

Resim Arama Motorlarının Sorgu Sözcük Sayısına Göre Performans Değerlendirmeleri

Performance Evaluation of Image Search Engines Based on the Number of Query Words

Erkan Emirzade

Yakın Doğu Üniversitesi Bilgisayar Mühendisliği Bölümü, Lefkoşa, KKTC.
erkan_emirzade@analiz-systems.com

Yıltan Bitirim

Doğu Akdeniz Üniversitesi Bilgisayar Mühendisliği Bölümü, Mağusa, KKTC.

yiltan.bitirim@emu.edu.tr

İçerik

- Giriş
- Yöntem
- Bulgular ve Yorum
- Sonuç
- İleriki Çalışmalar ve Öneriler

Giriş

- Bilgi çağında bilgiye hızlı erişmek daha çok önem kazanmıştır.
- İnternet, insanların ilk başvuru kaynağı olmayı şimdiden başarmıştır.
- Yeni kayıt ortamı olan İnternet hızlı gelişimini sürdürürken, bilgiye erişimde kullanılan arama motorları da kendilerini sürekli geliştirmeye devam etmektedir.
- İnternet büyüdükçe üzerinde saklanan resimler de artmaktadır
- Bu artışla birlikte ilgili resimlere erişim daha karmaşık bir hal almıştır.
- Bu çalışmada, duyarlılık ve normalize sıralama ölçütleri kullanılarak farklı kesme noktaları (ilk 10 ve ilk 20) için sorgu sözcük sayısındaki değişimin seçilen popüler resim arama motorları (Google, Yahoo ve MSN) üzerindeki etkisi araştırılmıştır.

Yöntem

- Artan sorgu sözcük sayısının resim arama motorları üzerindeki etkisini ölçmek için öncelikle üç popüler resim arama motoru seçilmiştir.
- Google (www.google.com)
- Yahoo (www.yahoo.com)
- MSN (www.msn.com)
- Wordtracker (<http://www.wordtracker.com>) - Haziran 2006

Tablo 1. Sorgu listesi

S. no.	Bir sözcüklü sorgular	S. no.	İki sözcüklü sorgular	S. no.	Üç sözcüklü sorgular	S. no.	Dört sözcüklü sorgular
1	car	11	50 cent	21	valentine day cards	31	halloween costumes for kids
2	map	12	britney spears	22	white roller skate	32	ray ban metal sunglasses
3	game	13	star wars	23	martin luther king	33	calvin klein deodorant perfume
4	tattoo	14	crazy frog	24	equipment travel bag	34	wood made musical instruments
5	flower	15	halloween costume	25	wild animal rescue	35	lord of the rings
6	beach	16	funny picture	26	wireless ethernet bridge	36	rosewood hand made furniture
7	simpsons	17	city map	27	bottled drinking water	37	remote control hobby toys
8	amazon	18	hard job	28	antic wall clocks	38	silver watch with leather
9	wallpaper	19	cat scan	29	white sleeveless shirt	39	hydro electric power plant
10	cat	20	football shoes	30	pink rose bouquet	40	pink high heeled shoes

Yöntem

- Her sorgu seçilmiş resim arama motoru üzerinde tek tek ve bağımsız olarak çalıştırılmıştır. Bütün sorguların araştırılması yaklaşık bir haftada bitirilmiştir.
- İnternet kullanıcılarının %80'i arama motoru sonuçlarının sadece ilk 20 ögesine (ilk iki sayfasına) bakmaktadır. (Jansen & Spink, 2003)
- “İlgili” ya da “İlgisiz”
- Duyarlılık ve normalize sıralama değerleri kesme noktası 10 ve kesme noktası 20 için teker teker hesaplanmıştır.

Yöntem

Duyarlılık = Erişilen ilgili öge sayısı / Erişilen toplam öge sayısı

Normalize Sıralama

$$S_{\text{norm}}(\Delta) = \frac{1}{2} \left(1 + \frac{S^+ - S^-}{S^+_{\text{max}}} \right)$$

Δ : Erişim çıktısı sıralaması;

S^+ : Erişim çıktısında ilgili ögelerin ilgisiz ögelerin önünde yer aldığı öge çiftleri sayısı;

S^- : Erişim çıktısında ilgisiz ögelerin ilgili ögelerin önünde yer aldığı öge çiftleri sayısı;

S^+_{max} : Mümkün olan en fazla S^+ .

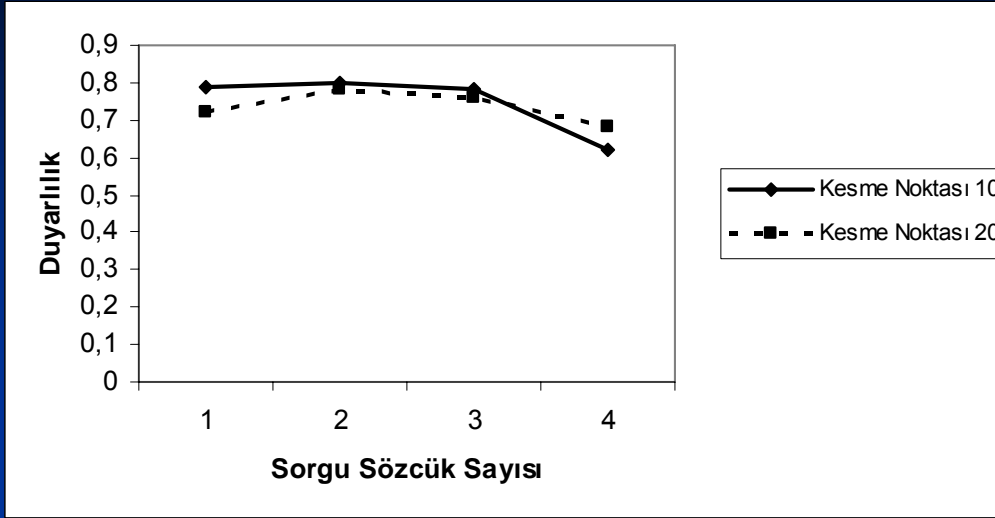
Bulgular ve Yorumlar

Tablo 2. Sorgu grupları için erişilen ilgili resim öge sayıları

Bir sözcüklü sorgu grubu				İki sözcüklü sorgu grubu				Üç sözcüklü sorgu grubu				Dört sözcüklü sorgu grubu			
Sorgu no	G	Y	M	Sorgu no	G	Y	M	Sorgu no	G	Y	M	Sorgu no	G	Y	M
S1	13	17	14	S11	19	19	20	S21	20	13	15	S31	14	4	11
S2	17	17	20	S12	20	20	20	S22	11	5	5	S32	19	12	15
S3	17	4	9	S13	18	18	15	S23	15	18	15	S33	16	16	20
S4	13	20	16	S14	20	20	19	S24	19	18	14	S34	9	4	0
S5	16	19	17	S15	15	15	9	S25	5	2	3	S35	13	17	9
S6	15	16	18	S16	10	11	8	S26	17	17	15	S36	11	19	0
S7	13	18	18	S17	19	19	19	S27	12	5	11	S37	12	5	3
S8	4	3	3	S18	7	5	3	S28	20	14	14	S38	17	13	19
S9	17	20	20	S19	12	12	12	S29	15	12	11	S39	11	5	5
S10	18	19	19	S20	17	17	8	S30	19	16	15	S40	13	1	6
Ort. (%)	72	77	77	Ort. (%)	78	78	67	Ort. (%)	77	60	59	Ort. (%)	68	48	44

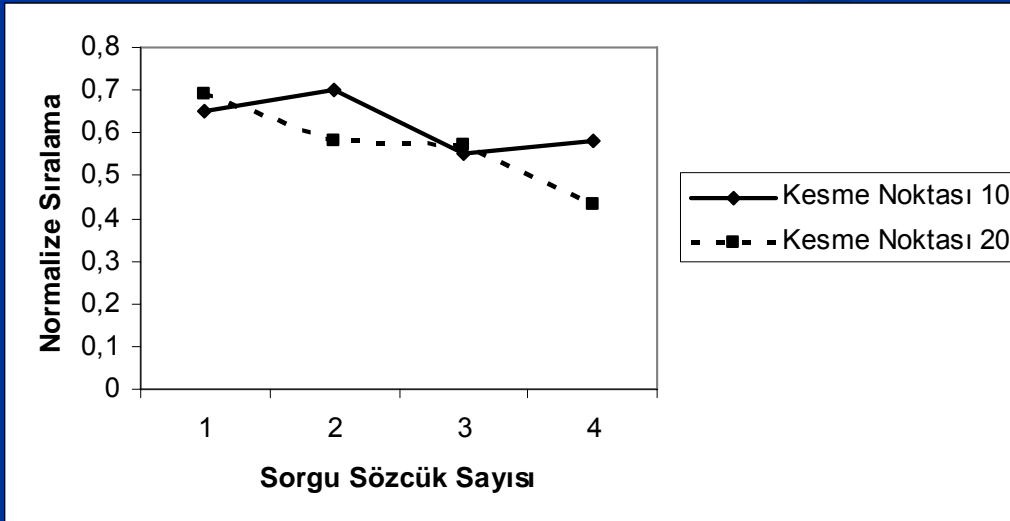
Nor: G: Google; Y: Yahoo; M: MSN.

Şekil 1. Resim arama motorlarının ortalama duyarlılık oranları



a) Google

Şekil 2. Resim arama motorlarının ortalama normalize sıralama değerleri:



a) Google

Şekil 1. Resim arama motorlarının ortalama duyarlılık oranları



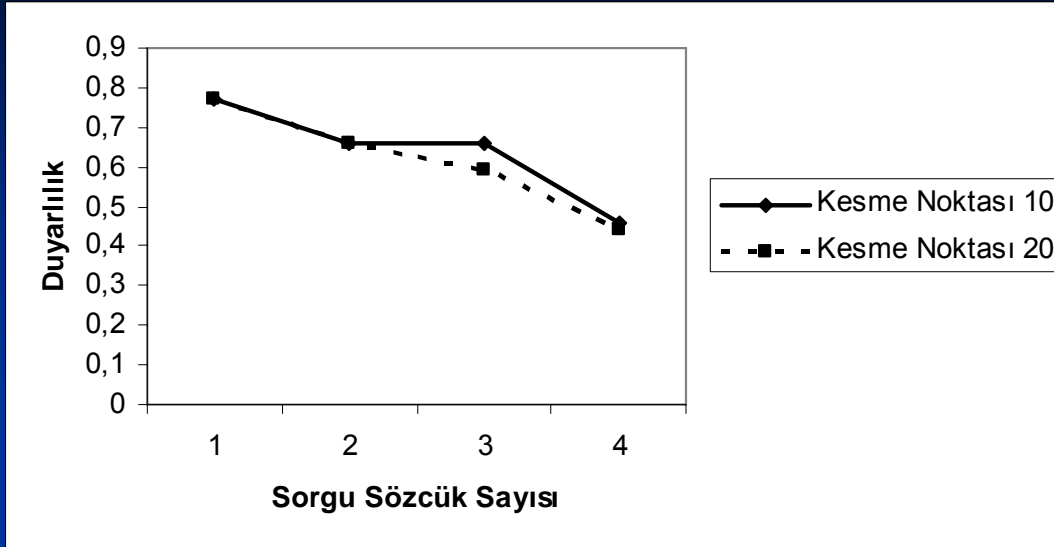
b) Yahoo

Şekil 2. Resim arama motorlarının ortalama normalize sıralama değerleri:

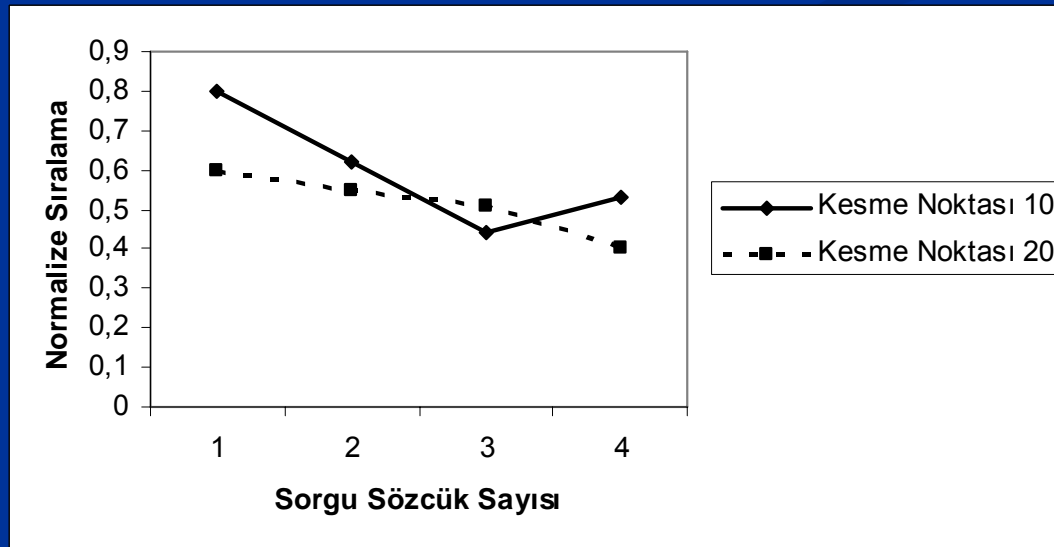


b) Yahoo

Şekil 1. Resim arama motorlarının ortalama duyarlılık oranları



Şekil 2. Resim arama motorlarının ortalama normalize sıralama değerleri:



Sonuç

- Google, Yahoo ve MSN resim arama motorlarında sorgularda yer alan sözcük sayısı arttıkça arama motorlarının duyarlılık performansları genelde düşmektedir
- Duyarlılık oranları kesme noktası 10 iken genelde daha yüksektir. Başka bir deyişle, erişilen resim sayısı arttıkça duyarlılık oranları da düşmektedir.
- Benzeri bir durum normalize sıralama oranları için de geçerlidir. Sorgu sözcük sayısı arttıkça arama motorlarının normalize sıralama performansları düşmektedir.
- Arama motorlarının kesme noktası 10'daki normalize sıralama oranları kesme noktası 20'dekilerden genelde daha yüksektir. Başka bir deyişle, erişilen resim sayısı arttıkça ilgili resimleri daha üst sıralarda gösterebilme başarısı düşmektedir.

Sonuç

- Metin türü veri tabanlarında sorgu sözcük sayısı arttıkça bilgi erişim performansı artarken, resim içeren veri tabanlarında bunun tam tersi bir sonuç elde edilmesi dikkat çekicidir.
- Bu sonuçlar resim arama motorlarının geliştirilmesi gerektiğini göstermektedir.

İleriki Çalışmalar ve Öneriler

- İleriki çalışmamızda, sorgu sayısı artırılacak, kesme noktaları dörde (5, 10, 15 ve 20) çıkarılacak ve değişik bilgi erişim sistemlerinin performansları arasında anlamlı derecede fark olup olmadığını test etmek için derinlemesine istatistiksel analizler yapılacaktır.
- Böylece, sorgu sözcük sayısı arttıkça resim arama motorlarının bilgi erişim performanslarının düşmesindeki etkenler daha ayrıntılı olarak incelenebilecektir.
- Resim arama motorlarının bilgi erişim performansını artırmak için resimleri tanımlamada daha kapsamlı üstverilerin (metadata) ve resim ontolojilerinin kullanılması (image ontologies) ve görüntü tanıma algoritmaları ile makine öğreniminin geliştirilmesi gerekmektedir.

Resim Arama Motorlarının Sorgu Sözcük Sayısına Göre Performans Değerlendirmeleri

Performance Evaluation of Image Search Engines Based on the Number of Query Words

Erkan Emirzade

Yakın Doğu Üniversitesi Bilgisayar Mühendisliği Bölümü, Lefkoşa, KKTC.
erkan_emirzade@analiz-systems.com

Yıltan Bitirim

Doğu Akdeniz Üniversitesi Bilgisayar Mühendisliği Bölümü, Mağusa, KKTC.

yiltan.bitirim@emu.edu.tr