

TÜRKÇEDE KONU BULMA VE İZLEME¹

Fazlı Can, Seyit Koçberber, Özgür Bağhoğlu,
Süleyman Kardeş, H. Çağdaş Öcalan, Erkan Uyar

Bilkent Bilgi Erişim Grubu
Bilgisayar Mühendisliği
Bilkent Üniversitesi

16 Mayıs, 2008

Özet

Konu bulma ve izleme (KOBİ) uygulamalarının amacı bir zaman sırası içinde gelen haberlerin konularına göre gruplanmasıdır. Bu alanda bilinen uygulamalara örnek olarak ticari olanlar arasında Google News ve NewsIsFree, araştırmaya yönelik olanlar arasındaysa Columbia University’de geliştirilen Newsblaster ve Michigan University’de geliştirilen NewsInEssence sayılabilir. Problemin bilgi erişim sistemlerindeki göre en önemli farkı gruplandırmanın konulara göre değil olaylara göre yapılmasıdır. Türkçe için KOBİ çalışmalarının gerçekleştirilebilmesi için bir deney derlemine ihtiyaç vardır. Öncü olma niteliği taşıyan bu çalışmada, 208305 Türkçe haber arasından seçilmiş 80 olay için 39 farklı kişi tarafından yapılmış konu izleme sonuçlarını içeren, standartlara uygun nitelikte bir deney derlemi gerçekleştirilmiştir. Derlemin haberleri beş farklı Türkçe haber kaynağının 2005 yılına ait, zaman damgası saat dakika ayrıntısında verilmiş, öğelerinden oluşmaktadır. Bu çalışmada KOBİ işlemlerinin iki önemli problemi olan yeni bir olayla ilgili ilk hikâyenin bulunması ve verilmiş bir haberle ilgili konunun izlenmesi üzerinde odaklanılmıştır. Çalışmada kök bulma ve durma listesi kullanımının başarımlar üzerindeki etkileri incelenmektedir. Yapılan araştırma, kök bulmada basit kelime kesme yöntemi kullanan KOBİ yaklaşımın, kelime yapısını dilbilgisi açısından inceleyen karmaşık bir kök bulma yöntemini kullanan yaklaşımla benzer sonuçlar verdiğini ve kesme listesinin seçiminin sonuçları etkileyebileceğini göstermektedir. Çalışma karar vermek amacıyla kullanılan iki farklı benzerlik hesaplama yönteminin sade bir yaklaşımla etkinliği arttıracak şekilde birleştirilebildiğini ve başarımların TDT (Topic Detection and Tracking) araştırmalarında başka diller için alınan sonuçlarla tutarlı olduğunu göstermektedir. Sunumda başka dillerde de kullanılabilir bir deney derlemi geliştirme yöntemi ve oluşturulan derlemin özellikleri de tanıtılacaktır. Bu çalışmanın sonuçları gerçek zamanda çalışmakta olan Bilkent Haber Portalı’nın geliştirilmesinde kullanılmaktadır.

Anahtar Sözcükler: İlk hikâyenin bulunması, konu bulma ve izleme (KOBİ), başarımlar ölçümü, KOBİ deney derlemi oluşturumu, Türkçe, Bilkent Haber Portalı.

¹ Ege Üniversitesi Uluslararası Bilgisayar Enstitüsü’nde 16 Mayıs 2008 tarihinde yapılacak sunumun özettir.