



ARAMA MOTORU BING'İN TEKİL-ÇOĞUL KELİMELER İÇİN GÖVDELEME MEKANİZMASI

STEMMING MECHANISM OF SEARCH ENGINE BING FOR SINGLE-PLURAL WORDS

Fatmana ŞENTÜRK^{1*}, Gürhan GÜNDÜZ¹

¹Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Pamukkale Üniversitesi, 20070, Denizli.
fatmanas@pau.edu.tr, ggunduz@pau.edu.tr

Geliş Tarihi/Received: 06.03.2013, Kabul Tarihi/Accepted: 15.05.2013
*Yazışılan yazar/Corresponding author

doi: 10.5505/pajes.2014.05025

Özet

Günümüzde, İnternet dünyasının gelişimi ile bilgiye ulaşmak hızlı ve basit bir hal almıştır. Bu devasa İnternet dünyasında istenilen bilgiyi bulabilmek için arama motorları geliştirilmiştir. Arama motorlarının nasıl çalıştığını bilmek aramaların daha etkili olmasını sağlayacaktır. Bu çalışmanın sağlayacağı sonuçlar ile kullanıcılar bilgilere daha hızlı ulaşabilirler. Bunun yanında araştırmacılar, arama motorları tarafından sağlanan sonuçları daha iyi yorumlayabilirler. Bu çalışmada, arama motoru Bing'in İngilizcedeki tekil-çoğul kelimeler için dokümanların indekseleme algoritması tahminlenmeye çalışılmıştır. Bunun için öncelikli olarak incelenecek kelimeler seçilmiştir. Bu kelimeler için sorgular oluşturulup, Bing'e gönderilmiştir. Bing'den dönen sonuçlar analiz edilmiştir. Bu sonuçlara göre, dokümanlar indekslenirken tekil-çoğul kelimeler için güçlü ilişkiler kurulduğu ve kelimenin kullanım sıklığının bu ilişkide etkili olduğu görülmüştür.

Anahtar kelimeler: Arama motoru, Bing, İndeksleme, Gövdeleme.

Abstract

Today, it has become really fast and easy to access information thanks to the developments in the internet world. Search engines have been developed to find desired information in this gigantic internet world. It is important to know how these search engines works so that searches are more efficient. This work provides results that could help end-users to access information faster. It could also help researchers to interpret results, provided by search engines, better. In this work we have tried to find the indexing algorithm of Bing for single-plural words. For this purpose, we first selected the suitable words and then send the constructed queries to the Bing. Afterwards, the returned results of the queries had been analyzed. We found out that there is a strong relation between single and plural form of the words in term of indexing by Bing. In addition, the frequencies of usage of the words also effect this relation.

Keywords: Search engine, Bing, Indexing, Stemming.

1 Giriş

Bilgi boyutlarının giderek artmasıyla, arama motorları ortaya çıkmıştır. Arama motorlarının kullanımı ve aranılan bilginin çeşitliliği bireylere göre değişmiştir. Arama motorları da, kullanıcıların istenilen bilgiye daha hızlı ulaşması için, verileri gruplandırmaya başlamışlardır. Günümüzde ise, arama motorları kullanıcıların aradıkları bilgiye hızlı ve kaliteli çözümler üreten ara yüzler haline gelmiştir. Ancak bilginin çokluğu ve kişilerin arayacakları bilgi hakkında tutarlı veriler kullanmaması bazı durumlarda kullanıcıları yanlış veya eksik sonuçlara yönlendirmektedir. Kullanıcıların zaman kaybetmeden kaliteli bilgiye en hızlı sürede erişmesi önemlidir.

Bugüne kadar, arama motorlarının kapsamı ve içeriği hakkında birçok çalışma yapılmıştır. Bu çalışmada, literatürdeki çalışmalar incelenmiş olup iki gruba ayrılmıştır.

İlk grup, kelime gövdeleme algoritmalarıdır. Kelime gövdeleme, bir kelimeye eklenmiş olan çekim eklerinin çıkartılması ile kelime kökünün bulunması işlemine verilen isimdir. Gövdeleme işlemi dillere göre farklılıklar göstermektedir. Örneğin, İngilizce gibi bir dil için yalnızca ekler sözlüğüne bakılarak bir gövdeleme sistemi gerçekleştirmek mümkündür. Ancak bu şekilde gerçekleştirilen bir gövdeleme algoritmasının Türkçe gibi sondan eklemeli bir dil için gerçekleştirmesi zordur. Bu yüzden bu çalışmada sadece İngilizce dili temel alınmıştır.

Kelime gövdeleme işlemi ile ilgili farklı dil yapılarına göre birçok algoritma geliştirilmiştir. İngilizce için bu algoritmaların ilki 1968'de Lovins tarafından yayınlanan gövdeleme algoritmasıdır [1]. Günümüzde ise, en çok kullanılan Porter gövdeleme algoritmasıdır [2].

İkinci grup ise arama motorlarının yapısıyla ilgili çalışmalardır. Arama motorları insanların arama alışkanlıkları hakkında bilgi verebilmektedir. Bu yüzden, arama motorları birçok çalışmada veri toplamak için kullanılmıştır [3, 4]. Ayrıca kişilerin arama alışkanlıkları farklı bölgelere göre analiz edilmiştir. Örneğin, Tawileh, 2010 yılı içerisinde Google arama verileri üzerinden Arap ülkelerinin arama alışkanlıklarını incelemiştir [5].

Arama yaparken birçok kişi tek bir kelimeye ya da sıralı bir kelime grubuna bağlı kalmakta olup, aranan bilgi aynı olmasına rağmen farklı arama sorguları oluşturabilmektedir. Bu sebeple; Srivasta, Kollipara, Tyagi ve Jaiswal daha önce kullanıcıların oluşturduğu özel sorguları öneren bir uygulama geliştirmiştir [6]. Ayrıca belirli bir kelime grubu için düzenli sırada arama yapılması ile farklı kombinasyonlar ile arama yapılması arasında dönen sonuçlar açısından farklılık görülmektedir. Bu farkı ortadan kaldırmak için, Raval ve Kumar arama yaparken ortaya çıkabilecek tüm kombinasyonları ele alan bir arayüz geliştirmiştir [7].

Birçok çalışmada, arama motorlarının güncelliği ve güvenilirliği ele alınmıştır. Kim ve Carvalho, arama motorlarının ilk sıralarda gelen sonuçlarının güncellenme sıklığını ele almıştır [8]. Kulasegarah, Harney ve Walsh

tarafından yapılan çalışmada, belirli bir hastalığa yönelik kelimeler aranmış ve dönen sonuçların kalitesi incelenmiştir [9].

Arama motorları, sorgu gönderildiğinde arama yapma süresi açısından da değerlendirilmeye alınmıştır. Edosomwan yaptığı çalışmada, büyük arama motorlarına belirli sorgular göndermiş ve dönüş süreleri açısından değerlendirmiştir [10]. Vaguhan ve Thewall'ın yaptığı çalışmada ise, Google arama motorunun hit sayısı kullanılarak çeşitli web sayfaları için indekslenmiş doküman sayısı bulunmuştur [11]. Ichikawa ve Uehara yaptıkları çalışmada, dağıtık arama motoru yapısını bulut teknolojisi ile entegre etmişlerdir [12].

Uyar tarafından benzer bir çalışma arama motoru Google için yapılmıştır. Bu çalışmada, İngilizce kelimeler için tekil-çoğul, bileşik kelimeler ve ek alan düzenli fiiller olmak üzere üç grup olarak ele alınmıştır. Her bir grup için farklı sorgu yapıları oluşturulmuş ve çıkan sonuçlar doğrultusunda Google arama motorunun gövdeleme algoritması tahminlenmeye çalışılmıştır [13].

Bu çalışmada ise, tekil-çoğul kelime grubu ele alınarak benzer bir sistem arama motoru Bing için yapılmıştır.

2 Yöntem

Öncelikli olarak bazı kelimeler incelenmiş ve yapısal benzerliklerine göre gruplara ayrılmıştır. Bu çalışmada, kullanılan kelime grubu tekil-çoğul kelime grubudur ve bu gruptan rastgele 25 adet kelime seçilmiştir. Sistemin gerçekleştirilmesi ile ilgili yapılan çalışmalar ise aşağıda anlatılmıştır.

2.1 Test Kelimelerinin Oluşturulması

Kök bakımından benzer dil yapısına sahip kelimeler belirlenmiştir. Örneğin, “benzene” ve “aerosol” kelimelerinin sadece tekil ve çoğul formları bulunmaktadır. Bu sebeple bu kelimeler tekil-çoğul kelime grubuna dahil edilmiştir. Test işlemleri için bu kelime grubundan rastgele 25 adet kelime seçilmiştir.

2.2 Sorgu Oluşturulması

Her bir kelime için arama yapıldığı zaman, o kelimenin tek bir formu ile ilgili linklerin dönmesi istenmektedir. O kelimeye ait diğer formları içeren linklerin dönmesi istenmemektedir. Örneğin, “aerosol” kelimesi için oluşturulan sorgu; “aerosol-aerosols” şeklindedir. Bu sorgu ile “aerosol” kelimesini içeren ilgili linkler dönerken, “aerosols” kelimesini içeren linkler dönmemektedir. Bu amaç doğrultusunda her bir kelime için uygun sorgular oluşturulmuştur.

2.3 Link Seçimi Algoritması

Her bir kelime için oluşturulan sorgular Bing'e gönderilmektedir. Seçilen kelime grubu için iki farklı link seçim algoritması kullanılmıştır.

İlk algoritmada, dönen sonuçlar içerisinde ilk 100 link seçilmiştir. Örneğin, “aerosol-aerosols” sorgusu sonucu dönen sonuçların ilk 100 tanesi alınmış ve kaydedilmiştir.

İkinci algoritmada ise, sıralamada daha gerilerde bulunan dokümanların getirilmesi amaçlanmıştır. Sıralamada daha geride olanlardan kastedilen, yaygınlığı düşük olan dokümanlardır. Bu metotla bulunmak istenen ise, Bing arama motorunun indeksleme yaparken yaygınlığı dikkate alıp almadığıdır. Bu amaç için sorgular oluşturulurken üç haneli rastgele sayılar kullanılmış ve sorgu sonucu dönen linkler arasından ilk 10 tanesi seçilmiştir. Bu işlem 10 farklı sayı için gerçekleştirilmiştir. Örneğin, “aerosol 918-aerosols” şeklinde

bir sorgu oluşturulmuş ve Bing'e gönderilmiştir. Dönen sonuçlar içerisinde ilk 10 tanesi alınmış ve kaydedilmiştir.

2.4 Link Kontrolü

Çalışma kapsamında, tüm sorgular Bing'e gönderildikten sonra dönen cevaplar, kontrol edilecek linkleri oluşturmaktadır. Bu linkleri kontrol etmekteki amaç ise o linkin, arama yapılan kelimenin farklı formları için indekslenip indekslenmediğini belirlemektir. Bu amaç doğrultusunda, her bir kelime için oluşturulan sorgular sonucu Bing'den dönen linkler ve kelimenin çoğul formu kullanılarak yeni sorgular oluşturulur. Bu sorgu Bing'e gönderilir, dönen sonuçlar arasında sorgulanan linkin olup olmadığına bakılır. Örneğin, “aerosol-aerosols” sorgusundan dönen linkler link1, link2, ..., link100 şeklindedir. Link kontrolü için “aerosols link1” şeklinde bir sorgu Bing'e gönderilir. Dönen sonuçlardan ilk 100 link alınır ve link1'in sorgu sonuçları arasında olup olmadığına bakılır. Eğer link1 var ise; link1 o kelimenin çoğul formu için indekslenmiştir, yok ise; link1 kelimenin çoğul formu için indekslenmemiş demektir. Bu işlem tüm linkler için gerçekleştirilmektedir.

3 Tekil-Çoğul Kelime Grubu Testleri

Kelime gövdeleme işleminin en çok yapılan kısmı, çoğul kelimelere gövdeleme işlemi yapılarak tekil formlarının oluşturulmasıdır. Tekil-çoğul kelimeler arası ilişkiler çok kuvvetlidir ve istisna durumlar İngilizce kelimeler için çok azdır. Arama motorlarının tekil çoğul kelimelerin gövdeleme işlemine karşı verdikleri cevap, kendi içerindeki gövdeleme mekanizması hakkında en fazla bilgiyi veren yapıdır.

Tekil-çoğul kelimeleri incelemek için öncelikli olarak kelime seçimi işlemi yapılmıştır. Çoğul eki almış formları dışında hiç bir ek almamış kelimeler gruplanmış ve bu grup içerisinde rastgele 25 adedi test için seçilmiştir. Test işlemi dört adımdan oluşmaktadır.

Birinci test adımında, seçilen kelimelerin her biri için “tekil-çoğul” şeklinde sorgular oluşturulmuş olup, Bing'de sorgulama yapılmıştır. Dönen sonuçlar içerisinde, ilk sıradan başlayıp ilk 100 tanesi kaydedilmiştir. Daha sonra bu 100 sonuç için, tek tek “çoğul link” şeklinde sorgular oluşturulup link kontrolü yapılmıştır. Eğer sorguyu oluşturan link, gelen ilk 100 link içerisinde var ise, bu link kelimenin çoğul formu için indekslenmiştir. Link yok ise, “indekslenmemiştir” sonucuna ulaşılmıştır ve ulaşılan sonuç kaydedilmiştir.

İkinci test adımında, seçilen kelimelerin her biri için “çoğul - tekil” şeklinde sorgular oluşturulmuştur. Birinci adımdakine benzer şekilde Bing'e gönderilip aynı sayıda sonuç kaydedilmiştir. Dönen linkler ile tekrardan “tekil link” şeklinde sorgu oluşturulup, Bing'den dönen sonuçlar değerlendirilerek kaydedilmiştir.

Tablo 1'de ilk iki sütun seçilen kelimelerin tekil ve çoğul formlarını göstermektedir. Tablonun üçüncü sütunu tekil kelime içeren dokümanların çoğul kelimeler için indekslenme oranını göstermektedir. Tablonun son sütunu ise, çoğul kelime içeren dokümanların tekil kelimeler için indekslenme oranını göstermektedir. Tablodan da anlaşıldığı üzere, genel olarak tekil kelimeler içeren dokümanların, çoğul kelimeler ile indekslenme oranı, çoğul kelime içeren dokümanların tekil kelimeler için indekslenme oranından daha yüksektir. Sorgu bakımından hiç bir fark olmamasına rağmen bazı kelimelerin indekslenme oranı yaklaşık % 10 iken, bazı kelimeler için bu oran yaklaşık % 90'dır. Bu oranlar arasındaki farklılıkların hem doküman hem de kelime bazlı olduğu düşünülmektedir. Yani doküman indekslenirken, çoğul kelime için indekslenme

bilgisi oluşmamış olabilir ya da kelimenin kullanım sıklığı göz önüne alınarak indekslenme yapılmış olabilir.

Üçüncü test adımında ise, popüler olmayan dokümanlar için de aynı oranda indekslenme yapıp yapılmadığını anlayabilmek için, oluşturulan sorguya 100 ile 999 arasında üç basamaklı bir sayı (ÜBS) eklenip, “tekil ÜBS-çoğul” şeklinde bir sorgu oluşturulmuştur. Her bir sorgu için, Bing’den dönen ilk 10 link kaydedilmiştir. Dönen linkler kullanılarak tekrar “çoğul link” şeklinde sorgular oluşturulup, benzer şekilde ilk 10 link değerlendirilmiş ve sonuçları kaydedilmiştir. Sorgu oluşturulurken kullanılan link, ikinci sorgu sonuçları arasında geliyorsa kullanılan link o kelimenin çoğul formu için indekslidir, gelmiyorsa indeksli değildir sonucuna ulaşılır. Bu işlem her bir kelime için 10 kez farklı sayılar ile tekrarlanmıştır.

Tablo 1: İlk iki test adımı için arama sonuçları.

Tekil Kelime	Çoğul Kelime	Tekil	Çoğul
		sorgular için döndürülen, çoğul kelimeler içeren doküman yüzdesi	sorgular için döndürülen, tekil kelimeler içeren doküman yüzdesi
benzene	benzenes	60	6
brickbat	brickbats	11	3
starling	starlings	83	4
fishmonger	fishmongers	37	37
parvenu	parvenus	26	4
nursery	nurseries	75	64
wellspring	wellsprings	73	43
mistress	mistresses	82	28
quoit	quoits	33	20
ruckus	ruckuses	73	7
aerosol	aerosols	80	36
broom	brooms	69	49
caisson	caissons	50	27
drawback	drawbacks	25	38
fusion	fusions	90	3
genius	geniuses	32	6
heirloom	heirlooms	81	59
mutt	mutts	71	63
nanny	nannies	82	71
pumpkin	pumpkins	72	55
rye	ryes	71	11
satellite	satellites	76	8
scope	scopes	79	47
tadpole	tadpoles	62	40
unction	unctions	27	13

Dördüncü test adımında ise, üçüncü test adımındaki benzer şekilde “çoğul ÜBS-tekil” şeklinde sorgu oluşturulup, Bing’e gönderilir. Yine dönen sonuçlar değerlendirilip sonuçları kaydedilir. Her bir kelime için benzer işlemler 10 kez farklı sayılar ile tekrarlanır.

Tablo 2’de ilk iki sütun seçilen kelimelerin tekil ve çoğul formlarını göstermektedir. Tablonun üçüncü sütununda, rastgele seçilmiş 100 ile 999 arasındaki üç haneli sayıyı ve kelimenin tekil formunu içeren dokümanların, çoğul kelime formu için indekslenme oranı gösterilmektedir. Tablonun son sütununda ise, rastgele seçilmiş 100 ile 999 arasındaki sayıyı ve kelimenin çoğul formunu içeren dokümanların tekil kelime formu için indekslenme oranı gösterilmektedir. Tablodan da anlaşılacağı gibi ÜBS ve tekil kelime formu içeren doküman

sayısının çoğul kelime formu için indekslenme oranı, çoğul kelimelere göre daha fazladır. Bu indekslenme oranı tekil kelime içeren dokümanlar için; % 6 ile % 46 arasında değişmekte, çoğul formunu içeren dokümanlar için ise; % 3 ile % 35 arasında değişmektedir. Bu oranlar sonucunda dokümanın indekslenme işleminin kelimeye ve dokümana bağlı olduğu düşünülmektedir.

Tablo 2: İlk iki test adımı için arama sonuçları.

Tekil Kelime	Çoğul Kelime	ÜBS içeren tekil kelimeler için döndürülen, çoğul kelimeler içeren doküman yüzdesi	ÜBS içeren çoğul kelimeler için döndürülen, tekil kelimeler içeren doküman yüzdesi
		benzene	benzenes
brickbat	brickbats	12	5
starling	starlings	26	5
fishmonger	fishmongers	34	22
parvenu	parvenus	16	5
nursery	nurseries	46	24
wellspring	wellsprings	31	31
mistress	mistresses	29	20
quoit	quoits	41	31
ruckus	ruckuses	27	7
aerosol	aerosols	20	11
broom	brooms	30	11
caisson	caissons	31	26
drawback	drawbacks	14	19
fusion	fusions	36	2
genius	geniuses	9	9
heirloom	heirlooms	31	20
mutt	mutts	25	17
nanny	nannies	45	17
pumpkin	pumpkins	37	17
rye	ryes	34	7
satellite	satellites	28	2
scope	scopes	18	13
tadpole	tadpoles	29	33
unction	unctions	42	2

Tekil-çoğul kelime grubu testleri sonucunda her bir kelime için farklı sonuçlar elde edilmiştir. Bazı kelimeler için indeksleme oranları % 80 civarındayken, bazı kelimeler için bu oran yaklaşık % 6’dır. Bu oranlar arasındaki farkın ilk sebebi, arama motoru tarafında indeksleme yapılırken tekil kelimeler için arama yapılırken arayan kişinin kelimenin çoğul formunu arıyor olabileceği düşünülmüş olabilir. Tablo 1 ve Tablo 2 incelendiğinde aynı kelime ile yapılan aramalar için bulunan indeksleme oranlarının değiştiği gözlemlenmektedir. Buna göre; örnek kelime için arama yapıldığı zaman ilk sıralarda gelen dokümanların, sonraki sıralarda gelen dokümanlara göre indeksleme oranı daha yüksek olabilir. Ayrıca bu oranların değişmesinin diğer bir sebebi, her dokümanın içeriğinin farklı olması olabilir. Yani bir kelimenin doküman için indekslenip indekslenmeyeceği bilgisi, dokümana göre farklılık gösterebilir. Bing’in tekil-çoğul kelimeler için hem kelime hem de doküman bazlı indeksleme yaptığı tahmin edilmektedir.

Genel olarak, hemen hemen tüm kelimeler için tekil kelime içeren dokümanların çoğul kelime formu içinde indekslenme

oranının, çoğul kelime içeren dokümanların tekil kelime formu için indeksleme oranından daha yüksek olduğu görülmüştür.

4 Sonuçlar

Bu çalışmada Bing'in tekil-çoğul kelime grubu için indeksleme algoritmasının nasıl davranış gösterdiği anlaşılmasına çalışılmıştır. Gerekli yapıların anlaşılabilmesi için, kelime grubu içerisinde 25 adet kelime çifti seçilmiştir.

Kelime grubu test işlemleri iki adımdan oluşmuştur ve test sonuçlarına göre, kelimenin tekil formunu içeren bir doküman, kelimenin çoğul formu için arama yapıldığında arama sonuçları içerisinde gelme ihtimali vardır. Bu ihtimal bazı kelimeler için % 85 civarındayken, bazı kelimeler için % 10 civarındadır. Benzer şekilde kelimenin çoğul formunu içeren bir doküman, tekil formu için arama yapıldığında arama sonuçları arasında gelme ihtimali % 60 ile % 6 arasında değişmektedir.

Genel olarak, testler ele alındığında, her kelime için farklı oranlar ortaya çıkmıştır. Farklı zaman dilimlerinde yapılan testler karşılaştırıldığında, oranlar arasında çok büyük değişimler gözlenmemektedir. Ayrıca arama sonuçlarında ilk sıralarda gelen dokümanlar için indeksleme oranının, son sıralarda gelen dokümanlara göre daha yüksek olduğu görülmüştür. Bu da dokümanların kelimeler için indekslenirken, farklı indeksleme oranlarının bulunduğu bilgisini ortaya çıkarmıştır. Örneğin, bir doküman kelimenin tekil formu için indekslenirken çoğul formu için indekslenemeyebilir, tam tersi bir durum da gerçekleşebilmektedir.

Bu çalışma, farklı dil yapısına sahip kelime grupları ve farklı diller açısından da değerlendirmeye alınabilir. Daha önce Google için yapılmış benzer çalışma ile indeksleme oranları bakımından karşılaştırılabilir. Özellikle diğer büyük arama motorları açısından da incelenip sonuçlar değerlendirilebilir. Ayrıca İngilizce dilindeki kelimelerin kullanım sıklıkları ele alınarak karşılaştırmalar da yapılabilir.

5 Kaynaklar

[1] Lovins, J. B., "Development of a Stemming Algorithm", Mechanical Translation and Computational Linguistics, Vol. 11, 22-31, 1968.

- [2] Porter, M., "An algorithm for suffix stripping", Program 14 (3) 130-137, 1980,
- [3] Vaughan, L. and Thelwall, M. "Scholarly Use of the Web: What are the Key Inducers of Links to Journal Web Sites?", Journal of the American Society for Information Science and Technology 54, 29-38, 2003.
- [4] Aguillo, I. F., Granadino, B., Ortega, J. L. and Prieto, J. A., "Scientific Research Activity and Communication Measured With Cybermetrics Indicators", Journal of the American Society for Information Science and Technology 57, 1296-1302, 2006.
- [5] Tawileh, W., "Exploring web search behavior of Arab internet users", 2011 International Conference on Innovations in Information Technology, 313-318, 2011.
- [6] Srivastava, V., Kollipara, S.V., Tyagi, V., Jaiswal, R.K., "Share-ken: A Way to Improve Web Search", 2010 International Conference on Recent Trends in Information, Telecommunication and Computing, 327-329, 2010.
- [7] Raval, V. and Kumar, P., "EGG (Enhanced Guided Google)- A Meta Search Engine for Combinatorial Keyword Search", Institute of Technology, Nirma University, Ahmedabad, 382-481, 2011.
- [8] Kim, J. and Carvalho, V. R., "An Analysis of Time-Instability in Web Search Results", 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, 466-478, 2011.
- [9] Odabaşoğlu, C., "İnternet Arama Motorları Analizi", Haliç Üniversitesi, Yüksek Lisans Tezi, 2009.
- [10] Edosomwan, J. and Edosomwan, T., "Comparative Analysis of Some Search Engines", South African Journal of Science, 2010.
- [11] Liwen Vaughana, L. and Thelwallb, M., "Search Engine Coverage Bias: Evidence and Possible Causes", Information Processing and Management 40, 693-707, 2004.
- [12] Ichikawa, Y. and Uehara, M., "Distributed Search Engine for an IaaS Based Cloud", 2011 International Conference on Broadband and Wireless Computing, Communication and Applications, 34-39, 2011.
- [13] Uyar, A., "Google stemming mechanisms", Journal of Information Science, 499-514, 2009.