerlendirilmektedir.

Bilimsel makalelerden anahtar kelime çıkarımı diğer belgelere kıyasla farklılık göstermektedir. Bir bilimsel çalışmanın yapısı, içerdiği bilgilerle hedeflenen alana özgü biçimlendirme tarzlarının kullanılması ve metnin özelleştirilmiş alt birimlere bölünmesiyle diğer belgelerden ayrılır. Bilimsel metinlerde belgenin özel alt birimlere ayrılması, bilgiye erişim süreçlerinde kullanıcıların arzuladığı bilgilere ulaşmayı kolaylaştırır (Dynch ve Wang, 2017). Güncel literatürde bilimsel makaleler için erişim noktaları oluşturma sürecinde veri kaynağı olarak belgenin kaynakça (C. Zhang vd., 2022) veya öz (abstract) (Bhowmik, 2008) gibi özel alt birimlere de odaklanıldığı görülmektedir.

Dergi makalelerinin özleri, okuyucunun ilgisini çekme konusunda kritik bir rol oynar ve makalenin önemli bölümlerinden birini teşkil eder (Šaupert vd., 2008). Makale özleri ve anahtar kelimeler, çevrimiçi veri tabanlarında makale hakkında ön değerlendirme yapılmasına olanak tanır ve kullanıcılara belgenin içeriği hakkında önemli bilgiler sunar (Hahs-Vaughn ve Onwuegbuzie, 2010; Jamar vd., 2014; Atanassova vd., 2016). Literatürdeki çalışmalar, makale özlerinden anahtar kelimeleri çıkarmayı amaçlayan algoritmaların özellikle İngilizce dili için tasarlandığını göstermektedir (Beliga, 2014). Ancak, İngilizce dışında yazılmış metinlerin özlerinin, İngilizce özlerle kıyaslandığında kuramsal ve retorik özellikler açısından önemli farklılıklar gösterdiği bilinmektedir (Alharbi ve Swales, 2011; Escudero ve Swales, 2011). İngilizce dışındaki dillerde yazılan özleri kullanarak anahtar kelime çıkarmada dil özelliklerini doğru yansıtabilmek için dilden bağımsız veya dile özgü algoritmaların kullanılması gerekmektedir (Daudaravicius, 2016). Bu, anahtar kelime atama sürecinde anlam bütünlüğünün korunması için kritik bir öneme sahiptir (Ercan ve Cicekli, 2007).

Literatür taraması sonucunda İngilizce haricinde dillerde yazılmış bilimsel çalışmalardan anahtar kelime çıkarımı için farklı metotların uyarlanması sonucunda geliştirilmiş metotlar tespit edilmiştir. Pala ve Cicekli (2007) Türkçe için uyarlanmış Keyphrase Extraction (Witten vd., 1999) algoritmasını kullanarak Türkçe belgeler için anahtar kelime çıkarımı yapmıştır. İngilizce'den farklı olarak, bir sondan eklemeli dil olan Türkçe'nin özelliklerini dikkate alarak algoritmanın eğitim kısmı için kullanılan kök ve durak kelime listelerini değiştirilmiştir. Bu uyarlanmış algoritma, Türkçe kelime köklerindeki çekim eklerini ayıklayarak, elde kalan kökleri anahtar kelime olarak belirlemiştir. Yöntem tek bir belgenin içeriğine odaklanmak yerine, belirli bir belge kümesinin özelliklerine dayalı olarak kelime köklerini tespit etmeye yöneliktir. Bu nedenle seçilen veri setinin anahtar kelime çıkarımı için hedeflenen konu ile yakından ilgili olduğu varsayımından hareket edilmektedir.

Çince, İngilizce' den farklı bir morfolojiye sahiptir ve bu dilde belirgin bir kelime sınırının olmaması belgelerden otomatik anahtar kelime çıkarımını zorlaştırmaktadır. Liu ve Zhai (2011), Çince metinlerden anahtar kelime çıkarımı için graf tabanlı bir çözüm olan TEXT-NET algoritmasını önermiştir. Bu yöntem, kelimeler arasındaki anlamsal benzerliklere dayanarak, incelenen belgeler için bir ağ grafiği oluşturur ve sonrasında belgenin istatistiksel özelliklerini ile ağın özelliklerini kullanarak, belgeye özgü anahtar kelimelerin çıkarımını gerçekleştirir. Li, Du ve Xing (2017), Çince bilimsel özetler için Textrank (Mihalcea ve Tarau, 2004) algoritmasına dayanan bir yöntem geliştirmiştir. Bu yöntemde, öncelikle bilimsel özetler Çince kelimelere ayrılır. Daha sonra özel anlam ifade etmeyen kelimeleri çıkarılıp, kalan kelimeler TextRank algoritması ile önem sırasına göre değerlendirilir. En önemli kelimeler, özetde bitişik olan kelimelerle birleştirilerek çoklu kelime adayları oluşturmak üzere seçilmiştir. Aday kelimeler, bir anahtar kelime sözlüğünde aranır ve bulunanlar anahtar kelime olarak kabul edilir. Sözlükte yer almayan adaylar için anahtar kelime olma olasılıkları, derlemde elde edilen kelime sıklığı bilgileri kullanılarak hesaplanır. Belirlenen bir eşik değerinin üzerindeki olasılığa sahip aday kelimeler anahtar kelime olarak seçilir.

Lingua franca olarak kabul edilen İngilizce dışında kalan diller için sınırlı kaynaklara sahip algoritmik dilbilimsel yöntemler geliştirilmiştir. Bu dillerde veri azlığı, araştırmacıların yapay sinir ağları modelleriyle veri analiz etmesi önünde büyük bir engel teşkil etmektedir. Yeterli eğitim verisinin olmaması düşük kaynaklı dillerde nöral tabanlı yaklaşımların uygulanmasını zorlaştırmakta ve doğru sonuçların elde edilmesini engellemektedir. Araştırmacılar, bu modelleri sınırlı kaynaklarla uyumlu hale getirme ve ince ayar yapma konusunda zorluklar yaşamakta, bu durum modellerin genel performansını ve genelleştirilebilirliğini olumsuz etkilemektedir. Özellikle Farsça gibi morfolojik açıdan zengin fakat sınırlı eğitim verisi içeren dillerde, dilin kelime oluşumundaki kök ve ekler nedeniyle düşük frekansa sahip farklı kelime biçimleri ön işlemde geçirilmesi gerekmektedir. Mohseni ve Faili (2020), bilimsel metinlerden kelime çıkarımı için LSTM kodlayıcı-kod çözücü ağları ile Byte Pair Encoding (BPE) yöntemini önermiştir. Bu yöntem, kelime boyutunun sabit olduğu durumlarda ve az kullanılan kelimelerin alt kelime birimlerine ayrılması gerektiğinde özellikle faydalıdır. BPE yaklaşımı, çok sayıda benzersiz kelimeyi alt kelime bileşenleri olarak temsil ederek işlemeyi kolaylaştırmaktadır.

İngilizce dışındaki dillerde yazılmış bilimsel metinleri kullanmak amacıyla dile özgü yöntemlerin yanı sıra dilden bağımsız, belge içeriğinin istatistiksel özelliklerine dayanan anahtar kelime çıkarımı yöntemleri geliştirilmiştir. Geliştirilen yöntemlerden birisi olan RAKE (Rapid Automatic Keyword Extraction) algoritması, dilbilimsel kurallar yerine istatistiksel kalıpları temel alarak dilden bağımsız yetenekler sunmaktadır (Rose vd., 2010). Çeşitli dillere uyum sağlayacak şekilde tasarlanan RAKE, bir metindeki kelimeler arasındaki istatistiksel eş-oluşumlar ve ilişkilere odaklanır. Bu yöntem tüm dillerde önem taşıyan ve bir belgenin ana temalarını anlamada kritik olan içerik zengini kelimelere yoğunlaşır. RAKE'nin belirgin bir özelliği, daha az bilgilendirici kelimeleri filtrelemek amacıyla durdurma kelimelerini kullanabilmesidir. Kelime ilişkilerine dayalı anahtar kelime belirleme işlevi dil farklılıklarından etkilenmeden tutarlı sonuçlar üretebilmektedir. Ayrıca RAKE'nin puanlama mekanizması metnin stili ve içeriğine uyum sağlayacak şekilde tasarlanmıştır.

Bu sayede farklı bağlamlarda kelimelerin önemini değerlendirme konusunda esneklik sunar. RAKE algoritmasının verimliliği ve basitliği, karmaşık dilbilimsel süreçleri bir kenara bırakarak basit istatistiksel analizleri tercih ettiği için, çeşitli belge türleri ve diller arasında metodolojik basitlik ve farklı dillere uyarlanabilirlik sağlamaktadır (Rose vd., 2010).

Geliştirilen bir diğer dil bağımsız algoritma YAKE (Yet Another Keyword Extraction) algoritmasıdır. Bu yöntem dile özgü araçlara gerek duymadan uygulanabilir (Campos vd., 2018). Bu algoritma, belgelerdeki önemli anahtar kelimeleri tespit etmek için istatistiksel ve yerel metin özelliklerinin birleşimini kullanır. Temel olarak, terim eş-oluşumu ve sıklık verilerine dayanan YAKE, bu özelliklerin dilden bağımsız doğası sayesinde herhangi bir metinden anahtar terimleri etkili bir şekilde çıkarabilir. YAKE'nin eğitim derleminden bağımsız olarak çalışabilmesi, özellikle sınırlı kaynaklara sahip diller için büyük bir avantaj sağlar ve eğitim için ihtiyaç duyulan veri setlerine olan ihtiyacı azaltacak biçimde tasarlanmıştır. Ayrıca, YAKE'nin tasarımı sadece dilden bağımsız değil, aynı zamanda uygulama alanına özgü kısıtlamalardan bağımsızdır, bu sayede belirli bir konuya özelleştirme gereksinimi olmaksızın farklı alanlarda ve disiplinlerde uygulanabilir. Diğer benzer algoritmalar dilbilimsel araçlar veya harici veri tabanları gerektirirken, YAKE bu tür bağımlılıklar olmadan çalışır ve bu, kaynakların kısıtlı olduğu veya hiç bulunmadığı ortamlarda kullanımını kolaylaştırır (Campos vd., 2020).

Mevcut literatür incelendiğinde, anahtar kelime çıkarımı konusunda çeşitli yaklaşımların geliştirildiği açıkça görülmektedir. Bu yaklaşımlar arasında dil bağımlı ve dil bağımsız yöntemler ile istatistiksel ve makine öğrenmesi tabanlı teknikler bulunmaktadır. Özellikle, Türkçe gibi morfolojik açıdan zengin diller için geliştirilen spesifik yöntemlerin yanı sıra, YAKE gibi dil bağımsız algoritmaların da kullanıldığı saptanmıştır. YAKE algoritmasının farklı dillerde uygulanabilirliği, etiketlenmiş veri seti gerektirmemesi ve mevcut bir Python kütüphanesinin bulunması gibi avantajları, bu çalışmada tercih edilmesini sağlamıştır.

Yöntem

Veri Toplama

Anahtar kelimesi olmayan bilimsel yayınlara anahtar kelime atanması için yürütülen çalışmanın veri kaynağı Türkiye’de yayımlanan KB alanın en köklü dergilerinden birisi olan TK dergisidir. Türkiye’de KB alanındaki akademik yayıncılık 1950’li yıllardan itibaren Türk Kütüphaneciler Derneği Aylık Haber Bülteni (1953-1961), Yeni Yayınlar (1956-1979), Milli Kütüphane Haberleri (1957-1972), Kütüphaneciliğin Sesi (1961-1962) ve Kütüphane Dünyası (1975-1979) gibi kısa ömürlü süreli yayınlarla basılı olarak kamuoyuna ulaşmıştır. TK dergisi de Cumhuriyetin bir asırlık ömründe Türkiye’deki KB araştırmacılarının bilimsel iletişimde kullandıkları dergilerin en uzun soluklularından birisi olmuştur (Atılğan, 2002). Cumhuriyet tarihinin önemli bir bölümüne tanıklık etmiş olan dergi, 1952-1986 yılları arasında Türk Kütüphaneciler Derneği Bülteni (TKDB) adıyla yayın yapmış; 1987’den itibaren Türk Kütüphaneciliği adını aldıktan sonra 1995 yılı Cilt 9 Sayı 2’den itibaren hakem değerlendirmesi süreçlerini işleten bir dergi haline gelmiştir (Tonta, 2002, s. 285).

Bu gelişmeler doğrultusunda dergi de açık erişim yayıncılığı (Open Journal Publishing) benimseyerek 1953 yılından günümüze fiziksel olarak yayımlanmış sayılarını sayısallaştırmış web sitesinde¹ kullanıcıların erişimine sunmuştur (Çapkın, 2009).

Çalışmamızda incelenen zaman dilimi (1995-1999), Türk Kütüphaneciliği dergisinin anahtar kelime kullanımına geçiş sürecindeki kritik bir dönemi temsil etmektedir. Bu dönem, derginin 9. Cilt 2. Sayısı ile 13. Cilt 4. Sayısı arasını kapsamakta olup, özellikle seçilmiştir çünkü 2000 yılından itibaren dergide anahtar kelimelerin zorunlu olarak kullanılmaya başlandığı tespit edilmiştir. Veri setimizi oluştururken, derginin web sitesindeki "Hakemli Yazılar" bölümünde yer alan ve belirtilen tarih aralığında yayımlanmış toplam 94 çalışma incelenmiştir. Ancak, bu çalışmaların sadece 58 tanesinin web sitesinde öz (abstract) bilgileri bulunmaktadır. Çalışmamızın amacı doğrultusunda, yalnızca anahtar kelimesi olmayan ve web sitesinde öz bilgisi yer alan bu 58 makale kapsam dahilinde değerlendirilmiştir. Bu seçim, hem anahtar kelime ataması için uygun bir örneklem oluşturmakta hem de dijital erişilebilirliği artırma hedefimize hizmet etmektedir. İncelenen 58 makale, belirtilen dönemdeki hakemli yazıların yaklaşık %62'sini temsil etmektedir.

Derlenen makalelerin özleri, derginin web sitesinden web kazıma yöntemi ile toplanmıştır. Bu süreçte, özel olarak hazırlanmış bir Python betiği kullanılmıştır. Bu betik, belirlenen URL'lerden öz metinlerini otomatik olarak çekip bir veri tabanında depolamıştır. Web kazıma işlemi sonrasında, sayısallaştırma sürecinde basılı formdan dijital ortama aktarılan metinlerde yazım hataları ve karakter eksiklikleri olup olmadığı manuel kontroller ile belirlenmiştir. Tespit edilen hataların giderilmesi için, açık kaynak kodlu kelime işlemcilerde kullanılan *hunspell* (Ooms, 2024) kütüphanesinin R dili için uyarlanmış paketi kullanılmıştır. Hazırlanan R betiği, Türkçe yazım denetimi gerçekleştirerek belirlenen yazım yanlışlarını tespit etmiş, ardından bu hatalar manuel olarak düzeltilerek dilbilgisi kurallarına uygun özler elde edilmiştir. Bu süreç, veri setimizin doğruluğunu ve güvenilirliğini sağlamak açısından kritik öneme sahiptir.

Veri Analizi

Anahtar kelimeler ataması için literatürde bulunan Türkçe veri setlerinden anahtar kelime çalışmaları incelenmiştir (Önal Süzek, 2017; Erzurumlu ve Akgul, 2023). Bu çalışmaların bilimsel makalelere anahtar kelime ataması yerine genel olarak Türkçe metinlere anahtar kelime ataması gerçekleştirmeyi hedeflediği belirlenmiştir. Bu çalışmalarda benimsenen ortak uygulama adımı olarak Türkçe metinlerin anahtar kelime çıkarımı için ön hazırlık aşamalarında Java tabanlı Türkçe doğal dil işleme paketi olan *Zemberek* (Akın, 2024) kullanıldığı belirlenmiştir. Bu bilgilerden hareketle anahtar kelime çıkarımında kullanılacak özleri işleyebilmek için tasarlanmış tek bir belge üzerinde işlem yapabilen dilden bağımsız bir algoritma olan YAKE (Campos vd., 2018) tercih edilmiştir. YAKE algoritmasının özlerde yer alan metinlere uygulanması için Python dili için hazırlanmış aynı isimli paket kullanılmıştır. Algoritmanın çalışma parametrelerini belirlemek için tekilleştirilen sözcüklerin *Zemberek* tarafından sadece isim (noun) olarak belirlenen sözcükler arasından seçilmiştir.

¹ <http://www.tk.org.tr>

Derginin yazım kılavuzunda "Her makalede en az 3 en fazla 5 anahtar sözcük bulunmalıdır" (Türk Kütüphaneciliği, 2024) ifadesine uygun olarak, maksimum anahtar kelime sayısı olan beş anahtar kelime atanması yapılmıştır. Anahtar kelimeler genellikle tek kelimedenden oluşurken, cümle öbekleri (keyphrase) iki veya daha fazla kelimedenden oluşan anlamlı ifadelerdir. Veri setindeki özlerin uzunlukları birbirinden farklı olduğundan, anahtar kelime atanması için tutarlılığı sağlamak adına tek kelimelik anahtar kelimeler tercih edilmiş, cümle öbekleri (keyphrase) kullanılmamıştır.

Anahtar kelime atama algoritmasının istatistiksel performansını değerlendirmek amacıyla, algoritmanın her bir makale için atadığı anahtar kelimelerin anlamlılığı *Anlamlılık Kontrol Oranı* (AKO) ile ölçülmüştür. AKO, her bir makale için atanmış anahtar kelimelerin, algoritmanın belirlediği parametrelere uygun biçimde oluşturulan anlamlı Türkçe kelimeler olup olmadığını göstermektedir. Bu oran araştırmacılar tarafından Türkçe'nin semantik anlam ve tutarlılık faktörleri dikkate alınarak atanan anahtar kelime seti içerisindeki anlamsız anahtar kelimelerin tüm anahtar kelimelere oranı olarak hesaplanmıştır. Algoritmanın anlamlı kelime çıkarım performansı *Ortalama Mutlak Hata* (OMH) (Mean Absolute Error – MAE) değeri kullanılarak test edilmiştir. Bu ölçüt, algoritmanın her bir makale için çıkardığı anahtar kelimelerin anlamlılık düzeyini ifade eder. Düşük ortalama mutlak hata değeri, algoritmanın yüksek doğrulukla anahtar kelime atadığını gösterirken, yüksek ortalama mutlak hata değeri, algoritmanın performansında iyileştirilmesi gereken alanları işaret etmektedir (Willmott ve Matsuura, 2005). Bu ölçüt farklı algoritmaların veya model parametrelerinin etkinliğini karşılaştırmada ve en uygun modelin seçiminde oldukça yararlıdır.

Algoritmanın bağlamsal geçerliliğini değerlendirmek üzere, bilgi ve belge bilimi alanında mezun olmuş ve aktif olarak bilgi merkezlerinde çalışan üç değerlendirici ile bir geçerlilik testi tasarlanmıştır. Değerlendiricilere makalelere ait özler ve URL adresleri verilmiş, birbirlerinden bağımsız olarak özleri okuyup anahtar kelime atanması yapmaları istenmiştir. Değerlendiricilere, özün yeterli bilgi sağlamadığını düşündükleri durumlarda makalenin tam metnine başvurabilecekleri belirtilmiştir. Bu süreçte, her bir değerlendirici için beşten fazla olmamak üzere ve farklı makaleler için tam metne başvurma durumu gerçekleşmiştir. Bu durum, değerlendiricilerin alan uzmanlığını ve örtük bilgilerini sürece dahil etmeleri açısından önemli görülmüş, ancak sınırlı sayıda olduğu için ayrı bir değerlendirmeye tabi tutulmamıştır. Üç değerlendiricinin atadığı anahtar kelimeler kategorik vektörlere dönüştürülmüş ve bu kelimeler arasındaki uyum, değerlendiriciler arası uyum AC1 (Gwet, 2008) katsayısına göre analiz edilmiştir. Bu katsayı, farklı özellik yaygınlığı senaryolarında dengeli şans anlaşmalarını dikkate alarak yanlılık riskini minimize eden bir istatistiksel ölçümdür. Özellikle pi, kappa (Fleiss ve Cohen, 1973) ve G-indeksi gibi geleneksel anlaşma ölçütlerinin paradoksal sonuçlar verebileceği durumlarda, AC1 katsayısı daha kararlı ve tutarlı sonuçlar sunmaktadır. Değerlendiriciler arasındaki uyum analizi için *irrCAC* (Gwet, 2019) R paketi kullanılmıştır. Daha sonra, algoritma tarafından atanan anahtar kelimeler dördüncü bir değerlendirici olarak kabul edilerek, bu kelimelerin uyum oranındaki değişiklikler şansa bağlı olarak test edilmiştir.

Tablo 1

Bağlamsal Geçerlilik Sonuçları Tablosu

	Anlaşma Olasılığı (Pa)	Şans Eseri Anlaşma Olasılığı (Pe)	Katsayı Katsayı Değeri	Katsayı Standart Hatası	Güven Aralığı	p- Değeri	Ağırlıklandırma
Üç Değerlendirici	0	0,005747126	0,00578	0	(-0,006, -0,006)	0	Ağırlıksız
Üç Değerlendirici + Algoritma	0	0,004310345	0,00433	0	(-0,004, -0,004)	0	Ağırlıksız

Hesaplanan AC_1 katsayısı (-0,00578) ise, şans düzeyinin altında bir anlaşma olduğunu ve hatta negatif değerlerle beklenen şans düzeyinden daha az anlaşma sağlandığını ortaya koymaktadır. Algoritma dördüncü bir değerlendirici olarak dahil edildiğinde, şans eseri anlaşma olasılığı bir miktar azalarak %0,00431'e düşmüştür. AC_1 katsayısı da -0,00433 olarak azalmış, bu durum makalelere atanan anahtar kelimeler üzerindeki anlaşmanın şans düzeyinin altında kaldığını tekrar göstermiştir. Bulgular, algoritmanın atadığı anahtar kelimelerin, değerlendiriciler arası anlaşmayı artırmak yerine azalttığına işaret etmektedir. Algoritmanın değerlendiricilerle olan uyumsuzluğu artmış, bu da algoritmanın anahtar kelime atamasının belirli bir veri kaynağına bağımlı olduğunu ve bu kaynağın içerdiği verilerin, alanın gerektirdiği bağlamsal ve mesleki derinliği tam olarak yansıtamadığını düşündürülebilir. Uzman değerlendiriciler, mesleki deneyimlerini ve alanın bağlamsal ihtiyaçlarını kullanarak daha kapsamlı ve alana özgü anahtar kelimeler atayabilmektedir. Elde edilen bulgular anahtar kelime atama süreçlerinde algoritmik yaklaşımların kullanılan veri kaynağına dayalı öğrenme süreçlerinden elde ettiği bilgilerle sınırlı kalabileceğini göstermektedir.

Sınırlılıklar

Güncel literatürde sıkça başvurulan denetimsiz anahtar kelime çıkarım yöntemlerinin birtakım sınırlılıkları bulunmaktadır. Özellikle bu yöntemlerde, anahtar kelime sayısı ve bir n-gramın (ardışık n adet kelimenin oluşturduğu dizilim) anahtar kelime olarak tanımlanabilecek maksimum uzunluğu, araştırmacının tercihlerine bağlı olarak belirlenmektedir. Ancak anahtar kelime sayısı ve anahtar sözcük öbeği uzunluğu gibi faktörler, dergilerin veya yayıncıların politikalarına ve araştırma disiplininin geleneksel kriterlerine göre değişiklik gösterebilmektedir. Örneğin, bazı dergiler veya yayıncılar, makalelerde kullanılacak anahtar kelime sayısını belirli bir aralıkta sınırlandırabilirken (örneğin 3-5 anahtar kelime), diğerleri daha esnek bir yaklaşım benimseyebilir. Benzer şekilde, farklı araştırma alanlarında, anahtar kelime olarak kabul edilebilecek ifadelerin uzunluğu da değişebilir. Bazı disiplinlerde araştırma alanına özgü kontrollü anahtar kelimeler yaygınken, diğerlerinde daha uzun anahtar sözcük

öbekleri tercih edilebilir. Bu nedenle, denetimsiz anahtar kelime çıkarım yöntemlerini kullanırken, araştırmacıların bu faktörleri göz önünde bulundurması ve çalışmanın bağlamına uygun seçimler yapması önemlidir. Ancak bu tercihlerin sübjektif olması ve standart bir yaklaşımın bulunmaması, yöntemlerin sınırlılıklarından biri olarak kabul edilebilir. Ayrıca, bu yöntemler genellikle kaynak metinden doğrudan anahtar kelime çıkarmak için tasarlanmıştır (Morozov vd., 2023). Oysa güncel akademik metinleri tanımlamak için kullanılan anahtar kelimeler, sıklıkla genelleştirilmiş kavramlar veya metinde geçen kavramların eş anlamlıları şeklinde ortaya çıkmaktadır (J. Zhang vd., 2016). Bu nedenle, algoritmik yaklaşımlar, kaynak metinde açıkça bulunmayan ancak araştırmayı tanımlayan ve kullanıcıların belirleyebileceği anahtar kelimeleri tespit etme konusunda yetersiz kalabilmektedir. Bu durum, algoritmik yaklaşımlarla atanan anahtar kelimelerin çalışmayı tanımlama konusunda birtakım sınırlılıklara sahip olduğunu göstermektedir.

Tartışma ve Sonuç

Çalışma kapsamında, anahtar kelimesi bulunmayan araştırma makalelerine Türkçe özetler kullanılarak anahtar kelime atama yapılması hedeflenmiştir. Dijital ortamda arşivlenmiş ve anahtar kelime eksikliği nedeniyle akademik arama motorları tarafından etkin bir biçimde erişilemeyen araştırma makalelerinin üst verilerinin zenginleştirilmesi amaçlanmıştır. Üst veriler, web içeriğini verimli bir şekilde tanımlamak, kategorize etmek ve erişilebilir hale getirmek için belge bilgilerini yapılandırarak erişimi artırır (Hawking ve Zobel, 2007). Ancak, doğru ve tutarlı üst verilerin oluşturulması her zaman zordur ve bu durum bilgi erişim etkinliğini doğrudan etkiler. Web içeriğinin dinamik doğası ve kullanıcıların farklılaşan ihtiyaçları, özellikle geçmişte dijital olarak oluşturulmayıp sonradan dijital ortama aktarılan belgeler için etkili üst verilerin oluşturulmasını ve sürdürülebilirliğini daha da zorlaştırmaktadır. İnternet bilgi kaynaklarını ve içeriklerini tanımlamak için standartlaştırılmış bir yaklaşım olarak üst veri kullanımı, belgelerin zenginleştirilmesi amacıyla sıklıkla başvurulan bir yöntemdir. Bu yöntem, internetin genişliği, dinamizmi ve çeşitliliğinden kaynaklanan dezavantajları en aza indirmek için kullanılır (Mahdi Taheri vd., 2014).

Önerilen yöntem sayesinde özellikle basılı formatta yayımlandıktan sonra dijitalleştirilerek web ortamında sunulan arşivlenmiş çalışmaların modern akademik arama motorlarının kullandığı yöntemlere uygun hale gelmesi sağlanabilir. Üst verilerin arama performansını artırmadaki etkinliği oldukça belirgin fark yaratmaktadır. Anahtar kelimeleri üst veri olarak eklemek, akademik arama motorlarının içeriği daha etkili bir şekilde tanımlamasına ve sınıflandırmasına yardımcı olarak akademik çalışmaların görünürlüğünü önemli ölçüde artırır (Beel vd., 2010). Çünkü akademik bir arama motoru, web’de karşılaştığı bir PDF’nin akademik bir makaleyi temsil edip etmediğini ve hangi kategoriye ait olduğunu belirlemek için varsa üst verilerini kullanır. Anahtar kelimelerin bu süreçte üst verilere dahil edilmesi yalnızca çalışmanın görünürlüğünü değil, aynı zamanda arama sorgularındaki alaka düzeyini artırarak arama amaçlarıyla daha yakından uyumlu sonuçları getirmesi sağlar (An ve Jung, 2021). Çalışma kapsamında anahtar kelime içermeyen hakemli makalelere anahtar kelime ataması yaparak üst veri eksikliğinden kaynaklı dezavantajların bir ölçüde giderilmesi sağlanmıştır. Algoritmanın atamış olduğu anahtar kelimelerin anlamlılık seviyeleri ise düşük OMH skoru

açısından bakıldığında kabul edilebilir bir seviyede olduğu ortaya koymaktadır. Akademik arama motoru optimizasyonu çerçevesinde ele alınan bu durum, geçmiş dönemdeki bilimsel makalelerin arama motorlarında daha üst sıralarda yer almasını ve Türkiye'de üretilen akademik bilginin ilgili araştırmacılar tarafından daha kolay erişilebilir hale gelmesini sağlayacaktır.

Doğru anahtar sözcük seçiminin bilgi alımını iyileştirdiğine dair literatürde destekleyici çalışmalar bulunmaktadır. Hartley ve Kostoff (2003), anahtar kelimelerin okuyucuların ilgili makaleleri bulmalarına ve indeksleyicilerin benzer konuları gruplandırmalarına yardımcı olduğunu belirtmiştir. Lu ve Kipp. (2014) yazar tarafından seçilen anahtar kelimelerin makale keşfinde önemli rol oynadığını göstermiş, bu kelimelerin bilimsel makale arama ve öneri sistemlerinde etkili olduğunu ortaya koymuştur. Ayrıca, Chamorro-Padial ve Rodríguez-Sánchez (2023) anahtar kelimelerin popülerliği ve özgünlüğü arasındaki dengenin makalelerin görünürlüğünü etkilediğini göstermiştir. Bu çalışmalar, doğru seçilmiş anahtar kelimelerin bilgi erişimini ve makalelerin keşfedilme olasılığını artırdığını desteklemektedir. Benzer şekilde, makale özlerinin de çalışmanın içeriğini en iyi biçimde yansıtması beklenmektedir (Hartley ve Cabanac, 2017). Hartley (2016) akademik makaleler için özün makalenin genel bir özetini sunmanın yanı sıra, makalenin bazı karakteristik özelliklerini de vurgulaması gerektiğini belirtmiştir. Bu bağlamda hem doğru seçilmiş anahtar kelimeler hem de iyi hazırlanmış özler, bir makalenin bilimsel literatürde keşfedilmesi, anlaşılması ve yayılması açısından kritik öneme sahiptir.

Farklı araştırma disiplinlerinde makalelerin tercih ettikleri öz biçimleri değişiklik göstermektedir. Makale özleri doğası gereği günlük konuşma dilinden, argo ifadelerden, kısaltmalardan ve deyimse fiillerden arındırılmış biçimde, resmi ve teknik kavramları içerecek şekilde yazılmaktadır (Kaplan vd., 1994, s. 406). Ayrıca biçimsel olarak kullanılan özler yapılandırılmış veya yapılandırılmamış olarak ikiye ayrılırken bilgi aktarma yaklaşımları olarak Gösterge, Bilgilendirici ve Bilgilendirici-Gösterge özler olarak üçe ayrılmaktadır. Veri setinde yer alan makale özlerinin yapılandırılmamış özler olarak tasarlandığı görülmektedir. Akademik literatürde kullanılan öz yazma biçimlerine göre incelenen özleri ağırlıklı olarak yapılandırılmamış bilgilendirici-gösterge özleri olarak tasarlandığı görülmektedir (Hyland ve Tse, 2005; Cutting, 2012). Gösterge özleri, ayrıntılı ayrıntılara girmeden makalenin içeriğine genel bir bakış sunan ve böylece okuyucuyu araştırmanın ampirik bulguları veya metodolojik özelliklerinden ziyade tematik kapsama yönlendiren bir kılavuz görevi görür. Bilgilendirici özler ise amaç, metodoloji, sonuçlar ve sonuçları kapsayan araştırmanın kritik ayrıntılarını rapor etmeyi amaçlar ve böylece çalışmanın katkılarının ve sonuçlarının kapsamlı bir özetini sağlamaktadır (C. Zhang ve Liu, 2011). Araştırma sonucunda incelenen makale özlerinde her iki yaklaşımın karması olan bilgilendirici-gösterge özlerin tercih edildiği belirlenmiştir.

Araştırmada tespit edilen, algoritmanın atadığı anahtar kelimeler ile uzman anahtar kelimeler arasındaki uyum eksikliğinin bir diğer nedeni, KB alanındaki akademik makalelerin özlerinde tespit edilen standart yazım eksikliğidir. Bitri ve Keseroğlu (2015), Türkçe KB alanındaki iki önemli dergide yayımlanmış özleri inceledikleri çalışmada, akademik yazım ölçütlerine yeterince özen gösterilmediğini tespit etmişlerdir. Çalışmaya konu olan TK dergisi ise 35. sayıdan itibaren yapılandırılmış öz sistemine geçiş yaparak bu durumu iyileştirme yönünde adım atmıştır. Bu tarihsel bağlam, algoritmanın çıkardığı anahtar kelimelerin neden

alan için çok spesifik olmadığını açıklamaktadır. Elde edilen bulgular, o dönemdeki akademik yazım pratiklerini yansıtmakta ve çalışmamızın sonuçlarını tarihsel perspektiften değerlendirmenin önemini vurgulamaktadır. "Kütüphane" ve "bilgi" gibi kelimelerin sıklıkla öne çıkması, bu dönemde yazılan makalelerin genel olarak alanın temel kavramlarına odaklandığını göstermektedir. Bu durum, aynı zamanda, o dönemde bilgi ve belge yönetimi alanında kullanılan terminolojinin zaman içinde nasıl evrildiğine dair ipuçları da sunmaktadır.

Çalışma kapsamında, Türkçe özlerin anahtar kelime atama sürecinde kullanılması özellikle tercih edilmiştir. Diller arasında çeviri yaparken nadiren mükemmel eşdeğerlik sağlanabileceği uzun zamandır kabul edilmektedir (Koller, 1995). Farklı diller arasında çeviri yapılırken baskın ve çevresel dil ayrımı gözetilmektedir. Akademik makalelerde baskın dil olarak kullanılan İngilizce için üretilen bilginin niteliğine bağlı olarak metin çeşitli değişikliklere uğramaktadır. Çeviri kuramlarına göre, metnin çevirisi "yerlileştirme" ve "yabancılaştırma" yaklaşımlarına göre şekillenmektedir. Yerlileştirme yaklaşımı, hedef kültürün kökenlerine tam olarak uyum sağlayarak çeviriyi görünmez kılmayı ve orijinal metnin kaynak kültürden izlerini maskeleymeyi amaçlar. Öte yandan, yabancılaştırma yaklaşımı, hedef dildeki baskın normlara karşı çıkmak pahasına, kaynak kültürün yapılarını, retorik biçimini ve üslubunu taklit etmeye çalışır (Chmutina vd., 2020). Bu bağlamda, çalışmada kullanılan yöntem, veri kaynağı olarak kullanılan kaynak özleri temel alan, ancak dilden bağımsız bir biçimde anahtar kelime atama işlemi yapabilen bir algoritma olarak tasarlanmıştır. Böylece, dile özgü üretilen bilgi ve kavramların farklı bir dile çevrilmesi esnasında karşılaşılan anlamsal eksiklik ve hataların önüne geçilmesi hedeflenmiştir. Çalışmada elde edilen anlamlılık kontrol oranları ve OMH değerlerinin düşük oluşu, algoritmanın seçilen kriterleri yerine getirmede oldukça başarılı olduğunu ortaya koymaktadır.

Çalışma sonucunda, algoritmanın anahtar kelime atama sürecinde seçtiği anahtar kelimeler ile meslek profesyonellerinin seçmiş olduğu anahtar kelimeler arasında düşük seviyede bir uyum tespit edilmiştir. Bu durum, algoritmanın ve meslek profesyonellerinin farklı bilgi kaynaklarına dayanmasından kaynaklanmaktadır. Algoritma, anahtar kelimeleri belirlerken yalnızca makale özlerindeki metinsel içeriği kullanmakta ve özlerin dilbilimsel özelliklerini dikkate almaktadır. Meslek profesyonelleri ise öncelikle özlere odaklanmakla birlikte, gerekli gördükleri durumlarda (her bir değerlendirici için en fazla beş makale olmak üzere) makalenin tamamını inceleme ve ilgili alana dair uzmanlık bilgilerini kullanma imkanına sahip olmuşlardır. Bu yaklaşım farkı, meslek profesyonellerinin alan uzmanlığını ve örtük bilgilerini sürece dahil etmelerine olanak sağlarken, algoritmanın sabit bir veri kaynağına (öz metinleri) bağlı kalmasına neden olmuştur. Dolayısıyla, algoritmanın kaynak olarak kullandığı metinler, meslek profesyonellerinin özlerden ve gerektiğinde makalenin bütününden ulaştıkları sonuçlara göre geliştirdikleri kavramlar ile tam olarak uyumlu olmayabilir. Bu durum, insan değerlendiricilerin bağlamsal anlayış ve alan bilgisinin, otomatik anahtar kelime atama süreçlerinde hala önemli fark yaratabileceğini göstermektedir.

Bu çalışmada kullanılan anahtar kelime atama yöntemi, etiketlenmiş veri seti gereksinimini ortadan kaldırarak daha geniş bir uygulama alanı sunmaktadır. Bu sayede, farklı araştırma alanları için arşivlenmiş makalelerin dijital ortamda üst verilerinin zenginleştirilmesi yoluyla erişimi artırmak mümkün olabilecektir. Çalışma, ayrıca İngilizce dışındaki anahtar kelime atama süreçlerinde kullanılan bir yöntem olarak literatüre katkı sunmaktadır. Aynı

zamanda, KB alanına özgü makale özlerini kullanması nedeniyle bu alana yönelik ilk çalışmalardan biri olarak değerlendirilebilir.

Geliştirilen algoritmanın uzmanlar tarafından kabul edilen ve tercih edilen anahtar kelimelerle daha uyumlu hale getirilmesi için, algoritmayı eğitmek üzere kullanılan verilerin bu uzmanların perspektifleri ve tercihleri dikkate alınarak hazırlanması önerilmektedir. Ayrıca, dergi politika belgelerinde, gelecekte bilgi erişim için yaygın kullanılacak algoritmik yaklaşımları dikkate alarak, algoritmik analize uyumlu yapılandırılmış öz kullanımının ve öz metnin uzunluğunun artırılması tavsiye edilmektedir.

İzin ve Katkı Bildirimleri

Etik Kurul İzni

Yazar makale için etik kurul onayı gerekmediğini beyan etmiştir.

Yazarlık Katkısı

Makale tek yazarlıdır.

Kaynakça

- Akın, A. A. (2024). Zemberek-NLP: Natural Language Processing tools for Turkish (Versiyon 0.17.1) [Java; Windows]. <https://github.com/ahmetaa/zemberek-nlp>
- Aldemir, A., ve Oğuz, E. S. (2006). Sayısal dijital kültürün korunması: Web arşivleme. *Türk Kütüphaneciliği*, 20(3), 283-312. <https://dergipark.org.tr/tr/pub/tk/issue/48940/624280>
- Alharbi, L. M., ve Swales, J. M. (2011). Arabic and English abstracts in bilingual language science journals: Same or different? *Languages in Contrast*, 11(1), 70-86. <https://doi.org/10.1075/lic.11.1.06alh>
- Allen, E. J., ve Weber, R. K. (2015). An exploration of indexed and non-indexed open access journals: Identifying metadata coding variations. *Journal of Web Librarianship*, 9(2-3), 65-84. <https://doi.org/10.1080/19322909.2015.1020185>
- An, S., ve Jung, J. J. (2021). A heuristic approach on metadata recommendation for search engine optimization. *Concurrency and Computation: Practice and Experience*, 33(3), e5407. <https://doi.org/10.1002/cpe.5407>
- Atanassova, I., Bertin, M., ve Larivière, V. (2016). On the composition of scientific abstracts. *Journal of Documentation*, 72(4), 636-647. <https://doi.org/10.1108/JDOC-09-2015-0111>
- Atılgan, D. (2002). Türk Kütüphaneciliği dergisinin yayın yaşamındaki yeri. *Türk Kütüphaneciliği*, 16(3), 248-256. <http://tk.org.tr/index.php/TK/article/view/918/911>
- Bayter, M. (2009). Türkçe web sitelerinin kataloglanmasının önemi. *Türk Kütüphaneciliği*, 23(3), 563-585. <https://dergipark.org.tr/en/pub/tk/issue/48915/623745>
- Beel, J., Gipp, B., ve Wilde, E. (2010). Academic Search Engine Optimization (ASEO): Optimizing Scholarly Literature for Google Scholar & Co. *Journal of Scholarly Publishing*, 41(2), 176-190. <https://doi.org/10.3138/jsp.41.2.176>

- Beliga, S. (2014). Keyword extraction: A review of methods and approaches. University of Rijeka, Department of Informatics, Rijeka, 1(9). http://langnet.uniri.hr/papers/beliga/Beliga_KeywordExtraction_a_review_of_methods_and_a_pproaches.pdf
- Beliga, S., Meštrović, A., ve Martinčić-Ipšić, S. (2016). Selectivity-based keyword extraction method. *International Journal on Semantic Web and Information Systems*, 12(3), 1-26. <https://doi.org/10.4018/IJSWIS.2016070101>
- Bhowmik, R. (2008). Keyword extraction from abstracts and titles. *IEEE SoutheastCon 2008*, 610-617. <https://doi.org/10.1109/SECON.2008.4494366>
- Binici, K. (2019). Makine öğrenmesi yaklaşımıyla e-belgelere standart dosya plan numaralarının otomatik olarak atanması üzerine bir çalışma. *Bilgi Yönetimi*, 2(2), 116-126. <https://doi.org/10.33721/by.654464>
- Bitri, E., ve Keseroğlu, H. S. (2015). Türk Kütüphaneciliği ve Bilgi Dünyası dergilerinin özlerine eleştirel bir bakış. *Türk Kütüphaneciliği*, 29(2), 241-257. <https://dergipark.org.tr/pub/tk/issue/48742/620117>
- Blank, I., Rokach, L., ve Shani, G. (2016). Leveraging metadata to recommend keywords for academic papers. *Journal of the Association for Information Science and Technology*, 67(12), 3073-3091. <https://doi.org/10.1002/asi.23571>
- Brienza, C. (2012). Opening the wrong gate? The academic spring and scholarly publishing in the humanities and social sciences. *Publishing Research Quarterly*, 28, 159-171. <https://doi.org/10.1007/s12109-012-9272-5>
- Campos, R., Mangaravite, V., Pasquali, A., Jorge, A. M., Nunes, C., ve Jatowt, A. (2018). YAKE! Collection-Independent Automatic Keyword Extractor. İçinde G. Pasi, B. Piwowarski, L. Azzopardi, ve A. Hanbury (Ed.), *Advances in Information Retrieval* (C. 10772, ss. 806-810). Springer International Publishing. https://doi.org/10.1007/978-3-319-76941-7_80
- Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., ve Jatowt, A. (2020). YAKE! Keyword extraction from single documents using multiple local features. *Information Sciences*, 509, 257-289. <https://doi.org/10.1016/j.ins.2019.09.013>
- Chamorro-Padial, J., ve Rodríguez-Sánchez, R. (2023). Attention–Survival Score: A Metric to Choose Better Keywords and Improve Visibility of Information. *Algorithms*, 16(4), 196. <https://doi.org/10.3390/a16040196>
- Chi, Y., Zhu, J., Huang, L., ve Xu, H. (2019). Concepts recommendation for searching scientific papers. *Cluster Computing*, 22(S4), 8669-8675. <https://doi.org/10.1007/s10586-018-1937-1>
- Chmutina, K., Sadler, N., Von Meding, J., ve Abukhalaf, A. H. I. (2020). Lost (and found?) in translation: Key terminology in disaster studies. *Disaster Prevention and Management: An International Journal*, 30(2), 149-162. <https://doi.org/10.1108/DPM-07-2020-0232>
- Conway, P. (2015). Digital transformations and the archival nature of surrogates. *Archival Science*, 15, 51-69. <https://doi.org/10.1007/s10502-014-9219-z>
- Cutting, D. J. (2012). Vague language in conference abstracts. *Journal of English for Academic Purposes*, 11(4), 283-293. <https://doi.org/10.1016/j.jeap.2012.05.004>
- Çapkın, Ç. (2009). Türk kütüphaneciler derneği web hizmetleri: 60. Yıla doğru yeni ufuklara. *Türk Kütüphaneciliği*, 23(2), 395-398. <https://dergipark.org.tr/en/download/article-file/815027>
- Çapkın, Ç. (2014). Bilimsel elektronik dergilerin kullanımı: Türk Kütüphaneciliği dergisinin işlem günlüğü analizi. *Türk Kütüphaneciliği*, 28(3), 308-319. <https://dergipark.org.tr/pub/tk/issue/48745/620207>

- Daudaravicius, V. (2016). A Framework for Keyphrase Extraction from Scientific Journals. İçinde A. González-Beltrán, F. Osborne, ve S. Peroni (Ed.), *Semantics, Analytics, Visualization. Enhancing Scholarly Data* (C. 9792, ss. 51-66). Springer International Publishing. https://doi.org/10.1007/978-3-319-53637-8_7
- Dişli, M., ve Tonta, Y. (2023). Veri olarak kültürel miras koleksiyonları. *Türk Kütüphaneciliği*. <https://doi.org/10.24146/tk.1317445>
- Dynich, A., ve Wang, Y. (2017). Analysis of novelty of a scientific text as a basis for assessment of efficiency of scientific activities. *Journal of Organizational Change Management*, 30(5), 668-682. <https://doi.org/10.1108/JOCM-10-2016-0226>
- Ercan, G., ve Cicekli, I. (2007). Using lexical chains for keyword extraction. *Information Processing & Management*, 43(6), 1705-1714. <https://doi.org/10.1016/j.ipm.2007.01.015>
- Erzurumlu, H. Y., ve Akgul, Y. S. (2023). Adaptive keyword extraction service for Turkish. İçinde K. Arai (Ed.), *Intelligent Systems and Applications* (C. 543, ss. 495-506). Springer International Publishing. https://doi.org/10.1007/978-3-031-16078-3_34
- Escudero, M. P., ve Swales, J. M. (2011). Tracing convergence and divergence in pairs of Spanish and English research article abstracts. *Ibérica*, 21, 49-70. <https://www.redalyc.org/pdf/2870/287023883004.pdf>
- Fleiss, J. L., ve Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33(3), 613-619. <https://doi.org/10.1177/001316447303300309>
- Gil-Leiva, I., ve Alonso-Arroyo, A. (2007). Keywords given by authors of scientific articles in database descriptors. *Journal of the American Society for Information Science and Technology*, 58(8), 1175-1187. <https://doi.org/10.1002/asi.20595>
- Gillis, R., Betz, S., Chan, J., Hatherill, J., Jay, S., Kosavic, A., Pritt, A., MacFarland, D., Maistrovskaya, M., Moore, A., Sprout, B., ve Stranack, K. (2019). Getting Found, Staying Found, Increasing Impact. *OJS på dansk*, 10(9). <https://doi.org/10.7146/ojssb.v10i9.116299>
- Grey, A., ve Hurko, C. R. (2012). So you think you're an expert: Keyword searching vs. Controlled subject headings. *Codex: The Journal of the Louisiana Chapter of the ACRL*, 1(4), 15-26.
- Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, 61(1), 29-48. <https://doi.org/10.1348/000711006X126600>
- Gwet, K. L. (2019). irrCAC: Computing Chance-Corrected Agreement Coefficients (CAC) (Versiyon 1.0) [R; Windows]. <https://CRAN.R-project.org/package=irrCAC>
- Hahs-Vaughn, D. L., ve Onwuegbuzie, A. J. (2010). Quality of abstracts in articles submitted to a scholarly journal: A mixed methods case study of the journal *Research in the Schools*. *Library & Information Science Research*, 32(1), 53-61. <https://doi.org/10.1016/j.lisr.2009.08.004>
- Hartley, J. (2016). What's new in abstracts of science articles? *Journal of the Medical Library Association*, 104(3), 235-236. <https://doi.org/10.3163/1536-5050.104.3.011>
- Hartley, J., ve Cabanac, G. (2017). Thirteen ways to write an abstract. *Publications*, 5(2), 11. <https://doi.org/10.3390/publications5020011>
- Hartley, J., ve Kostoff, R. N. (2003). How Useful are 'Key Words' in Scientific Journals? *Journal of Information Science*, 29(5), 433-438. <https://doi.org/10.1177/01655515030295008>
- Hawking, D., ve Zobel, J. (2007). Does topic metadata help with Web search? *Journal of the American Society for Information Science and Technology*, 58(5), 613-628. <https://doi.org/10.1002/asi.20548>

- Honey, S. L. (2005). Preservation of electronic scholarly publishing: An analysis of three approaches. *Portal: Libraries and the Academy*, 5(1), 59-75. <https://doi.org/10.1353/pla.2005.0005>
- Howcroft, G. (2007). A Beginner's Guide to Metadata and Keywords. *Editors' Bulletin*, 3(3), 75-77. <https://doi.org/10.1080/17521740701788437>
- Hyland, K., ve Tse, P. (2005). Hooking the reader: A corpus study of evaluative that in abstracts. *English for Specific Purposes*, 24(2), 123-139. <https://doi.org/10.1016/j.esp.2004.02.002>
- Jamar, N., Šauperl, A., ve Bawden, D. (2014). The components of abstracts: The logical structure of abstracts in the areas of materials science and technology and of library and information science. *New Library World*, 115(1/2), 15-33. <https://doi.org/10.1108/NLW-09-2013-0069>
- Kaplan, R. B., Cantor, S., Hagstrom, C., Kamhi-Stein, L. D., Shiotani, Y., ve Zimmerman, C. B. (1994). On abstract writing. *Text - Interdisciplinary Journal for the Study of Discourse*, 14(3). <https://doi.org/10.1515/text.1.1994.14.3.401>
- Kayaoğlu, H. D. (2007). Elektronik dergi kullanımı ile bilim dalı ilişkisi: Betimleyici bir inceleme. *Türk Kütüphaneciliği*, 21(4), 440-458. <https://dergipark.org.tr/tr/pub/tk/issue/48936/624246>
- Koller, W. (1995). The concept of equivalence and the object of translation studies. *Target. International Journal of Translation Studies*, 7(2), 191-222. <https://doi.org/10.1075/target.7.2.02kol>
- Küçük, M. E., ve Alır, G. (2003). Dijital koruma (arşivleme) stratejileri ve bazı uygulama örnekleri. *Türk Kütüphaneciliği*, 17(4), 340-356. <http://www.tk.org.tr/index.php/TK/article/view/299/291>
- Li, S.-Q., Du, S.-M., ve Xing, X.-Z. (2017). A Keyword Extraction Method for Chinese Scientific Abstracts. *Proceedings of the 2017 International Conference on Wireless Communications, Networking and Applications*, 133-137. <https://doi.org/10.1145/3180496.3180620>
- Liu, G., ve Zhai, Z. (2011). Research on keywords Extraction of Chinese documents based on TEXT-NET. *2011 International Conference on Electric Information and Control Engineering*, 6074-6077. <https://doi.org/10.1109/ICEICE.2011.5777961>
- Lu, K., ve Kipp, M. E. I. (2014). Understanding the retrieval effectiveness of collaborative tags and author keywords in different retrieval environments: An experimental study on medical collections. *Journal of the Association for Information Science and Technology*, 65(3), 483-500. <https://doi.org/10.1002/asi.22985>
- Lu, W., Huang, S., Yang, J., Bu, Y., Cheng, Q., ve Huang, Y. (2021). Detecting research topic trends by author-defined keyword frequency. *Information Processing & Management*, 58(4), 102594. <https://doi.org/10.1016/j.ipm.2021.102594>
- Mahdi Taheri, S., Hariri, N., ve Ramatollah Fattahi, S. (2014). Does discarding XML declarations and changing file extensions improve the indexability and visibility of metadata tag names in web search engines? *Journal of Information Science*, 40(6), 796-805. <https://doi.org/10.1177/0165551514550139>
- Mihalcea, R., ve Tarau, P. (2004). TextRank: Bringing order into text. İçinde D. Lin ve D. Wu (Ed.), *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (ss. 404-411). Association for Computational Linguistics. <https://web.archive.org/web/20240327095740/https://web.eecs.umich.edu/~mihalcea/papers/mihalcea.emnlp04.pdf>
- Mohamed, K. A. F. (2006). The impact of metadata in web resources discovering. *Online Information Review*, 30(2), 155-167. <https://doi.org/10.1108/14684520610659184>
- Mohseni, M., ve Faili, H. (2020). Title Generation and Keyphrase Extraction from Persian Scientific Texts. *2020 25th International Computer Conference, Computer Society of Iran (CSICC)*, 1-6. <https://doi.org/10.1109/CSICC49403.2020.9050113>

- Morozov, D. A., Glazkova, A. V., Tyutyulnikov, M. A., ve Iomdin, B. L. (2023). Keyphrase Generation for Abstracts of the Russian-Language Scientific Articles. *NSU Vestnik. Series: Linguistics and Intercultural Communication*, 21(1), 54-66. <https://doi.org/10.25205/1818-7935-2023-21-1-54-66>
- Moulaison, H. L., ve Million, A. J. (2015). E-publishing in libraries: The [digital] preservation imperative. *OCLC Systems & Services: International digital library perspectives*, 31(2), 87-98. <https://doi.org/10.1108/OCLC-02-2014-0009>
- Ooms, J. (2024). hunspell: High-Performance Stemmer, Tokenizer, and Spell Checker (Versiyon 3.03) [R; Windows]. <https://docs.ropensci.org/hunspell/> <https://ropensci.r-universe.dev/hunspell>
- Önal Süzek, T. (2017). Using latent semantic analysis for automated keyword extraction from large document corpora. *Turkish Journal of Electrical Engineering & Computer Sciences*, 25, 1784-1794. <https://doi.org/10.3906/elk-1511-203>
- Papagiannopoulou, E., ve Tsoumakas, G. (2020). A review of keyphrase extraction. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(2), e1339. <https://doi.org/10.1002/widm.1339>
- Rose, S., Engel, D., Cramer, N., ve Cowley, W. (2010). Automatic Keyword Extraction from Individual Documents. İçinde M. W. Berry ve J. Kogan (Ed.), *Text Mining* (1. bs, ss. 1-20). Wiley. <https://doi.org/10.1002/9780470689646.ch1>
- Šaupperl, A., Klasinc, J., ve Lužar, S. (2008). Components of abstracts: Logical structure of scholarly abstracts in pharmacology, sociology, and linguistics and literature. *Journal of the American Society for Information Science and Technology*, 59(9), 1420-1432. <https://doi.org/10.1002/asi.20858>
- Siddiqi, S., ve Sharan, A. (2015). Keyword and Keyphrase Extraction Techniques: A Literature Review. *International Journal of Computer Applications*, 109(2), 18-23. <https://doi.org/10.5120/19161-0607>
- Terra, A. L., Agustín Lacruz, C., Bernardes, Ó., Fujita, M. S. L., ve Bueno De La Fuente, G. (2021). Subject-access metadata on ETD supplied by authors: A case study about keywords, titles and abstracts in a Brazilian academic repository. *The Journal of Academic Librarianship*, 47(1), 102268. <https://doi.org/10.1016/j.acalib.2020.102268>
- Tonta, Y. (2002). Türk Kütüphaneciliği Dergisi, 1987-2001. *Türk Kütüphaneciliği*, 16(3), 282-320. <http://www.tk.org.tr/index.php/TK/article/view/929/922>
- Turney, P. D. (2002). Learning to extract keyphrases from text. arXiv preprint cs/0212013.
- Türk Kütüphaneciliği. (2024). Yazım Kuralları. İçinde Türk Kütüphaneciliği. <https://dergipark.org.tr/tr/pub/tk/writing-rules>
- Willmott, C., ve Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30, 79-82. <https://doi.org/10.3354/cr030079>
- Witten, I. H., Paynter, G. W., Frank, E., Gutwin, C., ve Nevill-Manning, C. G. (1999). KEA: Practical automatic keyphrase extraction. *Proceedings of the Fourth ACM Conference on Digital Libraries*, 254-255. <https://doi.org/10.1145/313238.313437>
- Yılmaz, Ö., ve Memişoğlu, S. P. (2019). Yükseköğretim kurumlarında akademik performans değerlendirmeye ilişkin akademisyen görüşleri. *Yükseköğretim ve Bilim Dergisi*, 3, 542-554. <https://dergipark.org.tr/tr/pub/higheredusci/issue/61500/918358>
- Zhang, C., ve Liu, X. (2011). Review of James Hartley's research on structured abstracts. *Journal of Information Science*, 37(6), 570-576. <https://doi.org/10.1177/0165551511420217>

- Zhang, C., Wang, H., Liu, Y., Wu, D., Liao, Y., ve Wang, B. (2008). Automatic keyword extraction from documents using conditional random fields. *Journal of Computational Information Systems*, 4(3), 1169-1180. <https://core.ac.uk/download/pdf/290479356.pdf>
- Zhang, C., Zhao, L., Zhao, M., ve Zhang, Y. (2022). Enhancing keyphrase extraction from academic articles with their reference information. *Scientometrics*, 127(2), 703-731. <https://doi.org/10.1007/s11192-021-04230-4>
- Zhang, J., ve Dimitroff, A. (2005). The impact of metadata implementation on webpage visibility in search engine results (Part II). *Information Processing & Management*, 41(3), 691-715. <https://doi.org/10.1016/j.ipm.2003.12.002>
- Zhang, J., Yu, Q., Zheng, F., Long, C., Lu, Z., ve Duan, Z. (2016). Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. *Journal of the Association for Information Science and Technology*, 67(4), 967-972. <https://doi.org/10.1002/asi.23437>