

Google Scholar ve Scirus Arama Motorlarında Türkçe Anahtar Sözcüklerle Yapılan Aramalar Üzerine Bir Değerlendirme*

An Evaluation of Google Scholar and Scirus Search Engines Using Turkish Search Queries*

Seda Kesen, Canan Şenol ve Zehra Yanar

Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü, 06800 Beytepe, Ankara.
{seda.kesen, canansenol, zehrayanar}@gmail.com

Öz: Bilgi miktarının hızla arttığı İnternet'te, farklı dillerde yaratılmış belgelerdeki bilgilere erişmek için arama motorları kullanılmaktadır. Bu çalışmanın amacı açık erişim bilgi kaynaklarına erişim sağlayan Google Scholar ve Scirus arama motorlarının özel Türkçe karakterleri (ç, ğ, ı, ö, ş, ü) doğru görüntüleyip görüntülemediklerini saptamak ve bu arama motorlarını erişilen belge sayısı yönünden değerlendirmektir. Her iki arama motorunda da özel karakterler alfabeğe uygun olarak görüntülenmiş, fakat Türkçe karakterler farklı yorumlanmıştır. Bu farklılıklar bilgi erişim açısından büyük sorunları da beraberinde getirmektedir. Türkçe açık erişim kaynaklarının üst verileri (metadata) Türkçe özel karakterlere uygun bir şekilde oluşturulmalı ve arama motorları Türkçe karakterleri destekleyecek biçimde geliştirilmelidir.

Anahtar sözcükler: Bilgi erişim, arama motorları, Google Scholar, Scirus, Türkçe, özel karakterler

Abstract: *Due to the rapid growth of information on the İnternet, search engines are used to getting information included in documents created in different languages. This paper aims to find out if the two search engines providing access to Open Access information sources, Google Scholar and Scirus, display the search results appropriately for special Turkish characters (ç, ğ, ı, ö, ş, ü) and to evaluate them on the basis of the total number of retrieved documents. Both search engines displayed the Turkish special characters correctly, but the search results differed. These differences create information retrieval problems for Turkish queries. Metadata of the Open Access Turkish*

information sources should be created using special Turkish characters and search engines should be developed to support them.

Keywords: *Information retrieval, search engines, Google Scholar, Scirus, Turkish language, special characters*

Giriş

Büyük miktarda bilgi barındıran World Wide Web'de, gereksinim duyulan bilgiye erişim sağlanması önemli bir sorundur. Arama motorları ağ üzerinde dizinledikleri kaynakları kullanıcı sorguları doğrultusunda çok kısa bir sürede tarayarak sonuçları ekrana getirmektedir. Gereksinim duyulan ilgili kaynağa hızlı ve doğru bir şekilde ulaşmak zorunlu hale gelmekte, bu da "bilgi erişim" kavramını ortaya çıkarmaktadır.

Bilgi erişim terimi belgeye ya da belgeyi temsil eden bir grup bilgiye ulaşabilme kapasitesine sahip bir sistemi tanımlamak için kullanılır (Lancaster ve Fayen, 1973). Bilgi erişim bilgi kaynaklarının temsil edilmesi (representation), depolanması, düzenlenmesi ve bilgi kaynaklarına erişim sağlanması ile ilgili bir kavramdır (Salton ve McGill, 1983).

Bir bilgi erişim sisteminin temel işlevi dermedeki ilgili belgelerin tümüne erişmek, ilgili olmayanları da ayıklamaktır. İdeal bir bilgi erişim sistemi ilgili belgelerin tümüne ve salt ilgili belgelere erişim sağlamalıdır (Tonta, 1995). Bilgi erişim sistemleri bilgi gereksinimini ifade eden sorguları alıp, dosya ve kayıtları işleyerek bazı belirli dosya ve kayıtları sorgulara karşılık olarak getiren sistemlerdir. Belirli kayıtların seçilmesi, sorgu ile kayıtlar arasındaki benzerliklerden faydalanılarak gerçekleştirilen bir işlemdir (Salton, 1989).

Bilgi erişim sistemleri bilgileri analiz etmek üzere tasarlanır, bilgi kaynaklarını işleme tabi tutar ve kullanıcıların isteğine karşılık gelen kaynakları sunar

* En İyi Öğrenci Bildirisi Birincilik Ödülü / Winner of the Best Student Paper Award, First Place.

(Chowdhury, 2004). Bilgi erişim sistemleri nesnel (objective) ve öznel (subjective) terimleri içerir. Nesnel terimler anlamsal içeriğin dışındadır. Bu terimler yazar adı, kaynağın adresi (URL) ve yayın tarihi gibi tartışma kabul etmeyen alanları içerir. Öznel terimler ise dokümanın konusunu yansıtmak üzere kullanılır. Bir dokümanın konusu ya da bunu tanımlamak için kullanılacak terimler üzerinde anlaşma olmayabilir (Gudivada, Raghavan, Groksy ve Kasanagottu, 1997).

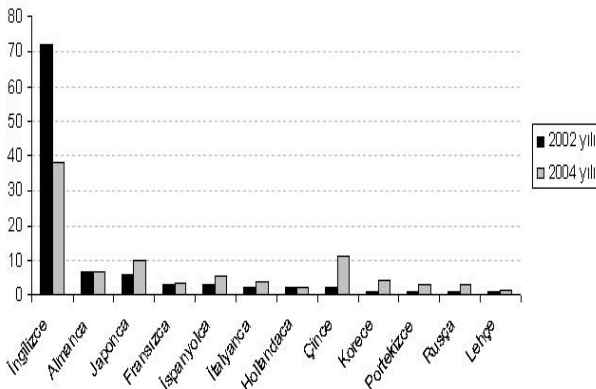
Arama motorları web üzerinde aranılan bilgilere ulaşmayı sağlayan bilgisayar yazılımlarıdır. Arama motorları bilgi erişim sistemlerini temel alır.

Web üzerinde ilgili kaynaklara ulaşmanın bir yolu web robotu (wanderer, worm, walker, spider veya knowbot) olarak bilinen yazılımlar kullanmaktır. Bu yazılımlar bir sorguyu alıp ilgili belgeleri bulmak için sistematik bir biçimde ağı tarayıp buldukları her belge için ilgililik değerini hesaplar ve ilgililik sırasına göre bir sonuç ekranı sunarlar (Gudivada, Raghavan, Groksy ve Kasanagottu, 1997).

Arama motorları üç önemli olanak sağlar. Bunlar:

- Araştırmacının arama yaptığı evrendeki web sayfalarını biraraya toplar ve ilgili sayfalara erişirir.
- Web sayfalarının içeriğini web üzerinde temsil eder.
- Erişim algoritması kullanarak arama sorgularıyla ilgili belgelere erişir ve araştırmacıya sunar (Gordon ve Pathak, 1999).

Arama motorları farklı dillerde pek çok kaynağı dizinlemektedir. 2002 yılında yapılan bir araştırmada Internet üzerinde yer alan belgelerin %72'sinin İngilizce, %7'sinin Almanca, %6'sının Japonca, %3'ünün ise Fransızca olduğu görülmüştür (OCLC, 2002). Ancak 2004 yılına gelindiğinde İngilizce kaynakların oranı neredeyse yarı yarıya (%38,3) düşmüş, Çince (%11,2), Japonca (%10) gibi diğer dillerdeki kaynakların oranı hızla yükselmeye başlamıştır. 2002 ve 2004 yıllarında Internet'te dil kullanım oranları Şekil 1'de verilmektedir (Translate, 2005).



Şekil 1. 2002 ve 2004 yıllarında Internet'te dil kullanım oranları (OCLC, 2002; Translate, 2005)

İki yıl içerisinde Web'de İngilizce kullanımı büyük bir düşüş gösterirken, başta Çince olmak üzere Uzak Doğu dillerinde bu oran artmıştır. Web'de erişim İngilizce odaklıyken zamanla çoklu dilde erişim artmıştır. Ülkeler anadilde erişime yönelmişlerdir.

Türkçe ise bu listelerde 2004–2005 yılında yapılan bir araştırmaya göre %0,7 oranıyla 16. sırada yer almıştır (Translate, 2005).

Önceki Çalışmalar

Kullanıcıların arama motorlarından etkin bir şekilde bilgi erişim sağlamaları için yapılan analiz çalışmaları arama motorlarının zayıf ve güçlü yönlerini tespit etmek için gereklidir. Bu tür çalışmalar arama motorlarının geliştirilmesi ve iyileştirilmesi için önemli bir adımdır.

Arama motorlarının performanslarına dair yapılan çalışmalar aşağıda özetlenmektedir.

Bir tez çalışmasında AltaVista, Excite, HotBot, Infoseek ve Northern Light arama motorlarının anma ve duyarlık açısından bilgi erişim performansları incelenmiş, ortalama duyarlık değerleri (yaklaşık %50) açısından adı geçen arama motorları arasında anlamlı bir farklılık olmadığı görülmüştür (Soydal, 2000). Ortalama anma değerlerinin ise %14 ile %31 arasında değiştiği gözlenmiştir.

Google Scholar ve Scirus arama motorları çeşitli yönlerden (indeksledikleri belge sayısı, kullanıcılara sundukları tarama seçenekleri, vd.) birbiriyle karşılaştırılmış, Scirus'un pek çok özellik bakımından Google Scholar'dan daha üstün olduğu görülmüştür (Notess, 2005).

Bilimsel bilgiye erişimde arama motorlarının getirdiği yenilikler, yapılan karşılaştırmalı bir çalışma çerçevesinde değerlendirilmiştir (Felter, 2005). Scirus'un sahip olduğu kurumsal kaynaklarla, kullanıcının aşına olduğu Google Scholar yapısının bütünleştirilmesi savunulmuş; bu sayede araştırmacılara faydalı olunacağı görüşü öne sürülmüştür.

Google Scholar, PubMed ve Scirus'un ele alındığı bir çalışmada Google Scholar ve PubMed karşılaştırılırken Google Scholar'a alternatif olarak Scirus önerilmiştir (Giustini ve Barsky, 2005). Scirus, Google Scholar'dan farklı olarak erişilen belgelerin yer aldığı kaynağı da (ScienceDirect, BioMedCentral, Beilstein) listelemektedir. Scirus'da konu kategorilerinin belirlenebilmesi bu arama motoruna esneklik sağlayarak ve kullanıcılara isteğe göre düzenleme imkânı sunarak duyarlık oranını artırmaktadır.

2005 yılında yapılan bir çalışmada Çin'e ait iki arama motoru (Baidu ve Sohu) ile Google ve AllTheWeb'in Çince sorguları nasıl algıladıkları karşılaştırılmıştır (Moukdad ve Cui, 2005). Çalışmada önce Çince'nin özellikleri tanıtılmış, daha sonra araştırma metodu tanımlanmıştır. Araştırma ölçütleri kelime bölümlendirilmesi, erişilen doküman sayısı, Çince karakterlerin kimliklendirilmesi ve doğru görüntülenmesi olarak belirlenmiştir. Dört arama motoruna yönlendirilen sorgular isim, fiil, sıfat gibi kelime özelliklerine göre

sınıflandırılmış ve sonuçlar karşılaştırılmıştır. Çince arama motorlarının kelime bölümlendirmesi, erişilen doküman sayısı ve karakterlerin görüntülenmesi konusunda daha üstün bir performans sergiledikleri görülmüştür.

Bir diğer çalışmada ise Rusça, Fransızca, Macarca ve İbranice'nin Web'de kullanımı araştırılmıştır (Bar-İlan ve Gutman, 2005). Bu çalışmada İngilizce olmayan diller için arama motorlarının kapasitesi ölçülmüştür. Çalışma kapsamına üç genel arama motoru (AltaVista, FAST ve Google) ile bazı yerel arama motorları alınmıştır. Sorguların çoğunda genel arama motorları İngilizce olmayan dillerdeki özel karakterleri görmezlikten gelmiş, hatta bazı işaretleri yok saymıştır.

“Tam Metin Arapça Veri Tabanlarından Bilgi Erişim” adlı çalışmada Arapça örneklerin erişimdeki etkisi araştırılmıştır (Moukdad ve Large, 2001). Arapça'nın sağdan sola yazılmasının AltaVista'da erişimde farklılık yaratıp yaratmadığı incelenmiş, örneklerinden ayrılan isimler üzerindeki arama sonuçları anma oranının azaldığını göstermiştir.

Sroka'nın (2000) hazırladığı “Lehçe Bilgiye Erişim için Web Arama Motorları” adlı çalışmada beş arama motoru değerlendirilmiştir. Arama motorlarından elde edilen kayıtlar duyarlılık (erişilen ilgili belgelerin erişilen tüm belgelere oranı), çakışma ve erişim süresine göre değerlendirmeye tabi tutulmuştur. Sonuç olarak beş arama motoru içinden Polski Infoseek ile Onet.pl arama motorlarının en yüksek duyarlığa sahip oldukları ortaya çıkmıştır. Belgelere en hızlı erişim sağlayan ve en kapsamlı arama motoru Polski Infoseek'dir.

Yapılan bir diğer araştırmada 2002 yılında Türkiye'de yaygın olarak kullanılan dört Türkçe arama motorunun (Arabul, Arama, Netbul ve Superonline) bilgi erişim performansları çeşitli ölçütlere göre değerlendirilmiştir. Değerlendirme ölçütlerinden biri de Türkçe karakter kullanımının erişim sonuçlarına etkileridir. Türkçe arama motorlarında “ç”, “ş”, “ü” gibi Türkçeye özgü karakterler kullanılarak yapılan aramalarda sorun olup olmadığı test edilmiştir. Bu çalışma sonucunda Türkçe karakterler kullanılarak yapılan aramaların farklı sonuçlar verdiği gözlenmiştir. Ayrıca aynı arama motorunda Türkçe karakter kullanılarak ve kullanılmadan yapılan farklı taramalarda erişilen belge sayılarının eşit olmadığı görülmüştür (Tonta, Bitirim ve Sever, 2002).

Türkçe kendine özgü özelliklerinden dolayı bu tür çalışmalarda yer verilmesi gereken bir dildir. Arama motorlarının arama yapılacak olan kelimenin kökünü alıp (gövdeleme) arama yapması yapı bakımından eklemeli (agglutinative) diller arasında yer alan Türkçe için önemlidir. Gövdeleme algoritmaları kullanılarak Türkçe derlemeler (corpora) üzerinde yapılan bilgi erişim performans değerlendirmelerinde anma ve duyarlılık değerlerinin %20-%25 civarında arttığı gözlenmiştir (Tonta, Bitirim ve Sever, 2002).

Bilgi erişimde gövdeleme algoritmaları kadar Internet üzerindeki bilgi kaynaklarının üst verilerinin (metadata) oluşturulması da etkilidir. Kullanıcıların bilgi gereksinimlerini karşılayacak belgelere erişimlerinin sağlanmasında üst veri önemli rol oynamaktadır. Üst verinin arama motorlarını geliştirenler ile web sitelerinin içeriklerini yaratanlar tarafından kullanılmasının web aramalarındaki etkinliği artıracak öne sürülmüştür (Thornely, 2000; aktaran: Al ve Küçük, 2003).

Amaç ve Yöntem

Bu çalışmada Türkçe kullanımı, açık erişim bilgi kaynaklarına erişim sağlayan arama motorlarından Google Scholar ve Scirus'da değerlendirilmiştir. Değerlendirme kapsamında Türkçe karakterlerin doğru görüntülenmesi ve erişilen belge sayısı ölçüt olarak belirlenmiştir. Türkçe karakterlerin doğru görüntülenmesi karakterlerin Internet tarayıcılarda alfabeğe uygun, okunaklı şekilde görüntülenmesi anlamına gelmektedir. Erişilen belge sayısı kriterinde ise taramalar hem Türkçe karakterlerle hem de bunlara en uygun İngilizce karakterlerle yapılmış ve erişilen belge sayıları karşılaştırılmıştır.

Google Scholar ve Scirus arama motorlarında Türkçe kullanımı üzerine yapılan bu çalışmada özel Türkçe karakterler (ç, ğ, ı, ö, ş, ü) içeren 18 Türkçe sorgu cümlesi kullanılmıştır (Tablo 1). Sorguların erişim sonuçlarının karşılaştırılabilmesi için harfler en uygun İngilizce karakterler kullanılarak değiştirilmiştir (ç→c, ğ→g, ı→i, ö→o, ş→s, ü→u). Örneğin, ilk sorgu cümlesi olan “direnc” “direnc”e çevrilmiştir. Bu kelimeler Google Scholar ve Scirus'da taramış ve sonuçlar tablolaştırılmıştır.

Tablo 1. Türkçe sorgu cümleleri

Sorgu cümleleri		
direnc	veri tabanı	şeffaf
çeviri	ıhlamur	kurtuluş
uçurum	donanım	bilişim
yanardağ	rejisor	üroloji
eğim	ödenti	telekomünikasyon
yoğunluk	terör	menisküs

Tarama sonuçları incelenirken her iki arama motoru tarafından erişilen ilk 20 kayıt incelenmiştir. Yirmi kayıttan az belgeye erişildiğinde ise belgelerin tümü dikkate alınmıştır.

Taramalar yapılırken dil kodlaması olarak Unicode (UTF-8) seçeneği belirlenmiştir. Çünkü Unicode diğer şifreleme yöntemlerine göre daha avantajlıdır. Eski şifreleme yöntemleri kendi aralarında çelişmektedir. İki farklı şifreleme, aynı sayıyı iki farklı karaktere vermiş olabilir ya da farklı sayılar aynı karakteri kodlayabilir (Unicode, 2007). Unicode, bilinen tüm modern ve eski dilleri, noktalamaları, bileşik karakterleri ya da matematik sembolleri kapsayan bir milyondan fazla karakteri

tanımlamaktadır. Ayrıca tanımlamada standart bir yapı sunması nedeniyle yeni ortaya çıkan karakterlerin tanımlanmasına imkân vermektedir (Alır, 2007).

Aşağıda uygulamanın yapılacağı arama motorları tanıtılmaktadır.

Google Scholar (GS)

Google Scholar (scholar.google.com) bilimsel literatürün kolay yoldan taranmasına imkân verir. Tek bir yerden, pek çok disiplin ve kaynak sorgulanabilir. Tarama kapsamına aldığı kaynaklar ile Google Scholar araştırmacılara dünya çapında araştırma konularıyla ilgili belgeleri sağlar (Google Scholar, 2007). Dizinlediği kaynak sayısı tam olarak bilinmemektedir. Bu konudaki çalışmalar Google Scholar Takımı tarafından sürdürülmektedir.²

Google Scholar dergiler, özler, gözden geçirilmiş makaleler, tezler, kitaplar, sunular ve teknik raporlar üzerinde arama yapar (Noruzi, 2005, s. 171).

Scirus

Scirus (www.scirus.com) web üzerindeki bilimsel, akademik, teknik ve tıbbi veriler içeren hakemli makaleler, teknik raporlar ve patentler üzerinde arama yapar. Bilim insanları ve araştırmacılar için tasarlanmıştır. Üç yüz milyonun üzerinde sayfanın bulunduğu ağı arar ve gerekli bilgileri sağlar. Scirus filtreleme özelliğine sahiptir. Bilimsel olmayan siteleri tarama dışında bırakır.

Scirus dünya çapında 104 milyon .edu, 26 milyon .org, 12,9 milyon .ac.uk, 25 milyon .com, 7,4 milyon .gov uzantılı ve 87 milyonun üzerinde diğer ilgili siteleri tarar (Scirus, 2007).

Google Scholar ve Scirus'ın Özelliklerinin Karşılaştırılması

Google Scholar ile Scirus'un benzer ve farklı yönleri aşağıda verilmektedir.

Benzerlikler:

- Basit ve gelişmiş arama imkânı sağlarlar;
- Aramayı tarihe ve konuya göre sınırlandırma seçeneği sunarlar;
- Her iki açık erişim arama motoru da yapılan atıfları görüntüler;
- Sonuç sayfasının farklı bir pencerede açılmasını sağlamak ve her bir sayfada gösterilecek sonuç sayısını belirlemek olanaklıdır;
- Büyük – küçük harf duyarlılığı yoktur;
- Boole işleçlerinin Türkçeleri geçerli değildir (VE, VEYA, DEĞİL);

- Kelimeler arasındaki boşluklar “AND” işleci olarak algılanır;
- Her iki arama motoru da bilgi kaynağının formatını (.pdf, .doc, .html) seçme imkânı sunar. Scirus ayrıntılı aramada seçeneklerle bu imkânı sağlarken, Google Scholar'da sorgu sözcüğüne “filetype” ya da “ext:” (pdf, doc, html) eklenerek bu işlem gerçekleştirilir.

Farklılıklar:

- Scirus bilgi kaynağının türünü (öz, makale, kitap vb.) seçme imkânı sunarken Google Scholar'da bu imkân verilmemektedir;
- Scirus seçilen veri tabanları ve dergi isimleri arasında seçim yaparak tarama imkânı sunar. Bu özellik Google Scholar'da bulunmaz;
- Google Scholar basit taramada sınırlandırma yapmaya olanak vermezken, Scirus'da basit taramada sınırlama yapmak mümkündür;
- Scirus tarama sonuçlarını belge ilgililiği ve tarihe göre sıralama seçeneği sunar. Varsayılan ilk değer ilgililik seçeneğidir;
- Google Scholar'da arayüz ve tarama dili için seçenekler bulunur. Arama sekiz dile yapılırken, arayüz dillerinde altı dil seçeneği daha mevcuttur.

Her iki arama motorunda kullanılan Boole işleçleri ve arama özellikleri Tablo 2'de verilmektedir.

Tablo 2. Google Scholar ve Scirus'da kullanılan Boole işleçleri ve arama özellikleri

İşleç/özellik	Komutlar	
	GS	Scirus
	Boşluk, AND,	
\cap (Kesişim)	and, +, &	Boşluk AND, and, +
U (Birleşim)	OR	OR, or
Yakınlık	Yok	NEAR, near
Tamlama özelliği	“”	“”, &
Dışlama özelliği	Yok	ANDNOT, andnot, -

- Kesişim: Arama sözcüklerinden hepsinin geçtiği belgelere erişilir;
- Birleşim: Arama sözcüklerinden herhangi birinin tek tek veya birarada geçtiği belgelere erişilir;
- Yakınlık: Sorgu cümlesinde yer alan kelimelerin ilgili belgelerde en az kaç kelime arayla geçmesi gerektiğini belirler;
- Tamlama özelliği: Arama sözcüklerinin yan yana geçtiği belgelere erişimi sağlar;

² “Ryan Google Scholar Team” üyesinden yazarlara gönderilen 25.01.2007 tarihli elektronik posta mesajı.

- Dışlama özelliği: İçerikte istenmeyen kelimelerin ANDNOT, NOT gibi işlemler kullanılarak sorgu sonuçlarında yer almaması sağlanır.

Bulgular ve Değerlendirme

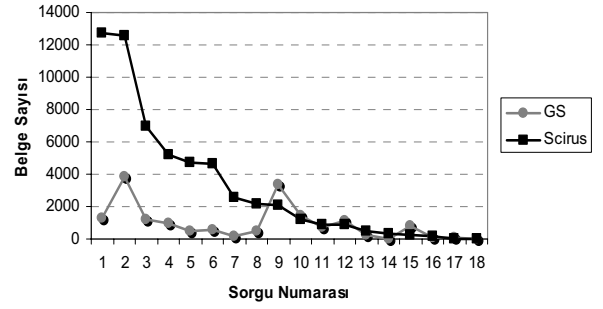
Seçilen kelime ve kelime grupları üzerinde Türkçe karakterleri değiştirmeden yapılan taramada erişilen belge sayıları Tablo 3'te verilmektedir.

Tablo 3. Türkçe karakterlerle ve onlara en uygun İngilizce karakterler ile yapılan taramalarda erişilen belge sayıları

no	Sorgu sözcükleri	Türkçe karakterlerle yapılan sorgular		İngilizce karakterlerle yapılan sorgular	
		GS	Scirus	GS	Scirus
1	çeviri/ceviri	1260	12.683	84	12.683
2	bilişim/bilisim	3860	12.524	194	12.524
3	donanım/donanim	1210	6934	80	2745
4	kurtuluş/kurtulus	946	5167	818	5167
5	veri tabanı/ veri tabani	448	4724	56	2260
6	terör/teror	590	4632	1140	4632
7	menisküs/meniskus	171	2544	2860	2544
8	telekomünikasyon/ telekomunikasyon	498	2179	77	2179
9	direnç/direnc	3390	2100	198	2100
10	yoğunluk/yogunluk	1480	1168	116	117
11	şeffaf/seffaf	688	875	27	875
12	üroloji/uroloji	1150	856	321	856
13	uçurum/ucurum	245	495	77	495
14	ödenti/odenti	11	301	17	301
15	eğim/egim	776	214	433	2235
16	yanardağ/yanardag	74	169	235	278
17	ıhlamur/ihlamur	50	36	91	380
18	rejisör/rejisor	5	24	0	24

Genel olarak, her iki arama motorunda yapılan sorgular sonucunda erişilen belge sayılarının farklı olduğu ve farklı belgelere erişildiği saptanmıştır. Sorgu sonuçlarına göre saptanan Türkçe karakterlerle yapılan taramalarda Scirus'un Google Scholar'a oranla daha fazla belgeye eriştiği Şekil 2'de daha net bir şekilde gözlenebilmektedir.

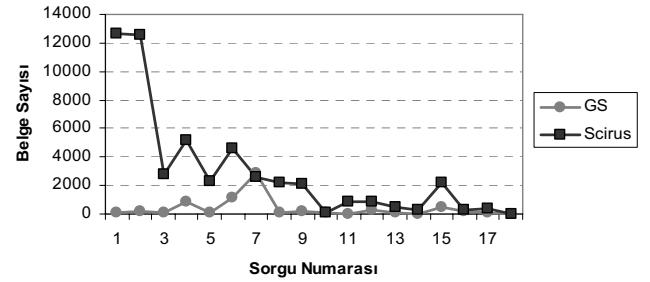
Google Scholar'da Türkçeye özgü karakterlerin İngilizleştirilmeden ("Anglicize") ve en uygun İngilizce karakterlere dönüştürülerek yapıldığı sorgular karşılaştırıldığında tutarlı bir davranış sergilenmediği gözlenmiştir. Sorgular İngilizleştirildiğinde bazı sorgular için erişilen belge sayısı artarken ("menisküs"- "meniskus" → 2689 artış), bazılarının azaldığı ("direnç" – "direnc" → 3192 azalma) gözlenmiştir. Bir sorgu için ("rejisör") hiçbir belgeye erişilememiştir.



Şekil 2. Türkçe terimler için erişilen belge sayıları

Scirus'da iki Türkçe karakter ("ğ", "ı") hariç erişilen belge sayılarının eşit olduğu ve aynı belgelerin aynı sıralama ile listendiği görülmüştür.

Karakterler İngilizleştirilerek yapılan taramalarda da Scirus'un erişilen belge sayısı açısından üstünlüğünün devam ettiği gözlenmiştir. Şekil 3'te bu durum açıkça görülmektedir.



Şekil 3. İngilizleştirilmiş terimler için erişilen belge sayıları

"ğ" ve "g" karakterini içeren anahtar sözcüklerle yapılan taramalar sonucunda her iki arama motoru da belgelere sorguda yazılan biçimiyle erişim sağlamıştır. Erişilen belgelerin sayısı açısından bir benzerlik görülmektedir. Google Scholar'da, sorgu sözcüklerinden "yanardağ" ve "yanardag" için erişilen belgelerin tümü yazar soyadlarından oluşmuştur. Scirus'da "ğ" için seçilen bir sözcüğün ("eğim") İngilizleştirilmesi ile elde edilen belge sayısı, özgün haline göre büyük artış göstermiştir. Bunun sebebi "egim" sözcüğünün EGIM (Enterprise Geographic Information Management) olarak işlem görmesidir.

"ö" ve ona en yakın İngilizce karakter olan "o" kullanılarak yapılan sorgularda Google Scholar'da sözcükler sadece yazıldıkları şekliyle yorumlanmıştır. Bunun yanında erişilen belge sayılarının da oldukça düşük olduğu gözlenmiştir. Karakterler İngilizleştirildiğinde iki sorgu için erişilen belge sayısı artarken ("teror", "odenti"), bir sorguda ("rejisör") hiçbir belgeye erişim sağlanamamıştır. Scirus'da ise "ö" ve "o" karakterleri ile

yapılan sorgularda yine aynı belgelere erişim sağlanmış, bu belgeler aynı sıra ile sunulmuştur.

“ş” ve “s” karakterlerini içeren sorgular her iki arama motoruna uygulandığında Google Scholar’daki tutarsızlığın devam ettiği saptanmıştır. “bilışim”, “kurtuluş” ve “şeffaf” taramalarında iki sorguda (“kurtuluş, şeffaf”) içerisinde hem “ş” hem de “s” karakterinin geçtiği belgelere erişim sağlanabilirken, birinde (“bilışim”) belgelere yazıldığı şekli ile erişim mümkün olmuştur. Ayrıca “ş” karakteri “s” karakterine dönüştürüldüğünde erişilen belge sayılarında azalma gözlenmiştir. “s” karakteri ile yapılan taramalarda ise yazıldığı şeklinden farklı sonuç veren tek sorgunun “bilışim” sorgusu olduğu görülmüştür. Scirus’da sorgu sözcüklerinde geçen “s” ve “ş” karakterleri sorgularda değişiklik yaratmamış, aynı belgelere aynı sıra ile erişim sağlanmıştır.

“ü” karakteri ile yapılan aramalarda Google Scholar sadece “ü” geçen sonuçlara erişmiş, “u” ile yapılan aramalar sonucunda ise “u” harfi geçen kelimeleri bulduğu gibi (meniskus) hem “u” hem “ü” geçenleri de bulmuştur (telekomünikasyon, uroloji). Scirus’daki sonuçlar ise her iki sorguda da aynı olmuştur.

Sonuç

Google Scholar ve Scirus arama motorlarında özel karakterlerin görüntülenmesi ve erişilen belge sayısı konusunda aşağıda sıralanan sonuçlar elde edilmiştir;

Her iki arama motorunda da çıkan sonuçlar alfabeyle uygun ve okunaklı bir şekilde görüntülenmiştir.

Erişilen belge sayılarında elde edilen sonuçlar arasında en dikkat çeken bulgu Scirus’un Google Scholar’dan daha fazla belgeye erişmesidir.

Google Scholar ile Scirus arasında Türkçeye özgü karakterleri yorumlama yönünden farklılıklar vardır. Anlaşıldığı kadarıyla Google Scholar girilen sorgu cümlelerinin hangi dilde olduğunu tahmin etmeye çalışmaktadır. Dolayısıyla Google Scholar sorguda sadece sorgu kelimelerinin geçtiği belgelere erişmekte, Türkçeye özgü karakterleri İngilizceleştirme yoluna gitmemektedir. Scirus ise “ç, ş, ö, ü” harflerini hem oldukları gibi hem de en uygun karakterlere çevirerek görüntülemiştir. Ancak “ğ” ve “ı” harflerini sadece tarama sorgusunda yer aldığı şekliyle sunmuştur.

Google Scholar sorgu sözcüklerinde geçen Türkçeye özgü karakterleri en yakın İngilizce karşılıklarına dönüştürmediğinden, örneğin, “ihlamur” ve “ihlamur” iki farklı sorgu olarak işlem görmekte ve farklı sonuçlara ulaşılmaktadır. Scirus 18 anahtar sözcükten 12’si için karakterler değiştirilmesine rağmen aynı sonuçları aynı sıralama ile ekrana getirmiştir (“ı” ve “ğ” hariç).

Kullanıcılar genellikle anadilde arama yapma eğilimindedirler. Bu sebeple üst veri oluşturulurken gerek web sitelerini tasarlayanların gerekse arama motoru geliştirenlerin kullanıcı odaklı düşünerek hareket etmeleri gerekmektedir. Türkçe açık erişim kaynaklarının üst

verileri oluşturulurken Türkçeye özgü karakterler dikkate alınmalıdır. Google Scholar ve Scirus gibi açık erişim bilgi kaynaklarına erişim sağlayan arama motorları da üst verilerin kaynağın diline uygun olup olmadığını kontrol edebilmelidir.

Arama motorlarını karşılaştırmaya yönelik çalışmalarda genellikle belgelerin ilgili olma durumu da değerlendirilmektedir. Fakat bu çalışmada erişilen belgelerin ilgili olma durumuna bakılmadığından kapsam dâhilindeki arama motorlarının birbirlerinden üstün oldukları söylenemez. Google Scholar ve Scirus arama motorları bilgi erişim performansları (ilgililik, anma, duyarlık ve çakışma) açısından da değerlendirilebilir.

Teşekkür

Bu çalışma için bizi cesaretlendiren ve çalışmamızı destekleyen hocalarımız Prof. Dr. Yaşar Tonta ile Araş. Gör. İrem Soydal’a teşekkür ederiz.

Kaynakça

- Al, U. ve Küçük, M. (2003). Üst veri standartları ve uygulamaları. *Hacettepe Üniversitesi Edebiyat Fakültesi Dergisi*, 20(1), 167-185.
- Alır, G. (2007). *Standartlar ve protokoller*. 17 Mayıs 2007 tarihinde <http://yunus.hacettepe.edu.tr/~gulbun/standartlar26122005.doc> adresinden erişildi.
- Bar-Ilan, J. ve Gutman, T. (2005). How do search engines respond to some non-English queries? *Journal of Information Science*, 31, 13-28. 20 Mayıs 2007 tarihinde <http://jis.sagepub.com/cgi/reprint/31/1/13> adresinden erişildi.
- Chowdhury, C.G. (2004). *Introduction to modern information retrieval* (2nd ed). London: Facet. 26 Ocak 2007 tarihinde <http://www.britishcouncil.org/lithuania-information-centre-collections-issue-2.doc> adresinden erişildi.
- Felter, L. M. (2005). Google Scholar, Scirus, and the Scholarly Search Revolution. *Searcher* 2(13), 43-48. 20 Mayıs 2007 tarihinde http://www.scirus.com/press/pdf/searcher_reprint.pdf adresinden erişildi.
- Google: *About Google Scholar*. (2007). 10 Mayıs 2007 tarihinde <http://scholar.google.com/intl/en/scholar/about.html> adresinden erişildi.
- Gordon, M. ve Pathak, P. (1999). Finding information on the World Wide Web: the retrieval effectiveness of search engines. *Information Processing & Management*, 35, 141-180. 21 Ocak 2007 tarihinde ScienceDirect 2007 veri tabanından erişildi.
- Gudivada, V.N., Raghavan, V.V., Groksy, W.I. ve Kananagottu, R. (1997 Eylül-Ekim). Information retrieval on the World Wide Web. *IEEE Internet Computing*, 1(5), 58-68. 20 Mayıs 2007 tarihinde <http://ieeexplore.ieee.org/iel1/4236/13574/00623969.pdf?tp=&arnumber=623969&isnumber=13574> adresinden erişildi.
- Guistini, D. ve Barsky, E. (2005). A look at Google Scholar, PubMed and Scirus: comparisons and recommendations. *Journal of the Canadian Health Libraries Association*, 26(3), 85-89. 19 Mayıs 2007 tarihinde <http://pubs.nrc-cnrc.gc.ca/jchla/jchla26/c05-030.pdf> adresinden erişildi.

- Lancaster, F.W. ve Fayen, E.G. (1973). *Information retrieval online*. Los Angeles, CA.: Melville Publishing Company.
- Moukdad, H. ve Cui, H. (2005). How do search engines handle Chinese queries? *Webology*, 2(3). 18 Mayıs 2007 tarihinde <http://www.webology.ir/2005/v2n3/a17.html> adresinden erişildi.
- Moukdad, H. ve Large, A. (2001). Information retrieval from full-text Arabic databases: can search engines designed for English do the job? *Libri*, 51, 63-74. 17 Mayıs 2007 tarihinde <http://librijournal.org/pdf/2001-2pp63-74.pdf> adresinden erişildi.
- Noruzi, A. (2005). Google Scholar: the new generation of citation indexes. *Libri*, 55, 170-180.
- Notess, G. (2005). Scholarly web searching: Google Scholar and Scirus. *Online*, 29(4), 39. 20 Mayıs 2007 tarihinde www.infotoday.com/online/jul05/OnTheNet.shtml adresinden erişildi.
- OCLC. (2002). *Country and language statistics*. 10 Mayıs 2007 tarihinde <http://www.oclc.org/research/projects/archive/wcp/stats/intnl.htm> adresinden erişildi.
- Salton, G. (1989). *Automatic text processing*. Boston, MA: Addison-Wesley.
- Salton, G. ve McGill, M.J. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.
- Scirus: *About Scirus*. (2007). 9 Mayıs 2007 tarihinde <http://www.scirus.com/srsapp/aboutus/> adresinden erişildi.
- Soydal, İ. (2000). *Web arama motorlarında performans değerlendirmesi*. Yayımlanmamış yüksek lisans tezi. Hacettepe Üniversitesi, Ankara.
- Sroka, M. (2000). Web search engine for Polish information retrieval: question of search capabilities and retrieval performance. *International Information & Library Review*, 32, 87-98. 9 Mayıs 2007 tarihinde ScienceDirect 2007 veri tabanından erişildi.
- Thornely, J. (2000). Metadata and deployment of Dublin Core at State Library of Queensland and Education Queensland, Australia. *OCLC Systems & Services*, 16(3), 118-129.
- Tonta, Y. (1995). Bilgi erişim sistemleri. *Türk Kütüphaneciliği*, 9, 302-314, 1995. 19 Mayıs 2007 tarihinde <http://yunus.hacettepe.edu.tr/~tonta/yayinlar/tkbes95.pdf> adresinden erişildi.
- Tonta, Y., Bitirim, Y. ve Sever, H. (2002). *Türkçe arama motorlarında performans değerlendirme*. Ankara: Total Bilişim. 19 Mayıs 2007 tarihinde <http://eprints.rclis.org/archive/00009697/01/tonta-bitirim-sever-arama-motorlari.pdf> adresinden erişildi.
- Translate to success: Internet language use statistics*. (2005). 13 Mayıs 2007 tarihinde <http://www.translate-to-success.com/internet-language-use.html> adresinden erişildi.
- Unicode: Evrensel kod nedir?* (2007). 16 Mayıs 2007 tarihinde <http://www.unicode.org/standard/translations/turkish.html> adresinden erişildi.