

Google Arama Motorunda İndekslenmiş Sayfaların Artırılması

Ali Çaylı¹, Adil Akyüz¹, Ercan Efe², Sait Üstün¹

¹ Kahramanmaraş Sütçü İmam Üniversitesi, Enformatik Bölümü, 46100, Kahramanmaraş

² Kahramanmaraş Sütçü İmam Üniversitesi, BAUM, 46100, Kahramanmaraş

alicayli@ksu.edu.tr, akuz@ksu.edu.tr, eefe@ksu.edu.tr, ustun@ksu.edu.tr

Özet: Günümüzde bilginin paylaşımı oldukça önemlidir. Bilgiye ulaşmak için yaygın olarak arama motorları kullanılmaktadır. Arama motorları sayesinde bilginin hangi kaynaktan geldiğinin tespit edilerek bilgiye ulaşılması son yıllarda yaygın olarak kullanılan yöntemdir. Bu sebeple üretilen bilgilerin arama motorları üzerinde indekslenmesi erişilebilir bilgi miktarının artması anlamına gelmektedir.

Bilgi üreten kurumların başında gelen üniversitelerin, ürettikleri bilgileri dünya ile paylaşması arama motorları sayesinde mümkün olmaktadır.

Bu çalışmada, Dünya ve Türkiye üniversitelerinin Google Arama Motorunda indekslenmiş sayfa sayıları php tabanlı bir yazılımla, data sunucuları taranarak hesaplanmış ve yetersiz olduğu görülmüştür ayrıca arama motorlarında indekslenmiş sayfa sayısının artırılması ile ilgili önerilerde de bulunulmuştur.

Anahtar Kelimeler: Bilgi, Arama Motoru, Google, Indexli Sayfa Sayısı, Web Sayfası

Increase of Indexed Page in Google Search Engine

Abstract: Today, it is very important to share knowledge. Search engines are used to reach knowledge. Therefore indexing produced information on search engine, lead to increase of reachable information.

Universities, the main knowledge producers, to share the produced information with world are possible by the search engines.

In this study, the page of Turkish and world universities indexed in Google Search Engine were investigated using php based software, by searching data servers, and found insufficient. Suggests for increasing indexed page in search engine was also given.

Keywords: Knowledge, Search Engine, Google, Indexed Page, Web Page

1.Giriş

Son yıllarda arama motorları kullanılarak bilgiye ulaşmak sıkça kullanılan bir yöntemdir. Artık aranan bir bilgiye arama motorlarında birkaç anahtar kelime girilerek dünya kütüphanesinde ulaşmak mümkün hale gelmiştir. Arama motorları belli aralıklarla “robot” yazı-

lımlarıyla web sitelerini tarayarak kendi veritabanlarına eklerler. Kullanıcılar arama motorlarının web sitelerinde bulunan anahtar kelime alanlarını kullanarak bu veritabanı üzerinde aranan bilginin hangi web sayfasında olduğunu öğrenerek bu sayfalara erişim sağlarlar. Arama motorları bilgi ile bilgiye erişmek isteyenler arasında bir köprü görevi görmektedir.

Arama motorları üzerinde kayıtlı olan (indekslenmiş) web sayfaları sayısı, sayfa sahibi olan kurumun bilgilerine erişimi artırmaktadır. Arama motorlarından aranan anahtar kelimelerle ulaşılabilen sayfa sayısını artırmak için öncelikli olarak bilgilerin web ortamına taşınması ve sonra da arama motorları robotlarının bu sayfalara erişiminin kolaylaştırılması gerekmektedir.

2. Arama Motorları

Arama motoru kısaca bilgiye erişme sistemi olarak tanımlanabilir [1]. Başlıca üç bileşenden oluşur. Bunlar web robotu, arama indeksi ve kullanıcı arabirimidir. İnternetteki sayfalar birbirine bağlantılar (linkler) vasıtasıyla bağlıdır. Bu sayede bir sayfadan diğerine geçmek mümkündür. Şu anda en çok bilinen ve kullanılan arama motorları google.com, yahoo.com, msn.com'dur. Bunların yanında İnternet üzerinde hizmet veren birçok arama motorları da aynı işlevleri yerine getirmektedir. Arama motorları, web robotu (crawler) yazılımlarla web sayfalarını incelerler ve içerisinde bulunan kelimeleri, cümleleri, resimleri vb diğer içeriği kendi veritabanları üzerine kaydederek bunlardan indeks oluştururlar. Bu işlemi belli aralıklarla sürekli yaparak güncel bir indekse sahip olurlar. Kullanıcılar ise istemci web arabirimini kullanarak veya izin verilen başka erişim yöntemleri ile arama motorlarının indekslerinden anahtar kelimeleri veya bir cümleyi girerek aradıkları bilgilerin nerede olduğu tespit ederler.

Arama motoru ilk olarak Montreal'deki McGill Üniversitesinde öğrenci olan Alan Emtage tarafından 1990 yılında yapılmıştır. Bu arama motoru FTP sitelerindeki dosya listelerini download ederek bunları aranabilir bir veritabanına kaydediyordu. Bu sistemde dosya içeriklerinde arama yapmak mümkün değildi. 1991 yılında University of Minnesota'dan Mark McCahill tarafından geliştirilen Gopher hizmeti kullanılmaya başlandı. Bu sistemle dosya içeriklerinde de tarama yapılabilirdi. 1993 yılında MIT'den Matthew Gray tarafından ilk web arama motoru olan Wandex geli-

tilmiştir. Bu dönemde hizmet veren ilk arama motorlarından bir diğeri olan Aliweb halen günümüzde de hizmet vermeyi sürdürmektedir. Tam metin arama yapabilen ilk arama motoru Webcrawler'da 1994'de hizmete girmiştir ve kullanıcıya web sayfası içeriğinde herhangi bir kelimeyi arayabilme imkânı vermiştir. Daha sonraları bu sistem Excite, InfoSeek, Inktomi, Northen Light ve Altavista gibi arama motorlarında da standart hale gelmiştir. 1990'lı yıllardan sonra en parlak yatırımlar arama motorları için yapılanlar olmuştur.

3. Metot

Bu araştırmada üniversitelerin Google arama motorunda indekslenmiş web sayfası sayısını tespit etmek amacıyla php tabanlı bir yazılım geliştirilmiştir. Yüksek Öğretim Kurulu web sitesinden alınan Türkiye Üniversiteleri web adresleri Mysql veritabanı sunucusu üzerine girilmiştir. Ayrıca Shanghai Jiao Tong University tarafından her yıl yapılan üniversite derecelendirme listesinde (Academic Ranking of World Universities) ilk 100 sırayı alan üniversitelerin web sayfaları da Mysql veritabanına aktarılmıştır [2]. Php tabanlı script programla bu üniversitelere ait web sayfası sayıları oy-oy.eu sitesinin XML formatlı sonuç çıktıları ayrıştırılarak yine Mysql veritabanına kaydedilmiştir. Her üniversitenin ana etki alanındaki web sayfası sayısı 3 farklı zamanda 43 farklı veri merkezi taranarak alınmıştır. Sonuçlar hesaplanırken elde edilen 43 farklı sistemden alınan sonucun aritmetik ortalaması alınmıştır.

4. Üniversitelerde Mevcut Durum

Yapılan bu taramalar sonucunda Türkiye'de en fazla indeksli web sayfası sayısına sahip üniversite 253,476 sayfa ile Bilkent Üniversitesi olduğu görülmüştür. Bilkent Üniversitesini Orta Doğu Teknik Üniversitesi (139,429) ve Ankara Üniversitesi (118,786) izlemektedir. Sayfa sayıları bakımından ilk ve son sıradaki 10'ar üniversite Çizelge 1 ve Çizelge 2'de verilmiştir. Türkiye'de tüm üniversitelere ait toplam web

sayfası sayısı yaklaşık 1.950.000'dir. Bu rakam Amerikan üniversiteleri için 1.490.000.000, Çin Üniversiteleri için 17.900.000'dir.

Sıra	Üniversite	Ortalama Sayfa Sayısı (Adet)
1	Bilkent Üniversitesi	253 476
2	Orta Dogu Teknik Üniversitesi	139 429
3	Ankara Üniversitesi	118 786
4	İstanbul Teknik Üniversitesi	91 083
5	Hacettepe Üniversitesi	89 081
6	Dokuz Eylül Üniversitesi	87 136
7	Inonu Üniversitesi	84 295
8	Kahramanmaraş Sutcu Imam Univ.	84 205
9	İstanbul Bilgi Univ.	79 283
10	Gazi Üniversitesi	79 145

Çizelge.1 Türkiye'de en fazla web sayfasına sahip ilk 10 üniversite

Sıra	Üniversite	Ortalama Sayfa Sayısı (Adet)
84	Izmir Ekonomi Üniversitesi	374
85	Yasar Üniversitesi	372
86	Girne Amerikan Üniversitesi	359
87	Ufuk Üniversitesi	347
88	Mimar Sinan Guzel Sanatlar Univ.	326
89	Uluslararası Kıbrıs Üniversitesi	270
90	Lefke Avrupa Üniversitesi	247
91	Kapadokya Meslek Yuksekokulu	174
92	Cag Üniversitesi	161
93	Mersin Deniz Ve Ticaret Meslek Yuksekokulu	131

Çizelge.2 Türkiye'de en az web sayfasına sahip son 10 üniversite

Sıra	Üniversite	Ortalama Sayfa Sayısı (Adet)
1	Massachusetts Inst Tech (Mit)	3 628 333
2	Harvard Univ	3 572 143
3	Pennsylvaniastate Univ	3 183 571
4	Stanford Univ	2 704 524
5	Univ California	2 305 476
6	Kyoto Univ	2 154 960
7	Univ Wisconsin	1 764 250
8	Univ Washington	1 703 067
9	Univ Illinois	1 645 714
10	Univ Arizona	1 347 381

Çizelge.3 En iyi 100 Üniversiteden ilk 10'unun indeksli web sayfası sayıları

Dünya sıralamasında ilk 100'de kabul edilen üniversitelerin ilk 10'unun web sayfası sayıları ise Çizelge 3'de verilmiştir. Buna göre; ilk sırayı 3,628,333 sayfa ile Massachusetts Inst Tech (MIT), ikinci sırada Harvard Üniversitesi (3,572,143), üçüncü sırada ise Pennsylvania State Üniversitesi (3,183,571) almıştır.

5. Sonuç ve Öneriler

Birçok üniversite web sayfası sadece resmi yazışmalar, duyurular, haberler ile buna benzer tanıtım faaliyetleri için aktif bir şekilde kullanılmaktadır.

Arama motorları sayesinde Dünya'nın herhangi bir noktasından ziyaretçiler günde defalarca web sayfalarına aradıkları bir bilgi için gelmektedirler. Bu ziyaretçilerin, aradıkları bilginin bir akademik kuruluş tarafından sunulmuş olmasından da oldukça memnuniyet duyacakları açıktır.

İnternet ortamında elde edilen bilgiler her zaman güvenilebilen bilgi değildir. Güvenilir ve doğru bilgiyi en iyi şekilde sunabilecek kurumlar üniversitelerdir. Üniversitelerimiz de bunu en hızlı, kolay ve ucuz bir şekilde web ortamını kullanarak sağlayabilirler.

İndeksli web sayfası sayısının artırılması için öncelikli yapılması gereken, çok sayıda yararlı bilgi sayfaları üretmektir. Bu amaçla üniversitelerde yayınlanan makaleler, kitaplar, ders notları, eğitim dokümanları gibi materyallerin web sayfalarına aktararak internet'e sunulması gerekmektedir.

Sayfalar bir yerden bağlantılı olmalıdır. Arama motorları bağlantıları izleyerek indeks oluşturduğundan hiçbir yerde bağlantı verilmemiş sayfalarınız varsa onlar indekslenemeyecektir.

Yeni açılan alt alanlara ayrılmış web siteleri arama motorlarına eklenerek hızlı bir şekilde indekslenmesi sağlanabilir. Bu amaç için arama motorları ana sayfalarında url ekleme bölümleri vardır.

Web sitesindeki çalışmayan bağlantılar günlük dosyasından tespit edilip bunlar düzeltilmelidir.

Web sitesindeki sayfaların yeri sık sık değiştirilmemelidir. Web sayfalarının Başlık bilgileri (META TAG) içerisinde sayfa ile ilgili açıklayıcı bilgiler ve anahtar kelimeler yazılmalıdır. Ayrıca başlık etiketi (<title>) bilgisi o sayfaya özgü olmalıdır. Tekrarlanan başlık etiketi bilgisinden kaçınılmalıdır.

SiteMap hizmeti kullanılarak web sunucu günlük dosyaları Google' a gönderilebilir [3].

Ayrıca bu hizmeti kullanarak sayfalarda var olan hatalı linkler ve diğer hatalar tespit edilebilir. Bu sayede bir web sitesindeki tüm sayfaların indekslenmesi sağlanabilir.

6. Kaynaklar

[1] The Free Encyclopedia Wikipedia Web Site, en.wikipedia.org/wiki/Search_engine, *Wikipedia*

[2] Shanghai Jiao Tong University, Web Site, ed.sjtu.edu.cn/ranking.htm, *Shanghai Jiao Tong University*

[3] Google Inc Web Site, <http://www.google.com/support/webmasters/?hl=tr>, *Google*