ÖZET

Bu çalışmada Türkçe arama motorlarının bilgi erişim performansları çeşitli ölçütlere göre değerlendirilmiştir. Ülkemizde yaygın olarak kullanılan Arabul, Arama, Netbul ve Superonline arama motorları üzerinde çeşitli türde 17 farklı soru için arama yapılmış ve bu sorulara karşılık erişilen "ilgili" ve "ilgisiz" belgelere dayanarak söz konusu dört arama motorunun çeşitli kesme noktalarındaki duyarlık ve normalize sıralama değerleri hesaplanmıştır. 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 sınanmıştır. Kruskal-Wallis ve Mann-Whitney istatistikleri kullanılarak arama motorlarının güncellik, duyarlık, normalize sıralama, kapsama ve yenilik oranlarının birbirinden farklı olup olmadığı test edilmiştir.

Araştırmadan elde edilen belli başlı bulgular şunlardır: Arabul, Arama, Netbul ve Superonline'ın eriştiği ortalama her altı belgeden birisi ölü bağlantı içermektedir. Netbul'un ölü bağlantı oranı diğer arama motorlarından daha düşüktür. Arama motorları bazı sorular için hiç bir belgeye ya da hiç bir ilgili belgeye erişememiştir. Erişilen ortalama her altı belgeden beşi ilgisizdir. Arama motorlarının ortalama duyarlık oranları %11 (Netbul) ile %28 (Arama) arasında değişmektedir (Superonline %20, Arabul %15). Arama, ilk 5 belgede Arabul ve Netbul'dan daha fazla sayıda ilgili belgeye erişmiştir. Arama motorları erişilen ilgili belgeleri erişim çıktılarının ilk sıralarında gösterme konusunda yeterince çaba sarfetmemektedirler. Arama motorlarının ortalama normalize sıralama değerleri %20 (Arabul) ile %54 (Arama) arasında değişmektedir (Superonline %37, Netbul %30). Arama, erişim çıktılarında ilgili belgeleri Arabul'dan ve Netbul'dan daha üst sıralarda göstermektedir. Duyarlık ile normalize sıralama değerleri arasında gözlenen güçlü pozitif ilişki, değerlendirilen belge sayısı arttıkça giderek zayıflamaktadır. Arama motorları, Web'de yaygın olarak kullanılan terimlerin geçtiği spesifik arama sorularında nispeten daha az başarı göstermişlerdir. 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 soruları daha iyi analiz etmek ve performansı artırmak için gövdeleme algoritmalarından yararlanmamaktadırlar. Türkçe arama motorlarında Türkçe karakter sorunu henüz çözülememiştir. Arama motorları Türkçe karakterler kullanılarak yapılan aramalarda farklı sonuçlar vermektedir. En sık aranan "mp3", "oyun", "sex", "erotik" ve "porno" soruları için Superonline'ın kapsama oranları daha yüksektir. Arama dışında diğer Türkçe arama motorlarının Türkiye adresli belgeleri/siteleri pek dizinlemedikleri ortaya çıkmıştır. Türkiye adresli belgeleri kapsamada Arama tartışmasız bir üstünlüğe sahiptir. En sık aranan sorularda hemen hemen tüm arama motorlarının yenilik oranları yüksektir. Aynı sorulara karşılık farklı arama motorları farklı ilgili belgelere erişmektedirler. HTML belgelerinde yer alan "anahtar sözcük" ve "tanım" üst veri (metadata) alanlarında geçen terimlerin bazı arama motorları (Netbul ve Superonline) tarafından dizinlendiği ve erişim amacıyla bu terimlerden yararlanılmadığı ortaya çıkmıştır.

Çalışmanın sonunda Türkçe arama motorlarının bilgi erişim performanslarını geliştirmek için bazı önerilere yer verilmektedir.