

6 SONUÇ VE ÖNERİLER

Bu çalışmada ülkemizde yaygın olarak kullanılan Arabul, Arama, Netbul ve Superonline'a çeşitli türde 17 soru yöneltilmiş ve bu sorulara karşılık erişilen “ilgili” ve “ilgisiz” belgelere dayanarak arama motorlarının çeşitli kesme noktalarındaki duyarlık ve normalize sıralama değerleri açısından performansları değerlendirilmiştir. Arama motorlarının dizinlenen belgeleri ne kadar sıklıkla ziyaret ettikleri ve güncelleştirdikleri erişim çıktılarında yer alan “ölü” (yani erişilemeyen) adreslerin sayısına bakılarak saptanmıştır. . Türkçe arama motorlarında en sık aranan beş sözcük ("mp3", "oyun", "sex", "erotik" ve "porno") dört arama motorunda aranmış ve her arama motorunun kapsama ve yenilik oranları bulunmuştur. Arabul, Arama, Netbul ve Superonline'ın belgeleri dizinlemek amacıyla "anahtar sözcük", "tanım" gibi HTML üst veri (metadata) alanlarından yararlanıp yararlanmadıkları iki küçük deneyle test edilmiştir. Aşağıda araştırma sonuçları kısaca özetlenmekte ve arama motorlarının performanslarını artırmak için bazı öneriler yer almaktadır.

Arama motorlarının eriştikleri ortalama her 6 belgeden birisi (%17) ölü bağlantı içermektedir. Arama motorlarının ölü bağlantı oranları %27 (Arama) ile %4 (Netbul) arasında değişmektedir (Superonline ve Arabul %19). Netbul ile diğer üç arama motorunun ölü bağlantı oranları arasındaki fark istatistiksel açıdan anlamlıdır. Diğer arama motorlarına göre Netbul’da dizinlenen belgeler dizinleme robotları tarafından daha sık aralıklarla ziyaret edilmekte ve erişilemez hale gelen (ölü) adresler daha hızlı güncelleştirilmektedir (bkz. 5.1).

Arama motorları dizinledikleri belgeleri daha sık aralıklarla ziyaret etmelidirler. Erişilemez hale gelen belgeler ya dizinlerden çıkarılmalı ya da bu belgelerin yeni adresleri hızla güncelleştirilmelidir.

Arabul 17 sorudan 6’sı (%35), Netbul ise 17 sorudan 1’i (%6) için hiç bir belgeye erişememiştir. Arama ve Superonline ise tüm sorular için en az bir belgeye erişmişlerdir. Toplam 17 sorudan 3’ünde (%18) hiç bir arama motoru ilgili belgeye erişememiştir. Arabul toplam 17 sorudan 11 (%65), Netbul 8 (%47), Superonline 5 (%29), Arama ise 4 (%24) soruda ilgili belgelere erişememiştir (bkz. 5.2.2.1).

Arama motorlarının bazı sorular için hiç bir belgeye ya da hiç bir ilgili belgeye erişememe nedenleri araştırılmalıdır. Bu sorunu çözmek için daha çok sayıda ve çeşitli belgeler/siteler dizinlenmeli ve erişim algoritmaları sıfır sonuç vermeyecek şekilde iyileştirilmelidir.

Dört arama motorunun soru başına eriştikleri toplam ilgili belge sayısı 10’dur. Arama soru başına ortalama yaklaşık 4, Superonline ise 3 ilgili belgeye erişmiştir. Arabul ve Netbul çok

sayıda soru için hiçbir ilgili belgeye erişemediklerinden soru başına erişilen ortalama ilgili belge sayıları düşüktür (1,5).

Arama motorlarının eriştikleri ortalama her 6 belgeden 5'i ilgisizdir. Arama motorlarının ortalama duyarlık oranları %28 (Arama) ile %11 (Netbul) arasında değişmektedir (Superonline %20, Arabul %15). Kesme noktası yükseldikçe, yani kullanıcının incelediği belge sayısı arttıkça, arama motorlarının ortalama duyarlık değerleri %50 oranında düşmüştür. Bu düşüş Arama'da daha belirgindir (%90). Arama'nın ilk 5 belgedeki ortalama duyarlık değeriyle (%40) Arabul (%16) ve Netbul'un (%13) ortalama duyarlık değerleri arasındaki fark istatistiksel yönden anlamlıdır (bkz. 5.2.2.2). Başka bir deyişle, ilk 5 belgede Arama, Arabul ve Netbul'dan daha fazla sayıda ilgili belgeye erişmektedir. Arama motorlarının daha yüksek kesme noktalarında eriştikleri ilgili belge sayıları ise birbirine benzemektedir.

Arama motorlarının her soruya karşılık az sayıda ilgili belgeye erişmesi dizinlenen toplam belge sayısının azlığından kaynaklanabileceği gibi kullanıcı arabirimlerinin etkin olmamasından ya da ileri tekniklere dayanan bilgi erişim algoritmaları kullanılmamasından da kaynaklanabilir. Arama motorlarının duyarlık değerlerinin düşük olma nedenleri ayrıntılı olarak incelenmelidir.

Arama motorlarının ortalama normalize sıralama değerleri %54 (Arama) ile %20 (Arabul) arasında değişmektedir (Superonline %37, Netbul %30). Arama, erişim çıktılarında ilgili belgeleri Arabul'dan ve Netbul'dan daha üst sıralarda göstermektedir. Arama'nın erişilen ilk 5, 10 ve 15 belgede kaydettiği ortalama normalize sıralama değerleri Arabul'unkilerden, ilk 10 belgede kaydettiği ortalama normalize sıralama değeri Netbul'unkinden daha yüksektir (bkz. 5.2.2.3).

Ortalama duyarlık değerlerinin yüksek olduğu aramalarda ortalama normalize sıralama değerleri de genellikle yüksektir (Pearson's $r = .61$, $p < .05$). Ancak değerlendirilen belge sayısı arttıkça duyarlık ile normalize sıralama değerleri arasındaki ilişki giderek zayıflamaktadır (bkz. 5.2.2.4).

Yapılan araştırmalarda duyarlık ve anma gibi geleneksel performans ölçütleriyle kullanıcı merkezli performans değerlendirmeleri arasında güçlü bir ilişki olmadığı ortaya çıkmıştır. Başka bir deyişle, bazen kullanıcılar az sayıda ilgili belgeye erişen ama bu belgeleri erişim çıktılarının üst sıralarında gösteren bilgi erişim sistemlerini daha başarılı bulabilmektedirler. Bu bakımdan arama motorlarının erişilen ilgili belgeleri ilk sıralarda gösterme konusunda daha fazla çaba sarfetmeleri gerekmektedir.

Arama motorlarının tüm sorular için kaydettikleri ortalama duyarlık değerleri % 0 ile %36 arasında değişmektedir. Hiç bir arama motoru 4. ("Türkçe arama motorlarında performans

değerlendirme”), 12. (“Demirel veya Sezer’in TEMA hakkındaki görüşleri”) ve 17. (“TBMM Başkanı Ömer İzgi”) sorular için ilgili belgeye erişememiştir. Arama motorları 2. (“barok müzik”), 7. (“DPT”) ve 11. (“Demirel veya Sezer”) sorularda nispeten daha yüksek sayıda ilgili belgeye erişmişlerdir (bkz. 5.2.2.5).

Arama motorlarının tüm sorular için ortalama normalize sıralama değerleri % 0 ile %66 arasında değişmektedir. Arama motorlarının ortalama normalize sıralama değeri açısından en başarılı oldukları sorular 7., 9. (“uzaylılar”) ve 13. (“uzay”) sorulardır.

Arama motorlarının hem ortalama duyarlık hem de ortalama normalize sıralama değerleri açısından en başarılı oldukları sorular ise 7., 8. (uzaylı”), 9. ve 11. sorulardır.

Arama motorları, Web’de yaygın olarak kullanılan “internet”, “etik”, “arama” vb. terimlerin geçtiği spesifik arama sorularında nispeten daha az başarı göstermiştir. Öte yandan tek sözcükten oluşan ya da “VEYA” işleci kullanılan sorularda, erişilen ilgisiz belge sayısı yüksek olmasına rağmen, arama motorları nispeten daha başarılı olmuştur. “VE” işlecinin kullanıldığı sorularda ise başarı oranı daha düşüktür. Arama motorları Boole işleçleri kullanılarak yapılan aramalarda genelde tutarlı sonuçlar vermektedir (bkz. 5.2.3).

Arama motorlarının “VE” işleci kullanılan sorulardaki başarı oranını yükseltmek için daha fazla belge dizinlemeleri gerekmektedir. Tek terimden oluşan sorularda başarı oranları, kullanıcının bu terimleri hangi bağlamda aradıkları belirlenmeye çalışılarak artırılabilir. Web’de yaygın olarak kullanılan terimlerin geçtiği soruları kullanıcının daha spesifik yapmasına olanak verilmelidir. Bir başka yöntem ise, kullanıcıların geçmişte tek sözcükten oluşan bu tür soruları aradıklarında daha çok hangi bağlantıları tıklamış oldukları bilgisine dayanan sezgisel (heuristic) bilgi erişim algoritmaları geliştirmektir.

Arama motorları kullanıcılar tarafından girilen soruları daha iyi analiz etmek ve performansı artırmak için gövdeleme algoritmalarından yararlanmamaktadır. Özellikle Türkçe sözcüklerle yapılan aramalarda gövdeleme algoritmalarının kullanılması arama motorlarının bilgi erişim performansını artırabilir. Bu nedenle Türkçenin dilbilgisi özelliklerini de dikkate alan gövdeleme algoritmaları geliştirilmeli ve kullanılmalıdır.

Türkçe karakter sorunu henüz çözülememiştir. Arama motorları Türkçe karakterler kullanılarak yapılan aramalarda farklı sonuçlar vermektedir (bkz. 5.2.3). Arama motorlarının Türkçe karakter sorununa farklı yaklaşımları kullanıcılar açısından bazı olumsuzluklar yaratmaktadır. Çoğu kullanıcı bu durumun genellikle farkında değildir. Bu bakımdan, kullanıcılar Türkçe karakter kullanımı nedeniyle erişilen belgeleri değerlendirirken zorlanmaktadırlar. Türkçe arama motorlarında yapılan aramaların büyük bir çoğunluğu Türkçeyi kullandıklarından bu sorunun bir an önce çözülmesi gerekmektedir. Web’deki

Türkçe içerik miktarının giderek arttığı düşünülecek olursa, artık arama sorusunda yer alan Türkçe karakterlerin en yakın İngilizce karakterlere çevrilmesi gibi basit yaklaşımlar yerine, gerek Türkçenin dil özelliklerini gerekse kullanıcıların arama davranışlarını da dikkate alan yaklaşımlar yeğlenmelidir.

Türkçe arama motorlarında en sık aranan sözcüklere (“mp3”, “oyun”, “sex”, “erotik” ve “porno”) karşılık erişilen belgelerin büyük bir çoğunluğu (%86) Türkiye adresli değildir. Superonline alan adı “.tr” ile bitmeyen belgelere erişmede tartışmasız bir üstünlüğe sahiptir. Bunda Superonline’ın AltaVista ile işbirliğinin büyük payı olduğu kanısındayız. Türkiye adresli en fazla belgeye erişen arama motoru ise Arama’dır.

Superonline arama motorları arasında en yüksek ortalama kapsama oranına sahiptir. En sık aranan sorulara karşılık erişilen ortalama her 5 ilgili belgeden 2’sine Superonline tarafından erişilmiştir (bkz. 5.3.1.1). Superonline’ı Arabul, Arama ve Netbul izlemektedir. Türkiye adresli belgelerde ise Arama en sık aranan beş soru için ilgili belgelerin büyük bir çoğunluğunu kapsamaktadır. Diğer arama motorlarının Türkiye adresli belgeleri kapsama oranları ihmal edilebilir düzeydedir (bkz. 5.3.1.2).

Arama ve Arabul’un en sık aranan sorular için yenilik oranları %100 civarındadır. Yani, bu arama motorlarının en sık aranan sorulara karşılık eriştikleri belgelerin hemen hemen tümü “yeni”dir. Superonline’ın ortalama yenilik oranı %95 civarındadır. Netbul’da ise bu oranlar %30 civarındadır (bkz. 5.3.2.1).

En sık aranan sorular için Türkiye adresli “yeni” belge bulmada Arama tartışmasız bir üstünlüğe sahiptir. Superonline en sık aranan beş sorudan dördünde Türkiye adresli hiç bir belgeye erişememiştir. En sık aranan sorular için Arabul ve Netbul az sayıda yeni belgeye erişmişlerdir (bkz. 5.3.2.2).

Arama motorları tarafından en sık aranan sorular için bulunan Türkiye adresli ve Türkiye adresli olmayan hemen hemen bütün belgeler yenidir. Bu durum, arama motorlarının eriştikleri ilgili belgeler arasında çok düşük bir çakışma olduğunu göstermektedir. Bir başka deyişle, aynı sorular için her arama motoru birbirinden oldukça farklı belgeler dizinlemekte ve doğal olarak farklı ilgili belgelere erişmektedir.

Bu sonuçlar belirli bir konuda ilgili belgelerin tümüne erişebilmek için birden fazla Türkçe arama motoru üzerinde arama yapılması gereğini ortaya çıkarmaktadır. Bazı arama motorlarının çok az sayıda Türkiye adresli belge dizinledikleri görülmektedir. Türkiye’yle ilgili bazı sorularda (“Atatürk ve Fikriye Hanım”, “Ömer İzgi” vb. gibi) arama motorlarının nispeten daha başarısız olmalarının nedenlerinden birisi de kanımızca budur. Türkçe arama motorlarının dizinledikleri Türkiye adresli belge sayıları artırılmalıdır.

Arama ve Arabul arama motorları HTML belgelerinde yer alan “anahtar sözcük” ve “tanım” üst veri (metadata) alanlarında geçen terimleri dizinlememekte ve erişim amacıyla bu terimlerden yararlanmamaktadır. Netbul ve Superonline’ın ise bu alanları dizinledikleri ve erişim amacıyla kullandıkları ortaya çıkmıştır (bkz. 5.4). İlgili alanlara girilen bilgilerin arama motorlarını yanıltmak amacıyla zaman zaman kötüye kullanıldığı bilinen bir gerçektir. Ancak arama motorlarının bu tür kötüye kullanımları eleyecek daha akıllı arama algoritmaları geliştirmeleri ve bu alanlarda yer alan erişim açısından değerli bilgilerden yararlanmaları gerektiği kanısındayız.