

Geçici Bilgi İhtiyacının Giderilme Sürecinde Kullanıcı Okuma Davranışlarının İncelenmesi

Examination of User Reading Behaviors in the Process of Meeting Ad-hoc Information Need

Veli Özcan Budak* 

Öz

Amaç: Kısa süreli/geçici olarak ortaya çıkan bilgi ihtiyacımızı gidermek amacıyla sıklıkla başvurduğumuz arama motorları gibi internet tabanlı araçlar sayesinde, bilgiye erişme işlemi hız ve pratiklik kazanmıştır. Bu tür araçlar literatürde “Bilgi Erişim Sistemleri (BES)” kategorisinde değerlendirilen sistemlerdir. Bu çalışmada, Kırklareli Üniversitesi’nde BES mantığıyla çalışan bir altyapı üzerinde, kullanıcıların geçici bilgi ihtiyaçlarını giderirken gerçekleştirdikleri arama etkileşimlerinin incelenmesi ve okuma davranışlarının ortaya çıkartılması amaçlanmıştır. **Yöntem:** Çalışma kapsamında, web kullanıcılarının arama etkileşimleri altı haftalık süre boyunca kayıt altına alınmış olup, toplam 6504 satırlık bir veri seti oluşturulmuştur. Bu veri seti, okuma davranışlarının farklı boyutlardan incelenebilmesi amacıyla dört farklı şekilde organize edilmiştir. Sonrasında, oluşturulan bu veri setleri üzerinden kullanıcıların ziyaret ettikleri web sayfalarındaki kaynak inceleme süreleri ile web sayfalarının kelime miktarları arasındaki ilişkiler tespit edilmeye çalışılmıştır. Ek olarak, kelime başına okuma süresi ile ziyaret edilen sayfalardaki kelime miktarları arasında ne tür bir ilişki bulunduğu da araştırılmıştır. Bu iki farklı ilişki incelemesi, bir bilgi ihtiyacını gidermek için aynı sayfaya yapılan ilk ve ikinci ziyaretler özelinde de gerçekleştirilmiştir. Analizlerin tamamında parametrik olmayan test yöntemlerinden faydalanılmıştır. **Bulgular:** Mann-Whitney U testi ile yapılan analizler, kullanıcıların ilk kez karşılaştıkları sayfalarda daha fazla zaman geçirdiklerini (Z: -4,568; $p < 0,05$) ve ikinci ziyaretlerindeyse, hızlı bir şekilde sayfayı tarama eğiliminde olduklarını ortaya çıkartmıştır (Z: -4,889; $p < 0,05$). Diğer taraftan, Spearman Korelasyon test sonuçları, kelime miktarı ile kullanıcıların sayfa inceleme süreleri arasında pozitif yönde güçlü bir ilişkiyi ortaya çıkarmıştır (r_s : 0,772 ve r_s : 0,730; $p < 0,05$). Aynı test doğrultusunda tespit edilen diğer bir bulgu ise, kelime miktarı ile kelime başına okuma süreleri arasındaki negatif yönde güçlü bir ilişkidir (r_s : -0,775 ve r_s : -0,808; $p < 0,05$). **Sonuç:** Kullanıcıların, kelime miktarının artması durumunda tek tek her bir kelimeyi okumak yerine sayfayı tarama davranışına yöneldikleri tespit edilmiştir. Bu davranışın, aynı sayfaya yapılan ikinci ziyaretlerde de benzer şekilde ortaya çıktığı gözlemlenmiştir. Çalışmada ortaya çıkarılan tüm bulgular, BES’lere yönelik olarak “kaynak inceleme süresi temel alınarak ilgili bilgi kaynaklarının bulunması” temasıyla yapılması muhtemel ileriki çalışmalar için oldukça dikkat

* Kırklareli Üniversitesi, Bilgi İşlem Daire Başkanlığı, Kırklareli, Türkiye. E-posta: veliozcanbudak@gmail.com
Kırklareli University, Information Technologies Office, Kırklareli, Turkey. E-mail: veliozcanbudak@gmail.com

Geliş Tarihi –Received: 21.06.2021

Kabul Tarihi – Accepted: 25.10.2021

Yayımlanma Tarihi – Published: 30.12.2021

gerektiren bir noktaya ışık tutmaktadır. **Özgünlük:** Çalışma kapsamında, kullanıcı okuma davranışlarındaki değişkenlik, kaynak inceleme süresi verisinin gözlemlenmesiyle gözler önüne serilmektedir. Bu çalışma, BES'lerin değerlendirilmesine ve geliştirilmesine yönelik yapılacak ileriki çalışmalarda, kaynak inceleme süresi verisinden faydalanmak istendiğinde daha fazla dikkatin gerektiğini güçlü bulgularla ortaya koymaktadır.

Anahtar kelimeler: Bilgi erişim, Kaynak inceleme süresi, Kelime başına okuma süresi, Okuma davranışları

Abstract

Objective: The process of accessing information has gained speed and practicality thanks to internet-based tools such as search engines, which we frequently use to meet our short-term/ad-hoc information needs. Such tools are systems considered in the "Information Retrieval Systems (IRS)" category in the literature. This study aims to examine the search interactions of users while they meet their ad-hoc information needs and to reveal their reading behaviors on an infrastructure working with the logic of IRS at Kırklareli University. **Method:** Within the scope of the study, the search interactions of web users were recorded for six weeks, and a dataset of 6504 rows was created. This dataset was organized in four different ways in order to examine reading behaviors from different dimensions. Afterward, it was tried to determine the relationships between the dwell times on the web pages visited by users and the number of words on the web pages through these datasets. In addition, the relationship between the reading time per word and the number of words on the pages was also examined. These two different relationship analyses were also carried out for the first and second visits to the same page to meet an information need. Non-parametric test methods were used in all of the analyses. **Findings:** Analysis with the Mann-Whitney U test revealed that users spend more time on the pages they encounter for the first time ($Z: -4,568; p < 0,05$) and tend to scan the page quickly on their second visit ($Z: -4,889; p < 0,05$). On the other hand, Spearman Correlation test results revealed a strong positive relationship between the number of words and the dwell time ($r_s: 0,772$ and $r_s: 0,730; p < 0,05$). Another finding in the direction of the same test is a robust negative relationship between the number of words and the reading times per word ($r_s: -0,775$ and $r_s: -0,808; p < 0,05$). **Implications:** It has been determined that if the amount of words increases, users tend to scan the page instead of reading each word one by one. It has been observed that this behavior occurs similarly on second visits to the same page as well. All the findings revealed in the study shed light on a point that requires great attention for possible future studies with the theme of "finding relevant information sources based on the dwell time" for IRS. **Originality:** Within the scope of the study, the variability in user reading behaviors is revealed by observing the dwell time data. This study demonstrates strong findings that more attention is required when benefiting from dwell time data in future studies on the evaluation and improvement of IRSs.

Keywords: Information retrieval, Dwell time, Reading time per word, Reading behaviors

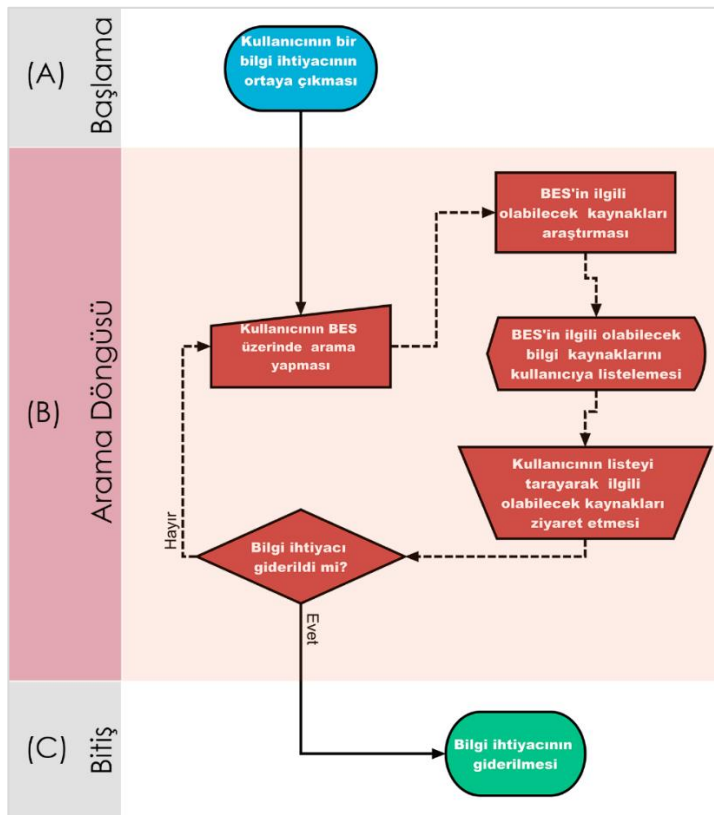
Giriş

Bir konu hakkında fikir sahibi olmak, karar almak ya da herhangi bir problemi çözmek gibi gündelik yaşamımızda sürekli gerçekleştirdiğimiz eylemler için ihtiyaç duyduğumuz en temel

şey bilgidir. Uçak ve Güzeldere (2006), bilginin öneminin giderek arttığını ve bu doğrultuda, mümkün olan en hızlı şekilde gerekli tüm bilgiye erişmenin önemli olduğunu ifade etmiştir. Geçmişte, çoğunlukla kütüphaneler gibi mekâna bağımlılığı gerektiren ortamlardan elde edilebilen bilgi, geliştirilen internet tabanlı erişim araçları sayesinde artık dijital ortam üzerinden de erişilebilir bir özellik kazanmıştır. Günümüzde bir bilgi ihtiyacının giderilmesi noktasında en çok kullanılan dijital ortam araçlarının Google, Yandex, Bing, vb. arama motorları olduğu rahatlıkla söylenebilir. Bu araçlar, literatürde Bilgi Erişim Sistemleri (BES) olarak tanımlanmaktadır. BES, “bilgiyi içine alan, araştırılabilir bir biçime dönüştüren ve kullanıcının bilgi aramasına ve erişimine imkân verecek bir ara yüz sağlayan sistem” şeklinde ifade edilmiştir (Kowalski, 2011, s. 1). Genel itibariyle, BES üzerinde bir bilgi ihtiyacının giderilme süreci, kullanıcının herhangi bir bilgiye duyduğu ihtiyaçla başlamaktadır (Şekil 1-A). Bu süreç, kullanıcının BES üzerinde çeşitli terimlerle gerçekleştirdiği arama eylemi doğrultusunda, ihtiyaca özel çeşitli bilgi kaynaklarının yine BES tarafından kullanıcıya listelenmesi ve kullanıcının ilgisini çeken kaynakları incelemesi ile sürdürülmektedir (Şekil 1-B). Şekil 1’de yer alan “(B) Arama Döngüsü” bölümü, kullanıcının ihtiyacını giderme durumuna bağlı olarak tekrarlı bir şekilde (farklı arama terimlerinin kullanılmasıyla) devam ettirilebilmektedir. Kullanıcı aradığı bilgiye yönelik olarak yeterli tatmine ulaştığı zaman süreç sonlanmaktadır (Şekil 1-C).

Şekil 1

Bilgi ihtiyacının giderilme süreci



Zhai ve Massung (2016), BES’leri “kısa vadeli” ve “uzun vadeli” olarak iki farklı biçimde sınıflandırmıştır. Kısa vadeli terimi, kullanıcının geçici bir bilgi ihtiyacına (*ad-hoc information need*) işaret etmekte olup, Şekil 1’de aktarılmış olan süreç adımlarını içermektedir. Geçici bilgi

ihtiyacının giderilmesinde, temel olarak, kullanıcının bir bilgi ihtiyacını temsil eden bir kaç kelimeye karşılık olarak BES'in ilgili tüm bilgi kaynaklarını bulması beklenmektedir (Kwok, Grunfeld ve Deng, 2007, s. 406). Uzun vadeli terimi ise sistemin, kullanıcının belirli bir süredir var olan bilgi ihtiyacını karşılamaya devam etmesi durumuna işaret etmektedir (Zhai ve Massung, 2016, s. 75). Kısa vadede, BES, kullanıcının bilgi ihtiyacını belirtmesiyle anlık olarak aktifleşirken, uzun vadede sistem sürekli aktif rolde olup, kullanıcının daha önceden belirttiği bir bilgi ihtiyacına yönelik yeni bir kaynak ortaya çıktığında kullanıcıyı bilgilendirmektedir. Uzun vadeli BES'lere öneri (*recommendation*) sistemleri örnek verilebilir. Buna ek olarak, spam filtreleme sistemleri de bir bakıma gereksiz bilginin filtrelenmesini sürekli olarak sağladığından uzun vadeliye örnek gösterilebilir. Diğer taraftan, arama motorları, multimedya BES'leri ve soru cevaplama (*question answering*) sistemleri ise kısa vadede örnek gösterilebilir. Bu çalışmanın birinci odağında, kısa vadeli BES'ler yer almaktadır.

Geçici bilgi edinme davranışının tüm dünyada yoğun bir şekilde gerçekleştirildiği (Manning, Raghavan ve Schütze, 2008, s. 15) ve bu türdeki bilgi ihtiyacının uzun vadede daha sık ortaya çıktığı ifade edilmiştir (Zhai ve Massung, 2016, s. 75). Öyle ki, arama motorları bazında 2021 yılı Şubat ayında market payının %86,6'sına sahip olan Google'ın (Johnson, 2021), 2012 yılında bir trilyonun üzerinde olan arama miktarını, 2021 yılı itibarıyla üç trilyonun üstüne çıkardığı ve günümüz itibarıyla arama miktarının saniyede 40,000 civarında olduğu raporlanmıştır (Internet Live Stats, 2021). Bu rakamlar göz önüne alındığında, dijital ortam üzerinden bilgi edinme çabamızın oldukça fazla olduğu rahatlıkla ifade edilebilir. Bu noktada, her ne kadar BES'ler sayesinde mekândan bağımsız olarak bilgiye erişme süreci pratikleşmiş olsa da, bilgi miktarının sürekli artış göstermesi, dikkat edilmesi gereken bir dezavantajı da ortaya çıkarmıştır. Bu dezavantaj, kullanılan bilgi erişim aracının bilgi ihtiyacını karşılamada ne kadar kabiliyetli olduğu ile ilgilidir. Günümüzde bu kabiliyetin ölçülmesinde çoğunlukla Cranfield paradigması takip edilmektedir. Bu paradigma temel olarak, bir sorguya karşılık olarak ilgili dokümanların insan değerlendiriciler tarafından uzun zaman ve emek gerektiren çalışmalar doğrultusunda belirlendiği ve bu doğrultuda her bir doküman-sorgu çifti için ilgililik değerlendirmeleri, sorgular ve dokümanlardan oluşan bir referans koleksiyonunun oluşturulmasını kapsamaktadır (Baeza-Yates ve Ribeiro-Neto, 2011, ss. 132-134). Cyril Cleverdon isimli bir kütüphanecinin önderlik ettiği bu paradigmada, ortaya çıkarılan referans koleksiyonları, BES'lerin çeşitli metrik ve istatistik ölçümler doğrultusunda bilgi bulma başarısının ölçümünde kullanılmaktadır. Bu paradigma temel alınarak yapılan değerlendirmelerin dezavantajı, bilgi ihtiyacının sabit olarak düşünülmesi ve insan değerlendiricilerin ilgili bilgi ihtiyacına göre tüm kullanıcı evrenini temsil ettiğinin varsayılmasıdır (Baeza-Yates ve Ribeiro-Neto, 2011, ss. 133-134). BES'lerin değerlendirilmesinde faydalanan diğer bir yöntem ise "günlük tutma (*logging*)"dır (Croft, Metzler ve Strohmman, 2009, s. 305). Bu yöntem, kullanıcıların BES üzerindeki arama etkileşimlerinin sürekli olarak kayıt altına alınmasını ve bu kayıtların çeşitli şekillerde analiz edilmesini kapsamaktadır. Bu haliyle, referans koleksiyonlarına nazaran çok daha fazla veri ile kullanıcı popülasyonunun sürekli değişen bilgi ihtiyacını takip ederek değerlendirme imkânı tanımaktadır. Buna ek olarak, kullanıcının arama eylemine müdahale edilmeden veri setinin oluşturulabilmesi de diğer bir avantajdır. Günlük tutma işlemi kullanıcının arama yaparken faydalandığı arayıcıya kurulabilecek bir uygulama vasıtasıyla yapılabildiği gibi (Croft, Metzler ve Strohmman, 2009, s. 306), BES'in barındırıldığı sunucu tarafında da yapılabilmektedir.

Günlük tutma yönteminde tıklama verisi, sorgu terimleri, sonuç listesinin sıralaması ve zaman damgası gibi çeşitli veri türleri ile birlikte (Croft, Metzler ve Strohman, 2009, ss. 305-306) kullanıcının ilgili olarak düşündüğü kaynağı incelerken harcadığı zaman (*dwell time*) da kullanılabilir. Kullanıcının incelediği dokümanda ne kadar süre kaldığı bilgisini taşıyan bu veri türünün, BES'lerin kullanıcı aramalarına karşılık olarak sunduğu kaynakların, arama terimleriyle ne kadar ilgili olduğunun belirlenmesinde önemli bir göstere olduğu ifade edilmiştir (Guo ve Agichtein, 2012, s. 569; Croft, Metzler ve Strohman, 2009, s. 306). Buna ek olarak, çeşitli çalışmalarda BES performansının geliştirildiğine yönelik önemli bulgular da elde edilmiştir (Agichtein, Brill ve Dumais, 2019; White ve Kelly, 2006). Bu veri türü, çalışmada odaklanılmış olunan ikinci noktaya karşılık gelmektedir.

Bu çalışmada, kullanıcıların geçici bilgi ihtiyaçlarını karşılarken, inceledikleri kaynaklarda harcadıkları süreler ile okuma davranışlarının analiz edilmesi hedeflenmiştir. Çalışma bu yönüyle herhangi bir BES'in performans değerlendirmesini içermekten ziyade, kullanıcıların geçici bilgi edinme sürecindeki okuma davranışlarının analizi üzerinden BES'lerin nasıl daha iyi hizmet verebileceğine yönelik dikkat edilebilecek bulgulara erişme çabasını kapsamaktadır. Bu açıklamalar doğrultusunda, bu çalışmada, Kırklareli Üniversitesi bünyesinde hizmet veren 56 farklı birim web sitesinde bütünleşik çalışan site-içi arama araçlarında gerçekleştirilmiş kullanıcı arama etkileşimlerinin incelenmesi ve kullanıcıların okuma davranışlarının ortaya çıkartılması amaçlanmıştır. Site-içi arama araçları, sadece bilgi erişim amaçlı kullanılan arama motorları gibi kısa vadeli bir BES türü olarak düşünülmesi bile, Şekil 1'de aktarılan sürecin gerçekleştirildiği ve web siteleri içerisinde yer alan sayfalardaki bilgiye erişme imkânı tanıyan bir yapıya sahiptir. Bu sebeple, tıpkı kısa vadeli BES'ler gibi geçici bilgi edinme çabasına katkı sunmaktadır. Çalışmanın bu bölümünden itibaren, kullanıcılar tarafından yapılan bir arama sorgusu doğrultusunda site-içi arama araçları tarafından kullanıcıların karşısına getirilen bilgi kaynağı listesi, “**web sayfaları listesi**” ve kullanıcıların bu bilgi kaynaklarını okurken harcadıkları süreler ise “**kaynak inceleme süresi**” olarak ifade edilecektir. Çalışmanın ikinci bölümünde bu çalışmayla ilgili olan araştırmalara, üçüncü bölümünde çalışmada cevap aranan araştırma sorularına, dördüncü bölümünde çalışmanın yöntemine, beşinci bölümünde çalışmada ortaya çıkarılan bulgulara ve altıncı bölümde ise bulgularda tespit edilen bilgiler ışığında yapılan değerlendirmelere yer verilmiştir.

Literatür

Çalışmanın bu bölümünde, literatürde yapılmış olan ve bu çalışmayla ilgili olduğu düşünülen araştırmalar aktarılmıştır:

Bilgi arama sürecinde kullanıcı davranışlarıyla birlikte çeşitli faktörlerin bilgi kaynaklarının yararlılık durumlarının tespitinde kullanılıp kullanılmayacağı üzerine bir çalışma gerçekleştirilmiştir (Mao vd., 2017). Çalışma sonuçlarında, bir kullanıcının herhangi bir dokümanı inceleme süresiyle, doküman yararlılığı arasında pozitif yönde güçlü bir korelasyon olduğu ifade edilmiştir. Bu sonuç, bir kullanıcı ne kadar uzun süre bir dokümanı incelerse, incelenen dokümanın yararlı olma olasılığının o kadar fazla olduğuna işaret etmektedir.

Kaynak inceleme süresi üzerinden ilgili bilgi kaynaklarının tespit edilebilmesine odaklanılan bir diğer çalışma Kim, Teevan ve Craswell (2016) tarafından gerçekleştirilmiştir. Yazarlar, bir arama motoruna ait günlük dosyasını analiz ederek ilgili/ilgisiz dokümanların belirlenmesinde hangi mutlak kaynak inceleme süresinin sınır olarak ifade edilebileceğini araştırmışlardır. Araştırma sonuçlarında, ilgili/ilgisiz dokümanların tespit edilme sınırının 87 saniye olduğu ifade edilmiştir. Diğer bir ifadeyle, kaynak inceleme süresi 87 saniyeden fazla ise incelenen bilgi kaynağından memnun kalındığına, az ise de bilgi kaynağından memnun kalınmadığına işaret edilmiştir.

Akuma, Iqbal, Jayne ve Doctor (2016) tarafından gerçekleştirilmiş olan bir çalışmada ise, kontrollü bir ortamda kullanıcılardan bir arama aracı üzerinde çeşitli arama etkileşimleri gerçekleştirmeleri istenmiştir. Kullanıcılar arama işlemlerini gerçekleştirdikleri esnada kaynak inceleme süreleri kayıt altına alınmıştır. Bu veriye ek olarak, kullanıcılardan, okudukları içerikle aradıkları bilginin ilgili olup olmadığını oylamaları istenmiştir. Bu iki tür verinin kıyaslanması doğrultusunda, kaynak inceleme süreleri ile oylamalar arasında herhangi bir fark olmadığı ve kaynak inceleme sürelerinin bilgi arama sürecinde ilgili içeriği tespit etmede önemli bir gösterge olduğu belirtilmiştir.

Kişiselleştirilmiş bilgi erişim çerçevesinde doküman kullanışlılığı ile kaynak inceleme süresi arasındaki ilişkinin incelendiği bir başka çalışma Liu ve Belkin (2015) tarafından gerçekleştirilmiştir. 24 kullanıcının üç oturum üzerinden çeşitli dokümanları incelemeleri doğrultusunda bir rapor oluşturdukları bu çalışmada, kaynak inceleme süreleri ile incelenen dokümanların kullanışlılığı arasında güçlü bir ilişki olduğu tespit edilmiştir.

Web kullanıcılarının, Yahoo ana sayfasında bulunan farklı türdeki içeriklerle girdikleri bir aylık etkileşimleri üzerinden kaynak inceleme sürelerinin elde edildiği ve bu sürelerin, kişiselleştirilmiş içerik sunmaya yönelik olarak nasıl kullanılabilirliği üzerine yapılan bir diğer çalışma Yi, Hong, Zhong, Liu ve Rajan (2014) tarafından gerçekleştirilmiştir. Yazarlar, hem çevrimiçi hem de çevrimdışı olarak gerçekleştirdikleri çeşitli incelemelerde, kaynak inceleme sürelerinden faydalanarak BES'lerin daha iyi öneri performansı ortaya koyabileceklerine yönelik bulguları ortaya çıkarmışlardır.

Bilgi erişim sürecinde kullanıcıların inceledikleri web sayfalarının, aranan bilgiye yönelik olarak yeterli tatmini sağlayıp sağlamadığının belirlenmesine çalışılan bir diğer çalışma, Kim, Hassan, White ve Zitouni (2014) tarafından gerçekleştirilmiştir. Yazarların, belirli bir kaynak inceleme süresi eşik değeri üzerinden memnuniyet tespitinin, farklı içeriklerde benzer olmayabileceğini vurguladıkları bu çalışmada, memnuniyetin belirlenmesine yönelik olarak kaynak inceleme süresi dağılımlarının tahmin edilebileceği bir regresyon modeli önerilmiştir.

25 üniversite öğrencisinin test sürecine dâhil edildiği ve içinde kaynak inceleme süresiyle beraber farklı göstergelerin de yer aldığı bir çalışma da Balakrishnan ve Zhang (2014) tarafından gerçekleştirilmiştir. Çalışmada, yazarlar tarafından geliştirilmiş olan bir BES ile birlikte hazır bir referans koleksiyonundan faydalanılmıştır. Çalışma sonuçlarında, kaynak inceleme süresi ile birlikte kullanılan diğer göstergelerin ilgili bilgi kaynaklarının listelenmesinde olumlu performans ortaya koyduğu ortaya çıkartılmıştır.

Her biri 350 kelimedenden oluşan 15 farklı web dokümanının 13 öğrenci tarafından okunduğu ve oylandığı bir başka çalışma Akuma, Iqbal, Jayne ve Doctor (2014) tarafından gerçekleştirilmiştir. Çalışmanın amacı, kullanıcı oylarıyla hem kaynak inceleme süreleri hem

de fare aktiviteleri arasında bir ilişki olup olmadığının incelenmesi ve bu doğrultuda, belirli bir konuyla ilgili bilgi içeriklerinin tespitinde kullanıcının haberi olmadan kayıt altına alınan kaynak inceleme süresi vb. veri türlerinin kullanılıp kullanılmayacağını belirlenmesidir. Çalışma sonuçları, kullanıcıların ilgili olarak düşündükleri dokümanlarla kaynak inceleme süreleri arasında pozitif yönde korelasyon olduğunu ortaya çıkarmıştır.

Kullanıcıların ilgi duyduğu içerikleri tespit etmeye yönelik olarak yapılmış olan bir diğer çalışma, Núñez-Valdéz vd. (2012) tarafından öneri sistemleri temel alınarak gerçekleştirilmiştir. Yazarlar, kullanıcılardan inceledikleri içerikleri oylamalarını istemiş ve aynı zamanda kullanıcılar içerikleri incelerken etkileşimlerini kaydetmişlerdir. Çalışma sonuçlarında, kaydedilen veri türleri içerisinde yer alan kaynak inceleme süresi ile kullanıcıların ilgi durumları arasında doğru orantılı bir ilişki olduğu ortaya çıkartılmıştır.

Kullanıcıların geçmiş arama etkileşimleri doğrultusunda inceledikleri web sayfalarında ortaya çıkan kaynak inceleme sürelerinden faydalanılarak, web sayfalarının kişiselleştirilmiş bir biçimde sıralanmasına yönelik yeni bir algoritma önerisinin getirildiği bir diğer çalışma ise Xu, Jiang ve Lau (2011) tarafından gerçekleştirilmiştir. Yazarlar, geçmiş kaynak inceleme süreleri doğrultusunda yeni karşılaşılabilecek olan web sayfasındaki muhtemel kaynak inceleme süresinin belirlenebileceğini ifade etmişlerdir. Önerilen algoritma, çeşitli sorgulamalar doğrultusunda Google, Yahoo ve Bing arama motorlarından elde edilen sıralamalar ile karşılaştırılmış olup, yazarlar tarafından başarılı sonuçlar ortaya koyduğu ifade edilmiştir.

Aktarılmış olan çalışmalar, kaynak inceleme süresi verisinin bilgi ihtiyacının giderilme sürecinde oldukça önemli yer tuttuğunu göstermektedir. Bu veri türünün, hem BES'lerin performanslarının değerlendirilmesinde hem de kullanıcıların yeterli bilgiye ulaşımına ulaşmama durumlarının belirlenmesinde sıklıkla kullanıldığı görülmektedir. Bu çalışma, literatürde yapılmış olan çalışmalara ek olarak, kullanıcı okuma davranışlarına yönelik farklı ipuçlarını ortaya çıkarmaya ve BES'lerin daha iyi hizmet sağlayabilmesi için bu ipuçlarının nasıl değerlendirilebileceğine yönelik incelemeleri içermektedir.

Araştırma Soruları

Bu çalışma kapsamında kullanıcı okuma davranışları araştırılırken, cevabı aranmış sorular aşağıda aktarılmıştır:

1. Web sayfalarındaki kelime miktarları ile kullanıcıların kaynak inceleme süreleri ve kelime başına okuma süreleri arasında nasıl bir ilişki bulunmaktadır?
2. Kullanıcıların, web sayfalarına ilk ve ikinci temasları esnasındaki kaynak inceleme süreleri arasında nasıl bir ilişki bulunmaktadır?
3. Kullanıcıların, web sayfalarına ilk ve ikinci temasları esnasındaki kelime başına okuma süreleri arasında nasıl bir ilişki bulunmaktadır?

Yukarıda aktarılan sorularda “temas” kavramı, herhangi bir kullanıcının, herhangi bir arama eylemi doğrultusunda BES tarafından sunulan web sayfaları listesindeki bir web sayfasını, ilk kez ya da ikinci kez ziyaret etme eylemini ifade etmektedir. Buna ek olarak, dikkat edilmesi gereken bir diğer nokta ise, web sayfalarına gerçekleştirilen ziyaretlerin tamamının,

kullanıcıların geçici bilgi ihtiyaçlarını gidermek amacıyla gerçekleştirdikleri arama eylemleri sonrasında yaptıkları ziyaretler olmasıdır.

İlk temas olarak ikinci ve üçüncü soruda yer alan bu kavramın kullanılma sebebi, kullanıcıların ilk kez karşılaştıkları web sayfalarındaki kaynak inceleme sürelerinin, okuma davranışlarına yönelik olarak daha gerçekçi sonuçları ortaya çıkaracağı varsayımdır. Çünkü bir kullanıcı daha önce okuduğu bir içeriği tekrar okumaya karar verdiğinde, önceki deneyiminden kalan bilgiye sahip olması sebebiyle ikinci ziyaretinde kaynak inceleme süresinde düşüş olma ihtimali yüksektir. Diğer taraftan, “ikinci temas” ile araştırılmak istenen ise, ikinci ziyaretteki düşüş ihtimalinin doğru olup olmadığının belirlenmesidir. Dolayısıyla, kullanıcıların farklı temaslarındaki değişimlerin incelenebilmesi amacıyla ikinci ve üçüncü sorulara cevap aranmıştır.

Metot

Bu bölümde, kullanıcıların arama etkileşimlerini gerçekleştirdikleri yapı, veri toplama sürecinin nasıl yürütüldüğü, toplanmış olan verinin çalışmanın amacına uygun olacak şekilde nasıl organize edilerek analiz edildiği ve tüm bu süreçte var olan sınırlılıklar aktarılmıştır.

Veri Toplama Sürecinde Kullanılan Site-İçi Arama Araçları

Çalışmada kullanıcıların geçici bilgi ihtiyaçlarını giderdikleri yapı, tıpkı kısa vadeli BES’ler gibi görev yapan ve Kırklareli Üniversitesi’ne ait birim web sitelerinde bütünleşik olarak çalıştırılabilen site-içi arama araçlarıdır (Şekil 2).

Şekil 2

Örnek bir site-içi arama aracı

Şekil 2’de bir örneği görünen bu araçlar, ihtiyaç durumuna göre birimlerin kendi web sitelerine entegre edilebilmekte ve üniversitenin tüm birimlerinde yayınlanan web sayfaları içinde bilgi aramaya olanak sağlayabilmektedir. *PHP* yazılım diliyle hizmet vermekte olan birim web sitelerindeki bu araçlar, kullanıcıların, en az iki harften oluşan bir arama terimini yazıp, “Ara” butonuna tıklamasıyla sorgulamalar gerçekleştirmekte ve arama terimiyle ilişkili olan web sayfalarının bir listesini kullanıcıların karşısına getirmektedir. Bu çalışmada, toplam 56 tane farklı birim web sitesinde kullanılmakta olan site-içi arama araçlarından faydalanılmıştır.

Veri Toplama Periyodu ve Veri Setinin Özellikleri

Kırklareli Üniversitesi web kullanıcılarının, farklı birim web siteleri üzerinde gerçekleştirdikleri arama etkileşimleri 2020 yılı Şubat ve Mart ayları içinde toplam altı haftalık süre zarfında kayıt altına alınmıştır. Bu süreçte, her bir kullanıcının, yaptığı arama sonrasında incelediği web sayfalarındaki kaynak inceleme süreleri tekil olarak **saniye** cinsinden tutulmuştur. Yani, her bir incelenen web sayfası için birbirinden bağımsız olacak şekilde

kaynak inceleme süreleri kaydedilmiştir. Bu durum, aynı kullanıcının aynı arama sonrası yaptığı aynı sayfa incelemesinde bile bağımsız olarak ele alınmıştır. Kaynak inceleme sürelerinin doğru kaydedilmesi amacıyla, *Javascript* diliyle hazırlanmış bir kod, veri toplama süreci öncesinde tüm birimlerdeki web sayfalarına eklenmiştir. Böylelikle, herhangi bir arama doğrultusunda incelenmesi muhtemel web sayfaları, kullanıcıların kaynak inceleme sürelerinin takip edilebileceği bir yapıya dönüştürülmüştür. Ek olarak, *PHP* yazılım diliyle hazırlanmış olan bir kod, yine web sayfalarına veri toplama süreci öncesinde eklenmiştir. Bu kod sayesinde, ziyaret edilecek sayfaların kaç kelimedenden oluştuğu verisinin elde edilebileceği bir güncelleme sağlanabilmiştir.

Çalışmada cevap aranan sorulara yönelik olarak hem kaynak inceleme süresi hem de kelime miktarının kullanılmasıyla hesaplanıp kaydedilmiş olan bir diğer veri türü de “kelime başına okuma süresi”dir. Bu veri türü, her bir web sayfası ziyaretinde ortaya çıkan değerler üzerinden Denklem 1 temel alınarak veri toplama süreci sonrasında oluşturulmuştur.

$$\text{Kelime başına okuma süresi} = \frac{\text{Kaynak İnceleme Süresi}}{\text{Kelime Miktarı}} \quad (1)$$

Çalışmanın üçüncü bölümünde aktarılmış olan sorulardan hatırlanacağı üzere, “temas” kavramı iki farklı soruda yer almaktaydı. Bu sorulara yönelik olarak incelemeler gerçekleştirmek amacıyla, hem kullanıcıların IP adresleri (kişisel veri kapsamına girmesinden dolayı anonimleştirilmiştir) hem de temasın gerçekleştirildiği zaman kaydedilmiştir. Bu iki tür veri sayesinde, hangi kullanıcının ne zaman kaçınıcı teması gerçekleştirdiği tespit edilebilmiştir. Bu veri türlerine ek olarak, veri setinin düzenlenmesi sürecinde aynı kelime miktarına sahip olan sayfaların birbiriyle karışmaması için her bir sayfanın kimlik (ID) değeri de kaydedilmiştir. Sonuç itibariyle, aşağıda aktarılmış olan altı farklı veri türü veri setinde yer almıştır:

1. Ziyaret edilen sayfanın ID değeri
2. Anonimleştirilmiş IP adresi
3. Kelime miktarı
4. Kaynak inceleme süresi
5. Kelime başına okuma süresi
6. Ziyaret zamanı

Çalışmada, kullanıcıların tüm arama etkileşimleri MySQL veritabanına kaydedilmiş olup, veri toplama süreci sonrasında analiz için kullanılmıştır (veri kullanım izni alınmıştır). Veri toplama sürecinde oluşturulan veri seti toplam 6504 satırlık arama etkileşimlerini içermektedir.

Veri Setinin Düzenlenmesi ve Analiz Süreci

Kullanıcı okuma davranışlarının nasıl değişkenlik gösterdiğinin tespit edilebilmesi amacıyla, çalışmada toplanmış olan veri seti, birinci ve son iki araştırma sorusu için farklı bir şekilde organize edilmiştir. İlk soru için, veri toplama sürecinde en az 30 defa ziyaret edilmiş olan ve aralarında kelime miktarları bakımından fark bulunan web sayfalarındaki etkileşimler tespit edilerek iki farklı veri seti oluşturulmuştur. Bu veri setlerinin ilkinde, ziyaret edilmiş olan sayfaların kelime miktarları ile birlikte bu sayfalara yapılmış olan ziyaretlerdeki kaynak inceleme sürelerinin ortalamaları bulunurken; diğerinde ise, sayfaların kelime miktarları ile

birlikte bu sayfalara yapılmış olan ziyaretlerdeki kaynak inceleme sürelerinin ortanca değerleri tutulmuştur. Böylelikle, kelime miktarına bağlı olarak en az 30 ziyaret üzerinden elde edilmiş iki farklı kaynak inceleme süresi üzerinden değerlendirme imkânı oluşturulmuştur. Ortalama ve ortanca olmak üzere iki farklı türde kaynak inceleme süresinin kullanılmasının sebebi, kaynak inceleme sürelerinin uç değer taşıma ihtimalinin olmasıdır. Ortanca değerlerin uç değerlerden etkilenmemesi sebebiyle, ortalamanın bulunduğu veri seti haricinde farklı bir veri setinde yer almasına karar verilmiştir. Bu haliyle, ilk veri setinde (VS1) “kelime miktarı”, “ortalama kaynak inceleme süresi” ve Denklem 1 kullanılarak oluşturulan “kelime başına okuma süresi-ortalama” veri türleri yer alırken; ikinci veri setindeyse (VS2), “kelime miktarı”, “ortanca kaynak inceleme süresi” ve yine Denklem 1 kullanılarak oluşturulan “kelime başına okuma süresi-ortanca” yer almıştır. VS1, toplam 52 satırlık veri içerirken, VS2, 56 satırlık veriden oluşmaktadır. VS1’in daha az satır sayısına sahip olmasının sebebi ise, “kaynak inceleme süresi” temel alınarak uç değerlerden arındırılmış olmasıdır.

Çalışmada cevap aranan son iki soru için, web sayfalarında gerçekleştirilen temaslar göz önünde bulundurularak 6504 satırlık veri seti iki farklı grupta organize edilmiştir. İlk olarak, en az iki teması bulunan kullanıcıların yaptıkları arama etkileşimleri tespit edilmiştir (toplam 915 satır). Sonrasında, tespit edilen etkileşimler üzerinden kullanıcıların sayfalara yaptıkları ilk ve ikinci temaslarını (bazı kullanıcılar aynı, bazıları ise farklı sayfaları ziyaret etmişlerdir) içeren 545 satırlık iki ayrı veri seti oluşturulmuştur (915 satırlık veri seti “kaynak inceleme süresi” temel alınarak uç değerlerden arındırılmıştır. Arındırma sürecinde bir veri setinde bulunan uç değer, diğer veri setinde uç değer olarak tespit edilmese bile, aynı kullanıcıya ait bir etkileşim olduğundan silinmiştir). Bu iki veri seti, “kelime miktarı”, “kaynak inceleme süresi” ve “kelime başına okuma süresi” olmak üzere üç farklı veri türüne sahip olup, temsili bir örneği Tablo 1’de paylaşılmıştır.

Tablo 1

Son iki soruya yönelik hazırlanmış olan veri setlerinin temsili örneği

Sıra	Anonimleştirilmiş adresi *	IP	Sayfa değeri *	ID	Kelime Miktarı	Kaynak Süresi (saniye)	İnceleme	Kelime Okuma Süresi	Başına
1	A		100		30	20		0,66	
2	B		200		40	25		0,625	
3	C		100		30	15		0,5	
4	D		300		50	20		0,4	
5	E		400		60	10		0,16	
6	A		500		70	35		0,5	

* Bu veri türleri örnek olması amacıyla tabloya eklenmiş olup, analiz öncesinde her iki veri setinden temizlenmiştir.

Tablo 1 incelendiğinde, farklı kullanıcıların aynı sayfalara ziyaret gerçekleştirme (Sıra 1 ve 3) durumuyla birlikte, aynı kullanıcının farklı sayfalara ziyaret gerçekleştirme (Sıra 1 ve 6) durumu görülmektedir. Çalışmanın son iki sorusu için, bir kullanıcının aynı bilgi ihtiyacını giderirken gerçekleştirdiği temaslardaki değişimin tespit edilmesine odaklanılmıştır. Oluşturulan her iki veri setindeki etkileşimlerin sıralamaları ise birbiriyle aynıdır. Bir başka ifadeyle, bir kullanıcıya ait etkileşimin sıra numarası her iki veri setinde aynı şekilde tutulmuştur. Kullanıcıların daha önce ziyaret ettiği sayfayı tekrar incelemesinin, ziyaret edilen sayfanın aranan bilgiyle ilgili olduğunu gösteren güçlü bir işaret olduğu düşünülmektedir. Bu

sebeple, her iki veri setindeki etkileşimleri gerçekleştiren kullanıcıların, geçici bilgi ihtiyaçlarını ilgili sayfaları okuyarak giderdikleri varsayılmıştır.

Çalışma sorularının cevaplanması için oluşturulmuş olan tüm veri setlerindeki “kelime miktarı” veri türü bağımsız değişken olarak nitelendirilmiştir. Diğer değişkenler ise bağımlı değişken olarak çalışmada yer almıştır.

Analiz sürecinde, çalışmanın birinci sorusu için her iki veri seti üzerinden ayrı olacak şekilde korelasyon ölçümleri yapılmıştır. Diğer sorular içinse, bağımsız iki örneklem testleri arasında yer alan “Mann-Whitney U” testinden faydalanılmıştır. Çalışmada yapılan analizlerin tamamında parametrik olmayan test yöntemlerinden faydalanılmış olup, detayları çalışmanın “Bulgular” bölümünde açıklanmıştır. Son iki sorunun cevaplanmasında faydalanılan test yöntemi, aynı kullanıcı grubu kullanılmasına rağmen, bağımlı iki örneklem testlerinden biri olan “Wilcoxon Eşleştirilmiş İki Örnek Testi” yerine tercih edilmiştir. Bunun sebebi, her ne kadar aynı kullanıcı grubuyla, aynı bağımsız değişken üzerinden hareket edilmiş olsa bile, bu bağımsız değişken, her bir kullanıcı için kendi içinde farklılık göstermektedir. Diğer bir ifadeyle, kullanıcılar farklı kelime miktarlarına ait sayfaları ziyaret ederek, birbirinden farklı bir bağımsız değişken üzerinden hem ilk temas hem de ikinci temas için “kaynak inceleme süresi” ve “kelime başına okuma süresi” değerleri ortaya çıkarmışlardır.

Sınırlılıklar

Kullanıcıların farklı temaslarını tespit edebilmek amacıyla anonimleştirilerek kullanılmış olan IP adresleri, belirli bir süre zarfında sadece bir kişi için sabit tutulmakta ve bir süre sonra bu değer farklılaşabilmektedir. Bu durum, aynı kullanıcının takibini zorlaştırabilecek bir sınırlılık olarak çalışmada yer almıştır. Kullanıcıların arama etkileşimlerini gerçekleştirdikleri cihaz türlerinin hepsinin aynı olduğu varsayımı da yine başka bir sınırlılıktır. Öyle ki, web sayfalarındaki kaynak inceleme sürelerinin, ziyaretçilerin kullandıkları cihazların ekran boyutlarına göre değişme olasılığı bulunmaktadır. Benzer şekilde, kullanıcıların web sayfasını okumaya başlaması sonrasında web sitelerinin barındırıldığı sunucuda yaşanabilecek olası veritabanı iletişim kopukluğu da bir sınırlılıktır. Bu tür bir durum, kullanıcının kaynak inceleme süresi artsa bile veritabanına herhangi bir güncelleme yapılamayacağından dolayı hatalı kayıtlar oluşturma ihtimalini barındırmaktadır. Kullanıcıların yaptıkları aramalar sonrasında herhangi bir kaynağı incelerken dikkatinin dağılması ve bu doğrultuda kaynak sayfasını açık bırakması yüksek kaynak inceleme süresinin oluşmasına neden olabilir. Bu türde bir ihtimal de çalışma sınırlılıklarındandır. Ek olarak, veri toplama süreci boyunca kullanıcılar farklı tür ve sürüme sahip tarayıcılardan faydalanarak web sayfalarına erişim sağlamışlardır. Kullanılan tarayıcılardaki tür ve sürüm farklılıklarından dolayı “kaynak inceleme süresi” değişkenini elde etmek amacıyla hazırlanmış olan *Javascript* kodunun desteklenmemesi durumu da söz konusu olup, bu çalışma için bir başka sınırlılık olarak değerlendirilmiştir.

Bulgular

Bu bölümde, ilk olarak veri setlerinin genel özellikleri aktarılmıştır. Bu işlem sonrasında, analizler doğrultusunda ortaya çıkarılan tüm bulgular, çalışmada cevap aranmış olan sorular

temel alınarak bu bölüm altında organize edilerek paylaşılmıştır. Veri setlerinde yer alan veri türleri, bu bölüm altında aşağıda yer alan kısaltmalar ile ifade edilmiştir:

- Kelime Miktarı – KM
- Kaynak İnceleme Süresi – KİS
- Kelime Başına Okuma Süresi – KEBOS

Veri Setlerine Ait Betimsel İstatistik Bulguları

Çalışmanın ilk sorusunun cevaplanması amacıyla oluşturulmuş olan iki farklı veri setine ait betimsel istatistik bilgileri Tablo 2’de aktarılmıştır.

Tablo 2

Bağımlı ve bağımsız değişkenlerin ilk soru için hazırlanmış olan veri setlerindeki betimsel istatistikleri

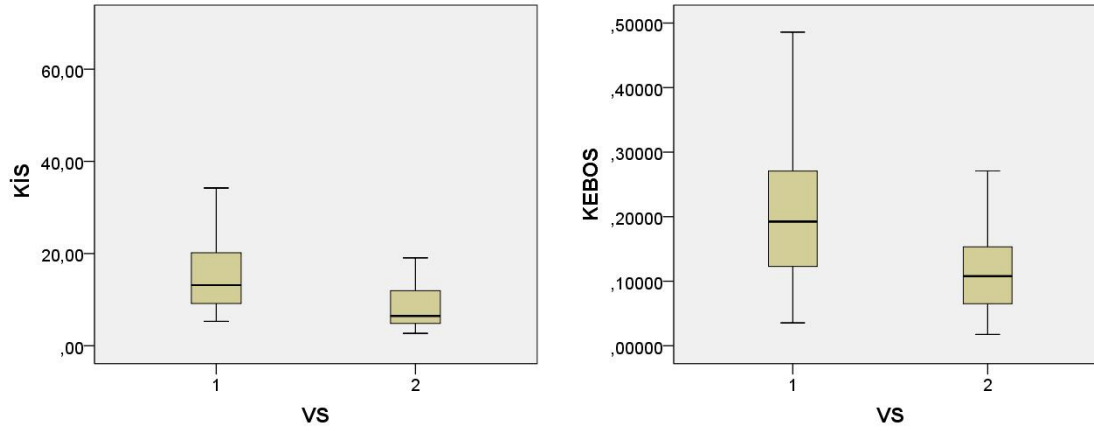
		Veri Setleri	
		VS1	VS2
KM	Özellikler		
	Gözlem Sayısı	52	56
	En küçük	20	20
	En büyük	525	525
	Ortalama	117,1	110,75
	Ortanca	68,5	64
	Standart Sapma	112,917	111,168
KİS	En küçük	5,307	2,7
	En büyük	62,505	23,85
	Ortalama	17,056	8,254
	Ortanca	13,152	6,45
	Standart Sapma	11,403	4,642
KEBOS	En küçük *	0,036	0,018
	En büyük	0,486	0,271
	Ortalama	0,211	0,114
	Ortanca	0,192	0,108
	Standart Sapma	0,113	0,061

* Bu değerlerin düşük olmasının sebebi, çalışmanın sınırlılıklarında da ifade edildiği üzere kullanıcıların ziyaret gerçekleştirdikleri tarayıcıların *Javascript* kodunu doğru çalıştıramaması, sunucu ile iletişimin kopması ya da kullanıcıların çok fazla kelime miktarı içeren web sayfalarında çok kısa süre kalmasıdır.

Tablo 2’deki KİS değerlerine odaklanıldığında, VS2’deki değerlerin VS1’e göre yarı yarıya az olduğu söylenebilir. Bu farklılık, KEBOS değerleri içinde benzer bir şekilde ortaya çıkmıştır (Tablo 2). Her iki veri setindeki bağımlı değişkenlere ait kutu grafikleri Şekil 3’de aktarılmıştır.

Şekil 3

VS1 ve VS2 bağımlı değişkenlerinin kutu grafikleri



KİS değerleri göz önünde alındığında, her iki veri setinde pozitif yönde bir çarpıklık söz konusudur (Şekil 3). KEBOS değerleri için bu durum normal dağılıma biraz daha yakın görünmesine rağmen (Şekil 3), hem KİS hem de KEBOS değerlerinin her iki veri setindeki Kolmogorov-Smirnov ve Shapiro-Wilk ölçümleri, dağılımların normal olmadığını ortaya çıkarmıştır (Tablo 3: $p < 0,05$).

Tablo 3

VS1 ve VS2 bağımlı değişkenlerinin normallik test sonuçları

	VS	Kolmogorov-Smirnov	Shapiro-Wilk
		p değerleri	
KİS	1	0,000	0,000
	2	0,000	0,000
KEBOS	1	0,027	0,010
	2	0,022	0,031

Hem Şekil 3 hem de Tablo 3'deki sonuçlar doğrultusunda, çalışmanın ilk sorusu için bağımlı ve bağımsız değişkenler arasındaki ilişkinin incelenmesi amacıyla parametrik olmayan Spearman Korelasyon testinin uygulanmasına karar verilmiştir.

Çalışmadaki ikinci ve üçüncü soruya yönelik oluşturulan veri setlerinde, KM değişkeni ile ilgili ortak olan temel özellikler Tablo 4'de aktarılmıştır.

Tablo 4

Çalışmanın ikinci ve üçüncü sorusuna yönelik oluşturulan veri setlerinde bağımsız değişken yönünden ortak olan özellikler

ÖZELLİKLER	DEĞERLER
En küçük değer	20
En büyük değer	5438
Ortalama	192,47
Ortanca	109
Standart Sapma	347,156
Gözlem Sayısı	545

Ziyaret edilen web sayfalarının en az 20 kelimedenden ve en fazla 5438 kelimedenden oluştuğu Tablo 4’de görülmektedir. KM değişkeni aynı olan gözlemler gruplandığında, 170 tane farklı kelime miktarına sahip web sayfasının ziyaret edildiği belirlenmiştir. Her iki veri setindeki KİS ve KEBOS bağımlı değişkenleri ile ilgili betimsel istatistikler Tablo 5’de paylaşılmıştır.

Tablo 5

İkinci ve üçüncü çalışma soruları için hazırlanmış olan veri setlerindeki bağımlı değişkenlerin betimsel istatistikleri

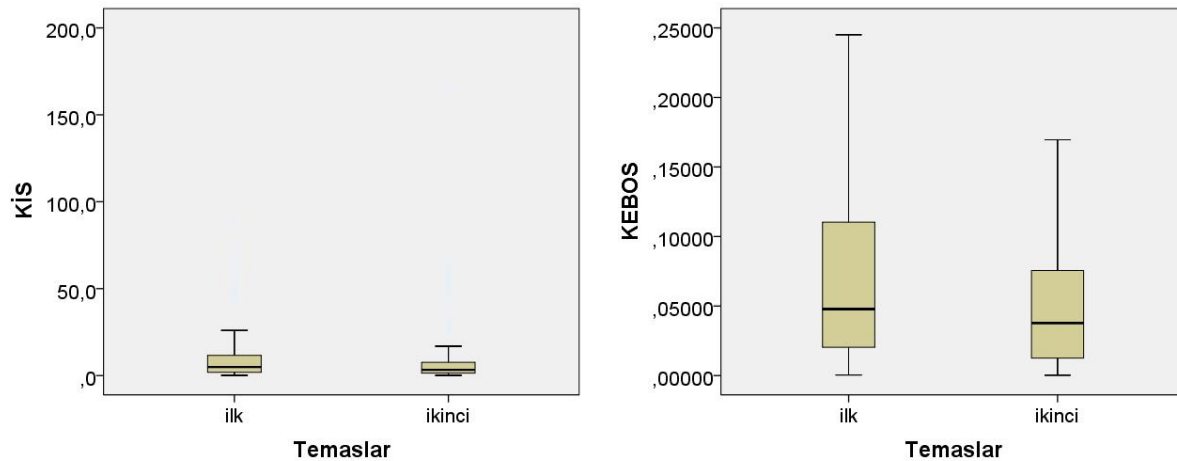
Özellikler	Veri Setleri			
	İlk Temas		İkinci Temas	
	KİS	KEBOS	KİS	KEBOS
En küçük	0,1	0,0004*	0,1	0,0002*
En büyük	123,3	0,245	164,7	0,17
Ortalama	9,975	0,070*	7,260	0,050*
Ortanca	4,9	0,048*	3,3	0,038*
Standart Sapma	14,541	0,062	12,914	0,045

* Bu değerlerin düşük olmasının sebebi, çalışmanın sınırlılıklarında da ifade edildiği üzere kullanıcıların ziyaret gerçekleştirdikleri tarayıcıların *Javascript* kodunu doğru çalıştıramaması, sunucu ile iletişimin kopması ya da kullanıcıların çok fazla kelime miktarı içeren web sayfalarında çok kısa süre kalmasıdır.

Tablo 3 incelendiğinde, KİS açısından ilk temastaki ortalama (9,975 saniye) ve ortanca (4,9 saniye) değerlerinin ikinci temastaki ortalama (7,260 saniye) ve ortanca (3,3 saniye) değerlerinden daha büyük olduğu görülmektedir. Aynı durum, KEBOS değişkeninde de söz konusudur (Tablo 5). Her iki veri setindeki bağımlı değişkenlere ait kutu grafikleri Şekil 5’de aktarılmıştır.

Şekil 4

Çalışmanın ikinci ve üçüncü sorusuna yönelik oluşturulan veri setlerinde bağımlı değişkenlerin kutu grafikleri



Şekil 5’de, her iki veri setinde bağımlı değişkenler yönünden pozitif yönde bir çarpıklığın olduğu görülmektedir. Bağımlı değişkenlere yönelik iki veri setinde de yapılmış olan Kolmogorov-Smirnov ve Shapiro-Wilk ölçümleri, değerlerin normal dağılım göstermediğini

ortaya çıkarmıştır ($p<0,05$). Bu doğrultuda, çalışmada cevap aranmış olan ikinci ve üçüncü soru için de parametrik olmayan test yöntemlerinin kullanılmasına karar verilmiştir.

Araştırma Sorularına Yönelik İstatistik Bulgular

Ortalama ve ortanca değerler üzerinden hazırlanmış olan VS1 ve VS2 veri setleri birinci araştırma sorusunun cevaplanması için kullanılmış olup, kullanıcıların ilk ve ikinci temaslarını içeren iki farklı veri setiyse, araştırmanın diğer sorularının cevaplanmasında kullanılmıştır.

1. Web sayfalarındaki kelime miktarları ile kullanıcıların kaynak inceleme süreleri ve kelime başına okuma süreleri arasında nasıl bir ilişki bulunmaktadır?

Bu soruya yönelik olarak her iki veri setinde yer alan bağımlı değişkenler ile bağımsız değişken arasında Spearman Korelasyon testleri gerçekleştirilmiş ve sonuçlar Tablo 6'da paylaşılmıştır.

Tablo 6

Birinci araştırma sorusuna yönelik bağımsız ve bağımlı değişkenler arasındaki korelasyon değerleri

		VS1		VS2	
	Özellikler	KİS	KEBOS	KİS	KEBOS
KM	Korelasyon	0,772	-0,775	0,730	-0,808
	Katsayısı				
	p değeri			0,000	
	Gözlem Sayısı		52		56

$p<0,05$

Morita ve Shinoda (1994) tarafından yapılan bir çalışmada, kullanıcıların okuma süreleri (KİS) ile bilgi kaynağı KM değerleri arasında 0,08'den küçük bir korelasyon ilişkisinin bulunduğu ortaya çıkartılmıştır. Tablo 6 incelendiğinde, ortalama KİS değerini temel alan VS1 için, KİS değerleri ile ziyaret edilen web sayfalarındaki KM değerleri arasında oldukça güçlü pozitif yönde bir ilişki bulunmaktadır (0,772). Aynı sonucun, ortanca KİS değerini temel alan VS2 için de benzer olduğu tespit edilmiştir (Tablo 6: 0,730). Her ne kadar Morita ve Shinoda (1994) tarafından yapılan çalışmada bu ilişki zayıf olarak ortaya çıkarılmış olsa da, bu çalışmada, KM ve KİS arasında oldukça güçlü bir ilişkinin olduğu ortaya çıkartılmıştır. Her iki veri seti için KEBOS değerleri temel alındığında, kullanıcıların KM değerinin artması durumunda sayfayı daha kısa süre inceledikleri belirlenmiştir (Tablo 6). Bu davranışın, VS2 de daha belirgin olduğu (Tablo 6: $-0,808 > -0,775$) ifade edilebilir. Genel itibariyle, KM değeriyle KİS değerleri arasında doğru orantılı bir artış olduğu belirtilebilir. Buna ek olarak, KM değerindeki artışın, kullanıcının sayfada yer alan kelimeleri daha hızlı okuyarak geçmesine sebep olduğu söylenebilir. Bir başka ifadeyle, kullanıcıların, KM değerinin fazla olması durumunda tek tek her bir kelimeyi okumak yerine kendileri için önemli olan bilgiyi bulmaya yönelik sayfayı tarama eğiliminde oldukları belirtilebilir. Morita ve Shinoda (1994) tarafından KİS ile KM arasındaki ilişki temel alınarak “göz gezdirme davranışı” şeklinde açıklanan bu eğilim, bu çalışmada, KEBOS değerlerinin KM ile negatif yönde güçlü bir ilişkide olması doğrultusunda belirginleştirilmiştir.

2. *Kullanıcıların, web sayfalarına ilk ve ikinci temasları esnasındaki kaynak inceleme süreleri arasında nasıl bir ilişki bulunmaktadır?*

Birinci araştırma sorusu, KM değerindeki değişimin kullanıcıların KİS ve KEBOS değerlerinde genel olarak nasıl bir değişime sebep olduğunu belirlerken, ikinci ve üçüncü sorular ise kullanıcıların bireysel olarak bilgi ihtiyaçlarını giderirken, farklı temaslarındaki KİS ve KEBOS değerlerindeki değişime odaklanmaktadır. Bu doğrultuda, ikinci araştırma sorusu için her iki veri setinde bağımsız değişken üzerinden ortaya çıkmış olan KİS değerleri Mann-Whitney U testi ile incelenmiş ve sonuçlar Tablo 7’de paylaşılmıştır.

Tablo 7

İkinci araştırma sorusuna yönelik olarak KİS değerlerinin karşılaştırılma istatistikleri

	Veri Setleri	Sıra Ortalaması	Sıra Toplamı	Mann-Whitney U	Z Değeri	Güç Değeri (1-β)	p
KİS	İlk Temas	589,05	321032,50	124777,500	-4,568	0,8836734	0,000
	İkinci Temas	501,95	273562,50				

n=545 ve p<0,05

Tablo 7’ye göre, kullanıcılar geçici bilgi ihtiyaçlarını giderirken, ilk kez karşılaştıkları web sayfalarında daha fazla kalmışlardır (Z: -4,568; p<0,05). Tablo 7’de yer alan “Sıra Ortalaması” sütununa odaklanıldığında, kullanıcıların, aynı bilgi ihtiyacı için aynı sayfayı ikinci kez ziyaret ettiklerinde sayfada daha kısa süre kaldıkları görülmektedir. Bu sonuç, korelasyon ölçümlerinde de ifade edilmiş olan “sayfayı tarama eğilimi” davranışını destekleyen önemli bir diğer bulgudur. Bir metni okurken harcanan zaman üzerinden, kullanıcının okuduğu metin ile aradığı bilgi arasında ilişki olup olmadığının belirlenemeyeceği ifade edilmiştir (Kelly ve Belkin, 2001, s. 409). Hatırlanacağı üzere, KİS değerleri üzerinden ilgili bilgi kaynaklarının tespit edildiği çeşitli çalışmalar araştırmanın ikinci bölümünde paylaşılmıştı. Bu noktada, Kelly ve Belkin (2001)’in değerlendirmesinin genelleyici olamayacağı söylenebilir, fakat Tablo 7’deki bulgularla da desteklenebilir. Öyle ki, bu çalışmada, kullanıcıların ikinci temaslarıyla birlikte bilgi ihtiyaçlarını giderdikleri varsayılmıştır. Kullanıcıların ikinci temaslarındaki KİS değerlerinde gözlenen düşüş, aynı kullanıcının bile ilgili bulduğu bir bilgi kaynağında farklı sürelerde okuma davranışı gösterebileceğini ortaya koymaktadır. Bu durumda, KİS değeri üzerinden bir bilgi kaynağının ilgili olarak nitelendirilebilmesinin oldukça zor olduğu söylenebilir. Bu açıklamayı destekleyici şekilde Kelly ve Belkin (2004) de KİS’in, bilgi ihtiyacıyla ilişki kaynakları belirlerken kullanılmasının dikkat gerektirdiğini ifade etmiştir. Bu doğrultuda, Kim, Teevan ve Craswell (2016) tarafından yapılan çalışmadaki gibi KİS değerine yönelik bir eşik değeri (87) belirlemenin, BES’lerin değerlendirilmesinde ve geliştirilmesinde hatalı sonuçlar ortaya çıkarabileceği rahatlıkla ifade edilebilir.

3. *Kullanıcıların, web sayfalarına ilk ve ikinci temasları esnasındaki kelime başına okuma süreleri arasında nasıl bir ilişki bulunmaktadır?*

Her iki veri seti için KEBOS değerleri Mann-Whitney U testi ile incelenmiş ve sonuçlar Tablo 8’de paylaşılmıştır.

Tablo 8

Üçüncü araştırma sorusuna yönelik olarak KEBOS değerlerinin karşılaştırılma istatistikleri

	Veri Setleri	Sıra Ortalaması	Sıra Toplamı	Mann-Whitney U	Z Değeri	Güç Değeri (1-β)	P
KEBOS	İlk Temas	592,11	322701,50	123108,500	-4,889	0,9214474	0,000
	İkinci Temas	498,89	271893,50				

n=545 ve p<0,05

Tablo 8 incelendiğinde, kullanıcıların ilk kez ziyaret ettikleri sayfalarda KEBOS değerlerinin daha fazla olduğu görülmektedir (Z: -4,889; p<0,05). İkinci ziyarette ortaya çıkan daha düşük değerlerin sebebi, “kullanıcıların bir önceki deneyimlerinden bilgi sahibi olmaları ve ikinci ziyaretlerinde unutmüş oldukları ya da doğru hatırlayamadıkları bilgiyi elde edip sayfadan ayrılmaları” şeklinde ifade edilebilir. Bu değişkenlik, ilk iki araştırma sorularında ortaya çıkarılan bulgulara benzer bir şekilde, kullanıcıların, ikinci temaslarında da sayfayı tarama eğiliminde olduklarını gösteren önemli bir bulgudur.

Sonuç ve Öneriler

Günlük olarak gerçekleştirdiğimiz en basit eylemlerde bile ihtiyaç duyduğumuz en temel şey bilgidir. Geçmişte kütüphaneler gibi mekâna bağımlılığı gerektiren ortamlardan elde edilen bilgi, günümüzde özellikle de internetin gücüyle kendisini daha hızlı ve kolay erişilebilir bir dijital ortama aktarmıştır. Diğer taraftan, bilgi miktarının sürekli artış göstermesinden dolayı ihtiyaç duyulan bilgiye hangi yollarla erişileceğinin kararını vermek oldukça karmaşıklaşmıştır. Bu noktada, internet ortamında hizmet sunan BES’ler en çok başvurulan bilgi erişim araçları olarak göze çarpmaktadır. Bu araçlar, hem kısa süreli/geçici bilgi ihtiyacının hem de uzun süreli bilgi ihtiyacının giderilmesi için kullanılan araçlardır. Bu çalışmada, geçici bilgi ihtiyacının giderilme sürecinde kullanıcı okuma davranışlarının nasıl değişkenlik gösterdiği araştırılmıştır. Çoğunlukla BES’lerin değerlendirilmesinde ve geliştirilmesinde kullanılan “kaynak inceleme süresi” değişkeninin, farklı miktarda kelime içeren web sayfalarına yapılan ziyaretlerde ne tür bir değişim gösterdiği incelenmiştir. Buna ek olarak, kaynak inceleme süresinin kelime miktarına bölünmesiyle oluşturulmuş olan “kelime başına okuma süresi” değişkeninin de, yine ziyaret edilen sayfalardaki kelime miktarlarıyla ne tür bir ilişki içerisinde olduğu araştırılmıştır. Bu incelemelerin yapılabilmesi için Kırklareli Üniversitesi bünyesinde hizmet veren toplam 56 tane web sitesindeki site-içi arama araçlarından faydalanılmıştır.

Bu çalışmadaki birinci araştırma sorusunda, kelime miktarı ile ortalama ve ortanca değerler üzerinden oluşturulmuş olan kaynak inceleme süresi ve kelime başına okuma süresi değerleri arasındaki ilişki incelenmiştir. Korelasyon testleri doğrultusunda ortaya çıkarılan bulgular, ortalama temel alınarak oluşturulan kaynak inceleme süreleri ile ziyaret edilen sayfalardaki kelime miktarları arasında pozitif yönde güçlü bir ilişkiyi göstermiştir ($r_s=0,772$). Aynı sonuç, ortanca ile oluşturulmuş olan kaynak inceleme sürelerinde de gözlemlenmiştir ($r_s=0,730$). Diğer taraftan, kullanıcıların, ortalama temel alınarak oluşturulmuş kelime başına okuma süreleri ile web sayfalarındaki kelime miktarları arasında negatif yönde güçlü bir ilişki

tespit edilmiştir ($r_s: -0,775$). Bu durum, ortanca temel alınarak oluşturulan kelime başına okuma süreleri için de benzer şekilde ortaya çıkmıştır ($r_s: -0,808$). Kelime başına okuma süresi değişkeni göz önüne alındığında, kullanıcıların, kelime miktarının artması durumunda tek tek her bir kelimeyi okumak yerine sayfayı tarama davranışına yöneldikleri rahatlıkla ifade edilebilir.

Çalışmanın ikinci ve üçüncü sorularında temel olarak, kullanıcıların, ihtiyaç duydukları bilgiyle ilgili olan web sayfalarına yaptıkları ilk ve ikinci ziyaretlerdeki/temaslarındaki kaynak inceleme süreleri ve kelime başına okuma sürelerinin ne derece değişkenlik gösterdiğine odaklanılmıştır. Bu değişkenliğin tespit edilebilmesi amacıyla, Mann-Whitney U testiyle kıyaslama işlemi gerçekleştirilmiştir. Kaynak inceleme sürelerine yönelik olarak gerçekleştirilen analiz sonuçları, kullanıcıların, ilk temaslarında daha uzun süre sayfada kaldıklarını göstermiştir (Z: -4,568; $p < 0,05$). Kelime başına okuma süreleri verisine yönelik olarak, ikinci temaslarda, sayfaların daha hızlı bir şekilde tarandığı gözlemlenmiştir (Z: -4,889; $p < 0,05$). İkinci ziyaretlerde tespit edilen kaynak inceleme süresi ve kelime başına okuma süresi değişkenlerindeki düşüşün, kullanıcıların önceki deneyimlerinden kalan bir bilgiye sahip olmalarından kaynaklandığı belirtilebilir. İlk temaslarda yeterli bilginin alınamaması ya da alınan bilginin unutulması sonrasında, ikinci temaslarda sadece eksik hissedilen bilgiye odaklanıldığı ve dolayısıyla sayfada daha az süre kaldığı söylenebilir.

Bu bilgiler ışığında, geçici bilgi ihtiyacının giderilmesi için hizmet veren sistem ve altyapılarda, kullanıcılar aradığı bilgiye erişirken, bilgi kaynağının mümkün olduğunca az miktarda kelimededen oluşturulmasının ve ilgili konuyla alakalı asıl önemli olan bilginin paylaşılmasının önemli olduğu düşünülmektedir. Bu sayede, kullanıcıların ikinci kez sayfaları ziyaret etme oranları düşürülerek ilk temaslarda tatmin edilmeleri sağlanabilir. Bu yöntemin, daha verimli bir bilgi erişim sürecinin oluşmasına da katkı sunacağı düşünülmektedir. Diğer taraftan, kaynak inceleme süresi ve kelime başına okuma süresi değerlerinin hem kelime miktarlarına bağlı olarak hem de aynı sayfaya yapılan ikinci temaslarda değişkenlik gösterdiği çalışma kapsamında ortaya çıkartılmıştır. Bu bulgu, “kaynak inceleme süresi temel alınarak ilgili bilgi kaynaklarının bulunması” temasıyla yapılması muhtemel ileriki çalışmalar için oldukça dikkat gerektiren bir noktaya ışık tutmuştur.

Bilgi ihtiyacı, kişiden kişiye değiştiği gibi ihtiyaç türünün uzun ya da geçici olma durumuna göre de farklılıklar içerebilir. Bu çalışmada ortaya çıkarılan bulguların tamamı geçici bir bilgi ihtiyacının giderilme sürecindeki etkileşimleri kapsamaktadır. Uzun vadeli bilgi ihtiyacına odaklanılarak gerçekleştirilecek olan çalışmalarda daha farklı sonuçların ortaya çıkması muhtemeldir. Uzun vadeli bilgi ihtiyacına, bir kişinin herhangi bir konu hakkında uzmanlaşmak istemesi doğrultusunda yapabileceği arama etkileşimleri örnek gösterilebilir. Uzmanlaşmak, zaman ve emek gerektireceğinden, uzun vadeli bir bilgi ihtiyacının ön plana çıkması ve bu durumda da, kullanıcının aradığı bilgiyle ilgili kaynakları incelerken daha uzun süre zaman harcaması yüksek ihtimaldir. Böyle bir senaryoda, kaynak inceleme süresinin bu çalışmadaki bulgulara benzer bir şekilde kelime miktarıyla doğru orantılı artış göstereceği, fakat kelime başına okuma süresindeyse, bu çalışmada ortaya çıkarılan değişkenliğin daha düşük olacağı tahmin edilmektedir. Sonuç olarak, bilgi gereksinimimizi gidermek amacıyla faydalandığımız BES’lerin bilgi ihtiyaçlarındaki farklılıklara mümkün olduğunca başarılı bir şekilde cevap verebilmesi önemlidir. Bu noktada, kullanıcı okuma davranışlarındaki

değişimlerin dikkatlice ele alınması ve bu değişimlerden ortaya çıkarılacak olan desenlerle değerlendirme ve geliştirme süreçlerinin yürütülmesinin önemli olduğu düşünülmektedir.

Kaynakça

- Agichtein, E., Brill, E. ve Dumais, S. (2019). Improving Web Search Ranking by Incorporating User Behavior Information. *SIGIR Forum*, 52(2), 11–18. <https://doi.org/10.1145/3308774.3308778>
- Akuma, S., Iqbal, R., Jayne, C. ve Doctor, F. (2016). Comparative analysis of relevance feedback methods based on two user studies. *Computers in Human Behavior*, 60, 138–146. <https://doi.org/10.1016/J.CHB.2016.02.064>
- Akuma, S., Jayne, C., Iqbal, R. ve Doctor, F. (2014). Implicit Predictive Indicators: Mouse Activity and Dwell Time. L. Iliadis, I. Maglogiannis ve H. Papadopoulos (Ed.), *Artificial Intelligence Applications and Innovations* (ss. 162–171). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-44654-6_16
- Baeza-Yates, R. ve Ribeiro-Neto, B. (2011). *Modern Information Retrieval: The concepts and technology behind search* (2. Baskı). Addison-Wesley Publishing Company.
- Balakrishnan, V. ve Zhang, X. (2014). Implicit user behaviours to improve post-retrieval document relevancy. *Computers in Human Behavior*, 33, 104–112. <https://doi.org/10.1016/J.CHB.2014.01.001>
- Croft, B., Metzler, D. ve Strohman, T. (2009). *Search Engines: Information Retrieval in Practice* (1. Baskı). Pearson.
- Guo, Q. ve Agichtein, E. (2012). Beyond Dwell Time: Estimating Document Relevance from Cursor Movements and Other Post-Click Searcher Behavior. *Proceedings of the 21st International Conference on World Wide Web*, 569–578. <https://doi.org/10.1145/2187836.2187914>
- Johnson, J. (2021). *Search engine market share worldwide*. Statista. <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>
- Kelly, D. ve Belkin, N. J. (2001). Reading Time, Scrolling and Interaction: Exploring Implicit Sources of User Preferences for Relevance Feedback. *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 408–409. <https://doi.org/10.1145/383952.384045>
- Kelly, D. ve Belkin, N. J. (2004). Display Time as Implicit Feedback: Understanding Task Effects. *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 377–384. <https://doi.org/10.1145/1008992.1009057>
- Kim, J. Y., Teevan, J. ve Craswell, N. (2016). Explicit In Situ User Feedback for Web Search Results. *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 829–832. <https://doi.org/10.1145/2911451.2914754>
- Kim, Y., Hassan, A., White, R. W. ve Zitouni, I. (2014). Modeling Dwell Time to Predict Click-Level Satisfaction. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, 193–202. <https://doi.org/10.1145/2556195.2556220>
- Kowalski, G. (2011). *Information Retrieval Architecture and Algorithms*. Springer US. <https://doi.org/10.1007/978-1-4419-7716-8>
- Kwok, K. L., Grunfeld, L. ve Deng, P. (2007). Employing web mining and data fusion to improve weak ad hoc retrieval. *Information Processing & Management*, 43(2), 406–419. <https://doi.org/10.1016/J.IPM.2006.07.008>
- Liu, J. ve Belkin, N. J. (2015). Personalizing information retrieval for multi-session tasks: Examining the roles of task stage, task type, and topic knowledge on the interpretation of dwell time as an

- indicator of document usefulness. *Journal of the Association for Information Science and Technology*, 66(1), 58–81. <https://doi.org/10.1002/asi.23160>
- Manning, C. D., Raghavan, P. ve Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.
- Mao, J., Liu, Y., Luan, H., Zhang, M., Ma, S., Luo, H. ve Zhang, Y. (2017). Understanding and Predicting Usefulness Judgment in Web Search. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1169–1172. <https://doi.org/10.1145/3077136.3080750>
- Morita, M. ve Shinoda, Y. (1994). Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval. *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 272–281.
- Núñez-Valdéz, E. R., Cueva Lovelle, J. M., Sanjuán Martínez, O., García-Díaz, V., Ordoñez De Pablos, P. ve Montenegro Marín, C. E. (2012). Implicit feedback techniques on recommender systems applied to electronic books. *Computers in Human Behavior*, 28(4), 1186–1193. <https://doi.org/10.1016/J.CHB.2012.02.001>
- Internet Live Stats. (2021). *Google Search Statistics*. <https://www.internetlivestats.com/google-search-statistics/>
- Uçak, N. Ö. ve Güzeldere, Ş. O. (2006). Bilişsel Yapının ve İşlemlerin Bilgi Arama Davranışı Üzerine Etkisi. *Türk Kütüphaneciliği*, 20(1), 7–28.
- White, R. W. ve Kelly, D. (2006). A Study on the Effects of Personalization and Task Information on Implicit Feedback Performance. *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, 297–306. <https://doi.org/10.1145/1183614.1183659>
- Xu, S., Jiang, H. ve Lau, F. C. M. (2011). Mining User Dwell Time for Personalized Web Search Re-Ranking. *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Three*, 2367–2372.
- Yi, X., Hong, L., Zhong, E., Liu, N. N. ve Rajan, S. (2014). Beyond Clicks: Dwell Time for Personalization. *Proceedings of the 8th ACM Conference on Recommender Systems*, 113–120. <https://doi.org/10.1145/2645710.2645724>
- Zhai, C. ve Massung, S. (2016). *Text data management and analysis: a practical introduction to information retrieval and text mining*. Association for Computing Machinery.