



Doğal Dilde Yapılan Türkçe Sorgularda Google Arama Motoru'nun Performans Değerlendirmesi

Performance Evaluation of Google Search Engine in Turkish Natural Language Queries

Meltem DİŞLİ

Makale Bilgisi / Article Information

Bu makaleye atıf yapmak için/ To cite this article:

Dişli, M. (2020). Doğal dilde yapılan türkçe sorgularda google arama motoru'nun performans değerlendirilmesi. *Bilgi Dünyası*, 21(2), 247-267. doi: 10.15612/BD.2020.554

Makale türü / Paper type: Hakemli / *Refereed*

Araştırma Makalesi / *Research Article*

Doi: 10.15612/BD.2020.554

Geliş Tarihi / Received: 24.08.2020

Kabul Tarihi / Accepted: 18.12.2020

Elektronik Yayınlanma Tarihi / Online Published: 31.12.2020

İletişim / Communication

Üniversite ve Araştırma Kütüphanecileri Derneği / University and Research Librarians Association

Posta Adresi / Postal Address: Marmara Sok. No:38/17 06420 Yenışehir, Ankara, TÜRKİYE/TURKEY

Tel: +90 312 430 03 61; Faks / Fax: +90 312 430 03 61; E-posta / E-mail: bilgi@bd.org.tr

Web: <http://www.bd.org.tr/index.php/bd/index>

Doğal Dilde Yapılan Türkçe Sorgularda Google Arama Motoru'nun Performans Değerlendirmesi

Meltem Dişli* 

Öz

Araştırmada Türkçe doğal cümlelerle yapılan aramalarda Google arama motorunun performansını değerlendirmek amaçlanmaktadır. Bu doğrultuda Google'a 2 farklı tipte (maddi sorular - diğer sorular) 20 soru sorulmuş ve ilk 10 sonuç duyarlık ve normalize sıralama değerlerine göre değerlendirilmiştir. Verilerin analizinde Mann-Whitney U, Spearman ve Ki-kare testleri kullanılmıştır. Araştırmanın sonucunda Google'ın Türkçe doğal dilde yapılan aramalara getirdiği sonuçların ilgililik düzeyi ortalama düzeyde bulunmuştur. Soru tipine göre ilgililik düzeylerinin farklılık göstermediği ancak ilgililik sıralamasının farklılık gösterdiği ortaya çıkmıştır. Maddi sorularda diğer sorulara göre ilgili dokümanlar daha üst sıralarda yer almıştır. Ayrıca Google ilgili dokümanları üst sıralarda göstermekte başarılı bulunmuştur. Çoğunlukla ilk 5 sonuçta ilgili sonuçlara ulaşılabildiği ortaya çıkmıştır. Soruların hemen hemen yarısında yanıt kutusu (snippet) gelmiş ve bunların büyük çoğunluğunun sorgu ile ilgili olduğu saptanmıştır. Sonuç sayfasında gelen videoların ise genellikle sorgu ile ilgisiz olduğu ortaya çıkmıştır. Elde edilen sonuçlarının Google'ın geliştirilmesi amacıyla yapılan çalışmalara fayda sağlayacağı düşünülmektedir.

Anahtar sözcükler: Bilgi erişim, Semantik arama, Google arama motoru, Duyarlık, Normalize sıralama, Normalize edilmiş anma değeri, Doğal dil.

* Arş. Gör., Hacettepe Üniversitesi, Edebiyat Fakültesi, Bilgi ve Belge Yönetimi Bölümü, meltem.disli@hacettepe.edu.tr

Performance Evaluation of Google Search Engine in Turkish Natural Language Queries

Meltem DİŞLİ* 

Abstract

The aim of the research is to evaluate the performance of the Google search engine on the queries made with Turkish natural sentences. In this direction, Google was asked 20 questions of 2 different types (factoid questions - other questions) and the first 10 results were evaluated according to precision and general normalized recall measure. Mann-Whitney U, Spearman and Chi-square tests were used in the analysis of the data. As a result of the research, the relevance level of the results brought by the Google in regards to Turkish natural language queries were found to be average. It was revealed that the relevance levels did not differ according to the question type, but the relevance ranking did differ. Compared to other questions, relevant documents ranked higher for factoid questions. In addition, Google has been successful in showing relevant documents at the top. It turned out that mostly in the first 5 results relevant results can be reached. A snippet was brought for almost half of the questions and the majority of these were determined to be relevant with the query. Videos brought up in the result pages were revealed to be mostly irrelevant to the query. It is thought that the results obtained will benefit the studies carried out to improve Google.

Keywords: Information retrieval, Semantic search, Google search engine, Precision, General normalized recall measure, Natural language.

* Research Assistant, Hacettepe University, Faculty of Letters, Department of Information Management, meltem.disli@hacettepe.edu.tr

Giriş

İnsanların gereksinim duyduğu bilgiye en hızlı şekilde erişmek istemesi arama motorlarının kullanımının yaygınlaşmasına ve günlük yaşamda öneminin artmasına sebep olmuştur. Arama motorları kullanıcı sorguları ile dizinlenen kaynakları karşılaştırarak kısa bir süre içinde kullanıcı ile bilgiyi buluşturan, bilgi erişim sistemlerini temel olarak oluşturulmuş bilgisayar yazılımlarıdır (Kesen, Şenol ve Yanar, 2008, s. 141-142). Bilgi erişim sistemleri ilgili dokümanların tümüne erişirken, ilgisiz dokümanları ayıklayarak yalnızca ilgili dokümana erişimi mümkün kılmalıdır (Tonta, 1995, s. 302). Mevcut arama motorlarında ise ilgili tüm dokümanlara erişmek mümkün olmamakla beraber ilgisiz dokümanlara da erişilmektedir. Arama motorlarında birçok bilgi erişim modeli kullanılsa da çoğunlukla kullanıcının sorgu terimleri ile dokümanlarda geçen sözcükler karşılaştırılmakta ve sözcüklerin kaç kere geçtiği dokümanın sorgu ile ilgi derecesini göstermektedir (Amasyalı ve Diri, 2005, s. 1). Arama motorları sözcük sıklıkları dışında ziyaret sıklıkları ve diğer sitelerden verilen referanslara göre internet sitelerini sorgu ile ilişkilendirerek sonuçları kullanıcıya sunmaktadır (Yozkat, 2017, s. 74). Kullanıcıya sorgu terimleri ile ilgili olabilecek dokümanlar belirli kriterlere göre sıralanarak gösterilmektedir. Ancak aradığı bilgiyi ilgili dokümanlardan bulup çıkarmak çoğunlukla kullanıcıya düşmektedir (Amasyalı ve Diri, 2005, s. 2). İnternette yer alan bilgi miktarının her geçen gün artması ve insanların bilgiye ulaşmakta daha da sabırsız hale gelmesi arama motorlarını uzun doküman listeleri sunmak yerine kullanıcı sorgusuna doğrudan yanıt verme çabasına doğru yönlendirmiştir. Bu durum soru-yanıt sistemlerini ortaya çıkarmış, arama motorları da soru yanıtlama hizmeti vermek için çalışmalarda bulunmuştur.

Doğal dille sorulan sorulara yanıt veren soru yanıtlama sistemleri Web’de yer alan bilgi miktarı arttıkça daha da önemli bir hale gelmiştir. Soru yanıtlama sistemi, doğal dilde sorulan sorulara yanıt vermeyi amaçlayan bir bilgi erişim sistemidir (Chaoticity, 2010). Bir soru yanıtlama sisteminde doğal dilde sorulan sorulara doğru yanıtlar verebilmek için yapay zeka, doğal dil işleme, istatistiksel analiz, örüntü eşleştirme, bilgi alma ve bilgi çıkarma teknikleri kullanılmalıdır (Dwivedi ve Singh, 2013, s. 418). Arama motorları da soru yanıtlama hizmetleri verebilmek için bu özelliklere sahip olmalıdır. Zadeh’e göre (2006, s. 163-164) arama motorlarının soru yanıtlama sistemlerine dönüştürülebilmesinin önünde dünya bilgisi, ilgililik ve algılama temelli bilgiden çıkarsama yapma olmak üzere 3 engel bulunmaktadır. Bu 3 engelin de altında yatan temel sorun doğal dil anlama sorunudur. Günümüzde arama motorları semantik arama özellikleri ile doğal dili daha iyi anlamlandırmakta ve soru yanıtlama hizmetlerini de geliştirmektedirler. Semantik arama motorları sonuçları yalnızca kullanıcı sorgusu ile dokümanlarda geçen terimlerin sıklıklarına göre getirmemekte, yapay zekâ sistemleri kullanarak kullanıcının ne aramak istediğini anlamaya çalışmaktadır. En çok kullanılan arama motoru olan Google (Alexa, 2020) da 2015 yılında geliştirdiği yapay zekâ sistemi ile doğal dili anlamaya bir adım daha yaklaşmıştır.

Google arama motoru Web'i daha iyi tarayıp indeksleyerek internetteki bilgilere daha güvenilir ve daha hızlı ulaşmayı vadeden algoritması PageRank ile ilk olarak 1998 yılında ortaya çıkmıştır (Brin ve Page, 1998, s. 107). PageRank akademik alanda kullanılan atıf analizi yöntemine benzer olarak Web'in atıf (link) grafiğini çıkarmakta ve sık kullanılan linklerin popüler sayfalara ait olduğu düşüncesiyle bu linkleri sonuçlarda üst sıralarda getirmektedir (Brin ve Page, 1998, s. 110). 2012 yılında ise Knowledge Graph özelliği ile Google soru yanıtlama hizmeti vermeye başlamıştır. Knowledge Graph ile Google yalnızca sorgu terimi ile eşleşen sözcükleri aramakla kalmamakta, aynı zamanda sorgu teriminin ne olduğunu da anlamaktadır. Google Knowledge Graph ile sorguları daha iyi anlayabilmekte, böylece konu ile ilgili özet bilgiler de gösterebilmektedir. Knowledge Graph sadece terimleri dizinlemekte, ayrıca tüm nesnelerin birbirleri ile ilişkilerini de tutmaktadır. Bu özellik Google'a aranılan kişi ile ilgili bilgileri sunarken, bağlantılı olduğu kişilerle ilgili bilgileri de sunmayı sağlamaktadır. Google Knowledge Graph ile kullanıcıların sorgularını da tutmakta, bir sorgu geldiğinde henüz sorulmadan bir sonraki soruyu tahmin ederek bunun cevabını da vermektedir (Singhal, 2012). Google 2015 yılında RankBrain algoritmasını kullanmaya başlayarak sorgulara çok daha iyi yanıtlar vermeyi hedeflemiştir. RankBrain, PageRank gibi Google'ın Hummingbird isimli genel arama algoritmasının bir parçasıdır. RankBrain Google'ın arama sonuçlarına yardımcı olmak için geliştirdiği bir yapay zekâ sistemidir. Makine öğrenimi sayesinde daha önceden yapılmış sorguları kullanarak yapılan yeni sorguları anlamaya çalışmakta ve birçok unsura bakarak sonuçları sıralamaktadır (Sullivan, 2016). Google hâlâ tam olarak doğal dil temelli aramalardan anlam çıkaramamaktadır. RankBrain'in yaptığı, sorgu ile ilgili yanıtları getirmek için ilişki veri tabanı ve benzer sorgular arasındaki ilişki vektörlerini kullanmaktır. Google'a henüz bilmediği bir soru sorulduğunda sorguyla en uyumlu bilgileri ilişki veri tabanını kullanarak bulabilmektedir. Google, doğal dili tam olarak anlayamasa da sürekli olarak öğrenmeye devam eden RankBrain algoritması ile semantik arama özelliği ve soru yanıtlama hizmetini her geçen gün daha da iyileştirmektedir.

Son yıllarda bilgi erişim sistemlerinde gerçekleşen gelişmeler insanların bilgi arama davranışlarını da değiştirmiştir. Artık insanlar bilgi gereksinimini karşılamak için doğru anahtar sözcüklerin seçilmesi veya gelişmiş arama stratejilerinin uygulanması gibi yöntemlerle uğraşmak istememekte; bilgiye doğrudan ve doğal cümleler ile arama yaparak ulaşmak istemektedirler. Bu sebeple doğal dilin arama motorları tarafından ne kadar etkili bir şekilde anlaşıldığının ve sorgulara ne kadar ilgili yanıtlar getirildiğinin analiz edilmesi arama motorlarının performanslarının artırılması açısından son derece önemlidir. Bunun yanı sıra doğrudan yanıt almak isteyen kullanıcılara uzun doküman listelerinden oluşan sonuçlar getirmek yerine yanıt kutuları ile doğru yanıtı gösteren arama motorları bilgi erişim sürecini oldukça kısaltmaktadır. Bu çalışmada Google'ın semantik arama özelliği ile soru yanıtlama hizmetini değerlendirmek amaçlanmaktadır. Çalışma kapsamında Google'ın Türkçe doğal cümlelerle sorulan soruları anlamlandırması ve ilgili sonuçları getirmesi, ilgili sonuçları üst sıralarda göstermesi ve yanıt kutularında doğru yanıtların verilmesi açılarından performansı değerlendirilmiştir.

İlgili çalışmalar

İnternetteki bilgi patlaması arama motorlarının önemini her geçen gün artırmaktadır. Arama motorlarının değerlendirilmesi ve sorunların saptanması, geliştirilebilmeleri için son derece önemlidir. Bu kapsamda ulusal ve uluslararası literatürde arama motorlarının çeşitli yönlerden performansını değerlendiren birçok çalışma yapılmıştır.

Arama motorlarının performansının değerlendirildiği bir çalışmada (Soydal, 2000) Alta Vista, Excite, Hotbot, Infoseek ve Northern Light adlı arama motorları karşılaştırılmıştır. Arama motorlarına 11 soru yöneltilmiş ve elde edilen ilk 20 sonuç anma, duyarlık ve çakışma oranları açısından değerlendirilmiştir. Arama motorlarının anma ve duyarlık değerleri düşük bulunmuş, ilk 10 sonucun ilk 20 sonuca göre daha ilgili dokümanlar getirdiği ve arama motorları arasında çakışma oranlarının düşük olduğu sonuçlarına ulaşılmıştır. Tonta, Bitirim ve Sever (2002) ise Arabul, Arama, Netbul ve Superonline arama motorlarının çeşitli ölçütlere göre performanslarını değerlendirmişlerdir. Bu kapsamda arama motorlarına 17 soru sorulmuş ve arama motorlarının çeşitli kesme noktalarındaki duyarlık ve normalize sıralama değerleri, ölü bağlantı sayıları, kapsama ve yenilik oranları belirlenmiştir. Kruskal-Wallis ve Mann-Whitney testleri ile arama motorlarının güncellik, duyarlık, normalize sıralama, kapsama ve yenilik oranlarının farklılık gösterip göstermediği araştırılmıştır. Arama motorlarında her 6 dokümandan 1'inin ölü bağlantı içerdiği, ortalama 6 dokümandan 5'inin ilgisiz olduğu, arama motorlarının ilgili dokümanları üst sıralarda göstermediği, arama motorlarının Türkiye adresli dokümanları çok dizinlemedikleri gibi sonuçlara ulaşılmıştır. Arabul, Arama ve Netbul Türkçe arama motorlarında gövdeleme algoritması kullanımının etkisini ölçmek üzerine yapılan bir çalışmada (Zeden, Ayalp, Bitirim ve Sever, 2006) Türkçe sözcükler gövdeli ve gövdesiz biçimlerde aranmıştır. Elde edilen sonuçlar ilgili ve ilgisiz olarak sınıflandırıldıktan sonra duyarlık ve normalize sıralama değerleri ölçülmüştür. Gövdeli sorguların duyarlık değerleri daha yüksek bulunmuş ve arama motorlarında gövdelemenin önemine vurgu yapılmıştır. Demirci, Kışmir ve Bitirim (2007) Türkçe dokümanlara erişim açısından Google, Yahoo, Msn, AlltheWeb ve Ask arama motorlarını duyarlık ve normalize sıralama değerlerine göre karşılaştırmıştır. Araştırma sonucunda Türkçe dokümanlara erişmekte en başarılı arama motorunun Google olduğu, yerel arama motorlarının ise Türkçe dokümanlara erişme performansının düşük olduğu ortaya çıkmıştır. Ayrıca araştırmada Google'ın hem duyarlık hem de normalize sıralama değerlerinde en yüksek performansı ilk 5 sonuçta gösterdiği belirlenmiştir. Kesen ve diğerlerinin (2008) çalışmasında Google Scholar ve Scirus arama motorlarının Türkçe karakterli aramalara getirdiği sonuçları değerlendirilmiş, anahtar sözcükler hem Türkçe karakterleri hem de en uygun İngilizce karakterleri ile aranmıştır. Google Scholar'da Türkçe karakterli aramalar ve İngilizceleştirilerek yapılan aramalarda erişilen dokümanlar arasında farklar bulunmakta; bazen Türkçe karakterli aramalarda daha çok sonuca ulaşıırken, bazen İngilizceleştirilen aramalarda daha çok sonuca ulaşılmıştır. Ancak Scirus'ta böyle bir fark 2 sonuç dışında gözlenmemiştir. Bir başka çalışmada (Yozkat, 2017) Google'da ilk sayfalarda

erişilen sonuçların güvenilirliği analiz edilmiştir. Bu kapsamda 3 anahtar sözcük belirlenmiş ve anahtar sözcüklere Google'ın getirdiği sonuçlar incelenmiştir. Araştırmada sonuç sayfasında öncelikle popüler sitelerin yer aldığı ve akademik anlamda güvenilir olmadıkları sonucuna ulaşılmıştır. Azeez (2017) çalışmasında Google, Yahoo ve Bing arama motorlarının Arapça aramalara yönelik olarak değerlendirilmesi için bir anket yapmış; Google'ın metin ve video aramalarında Yahoo ve Bing arama motorlarından daha yüksek performans gösterdiği, resim aramalarında ise Yahoo'nun daha doğru sonuçlar getirdiği sonuçlarına ulaşmıştır.

Arama motorlarını İngilizce dışındaki diller için değerlendiren Bar-Ilan ve Gutman'ın (2005) çalışmalarında AltaVista, Fast, Google ve yerel arama motorlarında Rusça, Fransızca, Macarca ve İbranice dillerinin özelliklerini temsil edebilecek anahtar sözcükler ile arama yapılmıştır. Araştırmada genel arama motorlarının dilin kendine has özelliklerini dikkate almadığı sonucuna ulaşılmıştır. Tümer, Shah ve Bitirim'in (2009) çalışmasında Google, Yahoo, Msn ve Hakia arama motorlarının semantik arama performansı değerlendirilmiştir. Araştırmada çeşitli konularda 10 soru sorulmuş ve ilk 20 sonuç ilgili ve ilgisiz olarak gruplandırılmıştır. Arama motorları çeşitli kesme noktalarında görelî anma ve duyarlık değerleri açısından karşılaştırılmıştır. Yahoo duyarlık oranında en iyi performansı gösterirken, Google görelî anma oranında en iyi performansı göstermiştir. Tüm arama motorlarının semantik arama özelliklerinin düşük olduğu sonucuna ulaşılmıştır. Bir başka çalışmada (Kumar ve Prakash, 2009) kütüphane ve bilgilendirme alanındaki terimlerden oluşan sorgulara yönelik olarak Google ve Yahoo'nun performansı karşılaştırılmıştır. Bu çalışmada arama motorlarına tek sözcüklük sorgular, basit çok sözcüklük sorgular ve karmaşık çok sözcüklük sorgular olmak üzere 3 tip soru sorulmuştur. Arama motorlarının sonuçları anma ve duyarlık değerleri ile karşılaştırılmıştır. Google'ın basit çok sözcüklük ve tek sözcüklük sorgular için anma ve duyarlık değerleri daha yüksek iken, Yahoo'nun karmaşık çok sözcüklük sorgular için anma ve duyarlık değerleri daha yüksek bulunmuştur. Benzer bir çalışmada (Usmani, Pant ve Bhatt, 2012) bilgisayar bilimleri alanıyla ilgili sorgularda Google ve Bing arama motorlarının gösterdiği performans karşılaştırılmıştır. Bu çalışmada tek sözcüklük sorgularda Google'ın anma ve duyarlık değerleri daha yüksek çıkmıştır. Basit çok sözcüklük ve karmaşık çok sözcüklük sorgular için ise duyarlık değerleri Bing'in daha yüksekken, görelî anma değerleri Google'ın daha yüksek bulunmuştur. Bilal (2012) arama motorlarının performansını çocukların bilgiye erişimi açısından değerlendirmiştir. Araştırma kapsamında yapılan 30 sorguya Google, Yahoo, Bing, Yahoo Kids ve Ask Kids'te getirilen sonuçlar karşılaştırılmıştır. Araştırma sonucunda doğal dilde yapılan sorgular için en iyi performansı Google'ın sergilediği ortaya çıkmıştır. Khan, Sangroha, Ahmad ve Rahman'ın (2014) çalışmasında semantik ve anahtar sözcük tabanlı arama motorlarının performansları değerlendirilmiştir. Bing, DuckDuckGo, Lexxe, Google ve Yahoo'nun değerlendirildiği çalışmada arama motorlarına 10 soru sorulmuş ve sonuçların duyarlık değerleri hesaplanmıştır. Bu çalışmada en fazla ilgili dokümana Bing arama motoru erişmiştir. Bing'i DuckDuckGo takip ederken, Google'ın duyarlık oranının da %60'larda olduğu görülmektedir. Ayrıca araştırma sonu-

cunda semantik arama motorlarının anahtar sözcük tabanlı arama motorlarından daha iyi performans gösterdiği sonucuna ulaşılmıştır. Semantik ve anahtar sözcük tabanlı arama motorlarının karşılaştırdığı bir başka çalışmada (Mala ve Lobiyal, 2016) arama motorlarına sorular sorulmuş ve sonuçlar ilgili/ilgisiz olmak üzere sınıflandırılmıştır. Anahtar sözcüğe dayalı arama motorlarında daha fazla ilgili dokümana erişilmiş, semantik aramanın geliştirilmesi gerektiği ortaya çıkmıştır. Google, Yahoo, Wikipedia, Halka, Bing ve DuckDuckGo arasında en yüksek duyarlık oranları Google ve DuckDuckGo arama motorlarında görülmüştür. Bitirim ve Görür (2017) Google, Yahoo, Bing ve Ask arama motorlarının Türkçe dokümanlara erişim açısından mevcut performanslarını, altı yıl önce ölçülen performansları ile karşılaştırarak değerlendirmektedir. Google'ın altı yıl öncekinden daha kötü bir performans gösterdiği, diğer arama motorlarının ise daha iyi olduğu sonucuna ulaşılmıştır. Tüm arama motorları en yüksek performansı ilk 5 sonuçta göstermiştir. Bir diğer çalışmada (Zhao, Zhang, Xia ve Le, 2019) ise Google soru yanıtlama kalitesi bağlamında değerlendirilmiştir. Çalışmada arama motorunun soru yanıtlama kalitesini değerlendirmek için bir ölçek geliştirilmiş ve Google tarafından getirilen yanıt kutuları incelenmiştir. Soruları 4 hedef türü (kişi, kurum, nesne, olay) ve 6 soru tipi (ne, ne zaman, nerede, kim, nasıl sorularından oluşan maddi sorular ve listeleme soruları) altında grupladıktan sonra sonuçlar bu türlere göre Anova ve Tukey testleri kullanılarak karşılaştırılmıştır. Google'ın kişi ile ilgili sorulara diğer hedef türlerine (kurum, nesne, olay) göre daha kaliteli yanıtlar getirdiği, soru tiplerinde ise en iyi yanıtları nerede sorularına getirdiği sonuçlarına ulaşılmıştır. Gul, Ali ve Hussain (2020) çalışmalarında Google, Yahoo ve Bing arama motorlarının yaşam bilimi ve biyotıp alanında yapılan sorgulara yönelik performanslarını değerlendirmişlerdir. Sorgu terimleri, Web of Science'ta bu alanlarda kullanılan anahtar sözcükler içinden seçilmiştir. Bir sözcüklük, 2 sözcüklük ve 3 sözcüklük sorgular gerçekleştirilmiştir. Sonuçların değerlendirilmesinde anma ve duyarlık parametreleri kullanılmıştır. En yüksek duyarlılık ve anma değerleri Google'da çıkmıştır.

Yöntem

Bu çalışmada Google'ın Türkçe doğal dilde sorulan soruları yanıtlayabilme performansı değerlendirilmektedir. Bilgi erişim sistemlerinin performans değerlendirmesinde anma, duyarlık ve posa ölçütleri kullanılmaktadır. Duyarlık (kesin isabet) erişilen ilgili dokümanların erişilen tüm dokümanlara oranını, anma (erişim isabeti) erişilen ilgili dokümanların koleksiyondaki tüm ilgili dokümanlara oranını, posa ise erişilen ilgisiz dokümanların koleksiyondaki tüm ilgisiz dokümanlara oranını vermektedir (Tonta ve diğerleri, 2002, s. 23-24). Arama motoru değerlendirmesinde anma değerinin ölçülmesi çok zordur. Çünkü milyonlarca dokümanın incelenip ilgili olup olmadığını saptamak uygulamada mümkün değildir (Soydal, 2000, s. 6). Posa değerinde de aynı şekilde koleksiyondaki tüm dokümanları inceleyip ilgisiz olanların saptanması teknik olarak imkânsızdır. Bu sebeple araştırma kapsamında Google'ın sorulan sorulara getirdiği sonuçlar duyarlık ölçütü ile değerlendirilmiştir. Araştırmada Google'a sorular sorulmuş; gelen sonuçlar

İlgili ve ilgisiz olmak üzere 2 kategoride sınıflandırılmıştır. Tüm sonuçları incelemek mümkün olmadığından araştırma ilk 10 sonuç ile sınırlandırılmıştır. Bu durumda duyarlık değeri hesaplanırken erişilen tüm dokümanlar 10 olarak belirlenmiştir.

Bir arama motorunun performansı eriştiği ilgili doküman sayısı kadar, bu dokümanları ilk sıralarda göstermesiyle de yakından ilişkilidir. Çünkü kullanıcı ilk sıralarda ilgili dokümanlara erişemezse sonraki sonuçlara bakmaktan vazgeçebilmektedir. Kullanıcı ilk sıralarda yer almadığı sürece ilgili dokümanlara erişse de fazla zaman kaybettiği için memnun olmayacaktır. Tam tersine, arama motoru çok az ilgili dokümana erişse de bu dokümanları ilk sıralarda sunduysa kullanıcının ihtiyacını kısa sürede karşılayacaktır. Arama motorlarının dokümanları ilgililik derecesine göre sıralama performansını değerlendirmek için normalize sıralama ölçütü kullanılmaktadır (Tonta ve diğerleri, 2002, s. 26). Normalize sıralama ölçütü diğer adıyla "genel normalize edilmiş anma değeri" (general normalized recall measure: R_{norm}) aşağıdaki formülle hesaplanmaktadır (Bollmann, 1983, s. 123; Akbulut, Tonta ve White, 2020, s. 966).

$$R_{norm}(\Delta) = 0.5 * [1 + (R^+ - R^-) / R^+_{max}]$$

$R_{norm}(\Delta)$ erişim çıktısı, R^+ ilgili dokümanların ilgisiz dokümanlardan önce gösterildiği doküman çiftleri sayısı, R^- ilgisiz dokümanların ilgili dokümanlardan önce gösterildiği doküman çiftleri sayısı, R^+_{max} ise maksimum R^+ sayısıdır. Bu araştırma kapsamında Google'ın ilgili dokümanları üst sıralarda gösterme performansını değerlendirmek amacıyla ilk 5 sonuç ve ilk 10 sonuçta normalize sıralama değerleri hesaplanmıştır. Son olarak normalize sıralama değerleri ile duyarlık değerleri arasındaki ilişki test edilmiştir.

Erişilen ilgili dokümanlar üst sıralarda yer alsın bile günümüzde kullanıcılar sonuçların içinden aradığı yanıtı bulmak istememekte, doğrudan sorusunun yanıtını görmek isteyebilmektedir. Bu sebeple Google'ın performans ölçümünde sonuçlarda yanıt kutusu (snippet) gelme oranı ve bu yanıt kutularının ilgililik düzeyi de incelenmiştir.

Çalışma kapsamında şu araştırma sorularına yanıt aranmıştır:

S1: Google Türkçe doğal dilde sorulan sorularda ilgili dokümanlara erişmekte başarılı mıdır?

S2: Google hangi tip sorularda (maddi sorular/diğer sorular) daha etkin performans göstermektedir?

S3: Google'ın sonuç sayfasında ilgili dokümanlar üst sıralarda gösterilmekte midir?

S4: Maddi sorularda diğer sorulara göre daha mı fazla yanıt kutusu gelmektedir?

S5: Google'ın soru yanıtlama hizmeti ilgili yanıtlar getirmekte başarılı mıdır?

S6: Sonuç sayfasında gelen videolar sorgularla ne kadar ilgilidir?

Araştırmamanın hipotezleri: H1: "Google Türkçe doğal dilde yapılan aramalarda maddi soruları yanıtlamakta daha başarılıdır."; H2: "Google ilk 5 sonuçta ilk 10 sonuçtan daha yüksek performans göstermektedir."; ve H3: "Google'ın soru yanıtlama hizmetindeki performansı düşüktür." şeklinde oluşturulmuştur. Bu araştırmada elde edilen sonuçların Google'ın performansının geliştirilmesi için yarar sağlaması beklenmektedir.

Sınırlılıklar

Araştırma kapsamında Google tarafından getirilen sonuçların hepsini değerlendirmek mümkün olmayacağından değerlendirme ilk 10 sonuç ile sınırlı tutulmuştur. İlk 10 sonucun tam metinleri incelenmiştir. Tam metinde soru ile ilgili bir bilgi vermek yerine ilgili bir sayfaya yönlendiren sonuçlar ilgisiz kabul edilmiştir. Google farklı zamanlarda yapılan sorgulara farklı sıralama ile sonuçlar getirebildiği için ilk sorguya getirilen sonuçlar dikkate alınmıştır. Sonuç sıralamasında en üstte çıkan reklamlar araştırmaya dâhil edilmemiştir.

Soruların Belirlenmesi

Google'a sorulmak üzere 20 soru belirlenmiştir. Soruların belirlenmesinde birkaç unsura dikkat edilmiştir. Öncelikle soruların farklı konuları ve farklı hedef türlerini (kişi, yer, nesne, olay) temsil etmesine özen gösterilmiştir.

Soru yanıtlama sistemlerinin genellikle birkaç sözcüklük ve tek bir doğru cevabı olan maddi soruları yanıtlamak için tasarlandığı bilinmektedir (Soricut ve Brill, 2004). Maddi soruların bile yanıtlanması bir arama motoru için kolay bir iş değildir. Oysaki insanlar yalnızca maddi sorular sormamakta; uzun listelerden oluşan yanıtlar bekledikleri listeleme soruları veya belirli bir cevabı olmayan daha karmaşık sorular da sorabilmektedirler. Araştırma kapsamında sorular belirlenirken yarısının maddi sorular, diğer yarısının ise listeleme veya belirli bir cevabı olmayan sorular olmasına dikkat edilmiştir. Böylece Google'ın Türkçe doğal dilde sorulan maddi sorular ile diğer sorulara getirdiği yanıtların karşılaştırılması mümkün olacaktır. Belirlenen 20 soru aşağıda yer almaktadır.

Maddi sorular:

1. Türkiye'nin 5. Cumhurbaşkanı'nın eşi kimdir?
2. MARC nedir?
3. Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü ne zaman kuruldu?
4. Mansur Yavaş 2014 seçimlerinde kaç oy almıştı?
5. Atatürk'ün annesi nerelidir?
6. Florak nedir?
7. Schengen Vizesi ile İngiltere'ye gidilir mi?
8. 18 yaşından önce içkili mekânlara girebilir miyim?
9. Isparta merkez ile Gönen arası kaç km?
10. Bugün Türkiye'de Covid-19 vaka sayısı kaç?

Diğer sorular;

11. Hangi üniversitelerde Bilgi ve Belge Yönetimi Bölümü var?
12. Kültürel miras konusunda yapılan ulusal akademik yayınlar?
13. Yükseliş dönemindeki Osmanlı Padişahları kimlerdir?
14. Düşük tansiyona ne iyi gelir?
15. Ankara'nın en yeşil semtleri?
16. Büyükler için evde yapılabilecek aktiviteler?
17. Covid-19 nedeniyle durdurulan uçuşlar hangi ülkelere geri açıldı?
18. Türkiye'de günümüzün popüler rapçileri kimlerdir?
19. Spor yapmadan kilo vermenin yolları nelerdir?
20. Atiye mi Hakan Muhafız mı?

Maddi soruların kısa ve tek bir doğru cevabı vardır, arama motorundan beklenen bu bilginin sonuçlarda getirilmesidir. Diğer sorular ise genellikle bir liste cevabı içeren veya doğru bir cevabı olmayan sorulardır. Örneğin; "Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü ne zaman kuruldu?" sorusuna "1972" cevabının getirilmesi beklenmektedir. "Hangi üniversitelerde Bilgi ve Belge Yönetimi Bölümü var?" sorusunda ise Bilgi ve Belge Yönetimi bölümlerinin web sayfalarını değil bölüm olan üniversitelerin listesini getirmesi beklenmektedir. Bu 2 sorunun ayrı ayrı incelenmesinin sebebi Google'ın maddi sorulara ve listeleme sorularına nasıl yaklaştığını görebilmektir. "18 yaşından önce içkili mekânlara girebilir miyim?" sorusunda da tek bir doğru yanıt vardır. 18 yaşından önce içkili mekânlara girmenin yasak olduğu bilgisinin Google tarafından getirilmesi beklenmektedir. Diğer sorularda yer alan "Atiye mi Hakan Muhafız mı?" sorusunun ise belirli bir cevabı bulunmamaktadır. Burada iki dizinin karşılaştırıldığı içeriklerin getirilmesi beklenmiştir. Bu 2 soruda da Google'ın doğal dil ile sorulan soruları ne kadar anladığını görmek amaçlanmıştır.

Verilerin Analizi

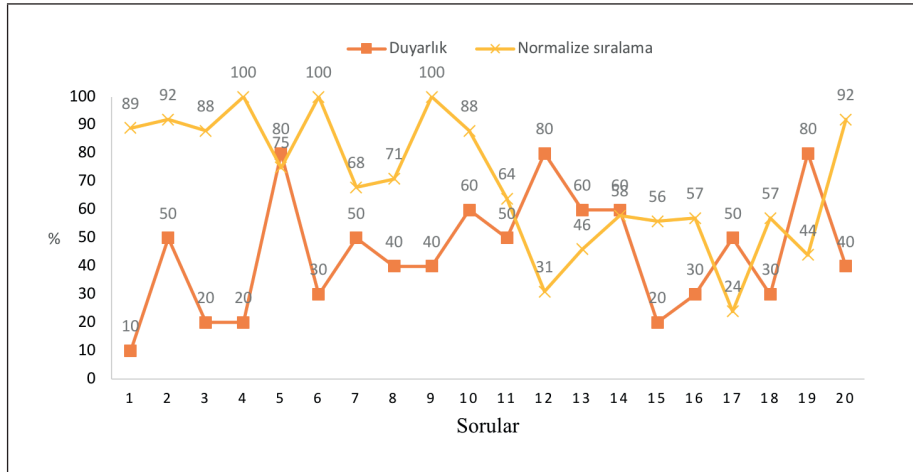
Araştırma kapsamında öncelikle arama motoruna sorulan sorulara getirilen ilk 10 sonucun tam metinlerine bakılarak ilgili ve ilgisiz olma durumları belirlenmiştir. Sonuç sayfasının en başında yanıt kutusu gelip gelmemesi, yanıt kutularının ilgililiği, sonuçlar arasında video olup olmaması, videoların ilgililiği ayrıca değerlendirilmiştir. Değerlendirmeler sonucunda elde edilen veriler öncelikle Excel'de tutulmuştur. Soruların duyarlık ve normalize sıralama değerleri hesaplanıp ayrı bir sayfada düzenlendikten sonra ilk 5 sonuç ve ilk 10 sonuç için duyarlık değerleri, normalize sıralama değerleri, soru tipi bilgileri, yanıt kutusu gelme oranları verilerin analizi için SPSS yazılımına aktarılmıştır.

Elde edilen sonuçların duyarlık ve normalize sıralama değerleri normal dağılım göstermemektedir. Bu sebeple verilerin analizinde normallik varsayımının karşılanma-

diği durumlarda kullanılan parametrik olmayan testler tercih edilmiştir (Ceylan, 2019, s. 12). Soru tiplerine göre duyarlık değerleri ve normalize sıralama değerlerinde bir farklılık olup olmadığı Mann-Whitney U testi ile ölçülmüştür. Duyarlık değerleri ile normalize sıralama değerleri arasındaki ilişki ise Spearman korelasyon testi ile belirlenmiştir. İlk 5 sonuç ve ilk 10 sonucun duyarlık değerleri ve normalize sıralama değerleri arasında farklılık olup olmadığı ile yanıt kutusu gelme oranlarının soru tiplerine göre farklılık gösterip göstermediği Ki-kare testi kullanılarak analiz edilmiştir.

Bulgular

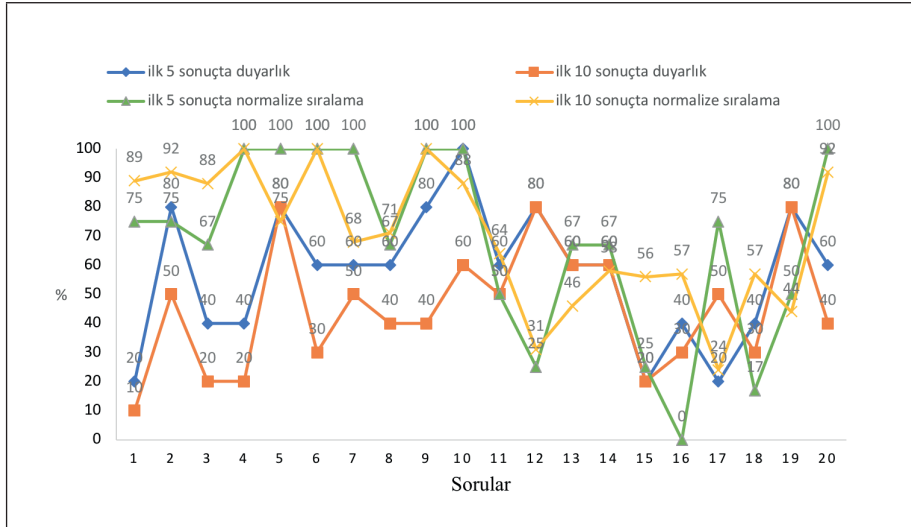
Bu bölümde Google'da yapılan 20 sorgunun sonuçları ile ilgili bulgulara yer verilmektedir. Araştırma kapsamında Google'a 20 soru sorulmuş ve her soru için ilk 10 sonuç ilgili ve ilgisiz olarak değerlendirilmiştir. Ulaşılan 200 sonucun 90'ı ilgili, 110'u ilgisiz bulunmuştur (duyarlık=%45). Sonuçların %13'ü (26) video'dur. Videoların duyarlık değeri ise %34 olarak ölçülmüştür. Elde edilen 200 sonucun yanı sıra sonuç sayfasında soruların %45'ine (9) gelen yanıt kutuları ayrıca analiz edilmiştir. Yanıt kutularının duyarlığı %89 olarak hesaplanmıştır. Aşağıda soruların duyarlık değerleri, normalize sıralama değerleri, soru tiplerine (maddi sorular/diğer sorular) göre farklılıklar, ilk 5 sonuç ve ilk 10 sonucun karşılaştırılması, yanıt kutuları ve videolar ile ilgili bulgulara yer verilmiştir.



Şekil 1. Sorulara Göre Duyarlık ve Normalize Sıralama Değerleri

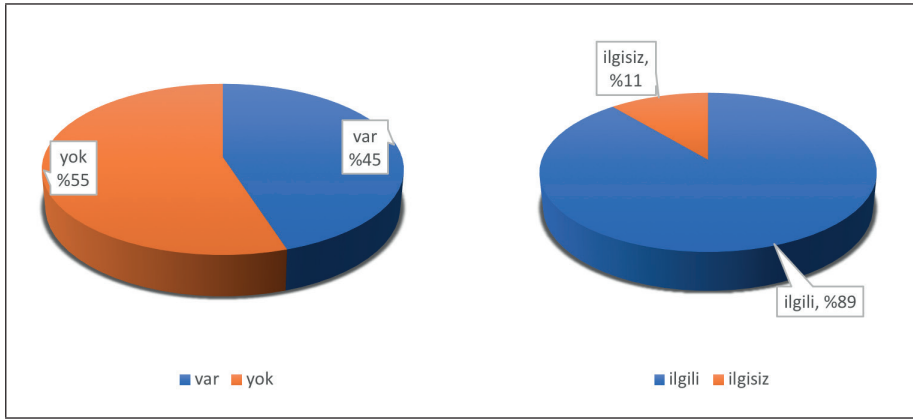
Şekil 1'de her bir sorunun duyarlık değeri ve normalize sıralama değerleri gösterilmiştir. Soruların duyarlık değerleri erişilen ilgili sonuçların toplam sonuca (10) oranını

vermektedir. Normalize sıralama ise ilgili dokümanların üst sıralarda bulunma oranlarını göstermektedir. Birinci soruda yalnızca 1 ilgili sonuç gelmiş, duyarlık değeri %10 olarak hesaplanmıştır ($1/10=0,10 \Rightarrow \%10$). En yüksek duyarlık değerleri ise 5., 12. ve 19. sorular için 80 olarak hesaplanmıştır ($8/10=0,80 \Rightarrow \%80$). İlgililik sıralamasında (normalize sıralama) ise en yüksek performanslar 4., 6. ve 9. sorularda gösterilmiş, en düşük performanslar 17. ve 12. sorularda gösterilmiştir. İlk 10 sorunun maddi sorular, 11.-20. soruların ise diğer sorular olduğu bilinmektedir. İki soru tipi arasında duyarlık değerlerinin çok farklı olmadığı şekilde görülmektedir. İstatistiksel açıdan da maddi sorular ile diğer soruların duyarlık değerleri arasında anlamlı bir fark bulunmamıştır ($U=36.000$, $p=0,28$). Normalize sıralama değerlerine bakıldığında ise soru tiplerine göre anlamlı bir farklılık olduğu saptanmıştır ($U= 6.500$, $p=0,001$). Maddi sorularda ortalama normalize sıralama değerlerinin daha yüksek olduğu görülmüştür. Duyarlığı en düşük çıkan 1. soruda Türkiye'nin 5. Cumhurbaşkanı'nın eşi sorulmuştur. Burada Google'dan istenen 5. Cumhurbaşkanı'nın Cevdet Sunay olduğunu bulup yanıt olarak eşinin adı "Atıfet Sunay"ı getirmesidir. Kısa ve net bir cevabı olmasına rağmen karmaşık bir soru olduğu için yalnızca 1 ilgili doküman getirilmiştir. Tek bir ilgili doküman getirilse de bu sorunun normalize sıralama değeri oldukça yüksektir (%89). En yüksek duyarlığa sahip sorulardan biri olan 5. soruda ise Atatürk'ün annesinin nereli olduğu sorulmuştur. İlk soruya benzer bir yapıda olmasına rağmen bu sorunun sonuçları genel olarak ilgili bulunmuştur. Bunun sebebinin Atatürk'ün annesinin Zübeyde Hanım olduğunu Google'ın daha kolay bir şekilde bulabilmesi olduğu düşünülmektedir. Duyarlığı yüksek çıkan diğer sorularda ise kültürel miras konusunda yapılan ulusal akademik yayınlar ve spor yapmadan kilo vermenin yolları aranmıştır. Net bir cevabı olmayan bu sorularda cümle yapılarının kolay olmasının duyarlığın yüksek çıkmasında etkili olduğu düşünülmektedir. Google bu sorularda ilgili dokümana erişmekte başarılı olsa da ilgili dokümanları üst sıralarda getirmekte başarılı değildir. Her 2 soruda da normalize sıralama değerleri düşüktür (12. soru = %31, 19. soru= %44). Duyarlık ve normalize sıralama değerleri arasında istatistiksel açıdan anlamlı olmayan negatif bir ilişki saptanmıştır ($\rho = -.427$, $p=0,061$). Yani erişilen ilgili doküman sayısı fazla olsa da ilgili dokümanların üst sıralarda gösterilme performansı düşük olabilmektedir. Tam tersi şekilde ilgili dokümana erişme performansı düşük olduğu durumlarda ilgili dokümanların üst sıralarda gösterilme performansı yüksek olabilmektedir.



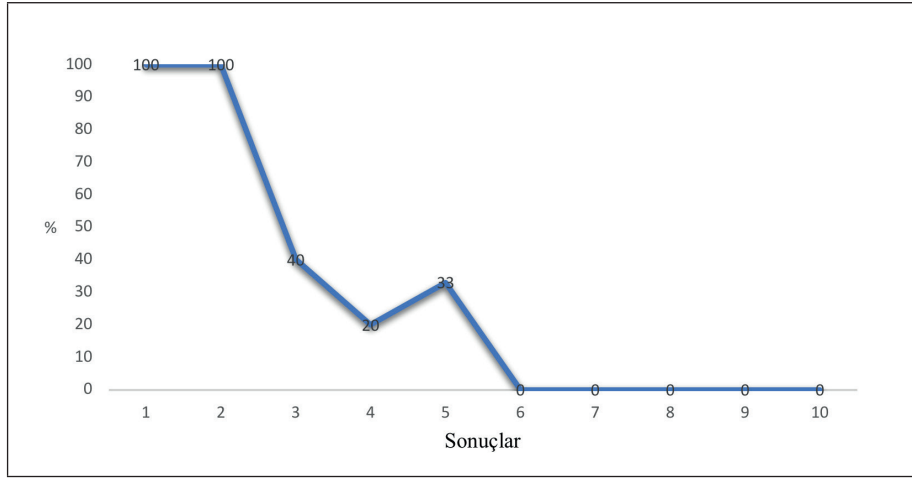
Şekil 2. İlk 5 Sonuç ve İlk 10 Sonuçta Soruların Duyarlık ve Normalize Sıralama Değerleri

Şekil 2 soruların ilk 5 sonuca ve ilk 10 sonuca göre duyarlık ve normalize sıralama değerlerini göstermektedir. İlk 5 sonuç ve ilk 10 sonucun duyarlık değerleri arasında istatistiksel açıdan anlamlı bir farklılık olmadığı sonucuna ulaşılmıştır ($\chi^2=36.317$, $df = 24$, $p=0,051$). Normalize sıralama değerlerinde de ilk 5 sonuç ve ilk 10 sonuç arasında farklılık bulunmamaktadır ($\chi^2=91.310$, $df = 84$, $p=0,274$). İlk 5 sonuçta da ilk 10 sonuçta da ilgili dokümanları üst sıralarda göstermekte başarı sağlanmıştır. Soruların 4'ünde ilk 5 sonuç ile ilk 10 sonucun duyarlık değerleri arasındaki farkın fazla olduğu belirlenmiştir (2, 6, 9, 10). Normalize sıralama değerlerinde ise büyük bir fark yoktur. Bu soruların hepsinin maddi sorular olması dikkat çekicidir. Google maddi sorularda ilgili dokümanları üst sıralarda göstermekte daha yüksek performans göstermektedir. İlk 5 sonucun ilk 10 sonuçtan daha düşük bir performans gösterdiği tek soru olan 17. soruda ise Covid-19 nedeniyle durdurulan uçuşların hangi ülkelere geri açıldığı sorulmuştur. Bu soru hem soru yapısı olarak karmaşık hem de ilgili yanıtların bulunması açısından zor bir sorudur. İlk 5 sonucun bu soru için duyarlık değeri düşük olsa da normalize sıralama değeri yüksektir (%75). İlk 10 sonuçta ise normalize sıralama değeri %24'e düşmektedir. İlk 5 sonuçta çok az ilgili dokümana erişilse de ilgili dokümanların üst sıralarda yer aldığı görülmektedir. Öte yandan ilk 10 sonuçta toplamda daha fazla ilgili dokümana ulaşılsa da ilgili dokümanların üst sıralarda yer alma oranı çok düşüktür.



Şekil 3. Sonuçlarda Yanıt Kutusu Gelmesi ve İlgililik Oranları

Şekil 3'te Google'ın sonuçlarında yanıt kutusu gelip gelmemesi ve yanıt kutularının ilgililiğine yönelik bulgular gösterilmektedir. Soruların %45'ine (9) yanıt kutusu geldiği, bunların ise %89'unun (8) sorgu ile ilgili olduğu sonucuna ulaşılmıştır. Maddi sorular (5) ile diğer sorular (4) yanıt kutusu gelmesi açısından karşılaştırıldığında anlamlı bir fark olmadığı görülmüştür ($\chi^2=1.667$, $p=0,197$). Maddi sorulara 5 yanıt kutusu gelmiş ve bunların duyarlılığı %80 olarak ölçülmüş, diğer sorulara ise 4 yanıt kutusu gelmiş ve duyarlılığı %100 olarak ölçülmüştür. Yanıt kutusu gelmesine rağmen ilgisiz olan 1 soru (9. soru) vardır. Arama motoruna Isparta merkez ile Gönen arası kaç km olduğu sorulmuştur. Bu soruda yanıtlanması istenen Isparta'nın Gönen ilçesi ile merkezi arası uzaklıktır. Ancak Google yanıt kutusunda Isparta Merkez ile Balıkesir Gönen arası uzaklığı göstermiştir. Aslında Google'ın yanıtı yanlış olmasa da bu soruda araştırılmak istenen Google'ın kullanıcının sormak istediğini ne kadar anladığını belirlemektir. Yanıt kutusu gelen diğer soruların 4'ü maddi soru (1, 3, 5, 7) diğer 4'ü ise listeleme sorularıdır (11, 13, 14, 19).



Şekil 4. Videoların Sonuçlara Göre Duyarlık Değerleri

Şekil 4 sonuç sıralamasına göre videoların duyarlılık değerlerini göstermektedir. Şekilden de çok net bir şekilde görüldüğü gibi ilk 5 sonuçta gelen videolar sorgu ile ilgili iken son 5 sonucun ilgisiz olduğu görülmektedir. Videoların duyarlılığı %34 olarak hesaplanmıştır. Ayrıca maddi soruların 5'ine (soru 1, 2, 5, 6, 10) toplam 14 video gelmiş ve duyarlılığı %21 olarak hesaplanmış, diğer soruların ise 4'üne (soru 13, 14, 19, 20) toplam 12 video gelmiş ve duyarlılığı %50 olarak hesaplanmıştır. Maddi sorulara gelen videoların sorgu ile ilgililik oranı diğer sorulardan daha düşüktür.

Tartışma

Araştırma kapsamında Google'ın Türkçe doğal dilde yapılan aramalarda performansını değerlendirmek amaçlanmıştır. İlk araştırma sorusu (S1) çerçevesinde Google'ın Türkçe doğal dilde sorulara erişimde ilgili dokümanlara erişmekte ortalama düzeyde bir başarı gösterdiği sonucuna ulaşılmıştır. 10 yıl önce yapılan bir çalışma (Tümer ve diğerleri, 2009) Google'ın semantik arama performansının düşük olduğunu göstermiştir. O yıllarda yapılan farklı çalışmalar da Google'ın karmaşık çok sözcüklü sorgularda Yahoo ve Bing arama motorlarından daha düşük performans gösterdiği sonuçlarına ulaşmıştır (Kumar ve Prakash, 2009; Usmani ve diğerleri, 2012). Öte yandan günümüze yaklaştıkça diğer arama motorlarıyla karşılaştırıldığı çalışmalarda Google'ın daha yüksek performans göstermeye başladığı görülmektedir (Khan ve diğerleri, 2014; Mala ve Lobiyal, 2016; Gul ve diğerleri, 2020). Google'ın performansının gittikçe daha fazla yükseldiği ancak bu yükselişin beklentinin altında olduğu düşünülmektedir.

İkinci araştırma sorusunun (S2) sonuçlarına göre Google maddi sorularda da diğer sorularda da ilgili dokümanlara erişmekte benzer bir performans göstermektedir. Google'ın maddi soruları yanıtlama performansının daha yüksek olduğuna yönelik hipotez (H1) reddedilmiştir. Maddi sorular ile diğer soruların duyarlık değerleri arasında anlamlı bir fark saptanmamıştır. Duyarlığı en düşük olan soru 1. soru iken en yüksek olanlar 5., 12. ve 19. sorular olarak belirlenmiştir. Yirmi yıl önce farklı arama motorlarının karşılaştırıldığı bir çalışmada ise kısa ve belirli bir cevabı olan maddi sorularda diğer sorulara göre daha üstün performans gösterildiği sonucuna ulaşılmıştır (Soydal, 2000, s. 87). Geçen sürede arama motorlarının performanslarının diğer sorular için de daha yüksek seviyeye ulaştığı düşünülmektedir. Ancak bu farklılık Google'ın diğer arama motorlarına göre bu soru tiplerinde daha yüksek performans göstermesinden kaynaklanıyor olabilir. Soru tiplerine göre Google ile farklı arama motorlarını karşılaştıran çalışmalar yapılmaması bu konuda daha net sonuçlar verecektir.

Araştırma sorusu (S3) kapsamında Google tarafından getirilen sonuçlarda ilgili dokümanların daha üst sıralarda olup olmadığı değerlendirilmiştir. Bu doğrultuda hem duyarlık hem de normalize sıralama değerleri ilk 5 sonuç ve ilk 10 sonuç için ayrı ayrı hesaplanarak karşılaştırılmıştır. İlk 5 sonuç ve ilk 10 sonucun duyarlık değerlerinde de normalize sıralama değerlerinde de istatistiksel açıdan anlamlı bir farklılık bulunmamaktadır. Google'ın ilk 5 sonuçta ilk 10 sonuca göre daha yüksek performans gösterdiğine ilişkin hipotez (H2) kabul edilmemiştir. Ancak duyarlık değerleri arasında genelde çok fark olmadığı, fazla fark olan sorularda da ilk 5 sonucun duyarlığının daha yüksek çıktığı görülmüştür. Bu sebeple Google'da ilgili sonuçlara ulaşmak için çoğunlukla ilk 5 sonuca bakmanın yeterli olduğu düşünülmektedir. Farklı araştırmalar arama motorlarının duyarlık değerlerinin ilk 5 sonuçta en yüksek olduğunu ve sonuç sayısı yükseldikçe duyarlık değerlerinin düştüğünü ortaya çıkarmıştır (Tonta ve diğerleri, 2002, s. 89; Demirci ve diğerleri, 2007; Tümer ve diğerleri, 2009). Soydal'ın (2000, s. 87) çalışmasında da arama motorları ilk 10 sonuçta ikinci 10 sonuca göre daha üstün performans göstermiştir. Arama motorlarının genel olarak ilgili dokümanları daha üst sıralarda göstermek için çabaladığı görülmektedir. Ayrıca maddi sorular ile diğer soruların normalize sıralama değerleri arasında istatistiksel açıdan anlamlı bir fark bulunmuştur. Maddi sorularda ortalama normalize sıralama değerleri daha yüksektir. Diğer bir deyişle Google'ın maddi sorularda ilgili dokümanları üst sıralarda gösterme performansı daha yüksektir. Maddi sorularda daha net yanıtlar içeren sonuçlar getirildiği için ilk sıralarda ilgili sonuçlara yer verildiği, diğer sorularda ise karmaşık yanıtlar beklendiği için ilgi sıralamasının daha düşük düzeyde olduğu düşünülmektedir.

Araştırma kapsamında yapılan aramaların yarıya yakınında sonuç sayfasının en başında yanıt kutusu gelmiştir. Dördüncü araştırma sorusu (S4) sonuçlarına göre yanıt kutularının maddi sorularda olduğu kadar Osmanlı padişahları, Bilgi ve Belge Yönetimi bölümleri gibi karmaşık olmayan listeleme sorularında da geldiği görülmektedir. Yanıt kutularının ilgililik düzeyi oldukça yüksek çıkmıştır.

Beşinci araştırma sorusu (S5) kapsamında yanıt kutularının ilgili sonuçlar getirmekte başarılı olduğu sonucuna ulaşılmıştır. Ancak yanıt kutularının ilgililik düzeyi yüksek çıksa da soruların yarısında bile yanıt kutusu gelmemesi dolayısıyla soru yanıtlama hizmeti performansının düşük olduğu hipotezi (H3) kabul edilmiştir. Google'ı soru yanıtlama kalitesi açısından değerlendiren bir çalışmada (Zhao ve diğerleri, 2019) Google'ın yanıt kutularında kişi ile ilgili sorulara diğer hedef türlerine (kurum, nesne, olay) göre daha kaliteli yanıtlar geldiği, soru tiplerinde ise en iyi yanıtların nerede sorularına getirildiği sonuçlarına ulaşılmıştır. Bu çalışmada sorular arasında böyle bir ayırım yapılmısa da kişi, kurum, olay veya kim, ne zaman, nasıl, nerede sorularına yanıt kutusu geldiği belirlenmiştir. Yanıt kutusu sayısı az olduğu için karşılaştırma yapılamamıştır. Ancak yanıt kutularının hangi tip sorularda daha ilgili yanıtlar getirdiğinin daha kapsamlı bir şekilde araştırıldığı çalışmaların yapılması önemli görülmektedir.

Son araştırma sorusu (S6) çerçevesinde Google tarafından sonuç sayfasında getirilen videoların da ilgililik düzeyleri incelenmiştir. Videoların genel olarak sorgu ile ilgisiz olduğu sonucuna ulaşılmıştır. Bu durumun videolarda üstveri etiketinin az olmasından kaynaklandığı düşünülmektedir. İlk 5 sonuçta gelen videolar çoğunlukla sorgu ile ilgili iken, son 5 sonuçta gelen videoların tamamı ilgisiz olarak belirlenmiştir. Öte yandan videoların ilgililiğinin soru tiplerine göre farklılık gösterdiği saptanmıştır. Diğer sorulara (listeleme ve karmaşık yanıtları olan sorular) gelen videoların maddi sorulara gelen videolardan daha ilgili olduğu görülmüştür.

Sonuç ve Öneriler

Arama motorları insanların gereksinim duyduğu bilgiye en hızlı şekilde erişmesini sağlayan bilgi erişim sistemleridir. Kullanıcının bilgiyi en kolay ve hızlı biçimde elde etme isteği arama motorlarını da kendilerini sürekli geliştirmeye zorlamıştır. Kullanıcılar arama motorlarında bilgiyi öncelikle doğal dilde arama eğilimindedirler. Çalışma kapsamında değerlendirilen Google, 2015 yılında doğal dili anlamaya yönelik olarak geliştirdiği yapay zekâ sistemi ile daha fazla başarı elde etmeyi hedeflemiştir. Bu çalışmada Google'ın Türkçe doğal dilde yapılan sorgularda gösterdiği performansını değerlendirmek amaçlanmıştır. Bu doğrultuda Google'a Türkçe doğal dilde 20 soru sorulmuş, ilk 10 sonuç duyarlık değerleri ve normalize sıralama değerleri dikkate alınarak değerlendirilmiştir. Araştırma kapsamında yapılan aramalarda duyarlık ve normalize sıralama değerleri arasında pozitif bir ilişki bulunmamıştır. Yani ilgili doküman sayısının fazla olması, ilgili dokümanların üst sıralarda yer aldığı anlamına gelmemektedir. Aynı şekilde erişilen ilgili doküman sayısı az olsa da ilgili dokümanlar üst sıralarda yer alabilmektedir. Kullanıcı ilgili dokümana üst sırada eriştiyse alttaki dokümanlara ihtiyaç duymayabilmekte ve tek bir dokümana erişilse bile üst sıralardaysa sonuçtan memnun olabilmektedir. Tam tersine, çok fazla ilgili doküman olmasına rağmen bu dokümanlar üst sıralarda yer almıyorsa kullanıcı ilgili dokümanı aramayı bırakabilmektedir. Bu sebeple arama motorlarının performansını değerlendirirken normalize sıralama değer-

leri de son derece önemlidir. Araştırma sonuçlarına göre Google duyarlık değerlerine göre yüksek bir performans göstermese de ilgili dokümanları üst sıralarda gösterme açısından başarılı bulunmuştur. Çalışmanın bulgularına dayanarak bazı öneriler sunulmuştur.

Araştırma sonuçlarına göre Türkçe doğal dilde yapılan aramalarda Google'ın getirdiği sonuçların iyileştirilmeye ihtiyaç duyduğu görülmektedir. Artık herkesin bir sebeple arama motorlarına gereksinim duyduğu göz önünde bulundurularak arama motorları doğal dili anlama yönünden geliştirilmeli, kullanıcının arama stratejilerini bilmesi beklenmemelidir. Türkçe doğal dil işleme, semantik arama, yapay zekâ konuları üzerine çalışmaların artırılması önerilmektedir.

Arama motorları her ne kadar ilgili dokümanları üst sıralarda göstermek için çabalsada bu kadar çok bilginin yer aldığı bilgi erişim sistemlerinde kullanıcının doğru ve ilgili bilgiye erişmesi zorlaşmaktadır. Google, kullanıcının bilgiye en hızlı ve kolay yoldan erişme ihtiyacını karşılayabilmek için soru yanıtlama hizmeti vermektedir. Soru yanıtlama hizmeti ilgililik açısından başarılı bulunmakla birlikte, birçok sorguda yanıt kutusu gelmediği görülmüştür. Bu hizmetin tüm sorgulara yanıt sunabilecek şekilde geliştirilmesi için çalışmalar yapılması önerilmektedir. Sonuç sayfasında gelen videoların ise aramalarla ilgisiz olduğu sonucuna ulaşılmıştır. Aramalarda getirilen video sonuçlarının iyileştirilmesi tavsiye edilmektedir.

Bu çalışmanın sonuçlarının Google'ın performansının geliştirilmesi açısından yapılacak olan gelecek çalışmalara fayda sağlaması beklenmektedir. Farklı sorgular ile gerçekleştirilen çalışmaların artması ve çeşitlenmesinin Google'ın daha iyi performans göstermesine katkı sağlayacağı düşünülmektedir. Daha fazla soruyla yapılan, sonuçları gerçek kullanıcıların değerlendirdiği daha geniş kapsamlı çalışmaların yapılması önemli görülmektedir.

Teşekkür

Önerileriyle çalışmanın geliştirilmesine katkıda bulunan Prof. Dr. Yaşar Tonta'ya çok teşekkür ederim.

Kaynakça

- Akbulut, M., Tonta, Y. ve White, H.D. (2020). Related records retrieval and pennant retrieval: An exploratory case study. *Scientometrics*, 122, 957–987. doi: 10.1007/s11192-019-03303-9
- Alexa. (2020). *Alexa top sites*. <https://www.alexacom/topsites> adresinden erişildi.
- Amasyalı, M. F. ve Diri, B. (2005). Bir soru cevaplama sistemi: Baybilmiş. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 1(1), 37-51. <https://dergipark.org.tr/tr/pub/tbbmd/issue/22240/238746> adresinden erişildi.

- Azeez, S. S. (2017). *An evaluation of three search engines (Google, Yahoo, Bing) based on Arabic user perception* (Yayımlanmamış Yüksek Lisans Tezi). Çankaya Üniversitesi, Ankara. <http://earsiv.cankaya.edu.tr:8080/xmlui/bitstream/handle/20.500.12416/1950/Azeez%2c%20Suad%20Shatti.pdf?sequence=1&isAllowed=y> adresinden erişildi.
- Bar-Ilan, J. ve Gutman, T. (2005). How do search engines respond to some non-English queries? *Journal of Information Science*, 31(1), 13-28. doi: 10.1177/0165551505049255
- Bilal, D. (2012). Ranking, relevance judgment, and precision of information retrieval on children's queries: Evaluation of Google, Yahoo!, Bing, Yahoo! Kids, and ask Kids. *Journal of the American Society for Information Science and Technology*, 63(9), 1879-1896. doi: 10.1002/asi.22675
- Bitirim, Y. ve Görür, A. K. (2017). A comparative evaluation of popular search engines on finding Turkish documents for a specific time period. *Tehnički vjesnik*, 24(2), 565-569. doi: 10.17559/TV-20140512181430
- Bollmann, P. (1983). The normalized recall and related measures. *ACM SIGIR Forum*, 17(4), 122-128. doi: 10.1145/1013230.511811
- Brin, S. ve Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7), 107-117. doi: 10.1016/S0169-7552(98)00110-X
- Ceylan, N. Ö. (2019). *Parametrik ve parametrik olmayan testler için güç karşılaştırmaları* (Yüksek Lisans Tezi). Selçuk Üniversitesi, Konya.
- Chaoticity. (2010). *Google as a question answering system*. <https://chaoticity.com/google-as-a-question-answering-system/> adresinden erişildi.
- Demirci, R. G., Kışmır, V. ve Bitirim, Y. (2007). An evaluation of popular search engines on finding Turkish documents. *Second International Conference on Internet and Web Applications and Services (ICIW'07)* içinde (s. 61). Morne. doi: 10.1109/ICIW.2007.15.
- Dwivedi, S. K. ve Singh, V. (2013). Research and reviews in question answering system. *Procedia Technology*, 10, 417-424. doi: 10.1016/j.protcy.2013.12.378
- Gul, S., Ali, S. ve Hussain, A. (2020). Retrieval performance of Google, Yahoo and Bing for navigational queries in the field of "life science and biomedicine". *Data Technologies and Applications*, 54(2), 133-150. doi: 10.1108/DTA-05-2019-0083
- Kesen, S., Şenol, C. ve Yanar, Z. (2008). Google Scholar ve Scirus arama motorlarında Türkçe anahatar sözcüklerle yapılan aramalar üzerine bir değerlendirme. *Bilgi Dünyası*, 9(1), 140-157. <http://bd.org.tr/index.php/bd/article/view/175> adresinden erişildi.
- Khan, J. A., Sangroha, D., Ahmad, M. ve Rahman, M. T. (2014). A performance evaluation of semantic based search engines and keyword based search engines. *2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)* içinde (ss. 168-173). Greater Noida. doi: 10.1109/MedCom.2014.7005997
- Kumar, B. T. S. ve Prakash, J. N. (2009). Precision and relative recall of search engines: A comparative study of Google and Yahoo. *Singapore Journal of Library & Information Management*, 38, 124-137.

- Mala, V. ve Lobiyal, D. K. (2016). Semantic and keyword based web techniques in information retrieval. *2016 International Conference on Computing, Communication and Automation (ICCCA)* içinde (ss. 23-26). Noida. doi: 10.1109/CCAA.2016.7813724
- Singhal, A. (2012). *Introducing the knowledge graph: Things, not strings*. <https://blog.google/products/search/introducing-knowledge-graph-things-not/> adresinden erişildi.
- Soricut, R. ve Brill, E. (2004). Automatic question answering: Beyond the factoid. *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004* içinde (ss. 57–64). Boston. <https://www.aclweb.org/anthology/N04-1008> adresinden erişildi.
- Soydal, İ. (2000). *Web arama motorlarında performans değerlendirme* (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara. <http://bby.hacettepe.edu.tr/yayinlar/70.pdf> adresinden erişildi.
- Sullivan, D. (2016). *All about the Google RankBrain algorithm*. <https://searchengineland.com/faq-all-about-the-new-google-rankbrain-algorithm-234440> adresinden erişildi.
- Tonta, Y. (1995). Bilgi erişim sistemleri. *Türk Kütüphaneciliği*, 9(3), 302-314. <http://www.tk.org.tr/index.php/TK/article/view/1010/1011> adresinden erişildi.
- Tonta, Y., Bitirim, Y. ve Sever, H. (2002). *Türkçe arama motorlarında performans değerlendirme*. Ankara: Total Bilişim. <http://yunus.hacettepe.edu.tr/~tonta/yayinlar/tonta-bitirim-sever-arama-motorlari.pdf> adresinden erişildi.
- Tümer, D., Shah, M. A. ve Bitirim, Y. (2009). An empirical evaluation on semantic search performance of keyword-based and semantic search engines: Google, Yahoo, Msn and Hakia. *2009 Fourth International Conference on Internet Monitoring and Protection* içinde (ss. 51-55). Venice. doi: 10.1109/ICIMP.2009.16
- Usmani, T. A., Pant, D. ve Bhatt, A. K. (2012). A comparative study of Google and Bing search engines in context of precision and relative recall parameter. *International Journal on Computer Science and Engineering*, 4(1), 21-34.
- Yozkat, S. A. (2017). Arama motoru sonuçları güvenilirliğinin analizi: Google örneği. *Yeni Medya Elektronik Dergisi*, 1(1), 72-81. <http://static.dergipark.org.tr/article-download/b3c9/be9f/55f6/imp-JA67PK24SM-0.pdf?> adresinden erişildi.
- Zadeh, L. A. (2006). From search engines to question answering systems: The problems of world knowledge, relevance, deduction and precisiation. E. Sanchez (Ed.). *Capturing Intelligence* içinde (ss. 163-210). New York: Elsevier. doi: 10.1016/S1574-9576(06)80011-0
- Zeden, R., Ayalp, E., Bitirim, Y. ve Sever, H. (2006). Duyarlılık ve normalize sıralama: Gövdelemenin Türkçe arama motorları üzerindeki etkisi. *The Fifteenth Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN)* içinde (ss. 215-222).
- Zhao, Y., Zhang, J., Xia, X. ve Le, T. (2019). Evaluation of Google question-answering quality. *Library Hi Tech*, 37(2), 312-328. doi: 10.1108/LHT-10-2017-0218.