

SANAL DÜNYADA DİLLERİN KULLANIM ORANLARI

LANGUAGES IN CYBER WORLD

Yrd.Doç.Dr.Ahmet TARCAN¹

ÖZET

İnternet doğduğu günden bu yana İngilizcenin hakim olduğu bir dünyadır. Son yıllarda diğer dillerinde katılımıyla sanal dünya çok dillilikle tanışmıştır. Makalede sanal dünyada dillerin kullanılma oranını tespit edilmeye çalışılmıştır. Örneğin kaç milyon kişi interneti Türkçe olarak yada Fransızca olarak kullanmaktadır. Günümüzde büyük bir Pazar haline gelen internette tüketiciye ulaşmanın önündeki en büyük engellerden biriside dil engelidir. İnternet ortamında kullanılan dillerin oranı e-ticaret ve internet reklamcılığı sektöründe çalışanlar için büyük önem taşımaktadır.

İnternet kullanıcılarının dil profili için Global Reach, Internet World Stats'ın 1996 yılından beri yapmış oldukları ölçümler esas alınmıştır. Global Reach ise verilerini Uluslararası Telekomünikasyon birliğinin (UIT) her ülkedeki kullanıcılara göre verdiği rakamlara dayandırmaktadır. Elde edilen veriler dil atlasları dikkate alınarak rakamlara dönüştürülmektedir

Bu çalışmada internet kullanıcılarının dil profili ile Arama motorlarında indekslenen 15 milyarın üzerindeki web sayfasının içerikleri ile ilgili sayısal rakamlar elde edilmeye çalışılmıştır. Bunun için dilbilimsel bazı ölçütlerden yararlanılmıştır. Web sitelerinin dil profili ise arama motorlarından elde edilen rakamlardan yola çıkılarak varsayımlar üretilmiştir. Her bir dilde kullanılan frekansı yüksek kelime, hece ve ekler tespit edilerek buna göre arama motorundan İngilizce, Fransızca vb. dillerdeki sayfalar görüntülenmiştir. (Örneğin İngilizcede the, Fransızcada de gibi)

Anahtar Kelimeler: İnternet, Diller, Web, Dilbilim, Görünmeyen Web

Abstract:

Internet has been an English speaking world since its origine. But in the latest years it is clear that multilinguism began to dominate the Internet world. This article studies the languages profiles in cyber world, in other words; how many people use Internet in French, in Turkish and so on. The numbers are mainly important for e-sellers and politics makers. Today there are several companies that studies Internet language demographics. The data of this article come from Global Reach, Internet World Stats and search engines. The number of websites in different languages change according to the kind of the search engine, the weeks and even according to the day.

The article sum up also some analyze techniques of web contents according their languages. It seems there are three linguistics ways to determine the number of WebPages by language: First by using the suffixed and prefixed of each language, second by typing the most frequently used words in a language, for example "the" in English, "de" in French and so on, the third is to use some specifics character's in each language. (vowels or consonants for example the Ç, İ Ğ in Turkish, oe in French, and so on.)

Keywords: Internet, Languages, Web, Linguistic, invisible Web

¹ Dicle Üniversitesi, Eğitim Fak.Öğretim Üyesi

GİRİŞ

İnternet üzerindeki sanal topluluklar günümüzde sosyal bilimcilerin en çok ilgisini çeken araştırma konularından birisidir. Bu toplulukların sosyo-ekonomik yapıları, kişilik özellikleri gibi konularda çok sayıda çalışma yapılmıştır.

Sanal toplulukların dilsel özellikleri ile ilgili de uluslar arası düzeyde bir çok araştırma ve proje yürütülmektedir. Bu çalışmaların bütün hedefi ağ üzerindeki nüfusun dil profilini belirlemektir. Bu amaçla üniversitelerde, Teknoloji enstitülerinde ve çok uluslu şirketler² bünyesinde farklı yapılarda ve düzeylerde internet üzerindeki her türlü değişiklik tespit edilmektedir.

Bütün bu araştırma ekiplerine ve dev şirketlerin yatırımlarına rağmen internet kullanıcılarının dil profili tam olarak belirlenememiştir. Teknik sorunlar aşıp kullanıcıların sayısı ve bağlandıkları ülkeler tespit edilse bile o anda ağ üzerinde bulunan kullanıcıların hangi dilde işlem yaptıklarını tespit etmek oldukça güçtür. Örneğin kullanıcı bir Alman web kanalının sunduğu chat odasında Türkçe görüşme yapıyor olabilir veya tersi.

Bütün bunlar tespit edilse bile elde edilen veriler sadece o an için geçerlidir. Çünkü her an dil profilleri farklı olan kitleler online olabilmekte veya hattan düşebilmektedir.

Dolayısıyla bugün elde edilen veri ve istatistikler yayınlandıklarında çoktan güncelliklerini, geçerliliklerini yitirmiş olmakta, ancak yaklaşık rakamlar olarak hizmete sunulabilmektedir.

Bugün internet ortamında var olan bilgilerin büyük bir kısmı İngilizcedir. İngilizcenin internette baskın dil olması Anglo-Saksonlar tarafından doğal karşılanırken, diğer dil topluluklarının mensupları özellikle Fransız ve İspanyollar bu durumu diğer dillere yapılan bir haksızlık olarak görürler ve İngilizcenin sanal dünyadaki hegemonyasına şiddetle karşı çıkarlar. Bu kesimler İnternetin Amerikalılar tarafından keşfedildiği savını da benimsemezler.

² “Dil Gözlem laboratuvarının” da müdürü olan Prof. Yoshiki MIKAMI'nin oluşturduğu ekip, çeşitli robot programlar sayesinde sanal dünyada her an kaydedilen dilsel gelişmeleri belirlemekte ve değerlendirmeye tabi tutmaktadır. Crawler Robots 10 milyarı aşkın web sitesini yılda en az bir kez ziyaret ederek listelemekte, LPIM adlı program ise hangi web sayfasının hangi dili, diyalektleri ile birlikte kullandığını çözmektedir. Söz konusu yazılımlar henüz deneme aşamasında olup proje devam etmektedir. <http://www.language-observatory.org/>

Ayrıca Global Reach adlı danışmanlık şirketi internet üzerindeki dillerin dağılımını ve istatistiklerini 1996 yılından beri takip etmektedir. <http://global-reach.biz/>

İnternet World Stats adlı şirkette yine dünyanın 233 bölgesinde internet kullanıcılarına ait istatistikleri araştıran önemli bir kuruluş. <http://www.internetworldstats.com/>

Onlara göre 1969'da Pentagonun resmi kurumlara ve üniversitelere ait bilgisayarları ağ üzerinde birbirine bağlama projesi internetin ilk adımı olabilir. Ancak bugün bildiğimiz anlamdaki World Wide Web ve HTML 1989/90 yılları arasında İsviçre'nin Cenevre şehrindeki CERN laboratuvarında üretilmiştir. Netscape'in atası sayılan MOSAIC adlı tarayıcı 1993 yılında ilk kez İsviçre'de denendi. Amerikalıların bunu alıp geliştirmesi ise daha sonraki yıllarda gerçekleşmiştir. Bu tespit göz önüne alındığında internetin Amerikalılar tarafından bulunduğu dolayısıyla İngilizcenin internette önceliğinin olduğu düşüncesi İngilizce konuşmayan ülkelerde pek kabul görmemektedir.

1. Araştırmanın Veri Tabanı ve Yöntemler

Çalışmanın veri tabanı iki ayrı bölüm halinde sınıflandırılmıştır:

1. İnternet kullanıcılarının dil profili

2. Arama motorlarında kayıtlı olan web sitelerinin dil profili

İnternet kullanıcılarının dil profili için Global Reach, Internet World Stats'ın 1996 yılından beri yapmış oldukları ölçümler esas alınmıştır. Global Reach ise verilerini Uluslararası Telekomünikasyon birliğinin (UIT) her ülkedeki kullanıcılara göre verdiği rakamlara dayandırmaktadır. Elde edilen veriler dil atlasları dikkate alınarak rakamlara dönüştürülmektedir.

Web sitelerinin dil profili ise arama motorlarından elde edilen rakamlardan yola çıkılarak varsayımlar üretilmiştir. Her bir dilde kullanılan frekansı yüksek kelime, hece ve ekler tespit edilerek buna göre arama motorundan İngilizce, Fransızca vb. dillerdeki sayfalar görüntülenmiştir.³ Web sayfalarının rakamları ile ilgili verilerin büyük çoğunluğu Google arama motorunun indekslerine dayandırılmıştır. Yahoo ve Altavista'nın indeksleri ise karşılaştırma gerektiğinde kullanılmıştır. Arama motorlarında bulunan "ileri arama seçenekleri" kullanılarak taramalar yapılmıştır. Sonuç sayfalarının dili bölümünde istenen dil örneğin İngilizce işaretlenmiş, sonuç sayfalarının biçimi bölümünde ise "tüm biçimler" seçeneği tıklanmıştır. (PDF, DOC, RTF, PPT XLS vb.)

³ Örneğin: THE tanımlığından hareketle yapılan aramada İngilizce web sayfalarının sayısı:

Google'da 856000000, Altavista'da 1030000000, Yahoo'da 998000000 olarak tespit edilmiştir.

Yine DE tanımlığından hareketle yapılan aramada Fransızca Web sayfalarının sayısı: Google'da 216000000, Yahoo'da 511000000 olarak tespit edilmiştir. Elif harfinden (I) hareketle yapılan taramada Arapça sayfaların sayısı 2.330.000, J (jav) harfinden hareketle yapılan taramada ise Arapça web sayfalarının sayısı 7810000 olarak tespit edilmiştir. Yahoo'da ise aynı harf ile yapılan taramada Arapça sayfaların sayısı 24 900 000 olarak belirlenmiştir.

⌋ (el) tanımlığından hareketle yapılan taramada 5,320,000 Arapça sayfa tespit edilmiştir.

Tarih bölümünde var olan seçeneklerden (son 3 ay, 6 ay vs.) “tüm zamanlar” veya “herhangi bir zaman” tıklanmıştır. “Geçiş yerleri Arama yapılacak alan/yer” bölümüne ise “Sayfanın Herhangi Bir Yerinde” seçeneği tercih edilmiştir. Bu alanın diğer seçenekleri ise şunlardır: “Sayfa Başlığı” “Sayfa Bağlantılarında” , “Sayfa İçeriğinde”.

Google’a dayandırılan rakamların büyük çoğunluğu 14 Aralık 2005/15 Ocak 2006 arasındaki çeşitli tarihlere aittir. Arama motorlarındaki rakamların neredeyse her gün değiştiği anlaşılmıştır.

İnternet ortamında var olan toplam web sayısı bilinmediğinden arama motorlarından elde edilen sayısal değerler yüzdelik dilimlere dönüştürülmedi. Ayrıca yeryüzünde konuşulan 6000 üzerindeki dilden kaçının şu anda internette ne oranda yer aldığı da bilinmemektedir.

2.İnternet Kullanıcılarının Dil Profili (Online Language Populations)

Günümüzde internet kullanıcılarının dil profillerini belirleyen çeşitli çalışmalar yapılmaktadır. Online olan kişi sayısına göre belirlenen rakamlar daha sonra ülkelerin IP çıkışlarına ve ana serverlerin sayaçları yardımıyla ülkelere göre ayırt edilebilmektedir. Bunun için çeşitli otomatik robot yazılımlar kullanılabilir .

Yinede bütün bu teknolojik imkanlar online kişilerin dillere göre dağılımını belirlemede çok fazla net ve kesin bilgiler içermemektedir. Bunu aynı tarihlerde farklı şirketler tarafından yapılan ve farklı çıkan rakamlardan tahmin edebilmekteyiz. Ağ üzerinde bulunan kişilerin (online populations) sayısını belirlemenin kendi içinde çeşitli zorlukları vardır. İnternet dilinde online olmak, o anda internete bağlı olmak, internet üzerinde herhangi bir işlem yapıyor olmak demektir. Dolayısıyla *online populations* kavramı web sayfalarını ziyaret edenleri, chat yapanları, e-posta kullananları, ağ üzerinde ticari işlem yapanları kapsamaktadır.

Peki online olan milyarlarca kişinin o anda hangi dili kullandıkları nasıl ayırt edilebilmektedir? Kullanıcıların dilleri hangi veriler temel alınarak belirlenebilir? Kullanıcıları tam olarak saymak ve daha sonra da dillerine göre ayırt etmek için ilk akla gelen belirleyici unsurlar şunlardır;

- IP numaralarının ülkelere göre dağılımı
- Web sitelerinin sayaçları
- Web sitelerinin dilleri

- Web sitelerinin ülkelere göre uzantıları (Fr.Tr vs)
- Elektronik postalarda kullanılan diller
- E-ticaretin yapıldığı diller
- Chat odalarında mevcut kişi sayıları ve dilleri
- Kullanıcıların ana dilleri veya o anda kullandıkları diller

Web sitelerinin sayıları dikkate alındığında siteler nasıl gruplandırılacak? Ziyaretçilerin o an hangi dili kullandıkları nasıl ayırt edilecek? Web siteleri iki dilli veya çok dilli ise durum daha da karışık hale gelecektir Eğer web sitelerinin dili temel alıyorsa o zaman Rusça bir sitede görsel öğelere (örneğin fotoğraflara) bakan ve Rusça bilmeyen kullanıcıları hangi kategoriye koymak gerekir?

3. Ülke Sembolleri

Tr. Fr. Uk. De .Fr. gibi web uzantıları da şu anda sanal dünyada milliyetçilik sembollerinin ötesinde bir değerlendirmeye tabi tutulamaz. Sadece ulusal varlık göstergesi olarak gözüken bu uzantılardan o sitenin dilini tahmin etmek mümkün. Ancak bu son derece sınırlı bir gerçeklik içerir. Örneğin uzantısı .tr olan siteler Türkçe yayın yapıyor olabilir, ancak uzantısı com olup Türkçe yayın yapan site sayısı uzantısı .tr olan site sayısından fazlaysa o zaman uzantılar dil profilini belirlemede çok önemli bir kriter olarak değerlendirilemez.

Bu uzantıları en çok kullanan ülkelerin başında İngiltere, Almanya, Fransa, Japonya gelmektedir. Google arama motorunun 05.01.2006 tarihindeki arşivleri esas alınarak yapılan ölçümlerde uzantılarına göre web sayfalarının sayıları şu şekildedir:

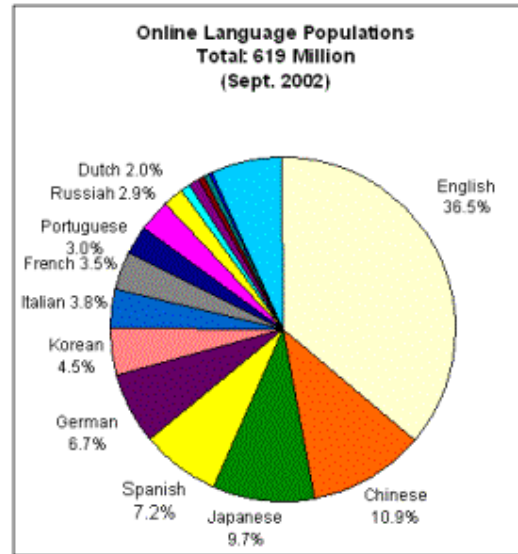
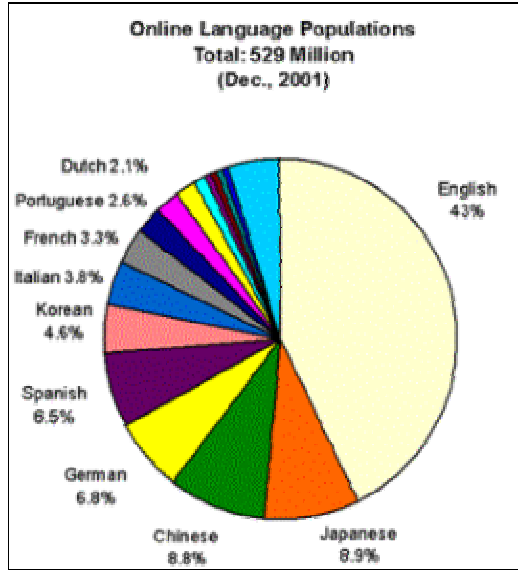
.uk(İngiltere uzantılı siteler)	1,970,000,000
.de (Almanya)	561,000,000
.fr (Fransa)	333,000,000
.jp (Japonya)	259,000,000
.cn(Çin basit ve eski)	59,100,000 +3,250,000
.es(İspanya)	182,000,000
.ca(Kanada-İngilizce)	1,330,000,000
.ca(Kanada-Fransızca)	100,000,000
.tr (Türkiye)	15,800,000

Ölçümlerde dillerin oranı kullanıcıların ana dillerine göre ölçülebilir. Ancak online olan kişilerin o anda hangi dili kullandıklarını tespit etmek son derece güçtür. Örneğin e-posta adresini Japon bir Web sayfasından alan bir kişi dış dünya ile haberleşmek için İngilizce kullanıyor olabilir veya ana dili Çince olan birisi İngilizce konuşmayı tercih ediyor olabilir.

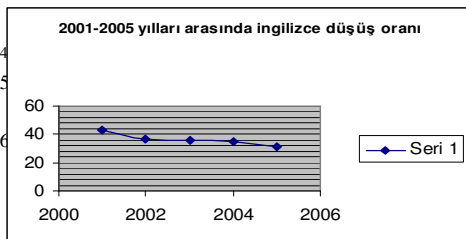
2000 yılında ABD'deki bir araştırma 45 milyon kişinin iş ortamı, internet vb. ortamlar da evde kullandıkları dilden daha farklı bir dil konuştuklarını ortaya koymuştur.⁴ Dolayısıyla internet üzerinde kullanıcıların kulanmış oldukları dili her zaman onların ana dili olarak görmek mümkün değildir.

Global Reach adlı danışmalık şirketi 1990'lı yıllardan beri internet kullanıcılarının dillerini belirlemek için önemli ve titiz çalışmalar yapmaktadır, ancak şirketin tespit ettiği rakamlar tam ve eksiksiz değildir, bu rakamlarda % 20 'lik gibi önemli bir yanılma payının olduğu iddia edilmektedir.⁵

Aşağıda Global-Reach'ın yıllara göre internet kullanıcılarını belirlediği rakamlar verilmektedir:



⁶ AusWeb



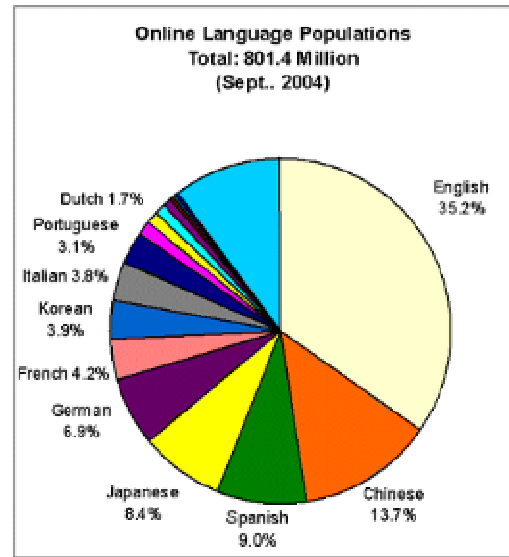
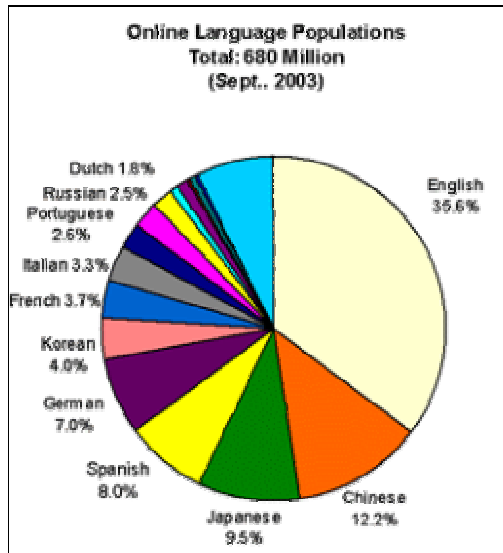
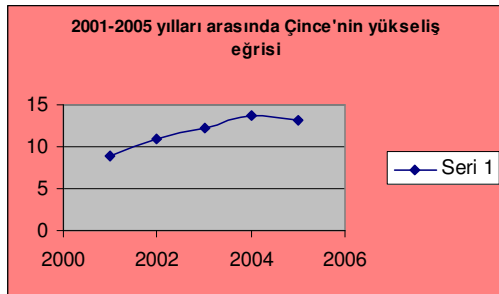
www.ausweb.gov.au/aw03/papers/edwards2/paper.html

es, <http://unesdoc.unesco.org/images/0014/001421/142186f.pdf>

www.ausweb.gov.au/aw03/papers/edwards2/paper.html

2001 ile 2002 yılları arasında en belirgin değişiklik İngilizce de görülmektedir. İngilizce bir yılda yaklaşık 7 puan gerilemiştir. 2001 yılından 2005 yılına kadar İngilizce de hiç yükselme görülmemesi dikkat çekicidir.

2001-2002 yılları arasındaki İngilizcenin % 43'ten % 36.5'e düşmesinin nedenlerinden biriside Japonca, Çince ve İspanyolcadaki önemli artışlar olabilir. Bu yıllar arasında Portekizce, İspanyolca, Japonca, Çince Fransızcanın oranlarında ilerleme, Korece, Almancada ise gerileme söz konusudur.



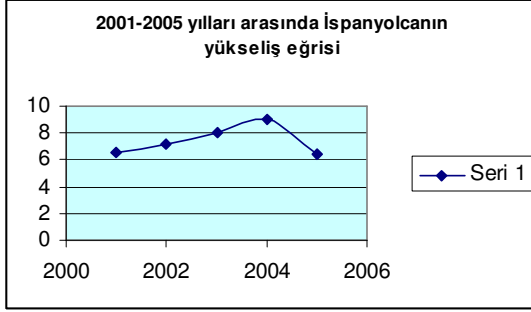
kaynak⁷

7

[http://tum.mgt.psu.ac.th/intro2IT/sectionThree/hitechcrime/webinfo/Global%20Internet%20Statistics%20\(by%20Language\).htm](http://tum.mgt.psu.ac.th/intro2IT/sectionThree/hitechcrime/webinfo/Global%20Internet%20Statistics%20(by%20Language).htm)

<http://www.cybertelecom.org/data/content.htm>

2003-2004 yıllarının karşılaştırılmasında en çok dikkati çeken İspanyolca, Fransızca ve Japoncadaki artıştır.



İngilizcenin 04 gibi küçük bir oranda gerilemesi 2001-2002 yılları arasındaki ile kıyaslanamayacak ölçüde önemsizdir. Rusçanın ise 2003 yılında sahip olduğu % 2.5 'luk payı kaybedip değerlendirmeye bile alınmayan sayısal bir değere düşmüş olması

dikkat çekicidir.

4.Web İçeriklerinin Dillere Göre Dağılımı

Web sayfalarının sayısı ve dili ile ilgili sayısal değerlerin çoğunluğu arama motorlarının indekslerine dayandırılmaktadır. Peki indekslere girmeyen web sayfaları nasıl sayılacak. Üstelik indekslere girmeyen sitelerin, indekslenen sitelere oranı ile ilgili elimizde net bir bilgi de yok.

İnternet üzerinde hangi dilde kaç web sitesi olduğunu tespit etmek, web sitelerini saymaktan ibaret olan basit bir işlem değildir. Bu çalışmaların kendi içinde çeşitli zorlukları vardır. Sanal dünyada kaç web sitesi olduğunu rakam olarak vermek çok belirleyici olmayabilir, aşağıda sözünü edeceğimiz içerik, etkinlik gibi kriterler de en az rakamlar kadar önemlidir.

Web sitelerinin durumu iki ayrı bağlamda değerlendirilebilir:

- 1.İçerik/Hacim/Byte
- 2.Etkinlik/Tıklanma Oranı

4.1.Web sitelerinin İçeriği

Şu an yapılan ölçümlerin web sitelerinin hacmini ve bilgi miktarını ölçmede çok hassas rakamlar verdiklerini söylemek oldukça güç. Sitelerin sayısını belirleyen yöntemler, bu sitelerin derinliğini tam olarak verememekte, alt linklerdeki bilgi miktarı konusunda son derece zayıf kalmaktadır. Çünkü durumuna göre içerik ve alt başlıklarda zengin ve hacimli bir site basit üç beş sayfalık bilgiden oluşan yüzlerce amatör siteye eşdeğer bir hacimde olabilir.

Ayrıca direk Com, Org, Net gibi uzantılarla yayın yapan bağımsız ciddi siteler Yahoo, Mynet, veya Üniversitelerin kendi üyelerine birkaç MB ayıran kişisel, çoğu kez güncellenmeyen, çok küçük ölçekli web sitelerinin aynı değerlendirmeye tabi tutulmaları ne denli yerinde olur? Örneğin www.yahoo.com ile www.dicle.edu.tr/cgi-bin/liste1.htm aynı kategorilerde değerlendirilmesine olanak yoktur.

4.2.Web Sitelerinin Etkinlikleri ve Tıklanma Oranları

Web sitelerinin etkinlikleri ekonomik olarak çeşitli sınıflandırmalara tabi tutulabilir. Örneğin yılda bir iki kez satış yapan bir site ile her gün milyarlarca \$ satış yapan bir site veya e-bankacılık sistemini uygulayan siteler arasında çeşitli kategoriler oluşturulabilir.

İngilizce içerikli web sitelerinin diğer dillerdeki sitelere oranla sayısının azalmakta olduğu belirtilmekte, ancak tıklanma oranı, işlem oranı gibi faktörlerde sitelerin etkinlik olarak aynı yapıda olmadıklarını göstermektedir. Örneğin günde 100.000 kez tıklanan bir site ile günde iki kez girilen bir site aynı kategorilerde sayılamaz. Sadece tıklanma oranını tespit etmek de yeterli olmayabilir, çünkü tıklanıp saatlerce online kalınan bir site ile tıklanıp birkaç saniye sonra kapatılan bir sitenin durumu da farklı bir değerlendirme gerektirir.

5.Arama Motorlarında Dil Profili

İnternet üzerinde gelişmiş ve basit düzeyde onlarca arama motoru mevcut. Şu anda en çok kullanılanları Google, Yahoo, MSN, Altavista⁸ olarak dikkati çekmektedir. Ancak toplam web sayfaları ile ilgili verdikleri rakamlar çok net ve sağlıklı görünmemektedir. Örneğin Google 8 milyarın üzerinde web sayfası arşivine sahip olduğunu iddia ederken Yahoo bu

⁸ Arama motorlarının bu yıl içinde 5 milyar dolar civarında önemli bir pazar payına sahip olacağı iddia edilmektedir. Sıradan kullanıcılara ücretsiz hizmet veren arama motorlarının çok saygın kurumlarla da iş yaptıkları bilinmektedir.Örneğin Google'un en saygın müşterilerinden birinin de Amerikan Merkezi Haber Alma Teşkilatı CIA olduğu iddialar arasındadır.
<http://www.lexpansion.com/compteur/compteur.asp?compteurid=689&redirUrl=http://www.lexpansion.com/art/2706.81652.0.html>
<http://abcnews.go.com/WNT/PersonOfWeek/story?id=131833&page=1>

rakamı 19,2⁹ milyar olarak açıklamaktadır. Ayrıca bazen tek bir arama motorunun verdiği rakamlarda çelişkili olabilmektedir. Google toplam 8 milyarın üzerinde web sayfası indekslediğini iddia etmekte ancak sadece İngilizce sayfaları tespit etmek için verilen herhangi bir anahtar kelime karşısında 8.5 milyar web sitesi görüntülenebilmektedir.¹⁰ Toplam web sayfasında dikkati çeken bu farklılıklar web sitelerinin dillerini belirlemede de çeşitli zorluklar oluşturmaktadır. Fransa Dil teknolojileri merkezi müdürü Jean Véronis aralarında Google, Yahoo ve MSN’inde bulunduğu bir çok arama motorunun gerçek dışı rakamlar vererek internet kullanıcılarını yanılttığını iddia etmektedir.¹¹ Google ise manipülasyon iddialarını reddederek kendi arama motorlarında listeleme işlemlerinin hiçbir manüel müdahale olmadan robot yazılımlar tarafından yapıldığını belirtmektedir.

İnternette alan adı tescillerini yenilemeyen veya başka sebeplerle kaybolan web sitelerini de kapsayan önemli bir arşivleme işlemini Wayback yapmaktadır. Wayback Machine 1996 yılından beri toplam 40 milyarın üzerinde (yayımdan kalkanlarda dahil) web sitesi arşivlediğini iddia etmektedir.¹²

15/16 Aralık .2005 tarihlerinde Google ve Yahoo arama motorlarının dillere göre verdikleri web sayfası sayıları:

	Google	Yahoo
İngilizce	8.560 000 000	9.980.000 000
Almanca	180.000 000	776.000 000
Arapça	12.300 000	24,900,000
Fransızca	216 000 000	557 000 000
İspanyolca	132.000 000	370 000 000
Türkçe	10.500 000	48 000 000

Görüldüğü üzere Google arama motorundaki İngilizce web sitesi sayısı tabloda görülen diğer dillerde mevcut web sitesi sayısından kat kat fazladır.

8.560 000 000 İngilizce web sitesine karşılık Google’un arşivinde diğer beş dilde mevcut olan web sitesi sayısı 550800000 civarındadır. Bu dillerin dışında kalan 5000’in

⁹ <http://aixtal.blogspot.com/2005/08/yahoo-19-milliards-de-pages.html>

¹⁰ Örneğin Google’da sadece İngilizce sayfalarda “the” tanımlığını görüntülendiğinde 8.560.000 000 sayfa çıkmaktadır. Rakam diğer dillerdeki sayfa ile toplandığında en az 11-15 milyar arasında rakamlar ortaya çıkmaktadır.

¹¹ Jean Véronis, <http://aixtal.blogspot.com/2005/08/yahoo-19-milliards-de-pages.html>

¹² <http://www.archive.org>

üzerindeki dillerde göz önüne getirildiğinde İngilizcenin Web ortamında çok ezici bir üstünlüğü olduğu açıktır.

Bir görüşe göre arama motorlarında taranmayan sitelerin oranı % 40¹³ civarındadır. Hatta Michael K. Bergman şu anda arama motorlarının taramış olduğu web sitelerinin buz dağının görünen kısmı olduğunu, derinlerde çok daha fazla sayıda web sitesi ve doküman olduğunu iddia etmektedir.¹⁴ Bütün bu görüşler doğrusa Google'un iddia ettiği 8.milyar web sitesi toplam web sitelerinin ancak % 60'ını oluşturmaktadır. Yahoo'nun verdiği 20 milyara yakın rakamı da en az 33 milyar olarak düşünmek gerekir.

Yahoo, Altavista ve Google'un dil seçenekleri bölümünde 30-50 arası dil bulunmaktadır. Google'un yaklaşık 100 dilde arama yapabildiği iddia edilmektedir.¹⁵

Arama motorlarının web sitelerinin dilleri hakkında verdiği rakamlar da değişkenlik arz etmektedir. Örneğin Google Fransızca sayfaları 216.000 000 civarlarında verirken, Yahoo ise 557.000 000 civarında vermektedir. Diğer diller ve arama motorları içinde aynı değişkenlikler geçerlidir. Google, Yahoo, Altavista arama motorları aynı dil için farklı farklı rakamlar verebilmektedir. Aynı arama motorundaki rakamlarda neredeyse her gün değişmektedir. Dolayısıyla her durumda rakamların tespit edildiği tarihleri belirtmek gerekir.

Genellikle Google'un rakamları Altavista ve Yahoo'nun çok gerisindedir. Bunun bir nedeni de arama motorlarının kuruluş tarihlerinden kaynaklanabilir, çünkü Yahoo 1994 yılından beri arşivleme işlemini yaparken, Larry Page ve Sergey Brin Google'u 1998 yılında kurmuştur. Altavista ve Yahoo'nun verdikleri rakamlar birbirine oldukça yakındır. Altavista'nın kuruluş tarihi ise 1995 'tir.

Arama motorlarının verdikleri rakamlardaki tutarsızlıkların çeşitli nedenleri olabilir, en önemli nedenlerinden birisi web'in yapısından kaynaklanmaktadır. Örneğin iki dilli bir site hem Fransızca hem de İngilizce sayfaları tararken karşımıza çıkabilir. Veya bir arama motoru sadece ana sayfaları tararken bir diğeri alt kategorilerdeki sayfaları da tarayabilir, dolayısıyla rakamların farklı çıkması neredeyse kaçınılmaz olmaktadır. Web sitesi sahipleri bazı arama motorlarına baş vururken diğerlerine başvurmamakta veya hiçbir arama motoruna kayıt yaptırmamaktadır. Arama motorlarından bazıları yüzlerce sayfalık Fransızca bilgi

¹³ <http://graal.ens-lyon.fr/~avernois/nte-cours3.html>

¹⁴ Bergman web'in yüzey kısmında yer alan 1 milyar dokümana karşılık derin web'de 555 milyar kişisel doküman olduğunu ileri sürmektedir. <http://www.press.umich.edu/jep/07-01/bergman.html>

¹⁵ <http://abcnews.go.com/WNT/PersonOfWeek/story?id=131833&page=1>

bulunan Frankofon bir siteyi, içindeki üç beş cümlelik Türkçe bir alıntıdan dolayı Türkçe siteler arasında görüntüleyebilmektedir. (Örn: www.espritcritique.org)

Rakamlardaki tutarsızlıkların bir diğer nedeni de istatistikleri hazırlayan kurumların yanlış davranmasıdır. Örneğin Almanların yaptığı bir çalışmada Almanca İngilizceden sonra ikinci dil olarak gözükmektedir. Başka çalışmalarda ikinci dil olarak Fransızca veya Çince çıkabilmektedir.

6.Değerlendirme/Yorum

İngilizce sayfaların sayısı Google'da 8.560.000 000 çıkarken yahoo'da 9.980 000 000, Altavista'da ise 10.300.000 000 çıkmaktadır. Arama motorlarına girmeyen, indekslenmeyen web sitelerini de sayarsak İngilizce web sitelerinin sayısının 11.000 000 000 ile 12.000 000 000 üzerinde olduğunu tahmin etmek yanlış olmaz. İstatistikler bu web sitelerinin büyük çoğunluğunun Kuzey Amerika (ABD, Kanada) olduğunu göstermektedir.

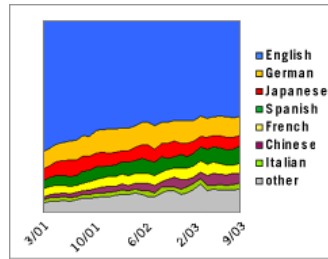
Çince ise eski ve basit Çince diye iki ayrı bölümde ele alındığında basitleştirilmiş Çincenin eski Çinceden daha çok internet ortamında yer aldığı anlaşılmaktadır. Basitleştirilmiş Çince 30.000 000 üzerinde gözükmektedir, eski Çince ise 4.000 000 üzerindedir. Dolayısıyla toplam Çince web sitesi sayısı 40.000 000 civarında tahmin edilmektedir. Çincenin internetteki yükseliş grafiği yüksek nüfusu ile ilgili olabileceği gibi, Batı ile soğuk savaş dönemi sonrasında girdiği sıcak ilişkilerin bir ürünü olarak da yorumlanabilir. Ancak Çin hükümetinin son zamanlarda interneti kısıtlayıcı hatta sansürleyici bazı yollara başvurması Çincenin gelecekte internetteki payını doğrudan etkileyebilir. Örneğin Çin hükümetinin Google ile arama motorundaki bazı kelimeleri sansürleme konusunda anlaşmış¹⁶ olması bu anlamda Çinlilerin internete güvenini azaltan dolayısıyla da Çince interneti zayıflatan bir unsur gibi gözükmektedir.

Fransızca ise dünyada yaklaşık 400 milyon civarında insan tarafından konuşulmasına rağmen internet dünyasındaki durumu pek iç açıcı değil. Toplam 220.000 000 civarında Fransızca web sayfası olduğu tahmin edilmektedir. Bunların önemli bir kısmı Fransa ve Kanada'daki Frankofon bölgeden yayınlanmaktadır. Fransa'daki internet kullanımının Frankofon Kanada'nın çok gerisinde olması Fransa da ki telefon ücretlerinin çok pahalı olması ile yakından ilgilidir. Ayrıca Belçika, Luxembourg, Cezayir, Tunus vb ülkelerde yaşayan kesimlerde Fransızcanın sanal dünyadaki varlığına katkıda bulunmaktadır.

¹⁶ Haluk Şahin, Ütopyadan cehenneme Google, Radikal Gazetesi, 27 Ocak 2006

Türkçe'nin Google arama motoruna kayıtlı web sitesi sayısı 10 500 000 civarındadır. Ancak bu rakama diğer Türk lehçeleri dahil değildir. Onlarla birlikte düşünüldüğünde Türkçenin aslında arama motorlarında görüntülediğinden daha büyük bir paya sahip olduğu düşünülebilir.

Arapça Ortadoğu ve dünyada önemli bir dil olmasına rağmen Araplar internete uyum konusunda çok yetenekli gözükmemektedir. 1 milyar civarında Arabın yaşadığı dünyada Arapçanın internet üzerindeki varlığı özellikle farklı bir alfabe kullanmasından ötürü oldukça sınırlıdır. Arapça olarak toplam 8.000.000 civarında web sitesinin olduğu tahmin edilmektedir. Arap alfabesi internette her zaman görüntüleme sorunu oluşturduğu için bu rakam muhtemelen daha da yüksek olabilir. Ancak arama motorlarının Farsça vb. bazı dilleri de (Arap Alfabesi ile yazıldığından) Arapça siteler olarak saymaktadır. Bu da Arapça siteleri olduğundan fazla gösteren önemli bir faktördür.



Google'un dillere göre kullanımı Eylül 2003¹⁷

Almanca Avusturya, Almanya ve İsviçre'nin bir bölümünde konuşulan bir dildir. Google araştırmanın yapıldığı tarih itibari ile toplam 180 000 000 Almanca web sitesi görüntülenmektedir.

Japonca sadece Japonya da konuşulan bir dil olmasına rağmen internette oldukça önemli bir paya sahiptir. Google 123 .000 000 civarında Japon web sitesi listelerken, yahoo ise 216 Milyon japon kaynaklı web sitesi indekslemektedir. Japonca dil karakterleri farklı olduğundan bu rakamın daha yüksek olduğu düşünülebilir.

İspanyolca, İspanya, Güney Amerika ülkeleri ve bazı ada devletlerinde konuşulan İspanyolca internette grafiği yükselen diller arasındadır. Google 132 000 000 İspanyolca web sitesi verirken, yahoo 370 000 000 İspanyolca web sitesi görüntülenmektedir.

¹⁷ http://www.google.com/press/zeitgeist/sep03_lang.gif
http://www.webservices.hu/languages_en/?s=default

SONUÇ

Verilerin çok hızlı değiştiği bu kaygan ortam da elde edilen rakamlar anlamsız gibi gözükse de başka referans kaynağı olmadığından politik ve ekonomik karar vericilerin stratejileri için son derece önemlidir.

Dünya nüfusunda önemli bir yere sahip olan İspanyolca, Çince, Arapça ve Hintçenin internet üzerinde hak ettikleri yeri almalarından sonra Almanca, Fransızca, Japonca gibi dillerin nüfusları kadar ancak internette yer alacaklarını bunun da şimdiki durumlarının çok gerisinde bir rakam olacağını tahmin etmek yanlış olmaz. Örneğin şu anda % 4 civarında olan Fransızca % 1 seviyesine inebilir.

Öyle görünüyor ki İngilizce her durumda sanal dünyada hegemonyasını devam ettirecek ve böylelikle gerçek dünyadaki konumunu daha da güçlendirecek, zayıf ve yok olmakta olan diller ise daha da çok gerileyecekler. Unesco bu dillerin internette var olmasını resmen desteklemesine rağmen bu dillerde kullanıcı veya webmaster bulmak her zaman o kadar kolay olmamaktadır.

Bazı dillerdeki karakterlerin e-posta adreslerinde kullanılmaması da İngilizce karşısında bu dillerin durumunu olumsuz yönde etkilemektedir. Örneğin Türkçe de ç ş ğ ü harfleri veya Fransızca da é, à ê gibi karakterler. Yapılan teknik çalışmalarla bu sorunun yakında çözüleceği düşünülmektedir.

Gelecekte sanal toplulukların dil özelliklerini tam olarak belirlemek devletlerin, kurumların, şirketlerin siyasi ve kültürel politikalarını yeniden yapılandırmalarında önemli veriler sağlayacaktır.

Günümüz dünyasında internet iş dünyası için büyük bir Pazar anlamına gelmektedir. (2004 yılında 6.8 Trilyon \$¹⁸) Ne var ki şu anda yeryüzünde yaşayan 6 milyar insanın ancak küçük bir kısmı, % 10 kadarı sanal dünyaya taşınmış durumda. Belki bir süre sonra internet ortamında online kişi sayısı bir kaç milyar daha artacak ve ortaya hiçbir ülke ve coğrafi bölgede bulunmayan büyük bir pazar çıkmış olacak. İnternet reklamcılığı da önümüzdeki yıllarda reklam piyasasındaki payını artıracak gibi gözüküyor. 2008 yılına kadar internette yayınlanan video reklamlarının toplam maliyetinin 1 milyar doları bulması bekleniyor.

¹⁸ <http://glreach.com/eng/ed/art/2004.ecommerce.php3>

İnternet üzerinden satış yapan uluslararası firmalar en çok aşağıdaki bilgilere ihtiyaç duymaktadırlar:

- 1-En çok e-satışın yapıldığı ülkeler
- 2-En çok tıklanan web sitelerinin dili
- 3-En çok e-satış yapan siteler
- 4-En çok e-satış yapan web sitelerinin dili
- 5-E-bankacılık dili
- 6-En çok tıklanan internet reklamlarının dili

Görüldüğü gibi dil bu büyük pazarda tüketicilere ulaşmanın önünde en büyük engellerden biri olmaya devam etmektedir. Bunun için iki temel çözüm gözükmemektedir:

1-Hızlı, tam, eksiksiz, pratik makine programları geliştirilinceye kadar sanal ortamdaki hedef kitleye uygun dillerle yayın yapmak Çünkü dünya da konuşulan 6000 dilde yayın yapmak bir web sitesi için çok fazla mümkün gözükmemektedir. Dolayısıyla e-satış yapan firmaların en çok kullanılan dilleri belirleyip onlara uygun, çok dilli web sayfaları yapmaları daha pratik gözükmemektedir.¹⁹ Zaten şu anda da bazı şirketler bunu yapmaktadırlar. Amerikan kökenli Amazon.com toplam 6 ayrı dilde satışlarını yapmaktadır, bu diller: İngilizce, Fransızca, Almanca, İspanyolca, Çince, Japonca.

2-Robot çeviri programları için veri tabanı oluşturmak, bütün dilleri kapsayacak bir otomatik program şimdilik zor görünüyor. 6000 dilden her biri için geri kalan 5999 dile çeviri yapabilecek ve bunu her dil için tek tek tekrarlayabilecek bir konfigürasyona ulaşmak çok zaman alacak gibi. Fransız Amerikan ortak şirketi olan SYSTRAN robot çeviri konusunda en önde gelen şirketlerden birisidir. Babel Fish Translate adlı yazılım İngilizce, Almanca, İspanyolca, Fransızca, İtalyanca ve Portekizce dillerinde çeviri yapabilmektedir. Son çıkan

SYSTRAN Web Translator 5.0 elerde bilgisayarlara kurulabilen, 29 \$ satışı yapılan İngilizce, İspanyolca, Fransızca, Almanca, İtalyanca, Portekizce, İsveççe, Dutç, Rusça, Japonca, Korece, Çince dillerinde web sayfalarını çevirebilen bir yazılım.

Bu tür programlarında farklı handikapları olduğunu unutmamak gerekir, adından da anlaşılacağı üzere makine çevirisi olduğu için istenilen seviyede kaliteli bir çeviri her zaman mümkün olamamakta ancak yazılımların gittikçe geliştiğini de unutmamak gerekir.

¹⁹ Online video ads, <http://www.redherring.com/article.aspx?a=14660>

İnternetteki dil profilleri ile ilgili veriler politikacıların da ilgisini çekmektedir. Örneğin; zaman içerisinde bütün insanlara internet yoluyla İngilizceyi öğretmek, hatta dil asimilasyonu uygulamak mümkün olabilir mi?

Bütün dilleri internete taşımak sonuç olarak dilleri konuşan toplumların sanal evrende tanınması anlamına mı gelecek, yoksa bu toplumlara da ağ üzerindeki büyük örümceğin çekim alanına sokacak bir araç mı olacak? Bütün bunları kestirmek şimdiden oldukça güç, ancak internetteki dil profili gelecekte politikacıların da en çok ihtiyaç duyabilecekleri veriler arasında önemini koruyacaktır. Eğer internet Televizyon gibi bir propaganda aracı olarak kullanılacaksa bundan herkes gibi politikanın iç ve dış aktörleri de yararlanmak isteyeceklerdir.

KAYNAKÇA

- FINKIELKRAUT, Alain Internet, l'Inquiétante extase, Mille et une Nuits, 2001
RONI AMELAN, From Information Society to knowledge society, The New Courier, October 2003, UNESCO p.32
ŞAHİN Haluk, Ütopyadan cehenneme Google, Radikal Gazetesi , 27 Ocak 2006
SUMY MOUHOUBI, Languages used on the web, The New Courier, November 2005, Unesco,p.62
SUNN Frank, Canavarın İnternetteki Sayısı: 666, Cep Kitapları, İstanbul , 2001
TARCAN Ahmet , İnternet ve Toplum, Anı Yayınları, Ankara, 2005

İnternet Kaynakları

- <http://www.internetworldstats.com/stats7.htm>
<http://www.language-observatory.org>
<http://www.greach.com>
<http://global-reach.biz/>
<http://global-reach.biz/globstats/evol.html>
<http://www.cymru.com/Bogons/>
www.internetworldstats.com
<http://public.web.cern.ch/Public/Welcome.html>
http://www.webservices.hu/languages_en/?s=default
<http://www.redherring.com/article.aspx?a=14660>
<http://unesdoc.unesco.org/images/0014/001421/142186f.pdf>
<http://www.google.com/intl/fr/options/>
<http://aixtal.blogspot.com/2005/01/web-google-le-compteur-bloqu.html>
<http://www.netz-tipp.de/sprachen.html>
<http://ausweb.scu.edu.au/aw03/papers/edwards2/paper.html>
<http://www.press.umich.edu/jep/07-01/bergman.html>
<http://www.espritcritique.org/0310/article5.html> (Jean Paul Marcotte)
<http://www.e-sosder.com/dergidetay.php?id=42>

<http://www.archive.org/>
<http://abcnews.go.com/WNT/PersonOfWeek/story?id=131833&page=1>
<http://yhoo.client.shareholder.com/press/overview.cfm>
<http://www.altavista.com/about/>
<http://www.lexpansion.com/compteur/compteur.asp?compteurid=689&redirUrl=http://www.lexpansion.com/art/2706.81652.0.html>

Kısaltmalar

WWW(web) :İnternet üzerinde dosya ve dokümanların aktarımına olanak sağlayan uygulama. Web internette en sık kullanılan uygulamalardan biridir.

HTML: (Hypertext Markup Language) Web sitelerini internet ortamına formatlayan dil.

LPIM: (Language Property Identification Module) :Dil Özelliklerini belirleme Modülü

CERN: (Centre Europeen de Recherche Nucleaire) : Avrupa Nükleer Araştırma Merkezi

IP: İnternet'te bulunan her bilgisayarın kendine ait bir adresi vardır. Bu adres İnternet üzerinde bilgisayarların birbirlerini tanımalarını ve iletişim kurmalarını sağlar ve numaralarla ifade edilir.IP adresleri noktayla ayrılmış 4 değişik rakam kümesinden oluşur ve 0-255 arasındaki rakamlardan oluşur. Örnek: 209.67.50.203. Alan adları ise bu rakamlara isimler vererek daha kolay hatırlanmasını sağlar. Her Alan Adı, yukarıda anlatılan tarzda bir rakamsal adrese bağlıdır. www.benimadresim.com yazdığınız zaman bilgisayar bu ismi rakamlara çevirir, İnternet üzerinde hangi bilgisayarda durduğunu kontrol eder ve trafiği bu adrese yönlendirir. <http://yardim.mynet.com/Help.aspx?id=393&page=det&SORU=578>