

Ses İçinde Kelime Aramak

Fast-Talk Communications firması, geliştirdiği Fonetik Arama teknolojisi sayesinde konuşma içeren ses kayıtlarındaki sözcük tabanlı aramalarınızı çok daha kolay ve hızlı gerçekleştirebilme imkanı sunuyor.



Bilgisayar üzerinde metin tabanlı uygulamalarla sıkça haşır neşir olan ve zaman zaman bilgisayarda kayıtlı onca sayfanın içinde özel bir konuyu bulmak isteyenler, Find (Bul) komutunun değerini gayet iyi bilirler. Bu komut sayesinde, binlerce sayfa metin arasında ilgilendiğiniz bir veya birkaç kelimeyi kolayca aramak ve sonuçların saniyeler içinde karşınıza dizilmesini sağlamak mümkün.

Ancak kelime bazlı arama işlemini metinler üzerinde uygulamak ne kadar kolaysa, aynı şeyi sesli olarak kaydedilmiş arşivler üzerinde uygulamak da bir o kadar zor ve zahmetli. Oysa dünya üzerinde ses ve video olarak arşivlenen bilgilerin ulaştığı akıl almaz boyutlar göz önüne alındığında, bu arşivlerin içeriğinde kelime bazlı arama yapabilecek hızlı ve güvenilir bir yöntemin gerekliliği açıkça ortaya çıkıyor. Televizyon ve radyoların yayın arşivlerinin taranmasından şirketlerin çağrı merkezi kayıtlarının kontrolüne, hatta sözcük bazlı otomatik yönlendirme sitelerine kadar böyle bir teknolojinin hayatı kolaylaştırabileceği bir çok uygulama alanı saymak mümkün.

İşte bu nedenle araştırmacılar, yaklaşık 10 yıldır ses kayıtları içinde sözcük bazlı arama yapabilmenin kolay uygulanabilir ve hızlı bir yolunu bulmak için çaba harcıyorlar. Bundan kısa bir süre öncesine kadar, bu işin gerçekleştirilebilmesi için uygulanan geçerli tek yolun LVCSR (Large Vocabulary Continuous Speech Recognition-Geniş Sözlük Yardımıyla Devamlı Ses Tanıma) temelli uygulamalar olduğu kabul ediliyordu. LVCSR, ya da bilinen diğer adıyla Speech to Text (Sesten Metne Dönüştürme), gerçek zamanlı olarak veya belli bir arşive ait olan konuşmaları veritabanında yer alan sözcüklerle sürekli olarak karşılaştırarak konuşmayı metin haline dönüştüren ve ardından gerekli yerlere indeks ve zaman etiketleri yerleştiren bir teknoloji. LVCSR teknolojisi uzun süredir gündemde olduğundan dolayı, bu teknolojiyi temel alan uygulamalara günümüzde yaygın olarak rastlanıyor. Hatta bu iş için kullanılan ticari geliştirme araçlarının yanında, İngiltere'deki Cambridge Üniversitesi'nin HTK (<http://htk.eng.cam.ac.uk/>) ve Amerika'daki Mississippi State Üniversitesi'nin ISIP ([te.edu/projects/speech/\) adını verdikleri geliştirme araçlarından ücretsiz olarak faydalanmak da mümkün.](http://www.isip.mssta-</p></div><div data-bbox=)

Ancak uzun zamana ve güçlü donanım profiline ihtiyaç duyan bu işlemin gerektirdiği yüksek maliyet, büyük ses arşivlerinde aranan verinin bulunması için gereken uzun süre ve kullanılan sisteme göre %50'ye kadar düşebilen doğruluk derecesi, tercih edilebilirliğini önemli ölçüde kısıtlıyor. Bu nedenle Fast-Talk Communications (<http://www.fast-talk.com>) adlı şirket, LVCSR adı verilen yönteme karşı ciddi bir alternatif niteliğinde olan Phonetic Searching (Fonetik Arama) yöntemini geliştirmekle ve yaygınlaştırmakla uğraşiyor.

Fonetik Arama yöntemi, konuşma içeren seslerdeki fonetik unsurları ayırmak ve algılamak için konuşma içeren veriyi önce detaylı bir ön işleme tabi tutuyor. Daha sonra akustik modeller ve fonetik gramerler yardımıyla ses içeriği taranarak bir phonetic search track (fonetik arama izi) oluşturuluyor. Fonetik arama izi, içeriğinde ait olduğu ses dosyasının yüksek oranda sıkıştırılmış fonetik karakterlerini, yani söyleniş biçimlerini barındıran bir dosya ve kolayca depolanabilme, paylaşılabilir ve farklı uygulamalar tarafından kullanılabilir özelliğine sahip. Fonetik arama izi bir kez oluşturulduktan sonra ait olduğu ses dosyasından bağımsız olarak saklanabildiği için, herhangi bir arama yapmak istendiğinde arama gerçek ses dosyasını

işin içine karıştırmadan bu iz üzerinden gerçekleştiriliyor. Kullanılan bu yöntem sayesinde, arama işlemi ortalama hıza sahip bir masaüstü bilgisayar sistemiyle bile gerçek zamanlı işlemin yaklaşık 36.000 katı hızında gerçekleştirilebildiği söyleniyor. Daha somut bir şekilde ifade etmek gerekirse; 30 saatlik bir ses kaydında aradığınız kelimeleri bulmak sadece 1 saniye alıyor. Bir saatlik ses kaydının işlenmesi gereken süreyse aynı sistemle yaklaşık 6 dakika.

Sistem, arama işlemi sırasında kelime bazlı olarak veya fonetik eşdeğerliğe göre arama yapılabilir. Örneğin bir ses kaydının içinde beyin anlamına gelen "brain" kelimesini kolayca aratmanız mümkün, ancak işletmeden işletmeye anlamına gelen ve Business to Business kelimesinin kısaltması olan B2B'yi aramak için ne yapacaksınız? Bu durumda B2B'nin okunuşuna dair fonetik karakterleri sisteme tanıtmamız yeterli. Örneğin "[B IY T UW B IY]" şeklinde yapılacak olan ve ardışık dizildiğinde B2B'nin okunuşundaki fonetik öğelere uygun ses veren bir aramayla, uygun sonuçların karşınıza gelmesini sağlayabiliyorsunuz. Bu yöntem geleneksel sestem metne çevrim yöntemlerinin karşılaştığı sözlükte bulunmayan kelime sorununa çözüm getirirken, aynı zamanda normal sözcüklerin fonetik karakterlerinin yansıtılmasıyla daha doğru sonuçlara ulaşılmasını da sağlayabiliyor. Üstelik geliştiricilerinin söylediğine göre bu yöntemin bir diğer avantajı da sadece kelimelerin değil; mırıldanılan müzik parçalarının veya tanımlanmış seslerin de ses kayıtları içinde aranmasına olanak sağlayacak bir yapıda olması.

Tabii Fonetik Arama sisteminin getirdiği bu değişik ve yenilikçi yaklaşıma rağmen henüz her şey mükemmel değil. Fast-Talk Communications'un Web sitesindeki açıklayıcı PDF dosyalarına bakılırsa, özellikle çoklu kelimelerin aranmasında her iki yöntemin de performansı -her ne kadar Fonetik Arama yöntemi sonuçları itibarıyla önde de olsa- ciddi oranda düşüyor. Özellikle de eşsesli olarak telaffuz edilen sözcükleri birbirinden doğru biçimde ayırabilmek hala, bu işi yapan yazılımlar için başlı başına bir problem.

Ancak yine de ortaya koyulan bu yeni yöntem, ses arşivlerinin içeriğinde arama yapabilmek için ortaya koyulan çalışmaların mükemmel sonuca ulaşabilmek için geleneksel kelime bazlı düşüncenin ötesine geçmeleri gerektiği fikrini güçlendiriyor. Fakat en dibe de şu notu düşmek lazım: İster LVCSR olsun ister Fonetik Arama yöntemi, desteklenen diller arasında maalesef şimdilik Türkçe'nin adı bile geçmiyor.

Levent Daşkiran

Kaynaklar
<http://www.fast-talk.com>
http://www.technologyreview.com/articles/wo_harney043003.asp

