

# Asya Ülke Kodları ve tr Uzantılı Web Sitelerinin Dil Özelliklerinin Analizi

Ahmet TARCAN<sup>1</sup>, Yoshiki MIKAMI<sup>2</sup>, Turrance NANDASARA<sup>3</sup>

<sup>1</sup> Dicle Üniversitesi, Eğitim Fakültesi, (Turkey)

<sup>2</sup> Nagaoka University of Technology (Japan)

<sup>3</sup> University of Colombo School of Computing (Srilanka)

tarcan@dicle.edu.tr, mikami@kjs.nagaokaut.ac.jp, nandasara@yahoo.com

**Özet:** Bu çalışmada Asya kıtasındaki yayın yapan sitelerinin dil analizleri yapılmıştır. Bu bağlamda .tr uzantılı web siteleri diğer ülke kodları ile birlikte çeşitli değerlendirmelere tabi tutulmuştur. Uzakdoğu, Ortadoğu ve Orta Asya'dan bir çok ülke kodunun dil analiz sonuçları çeşitli karşılaştırmalara tabi tutularak etkin diller tespit edilmeye çalışılmıştır. Haziran 2006 itibarıyla 6.010.337 olan .tr uzantılı web sitesi sayısının 2007 itibarıyla yükseldiği gözlenirken, UTF ve Latin kodlu Türkçe web sayfalarındaki bazı artış ve düşüşler dikkati çekmiştir. Bu artış ve düşüşlerin nedenleri üzerinde çeşitli analizler yapılmıştır. Çalışmada ALRN (Asian Language Resource Network ) projesi çerçevesinde oluşturulan dil tarama robotunun verdiği rakamlar esas alınmıştır. Araştırmanın veri tabanı Nagaoka Teknoloji Üniversitesindeki Dil Gözlem laboratuvarının 2006 ve 2007 yılları itibarıyla verdiği rakamlar karşılaştırılarak bir sonuca ulaşılmaya çalışılmıştır. Web ortamının değişken yapısı dikkate alındığında verilen rakamların yaklaşık sonuçlar olduğu açıktır.

**Anahtar Sözcükler:** Internet, Dil İçerikleri, Web Sayfası, tr.

## Asian Countries Domain Names And Linguistic Analyse Of Websites By tr Domain Name

**Abstract:** As it is known the country domain registration for Turkey is tr. This paper studies the linguistic profile of website by “.tr” domain name and analyse linguistic profile of Asian web. The webpages by tr domain names has been evaluated for their linguistic content between 2006 and 2007. The main goal of this study is to observe the linguistic proportions of webpages in Asia. Statistics acquired from the data of Language Observatory about languages percentages was compared according to years. The data of this study come from The Asian Language Resource Network (ALRN) Project from Nagaoka University of Technology. It should not be forgotten the difficulty to obtain the adequate numbers in the web. Thus the numbers given in this research are approximate results.

**Keywords:** Internet, Linguistic Content, Webpages

### 1. Giriş

Asian Language Resource Network (ALRN) projesi Asyada 43 civarında ülke kodunu gözlemleyerek hangi ülke kodundan hangi dillerde

yayın yapıldığını tespit etmeye çalışmaktadır. Araştırma com org net gibi genel alan adlarını ve birkaç ülke kodunu teknik sebeplerden ötürü kapsamamaktadır. Bu proje kapsamında

aralarında .tr'ninde bulunduğu 43<sup>1</sup> ülke koduyla yayın yapan websitelerini dil içerikleri açısından analize tabi tutan araştırma bu ülke kodlarına ait 55 milyon civarında web sitesi tespit etmiştir. Bu sitelerin kendi aralarındaki oranları, yayın yaptıkları ağırlıklı dilleri ve ana dillerine oranlarının karşılaştırılması ilginç sonuçlar vermektedir.

## 2.Asya Ülkeleri Alan Adları Dil Analizi

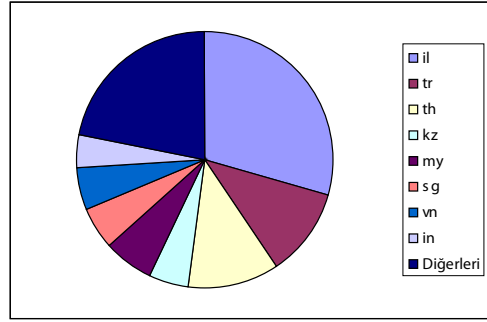
Araştırma kapsamındaki sitelerin yaklaşık % 80'ine yakını İsrail, Türkiye, Tayland, Kazakistan, Malezya Singapur, Vietnam gibi ülke kodlarından gelmektedir. Geriye kalan yaklaşık 36 ülke kodu 55 milyon web sitesinin % 20'sini oluşturmaktadır.

Ülke kodu	Toplam web sitesi sayısı	%
il	15 660 378	28
Tr	6 010 337	10.8
th	5 962 043	10.7
My	3 650 211	3
kz	2 938 978	2
Sg	2 900 064	2
Vn	2 848 682	2
In	2 466 562	2
GENEL TOPLAM	55 328 231	

**Tablo 1**

Görüldüğü üzere nüfuslarıyla karşılaştırıldığında kişi başına düşen websayfası sayısı İsrail (il) ülke kodunda diğer ülkelere oranla daha yüksek bir rakam oran göstermektedir.

1 Araştırma kapsamındaki ülke kodları: ae af az bd bh bn bt cy id il in iq ir jo kg kh kw kz la lb lk mm mv my np om pq ph pk ps qa sa sq sy th tj tm tp tr uz vn ye



**Şekil 1**

## 3.Asya Kıtasında İngilizce

Her geçen gün etkinliği daha da artan İngilizce kendini web ortamında daha da baskın bir şekilde hissettirmektedir. Asya ülke kodlarıyla oluşturulan neredeyes her yeni sitenin İngilizce versiyonu da tasarlanmaktadır. Asya ülkeleri alan adlarından toplam 18.247.410 websitesi İngilizce yayın yapmakta, Bu rakam Asya kıtasındaki toplam websitelerinin % 32 oluşturmaktadır.Yani Asyada her 100 siteden 32'si İngilizce yayın yapmaktadır.

İngilizce	18.247.410	% 32
İbranice	8 455 690	% 15
Thai	7 202 583	% 13
Rusça	3 697 809	% 6.6
Türkçe	3 657 589	% 6.6
Vietnamca	1 758 796	% 3
Arapça	964 575	% 1.7

**Tablo 2**

Toplam 346 dilin analize tabi tutulduğu bu araştırmada Asya ülke kodlarından yayın yapan sitelerde en büyük pay kolaylıkla tahmin edileceği üzere İngilizceye düşmektedir.

İngiltere'nin alan adı olan uk. ve usa. gibi siteler araştırmanın kapsamı dışında olduklarına göre ingilizce siteler hangi ülke kodlarını kullanmaktadır? Hangi ülkeler websitelerinde İngilizceye ağırlık vermektedirler?

Hemen hemen bütün ülke kodlarının istisnasız az ya da çok İngilizce içerik kullandıkları görülmektedir. Ancak İngilizce diline ait oranlar arasında çok yüksek farklar görülmektedir. Örneğin Singapur ülke kodundan yayın yapan sitelerinin İngilizce oranı % 82 civarında iken, sy uzantılı Suriye sitelerinde ise İngilizcenin oranı % 9 olarak tespit edilmiştir.

Asya kıtasında en çok İngilizce yayın yaptığı tespit edilen sitelerin şu uzantıları kullandıkları tespit edilmiştir:

Ülke kodu	İngilizce Websayfası sayısı	Toplam websayfası sayısı	İngilizce Oranı %
Sg	2390771	2900064	82
In	1836194	2466562	74
Mv	7630	10817	70
My	2435222	3650211	66
Ph	1096428	1399988	78
Th	1516381	5962043	25
Cy	187.325	380.333	49
Il	3484379	15660 378	20
Tr	1079361	6010337	17

**Tablo 3**

Toplam web içeriğinin % 49'u İngilizce olan Kıbrısın ülke kodu olan cy uzantısı ile ilgili diğer dil analizleri aşağıda görülmektedir:

İngilizce	187 325	% 49
Yunanca	101 663	% 26
Thai	57 337	% 15
Almanca	3096	% 0.8
Türkçe	275	% 0.072
Toplam	380.333	

**Tablo 4**

Asyada en az İngilizce içerik taşıyan ülke kodları ise Yemen (ye), Suriye (sy), Kuveyt (kw) olarak tespit edilmiştir.

Ülke kodu	İngilizce websayfası sayısı	Toplam websayfası	İngilizce Oranı
Ye	2723	28 988	% 9
Sy	1510	15 549	% 9
Kw	5266	44 265	% 11

**Tablo 5**

#### 4.Sonuç

Yukarıdaki rakamlardan İngilizce olarak yayın yapan alan adları toplam web sitesine olan oranıyla daha anlamlı olacaktır. Örneğin Moldovyanın az sayıda İngilizce siteye sahip olması tek başına anlamlı değildir. Toplam mv uzantılı sitelerin % 70 'i İngilizce içerikli websitelerinden oluşmaktadır. Öte yandan Türkiye (tr) uzantılı toplam sitelerde 1 milyonun üzerinde İngilizce websitesi olmasına rağmen toplamda İngilizce oranı Moldovya kökenli sitelerinin gerisinde kalmıştır.

#### Kaynakça

GARY Simons and al., Using Computers in / Linguistics: A Practical Guide, published by Routledge Newyork, 1998

PAOLILLO John et Autres, Mesurer la diversité linguistique sur Internet. Publications de l'UNESCO pour le Sommet mondial sur la société de l'information, Paris, 2005

RONI AMELAN, From Information Society to knowledge society, The New Courier, October 2003, UNESCO p.32

SUMY MOUHOUBI, Languages used on the web, The New Courier, November 2005, Unesco,p.62, 42

Language Observatory Laboratory, Nagaoka University of Technology, ALRN Project, Nagaoka Japan

### **Teşekkür**

Çalışmanın verilerini bizlere sağlayan Nagaoka Teknoloji Üniversitesi Dil Gözlem Laboratuvarına ve Japon Eğitim, Kültür, Spor, Bilim bakanlığına teşekkürlerimizi sunuyoruz..

### **Acknowledgement**

We would especially like to acknowledge and express our gratitude to Language Observatory Laboratory in Nagaoka University of Technology. The data come from Language Observatory Project sponsored by Ministry of Education, Culture, Sports, Science (MEXT) and Technology ,to which I am deeply grateful (MEXT).