

Bilgi Erişim Sistemleri

Information Retrieval Systems

Yaşar Tonta*

Öz

Bu çalışmada bilgi erişim sistemleriyle ilgili temel kavramlar incelenmektedir. Bilgi erişim sorunu kısaca bir göz atıldıktan sonra bir bilgi erişim sistemini oluşturan derme veri tabanı, belgelerin dizinlenmesi, sorgu cümlesi hazırlama, kullanıcı arabirimi gibi temel bileşenler kısaca açıklanmaktadır. Belli başlı erişim kuralları ve modeller (kesin çakışma ve olasılık kuramına dayanan erişim kuralları, Boole erişim kuralları, vektör uzayı modeli) özetlenerek "erişim isabeti" ve "kesin isabet" gibi erişim etkinliği ölçüleri incelenmektedir. Bilgi erişim sistemlerinde erişim etkinliğini artırmak için kullanılan ilgililik geribildirimi (relevance feedback) ve kümeleme (clustering) yöntemlerine de değinilmektedir.

Abstract

This paper examines the basic concepts in information retrieval systems. After the introduction of the information retrieval problem, it briefly reviews the major components of an information retrieval system such as the document database, indexing, query formulation, and the user interface. It summarizes the retrieval rules and models (exact match, Boolean, probabilistic, vector space) and describes a few measures of retrieval effectiveness such as precision and recall. Finally, it discusses more advanced retrieval methods (i.e., relevance feedback and clustering) which are used to improve the effectiveness of information retrieval systems.

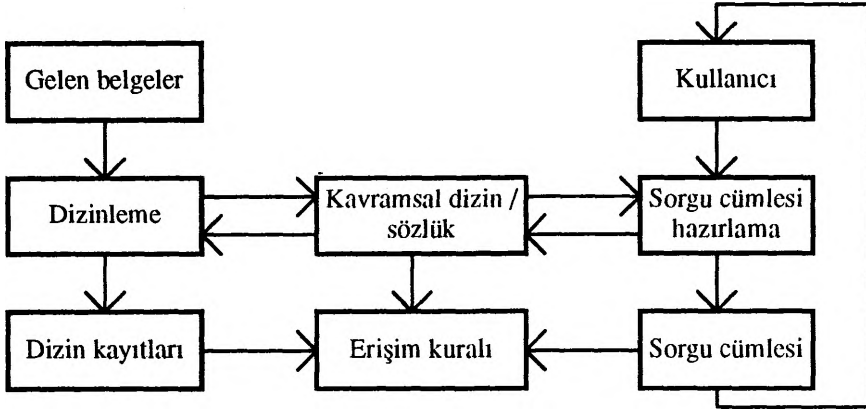
1. Bilgi Erişim Sistemlerine Genel Bakış

Bir bilgi erişim sisteminin temel işlevi dermedeki ilgili (relevant) belgelerin tümüne erişmek, ilgili olmayanları da ayıklamaktır/reddetmektir. İdeal bir bilgi erişim sistemi ilgili belgelerin tümüne ve salt ilgili belgelere erişim sağlamalıdır. Ama, aslına bakılırsa, ideal bir bilgi erişim sistemi yoktur. Bilgi erişim sistemleri ilgili belgelerin tümüne ve salt ilgili belgelere genellikle erişim sağlayamamaktadır. Kullanıcılar birkaç ilgili belgeye hızlı bir şekilde erişim sağlayan sistemlerle de tatmin olabilmektedirler.

* Doç. Dr. Yaşar Tonta H.Ü. Kütüphanecilik Bölümü Öğretim Elemanıdır.

Maron (1984) belge erişim sorununun daha ayrıntılı bir tanımını yapmakta ve geleneksel bir belge erişim sisteminin mantıksal organizasyonunu ana çizgileriyle vermektedir (çizim 1).

Çizim 1. Geleneksel Bir Belge Erişim Sisteminin Mantıksal Organizasyonu
(Kaynak: Maron, 1984: 155)



Çizim 1’den de görülebileceği gibi, dermeye eklenen her belgenin temel karakteristikleri (örneğin, yazar adı, kitap adı, konusu) dizinleme işlemi sırasında belirlenmektedir. Dizinciler her belgeye dizin terimleri ya da konu başlıkları vermek için kavramsal dizinlerden (thesauri) veya sözlüklerden (denetimli söz dağarcıklarından) yararlanabilirler. Sonuç olarak, her belge için daha sonra erişim sağlamak amacıyla bir dizin tutanağı (kayıd) oluşturulur.

Kullanıcılar da uygun arama terimlerini bu dizinleme kaynaklarına bakarak belirleyebilirler. İlk akla gelen terimlerin geçerliliğini kontrol ettikten ve yeni terimleri belirledikten sonra kullanıcı, sorgu cümlesi olarak sisteme sunmak üzere erişim açısından hangi terimlerin en uygun olduğuna karar verir. Ancak çoğu kullanıcıların bilgi gereksinimlerini daha iyi açıklamak için yararlanabilecekleri bu kaynaklardan haberleri yoktur. Bu da, kullanıcının seçtiği terimlerle sistemdeki terimlerin uyuşmaması dolayısıyla aramaların başarısızlıkla sonuçlanmasına yol açabilmektedir.

Bir bilgi erişim sisteminin veri tabanındaki bazı belgelere erişim sağlayabilmesi için iki koşul yerine getirilmelidir. Birincisi, dizinciler tarafından belgelere uygun dizin terimleri verilmelidir. İkincisi, kullanıcılar belgelere verilen bu dizin terimlerini doğru olarak tahmin edip sorgu cümlelerini ona göre girmelidirler. Maron (1984: 155) arama sürecini şöyle açıklamaktadır:

...gerçek arama ve erişim, dizin kayıtlarıyla resmi (formal) sorgu cümlesinin çakışmasıyla gerçekleşir. Çakışma ‘Erişim Kuralı’ (Retrieval Rule) olarak adlandırılan kuralı izler. Bu kural şöyle açıklanabilir: Herhangi bir resmi sorgu [cümlesi] için bu arama sorgusunda belirlenen kayıtların alt setinde yer alan dizin kayıtlarının tümüne ve salt bu dizin kayıtlarına erişim sağla.

Böylece, bir bilgi erişim sisteminin şu bileşenlerden oluştuğu ortaya çıkmaktadır:
(1) bir belge dermesi (ya da bu belgeleri temsil eden kayıtlar [surrogates]); (2)

kullanıcıların sistemle etkileşimini sağlayan bir kullanıcı arabirimi (user interface); (3) kullanıcıların sorgu cümleleinde yer alan terimlerle dermedeki belgelere verilen terimleri karşılaştırarak dermedeki ilgili belgeleri belirlemek için bir erişim kuralı. Kuşkusuz, bilgi gereksinimlerini karşılamak için sistemi kullanan bir kullanıcı grubunun gerekli olduğu da unutulmamalıdır.

Çevrimiçi (online) bilgi erişim sistemlerinin belli başlı bileşenleri daha ayrıntılı olarak aşağıda incelenmektedir.

2. Derme Veri Tabanı

Her bilgi erişim sistemi için bir derme ya da bu dermedeki belgeleri temsil eden kayıtlardan oluşan bir veri tabanının (database) varlığı bir önkoşuldur. "Belge" terimi burada en geniş anlamıyla kullanılmaktadır. Bilgi taşıyan herşey belgedir (kitaplar, kasetler, üç boyutlu materyaller, elektronik dosyalar, vd.). Veri tabanında belgeleri temsil eden kayıtlar depolanmış olabileceği gibi bu belgelerin tam metinleri de depolanmış olabilir.

3. Belgelerin Dizinenmesi

Belgeleri ya da dermedeki belgeleri temsil eden kayıtları içeren bir veri tabanı yaratmak için her belgenin özellikleri belirlenip kaydedilmelidir. Dizinleme olarak bilinen bu işlem ya entellektüel olarak ya da otomatik olarak gerçekleştirilebilir. Entellektüel dizinlemenin yapıldığı bir ortamda profesyonel dizinciler belgelerin tanımlayıcı (descriptive) ve konusal karakteristiklerini saptarlar ve her belge için bir kayıt oluştururlar.

Çizim 1'de de görüldüğü gibi, dizinciler kavramsal dizin, sözlük ve denetimli söz dağarcığı (controlled vocabulary) listesi gibi standart kaynaklardan yararlanabilirler. Anglo-Amerikan Kataloqlama Kuralları 2 (AAKK2) ve Kongre Kütüphanesi Konu Başlıkları Listesi belgelerin tanımlanmasında ve konusal analizinde kullanılan kaynaklara örnek olarak verilebilir. Bu kaynaklardan yararlanan dizinciler belgelerin özelliklerini kaydeder ve her belgeye konu başlıkları verirler. Kaydedilen bu tanımlayıcı ve konusal bilgiler belgeyi temsil eden kaydı oluşturur. Bu kayıt daha sonra erişim noktaları sağlamak ve dolayısıyla ilgili belgeye erişmek için kullanılır.

Erişim sağlamak amacıyla dizin kayıtları yaratmak için otomatik dizinleme yöntemi de kullanılmaktadır. Bu yöntemde bir makina belgelerin karakteristiklerini tanıyarak kaydedecek şekilde programlanır. Otomatik dizinlemede konusal analiz için belgelerin tam metinlerindeki ya da özetlerindeki anahtar sözcük (keyword) ve terimlerden yararlanır. Erişim için yararsız olan ("bir", "bu", "şu", "ve", "veya" gibi) sözcükler dikkate alınmaz. Erişimde kullanılan terimler sözlüğünün hacmini azaltmak için genellikle terimlerin kökleri esas kabul edilir. Bu işlem değişik biçimdeki anahtar sözcüklere erişim sağlamak bakımından yararlıdır.

Dizin kayıtları yaratıldıktan sonra belge veri tabanı kullanıcıların sorgulamasına hazır hale gelir. Dizin kayıtlarını içeren bir veri tabanı yaratarak bir bilgi erişim sistemi tasarılmanın ana nedeni kuşkusuz potansiyel kullanıcıların bilgi gereksinimlerini karşılamaktır. Şimdi de kullanıcıların bilgi erişim sistemlerine nasıl yaklaştıklarına ve sistemde nasıl arama yaptıklarına bir göz atalım.

4. Sorgu Cümlesi Hazırlama Süreci

Bu süreç, sorgulama işleminde kullanılacak olan sorgu cümlesinin oluşturulmasını içermektedir. Sanıldığıının aksine bu, zor bir iştir. İyi tanımlanmış arama cümleleri kullanıcıların bilgili olmalarını gerektirmektedir. Oysa kullanıcılar arama yaptıkları konularda tam olarak ne aradıklarını tanımlayacak kadar bilgili olmayabilirler. Eğer tam olarak ne aradıklarını tanımlayabilselerdi belki de sistemi kullanmalarına gerek kalmayacaktı. Hjerpe bu durumu bilgi erişimin temel paradoksu olarak değerlendirmektedir: “Hakkında bilgi bulmak için bilmediğin bir şeyi tanımlama gereği” (Hjerpe, 1986; Larson, 1991: 147’den aktarma). Bu sorun bir bakıma “sözlük” sözcüğünün anlamını bilmeyen bir kimsenin çaresizliğine benzetilebilir.

Bilgi erişim sistemlerini ilk kez kullananlar genellikle ihtiyatlı davranıp sisteme kapsamlı arama sorguları girerler. Bu kullanıcılar, aramanın başlangıcında veri tabanının özelliklerini (örneğin, kayıt sayısı ve dermenin hangi konularda yoğunlaştığı gibi) bilmediklerinden bilgi erişim sistemiyle ilgili kafalarındaki zihinsel modelleri gerçeğe bağdaştırmaya çalışırlar. Bazen de bunun tersi olur ve kullanıcılar sisteme çok ayrıntılı sorular yöneltebilirler. Ne kadar dar ya da kapsamlı olursa olsun, kullanıcılar bilgi erişim sisteminin her türdeki sorgulamaları yanıtlaması gerektiğini düşünürler.

Çizim 1’den de görüleceği gibi, sorgu cümlesi hazırlama aşamasında sözlüklerden, kavramsal dizinlerden, basılı el kitaplarından ve konu başlıkları listelerinden yararlanılabilir. Ayrıca, bazı sistemlerde bu aşamada kullanıcıların işini kolaylaştırmak amacıyla çevrimiçi yardım sunulmaktadır.

5. Sorgu Cümlesi

Kullanıcı, bilgi isteğini doğal dil kullanarak tanımladıktan sonra sıra sorgu cümlesinin sisteme girilmesine gelir. Bu sorgu cümlesinin söz dizimi (sentaks) sistemden sisteme değişebilir. Kullanıcılar sorgu cümlelerini girmek için çoğu kez komut ve sorgu dillerinin katı söz dizimi kurallarına uymak zorundadırlar. Öte yandan, bu tür sistemlerin çok azı doğal dilde girilmiş sorgu cümlelerini kabul ederler.

Sorgu cümlesi hazırlamak kolay bir iş değildir. Kullanıcılar komut dillerinin varlığından haberdar olmak ve gerekli komutları bilmek zorundadırlar. Ayrıca, kullanıcıların, her sorgu dilinin özelliklerine göre sorgu cümlesinin nasıl oluşturulduğuna ilişkin entellektüel bir kavrayışa sahip olmaları gereklidir. Örneğin, Boole mantığını kullanarak karmaşık sorgu cümleleri oluştururken çoğu kullanıcılar zorlanmaktadır.

6. Kullanıcı Arabirimi

Her sistemde kullanıcıların girdiği sorgu cümlelerini kabul ederek bu sorgulamaları arama ve erişim sisteminin “anlayabileceği” biçime dönüştüren bir kullanıcı arabirimi vardır. Bir başka deyişle, sistemle kullanıcı arasındaki iletişim kullanıcı arabirimi aracılığıyla gerçekleşir.

Bir kullanıcı arabiriminin temel işlevleri daha ayrıntılı olarak şöyle sıralanabilir: a) kullanıcıların, ya doğal dili veya sistemce sağlanan sorgu dilini kullanarak sorgu cümlelerini girmelerine olanak sağlamak; b) kullanıcının girdiği sorgu cümlesini değerlendirmek (örneğin, sorgu cümlesinin sorgu dili söz

dizimine uygunluğunu kontrol etmek, anahtar sözcüklerin köklerini bulmak gibi); c) kullanıcının girdiği sorgu cümlesini bilgi erişim sisteminin anlayacağı biçime dönüştürmek ve bunun sisteme girilmesini sağlamak; d) erişim sonuçlarını göstermek; e) erişilen kayıtların ilgili olup olmadıklarına ilişkin kullanıcıdan geribildirim (feedback) elde etmek ve buna göre özgün sorgu cümlesini yeniden değerlendirmek; ve, f) sistem, veri tabanı ve sistemin kullanımı hakkında yararlı bilgiler sağlamak.

Sorgu cümlelerini tanımlayarak bilgi erişim sistemine girmek ve sistemi kullanmak için kullanıcıların yararlanabilecekleri birçok kullanıcı arabirimi türü vardır: mönüye veya komut dillerine dayalı kullanıcı arabirimleri, grafik kullanıcı arabirimleri, boşluk doldurma türü kullanıcı arabirimleri gibi (Shneiderman, 1986; Bates, 1989). Kuşkusuz, günümüz bilgi erişim sistemlerinde bazı kullanıcı arabirim türleri (mönü, komut dilleri) diğerlerine oranla daha yaygın olarak kullanılmaktadır. Konuşmayı (ses) girdi olarak kabul eden kullanıcı arabirimi türü ise yeni yeni kullanılmaya başlanmıştır. Ancak hangi tür kullanıcı arabirimi türü olursa olsun, bir kullanıcı arabiriminden yararlanmak gerektiğinde her zaman için bir “öğrenme eğrisi” söz konusu olacaktır. Bir başka deyişle, bilgi erişim sistemleriyle başarılı bir şekilde iletişim kurarak sorgulama yapmak ve istenen bilgilere erişim sağlamak için kullanıcıların arabirimlerin nasıl kullanıldığını öğrenmeleri gerekmektedir.

Görüldüğü gibi, kullanıcı arabirimi veri tabanındaki zengin bilgilere erişim sağlayan bir araçtır. Kullanıcıları ilgilendirdiği kadarıyla bu araç, kullanıcıların deneyimlerine, istedikleri bilgi miktarına, veri tabanının ya da sorgu dilinin karmaşıklığına bakmaksızın herkesin bu bilgi kaynaklarından yararlanmasına olanak sağlamalıdır. Mooers yasası kullanıcı arabirimleri için de geçerlidir: “Bir kullanıcı için bilgi edinmek bu bilgiye sahip olmaktan daha zahmetli ve sıkıntı verici hale geldiği anda bilgi erişim sistemleri *kullanılmamaya* başlar” (Mooers, 1960: ii).

“Kötü bir biçimde tasarlanmış bir kullanıcı arabirimini kullanmak, kullanıcılar için bu arabirimi kullanmamaktan daha zahmetli ve sıkıntı verici hale geldiği anda bilgi erişim sistemleri *kullanılmamaya* başlar” şeklinde bir yargıya varmak belki de fazla abartma sayılmamalıdır.

7. Erişim Kuralları

Bilgi erişim sürecinde belirleyici olan nokta, kullanıcıların girdiği sorgulama terimlerinin, erişim sağlamak amacıyla, sistem tarafından yorumlanmasıdır. Kullanıcıların sorgu cümlelerinde yer alan terimlerle belgelerin dizin kayıtlarında yer alan terimler karşılaştırılır. Arama sonucunun kalitesinin büyük ölçüde bu çakışma işleminde kullanılan erişim kurallarına bağlı olduğu açıktır. Çünkü hangi kayıtlara erişilip hangi kayıtlara erişilmeyeceğini erişim kuralları belirler. Ancak, ilk olarak erişim kurallarının uygulanmasından önce gelen bir yöntemi, *kümeleme yöntemi*ni incelemekte yarar vardır.

7.1. Bilgi Erişim Sistemlerinde Kümeleme Yönteminin Kullanılması

Daha önce gerçekleştirilen bilgi erişim deneylerinde erişimden önce belgeleri kümelemenin / sınıflandırmanın daha etkili bir yöntem olacağı belirtilmiştir. Buna göre, eğer birbirlerine konu yönünden benzer belgeleri kümelemek mümkün olabilirse, o zaman dermedeki tüm ilgili belgelere erişmek için sorgu cümlesinin gösterimiyle (representation) sadece benzer kümelerin gösterimlerini karşılaştırmak yeterli olurdu. Bir başka deyişle, sorgu cümlesinin gösterimiyle dermedeki her bir belgenin gösterimini karşılaştırma gereği kalmayacaktı. Bu da, kuşkusuz, daha az işlem gerektireceğinden ve erişim hızını artıracığından çekici gözüktü.

Van Rijsbergen (1979: 45) kümeleme yöntemini destekleyen temel varsayımı -ki bunu "kümeleme hipotezi" olarak adlandırmaktadır- şöyle vurgulamaktadır: "birbiriyle yakından ilişkili belgelerin aynı sorgulamalar için ilgili olarak kabul edilme eğilimi vardır." Kümeleme hipotezinin doğruluğu kanıtlanmıştır. Bir bilgi erişim sisteminin etkinliğinin belge kümeleme yöntemlerinin yardımıyla benzer belgeleri gruplayarak artırılabilceği deneylerle kanıtlanmıştır (Van Rijsbergen, 1979). Belge kümeleme yöntemleri hem ekonomik olup hem de belli bir sorgulama için erişilen belge sayısını artırmaktadır. Belgeler bir kez kümelendikleri zaman bu belgelere tek tek erişim sağlamak gerekmemektedir. Kümelemenin bilgi erişim sistemlerinde kullanımı hakkında Van Rijsbergen (1979) ve Salton (1971b) daha ayrıntılı bilgi vermektedirler.

"Küme" burada birbirine benzeyen belgeler grubu anlamına kullanılmaktadır. Belli bir kümede yer alan belge sayısı ilgili dermenin özelliklerine ve kullanılan kümeleme formülüne bağlıdır. Çok çeşitli konularda belgeler içeren dermelerde daha küçük kümeler oluşma eğilimi gözlenirken, belli bir konuda belgeler içeren dermelerde sayıca daha az ancak daha büyük kümeler ortaya çıkmaktadır. Kümeleme için kullanılan formül küme sayısını ve büyüklüğünü etkileyebilir. Örneğin, kütüphane ve bilgi araştırmaları konusundaki 30.000'in üzerindeki bir derme için 8400 civarında küme yaratılmıştır (Larson, 1989).

Belge kümeleme, kümelenecek belgeler arasındaki benzerliğin ölçülmesine dayanır. Geçmişte, Kosinüs, Dice ve Jackard katsayıları gibi değişik benzerlik ölçümlerine dayalı birçok kümeleme formülü geliştirildi (Salton ve McGill, 1983; Van Rijsbergen, 1979). Kitap adlarındaki terimler, konu başlıkları ve belgelerin tam metinleri birbiriyle yakından ilişkili belgeleri kümelemek için en sık olarak kullanılan 'nesne'lerdir. Bir başka deyişle, adlarında aynı anahtar sözcükler bulunan ya da benzer konu başlıkları verilmiş iki belge varsa, bir kümeleme formülü bu iki belgeyi bir araya getirir/getirmelidir.

Daha yakın bir tarihte Larson (1991) benzer belgeleri kümelemek için sınıflama numaralarını başarıyla kullanmıştır. Larson'a göre erişim sağlamak amacıyla bilgi erişim sistemlerinde sınıflama numaralarının kullanımı sınırlıdır. Bir belgeye verilen sınıflama numarası genellikle herhangi bir anahtar sözcük gibi görülmektedir. Arama işlemi sırasında benzer sınıflama numarası verilmiş belgelere tek tek bakılmaktadır. Oysaki benzer ya da aynı sınıflama numarası verilmiş belgeler büyük bir olasılıkla aynı sorulara yanıt olabilir. "Sınıflama, tıpkı konu başlıkları gibi, bir yapıt hakkında konu başlıklarında açıkça belirtilmemiş olan, konusal bakış açısı sağlar" (Larson, 1991: 152; ayrıca bkz. Chan, 1986, 1989; Svenonius, 1983; Shepherd, 1981, 1983). Kullanıcıların raflarda arama yaparken sergiledikleri arama davranışları da yukarıdaki düşünceleri desteklemektedir. Sınıflama bilgisi günümüzdeki çevrimiçi kataloglarda daha akıllı bir biçimde kullanılabilir (Hancock-Beaulieu, 1987, 1990).

“Sınıflamaya dayalı kümeleme yöntemi” erişim sürecindeki etkinliği artırmak için kullanılabilir. Bir dermedeki belgelere sınıflama numarası verilmişse, aynı numara verilmiş olan belgeler en sık kullanılan konu başlıklarıyla birlikte biraraya getirilebilir. Böylece, ne zaman bir sorgu cümlesi bu kümedeki belgelerin gösterimiyle çıkarsa bu belgelere tek bir grup olarak erişilir.

7. 2. Erişim Kurallarının İncelenmesi

Sorgu cümlesindeki terimlerle dizin terimleri arasında çakışma olup olmadığı; belirlemek için kullanılan birçok erişim kuralı vardır. Blair (1990) 12 değişik erişim kuralı (“model”) listelemekte ve her bir kuralı ayrıntılı olarak açıklamaktadır.¹ Blair’in ayrıntılı olarak açıkladığı erişim kurallarının kısa bir özeti Tablo 1’de verilmektedir. Bu tablo Blair’in (1990) yapının ikinci bölümünden derlemiştir.

Tablo 1’de listelenen erişim kuralları kabaca üç grup altında sınıflandırılabilir: 1) sorgu cümlesindeki terimlerle dizin terimleri arasında kesin çakışma (exact match) gerektiren erişim kuralları ve Boole erişim kuralları (model 1-4, 9-12); 2) olasılık kuramına dayalı erişim kuralları (model 5-7); ve 3) vektör uzayı modeli (model 8).

İlk grupta sorgu cümlesindeki terimler ve dizin terimleri ikilidir (binary). Yani bir terim sorgu cümlesinde (ya da belgenin dizin kaydında) ya vardır ya da yoktur. Erişim için her terim eşit derecede önem taşır. Birden çok terimden oluşan sorgu cümleleri için eşik değerleri oluşturulabilir (model 3 ve 4). Sorgu cümlesindeki terimler kavramsal dizinlerden alınan ilgili terimlerle genişletilebilir (model 11 ve 12). Erişilen kayıtlar kabaca (‘erişildi’ ya da ‘erişilmedi’ şeklinde) sıralanabilir (model 1-3, 12). Ya da erişilen kayıtlar sorgu cümlesinde ve dizin kaydında mevcut çakışan terim sayısına göre sıralanabilir (model 4). Sorgu cümlesindeki terimler arasındaki ilişkiler Boole mantığı kullanılarak tanımlanabilir (örneğin, dizin terimlerinde sadece A ve B terimleri olan belgelere eriş) (model 9 ve 10). Boole mantığına dayalı modelin “bilgisayara dayalı bilgi erişim sistemleri için en gözde erişim tasarımı” olduğu sanılmaktadır (Blair, 1990: 44).

İkinci gruptaki erişim kurallarında sorgu cümlesindeki terimlerin (model 5), dizin kaydındaki terimlerin (model 6), ya da hem sorgu cümlesindeki hem de dizin kaydındaki terimlerin (model 7) ağırlıklı (weighted) olması gereklidir. Bir başka deyişle, erişim amacıyla kullanılacak bir terimin erişim için ne kadar önemli olduğu kullanıcı tarafından belirlenebilir. Erişilen kayıtlar sorgu cümlesindeki terimlerle dizin kaydındaki terimler arasındaki çakışmanın gücüne göre sıralanır. Bu kategorideki erişim kuralları olasılık kuramına dayalı erişim modelleri olarak bilinir.

Üçüncü gruptaki vektör uzayı modeli (model 9) bir bakıma model 7’ye benzemektedir. Çünkü hem sorgu cümlesindeki terimler hem de dizin kaydındaki terimler ağırlıklıdır. Ancak bu modelde gerek sorgu cümlesindeki gerekse dizin kaydındaki terimler n -boyutlu bir uzaydaki vektörler olarak işlem görür. Terimler arasındaki çakışmanın gücü (yani sıralama) sorgu cümlesi ve dizin kaydındaki terimler arasındaki açının kosinüsü hesaplanarak belirlenir. Vektör uzayı modelini kullanan erişim sistemleri 1960’lı yılların başından beri kullanılmaktadır (örneğin, SMART sistemi).

¹ Erişim teknikleriyle ilgili yetkin bir çalışma için ayrıca bkz.: Belkin ve Croft (1987).

Tablo 1. Erişim Kurallarının Özeti
Kaynak: Blair (1990). İkinci Bölüm.

<i>Model</i>	<i>Arama isteği</i>	<i>Belgeler</i>	<i>Erişim Kuralı</i>
1	Tek sorgu terimi	Belgelere bir ya da daha fazla dizin terimi verilir	Eğer sorgudaki terim belgeye verilen terimlerle çıkışıyorsa o zaman bu belgeye erişilir
2	Sorgu terimleri seti (yani birden fazla)	Dizin terimleri seti	Sorgudaki tüm terimler belgenin dizin kaydında da varsa belgeye erişilir
3	Sorgu terimleri seti atı bir de "eşik" değeri	Bir veya daha fazla dizin terimleri seti	Eşik değerini aşan sayıda terim sorgudaki terimlerle çıkışıyorsa belgeye erişilir
4	Model 3 ile aynı	Model 3 ile aynı	Belirlenen sayıda terimden daha fazla çıkışma gösteren belgeler çıkışma sırasına göre azalarak belirlenir
5 Ağırlıklı sorgular	Her biri pozitif değere sahip sorgu terimleri seti	Model 3 ile aynı	Belgeler hem sorgu cümlesindeki hem de dizin kaydındaki terimlerin toplam ağırlıklarına göre azalarak sıralanır
6 Ağırlıklı Dizinleme	Sorgu terimleri seti	Her biri pozitif değere sahip dizin terimleri seti	Belgeler hem sorgu cümlesindeki hem de dizin kaydındaki terimlerin toplam ağırlıklarına göre azalarak sıralanır
7 Ağırlıklı Sorgular ve Dizinleme	Model 5 ile aynı	Model 6 ile aynı	Belgeler sorgudaki terimin ağırlığıyla aynı terimin dizin kaydındaki ağırlığının çarpımlarının toplamına göre sıralanır
8 Kosinüs Kuralı	Model 5 ile aynı	Model 6 ile aynı	Hem sorguda hem de bir dizin kaydında mevcut ortak terimlerin ağırlıkları aynı vektörler olarak alınır. Erişilen bir belgenin değeri iki vektör arasındaki açının kosinüsüdür.
9 Boole Mantığına Göre Hazırlanmış Sorgu Cümleleri	Sorgu cümleleri sorgu terimlerinin VE, VEYA veya DEĞİL gibi Boole işleçlerinin (operators) bileşiminden oluşur	Bir ya da daha fazla dizin terimleri seti	i) VE: Sadece sorgu cümlesindeki terimlerle tümüyle çıkan belgelere eriş ii) VEYA: Sorgu cümlesindeki herhangi bir terimle çıkan belgelere eriş iii) DEĞİL: Sorgu cümlesinde DEĞİL işleciyle verilen terimler dışındaki tüm belgelere eriş
10 Tam Metin Erişim	Model 9 ile aynı	(Erişimsiz sözcükler hariç) belgelerin tam metinleri üzerinde arama yapılabilir	Model 9 ile aynı. Yakınlık işleçleri (proximity operators) de kullanılabilir
11 Basit Kavramsal Dizin	Tek terimler	Bir ya da daha fazla dizin terimleri seti	(Çevrimiçi) bir kavramsal dizinden bakılarak sorgu cümlesindeki terime anladış terim(ler) eklenir
12 Ağırlıklı Kavramsal Dizin	Tek terimler	Bir ya da daha fazla dizin terimleri seti	(Çevrimiçi) bir kavramsal dizinden bakılarak sorgu cümlesindeki terime belli bir eşik değerinin üzerindeki (ağırlıklı) anladış terimler ayrı olarak (disjunctively) eklenir. Eşik değeri kullanıcı tarafından belirlenebilir.

Buraya dek geleneksel bir bilgi erişim sisteminin temel bileşenlerini belge veri tabanı, sorgu cümlesi hazırlama, kullanıcı arabirimi ve erişim kuralları açısından inceledik. Hangi erişim kuralı kullanılırsa kullanılsın bir bilgi erişim sisteminin son amacı kullanıcının bilgi gereksinimlerini en iyi şekilde karşılayacak kayıtlara erişmektir. Bu nedenle, kullanıcı için en önemli nokta elde edilen sonuçlardır (yani erişimin etkinliği). Aşağıda erişim etkinliğiyle ilgili başlıca ölçüler (measures) incelenmektedir.

8. Erişim Etkinliği Ölçüleri

Bilgi erişim sistemlerinin etkinliğini değerlendirmek amacıyla birçok değişik ölçü kullanılmaktadır. Erişim etkinliğinin ve başarısızlıkların araştırılmasında yaygınlıkla kullanılan **erişim isabeti** ve **kesin isabet**² gibi birkaç ölçü aşağıda tartışılmaktadır. Literatürde önerilen diğer erişim etkinliği ölçüleri pek seyrek kullanıldığından burada incelenmeyecektir.

Çevrimiçi bilgi erişim sistemleri çoğu kez ilgisiz belgelere erişirken aynı zamanda ilgili olan bazılarını da erişememektedir. Blair (1990: 73-74) erişim sürecini şöyle özetlemektedir:

Bilgi erişim temelde bir deneme-yanılma süreci olduğundan, bilgi erişim sistemindeki belgelere erişmek için yapılan hemen hemen her aramanın, yararlı (yani ilgili) belgelerin yanı sıra değişen oranlarda yararsız (yani ilgisiz) belgelere erişmesi beklenbilir. Arama sürecindeki bu belirsizliğin yol açtığı bir diğer sorun da şudur: bir veri tabanından yararlı belgelere erişilse bile, kullanıcının çok ısrarlı çabalarına karşın, daha birçok yararlı ye erişilemeyebilir. Sonuç olarak, herhangi bir aramadan sonra, veri tabanındaki belgeler dört değişik biçimde sınıflandırılabilir:

Erişilen ve ilgili (yararlı)
Erişilen ve ilgisiz (yararsız)
Erişilmeyen ve ilgili [kaçan]
Erişilmeyen ve ilgisiz.

Blair bu dört sınıfta yer alan belgeleri aşağıdaki çizimle göstermektedir:

Çizim 2. Arama Sonucunun (Çıktının) Gösterimi
(Kaynak: Blair, 1990: 76)

	İLGİLİ	İLGİSİZ	
ERİŞİLEN	x	u	ERİŞİLEN TOPLAM BELGE SAYISI= n_1
ERİŞİLMEYEN	v	y	
	DERMEDEKİ TOPLAM İLGİLİ BELGE SAYISI= n_2		

Çizim 2'ye dayanarak aşağıdaki erişim etkinliği ölçüleri tanımlanmıştır:

² Bu terimler henüz Türkçede genel kabul görmüş değildir. Bu çalışmada "kesin isabet" ve erişim isabeti", sırasıyla, İngilizcedeki "precision" ve "recall" terimlerinin karşılığı olarak kullanılmıştır (Alkan, 1994: 254-265). Bir başka çalışmada "precision" için "duyarlılık", "recall" için ise "anma" terimleri yeğlenmiştir (Köksal, 1987; ayrıca bkz. Köksal, 1979).

$$Kesin\ isabet = x / n_1$$

$$Erişim\ isabeti = x / n_2$$

$$Posa = u / (u + y)$$

Formüllerde yer alan x , erişilen ilgili belge sayısını; n_1 , erişilen toplam belge sayısını (Çizim 2'de $x + u$); n_2 , dermedeki toplam ilgili belge sayısını (Çizim 2'de $x + v$); u , erişilen ilgisiz belge sayısını; y ise erişilmeyen ilgisiz belge sayısını ifade etmektedir.

Kesin isabet ve *erişim isabeti* ölçüleri bilgi erişim sistemlerinde erişim etkinliğini değerlendirmede genellikle birlikte kullanılır. "*Kesin isabet*, erişilen ilgili belge sayısının erişilen toplam belge sayısına oranıdır" (Van Rijsbergen, 1979: 10). Örneğin, belirli bir sorgulama için bilgi erişim sisteminin iki kayda erişim sağladığını düşünelim (n_1). Kullanıcının erişilen bu iki kayıttan bir tanesini ilgili (x) bulunduğunu varsayalım. Bu durumda bu arama sorusu için kesin isabet oranı %50 olurdu (x / n_1).

Erişim isabeti değerini hesaplamak kesin isabet oranını hesaplamaktan çok daha güçtür. Çünkü erişim isabeti oranını bulmak için kullanıcının ilk aramasında erişilemeyen diğer ilgili kayıtlara da erişilmesi gerekmektedir (Blair ve Maron, 1985: 291). "*Erişim isabeti* erişilen ilgili belge sayısının dermedeki toplam (hem erişilen hem de erişilemeyen) belge sayısına oranıdır (Van Rijsbergen, 1979: 10). Yukarıda verdiğimiz örneği ele alalım. Kullanıcı erişilen iki kayıttan birisini ilgili buldu. Diyelim ki daha sonra dermede ilk aramada erişilemeyen üç ilgili kayıt (v) daha bulundu. Yani sistem, veri tabanındaki dört (n_2) ilgili kayıttan sadece bir tanesine erişim sağlamış oldu. Bu durumda bu arama sorusu için erişim isabeti oranı %25'e eşit olurdu (x / n_2).

Blair ve Maron (1985: 290) erişim isabeti ölçüsünün, sistemin tüm ilgili belgelere ne derece sağlıklı erişim sağladığını, öte yandan kesin isabet ölçüsünün ise, sistemin salt ilgili belgelere ne derece sağlıklı erişim sağladığını ölçtüğünü söylemektedirler.

Posa, erişim etkinliğini değerlendirmede kullanılan bir diğer ölçüdür. *Posa*, erişilen ilgisiz kayıtların (u) dermedeki tüm ilgisiz kayıtlara ($u + y$) oranı olarak tanımlanabilir. *Posa* oranı "bir sistemin ilgisiz belgeleri ne derece sağlıklı olarak reddettiğini ölçer" (Blair, 1990: 116). Daha önce kullandığımız örneği *posa* oranını açıklamak için de kullanabiliriz. Kullanıcının erişilen iki kayıttan bir tanesini ilgili bulunduğunu varsaymıştık. Daha sonra yapılan aramada ilk aramada erişilemeyen üç kayıt daha belirlenmişti. Bu dermenin tamamında dokuz kayıt (dördü ilgili beşi ilgisiz) olduğunu varsayalım. Kullanıcı dermedeki toplam beş ilgisiz kayıttan ($u + y$) bir tanesine (u) eriştiğinden, bu arama için *posa* oranı %20 olarak gerçekleşir ($u / (u + y)$).

9. İlgililik Geribildirimi Kavramları

Çalışmamızın bundan önceki bölümlerinde, bir bilgi erişim sisteminde kullanıcıların sistemle etkileşimine olanak veren bir tür kullanıcı arabiriminin varlığına (bkz. altbölüm 2) ve bu arabirimin temel işlevlerine (bkz. altbölüm 6) değinmiş, bu işlevlerden birisinin de kullanıcı ile bilgi erişim sistemi arasında değişik biçimlerde geribildirimi mümkün kılması gerektiğini vurgulamıştık.

Kullanıcılar aradıklarını nadiren bir kerede bulduklarından, geribildirim işlevine daha ayrıntılı olarak değinmekte yarar vardır. Erişim kuralları kendi başına (bizatihi olarak) erişilen kayıtların kullanıcının istediği kayıtlar olmasını garanti edemez. Kullanıcı arabirimi, kullanıcının bir sonraki aşamada ne yapması gerektiğini söyler ya da sistem tarafından yaratılan geribildirim mesajlarıyla alternatif stratejiler önerebilir (örneğin, yardım ekranları, aramanın statüsü, ne yapılması gerektiği vb. gibi). Daha da önemlisi, sistem, erişilen örnek kayıtlar ışığında kullanıcıya, sorgulamasında değişiklik yapma olanağı vererek bir sonraki erişim işleminde başarı düzeyinin daha da artmasını sağlar (Van Rijsbergen, 1979). Bazı sistemler ilk denemede erişilen kayıtların kullanıcı tarafından ilgili bulunup bulunmadığına bakarak özgün arama sorusunda otomatik olarak bazı değişiklikler yapabilir. Bu durum "ilgililik geribildirimi" (relevance feedback) olarak bilinmektedir. şimdi ilgililik geribildirimi sürecini biraz daha yakından inceleyelim.

Swanson (1977) literatüre geçmiş bazı bilgi erişim deneylerini ve bu deneylerde kullanılan ölçüleri incelemiştir. Swanson'a göre bilgi erişim sistemleri deneme-yanılma sürecini kolaylaştıracak ve sorgulamanın düzeltilmesine olanak verecek biçimde tasarlanmalıdır (s. 142). Van Rijsbergen (1979) de aynı görüşü paylaşmakta ve otomatik bir bilgi erişim sistemiyle karşılaşan bir kullanıcının bir kerede bilgi gereksinimini tanımlayabilmesinin pek mümkün olmadığını belirtmektedir. Van Rijsbergen'e göre, kullanıcı, sorgulama hakkında edindiği geribildirim ışığında sorgulamasını hazırlayacağı bir deneme-yanılma sürecine girmek isteyecektir (s. 105).

Van Rijsbergen (1979), sorgulamalarını hazırlarken kullanıcılara yardımcı olabilecek türden bilgileri de listelemektedir (kullanıcıların sorgu terimlerinin veri tabanında olup olmadığı, belirli bir arama için küçük bir örneklem alınarak erişilebilecek kayıt sayısının belirlenmesi, ve, daha etkin tarama sonuçları elde etmek için kullanılması gereken ilgili arama terimlerinin kullanıcıya bildirilmesi gibi).

İlgililik geribildirimi, ilk aramada elde edilen sonuçlar ışığında, kullanıcıya sorgu cümlesinde etkileşimli olarak değişiklik yapma olanağı vererek deneme-yanılma sürecini kolaylaştıran tekniklerden birisidir. Aşağıdaki alıntı ilgililik geribildirim sürecini çok güzel özetlemektedir:

Çoğu bilgi sistemi kullanıcıları için ilk soru cümlesi hazırlama (formulate) sürecinin saydam olmadığı çok iyi bilinir. Çoğu kullanıcılar, dermenin yapısı ve erişim ortamı hakkında ayrıntılı bilgi sahibi olmaksızın, erişim amacıyla çok iyi tasarlanmış arama soruları hazırlamakta özellikle zorlanmaktadırlar. Bu durum, ilk erişim işleminin...sadece bir deneme olarak görülmesi gerektiğine işaret etmektedir. Daha sonra, başlangıçta erişilen bu kayıtların ilgili olup olmadığı incelenerek, bunu izleyen arama işlemleri sırasında diğer yararlı kayıtlara erişmek umuduyla, yeni ve geliştirilmiş sorular hazırlanabilir (Salton ve Buckley, 1990: 288).

SMART bilgi erişim deneylerinde ilk kez ilgililik geribildirimi tekniğinin kullanılmasından bu yana neredeyse 25 yıl geçmiştir (Salton, 1971b). İlk ilgililik geribildirim deneyleri erişim performansının çok yüksek olduğu küçük dermeler üzerinde gerçekleştirildi (Rocchio, 1971a; Salton, 1971a, 1971b).³ İlgililik geribildirim tekniğinin erişim performansını önemli ölçüde artırdığına tanık olundu. Salton ve Buckley (1990) deneysel amaçlarla değişik konularda altı ayrı belge dermesi

³ Çevrimiçi kataloglarda ilgililik geribildirim tekniğinin kullanımı hakkında bkz.: Porter ve Galpin, 1988; Walker ve de Vere, 1990; Larson, 1989, 1991; Walker ve Hancock-Beaulieu, 1991).

kullanarak 12 farklı geribildirim yöntemini inceleyip değerlendirdiler. Deneylerinde kullandıkları dermelerin büyüklükleri 1400 ile 12.600 belge arasında değişmekteydi. İlgililik geribildirim yöntemlerinin erişim performansını %47 ile %160 arasında artırdığı gözlemlendi.

İlgililik geribildirimi tekniği hakkında daha ayrıntılı bilgi ve formüller ve bu tekniğin matematiksel açıklaması Salton ve Buckley (1990), Rocchio (1971) ve Ide'nin (1971) makalelerinde bulunabilir.

10. Sonuç

Bu çalışmada bir bilgi erişim sisteminin temel bileşenleri incelenmiştir. Dizinleme, soru cümlesi hazırlama süreçlerinin önemi tartışılarak kullanıcı arabirimlerinin ve erişim kurallarının bilgi erişimde oynadığı roller açıklanmıştır. Kesin isabet ve erişim isabeti gibi belli başlı erişim ölçüleri örnekler verilerek anlatılmış ve kümeleme ve ilgililik geribildirimi gibi erişim etkinliğini artıran daha ileri bilgi erişim yöntemlerine de kısaca değinilmiştir.

11. Kaynakça

- Alkan, Nazlı. (1994). "Bilgi taramalarının nitelik açısından değerlendirilmesinde 'Kesin İsalet' ve 'Erişim İsalet' oranları," *Türk Kütüphaneciliği* 8(4): 254-265.
- Bates, Marcia J. (1989). "The design of browsing and berrypicking techniques for the online search," *Online Review* 13(5): 407-424.
- Belkin, Nicholas J. ve W. Bruce Croft. (1987). "Retrieval techniques," *Annual Review of Information Science and Technology* 22: 109-145.
- Blair, David C. (1990). *Language and representation in information retrieval*. Amsterdam: Elsevier.
- Blair, David C. ve M.E. Maron. (1985). "An evaluation of retrieval effectiveness for a full-text document-retrieval system," *Communications of the ACM* 28: 289-299.
- Chan, Lois Mai. (1986). "Library of Congress Classification as an online retrieval tool: potentials and limitations," *Information Technology and Libraries* 5: 181-192.
- Chan, Lois Mai. (1989). "Library of Congress class numbers in online catalog searching," *RQ* 28(4): 530-536.
- Hancock-Beaulieu, Micheline. (1987). "Subject searching behaviour at the library catalogue and at the shelves: implications for online interactive catalogues," *Journal of Documentation* 43: 303-321.
- _____. (1990). "Evaluating the impact of an online library catalogue on subject searching behaviour at the catalogue and at the shelves," *Journal of Documentation* 46(4): 318-338.
- Hjerrpe, Roland. (1986). "Project HYPERCATalog: visions and preliminary conceptions of an extended and enhanced catalog," *Intelligent information systems for the information society* içinde. Ed. by B.C. Brookes. New York: North-Holland Publishing Co.
- Ide, E. (1971). "New experiments in relevance feedback." Bkz.: Salton (1971b: 337-354).
- Köksal, Aydın. (1979). "Bilgi erişim sorunu ve bir belge dizinleme ve erişim dizgesi tasarımı ve gerçekleştirimi". (Yayınlanmamış) Doçentlik Tezi. Hacettepe Üniversitesi.
- _____. (1987 Ocak-Mart). "Bilgi erişim sorunu ve bilgi erişim dizgelerine ilişkin temel kavramlar", *Bilişim* s.18-25.
- Larson, Kay R. (1989). "Managing information overload in online catalog subject searching", *ASIS '89 Proceedings of the 52nd ASIS Annual Meeting Washington, DC, October 30-November 2, 1989* içinde (129-135). Ed. by Jeffrey Katzner and others. Medford, NJ: Learned Information.

- _____. (1991). "Classification clustering, probabilistic information retrieval and the online catalog," *Library Quarterly* 61(2): 133-173.
- Maron, M.E. (1984). "Probabilistic retrieval models", Dervin, Brenda and Melvin J. Voigt, (eds.) *Progress in communication sciences*, vol. 5 içinde (145-176). Norwood, NJ: Ablex.
- Mooers, Calvin N. (1960 July). "Mooers Law or, why some retrieval systems are used and others are not," (Editorial) *American Documentation* 11: ii.
- Porter, Martin ve Valerie Galpin. (1988). "Relevance feedback in a public access catalogue for a research library: Muscat at the Scott Polar Research Institute," *Program* 22(1): 1-20.
- Rocchio, J.J.Jr. (1971). "Relevance feedback in information retrieval." Bkz.: Salton (1971b: 313-323).
- Salton, G. (1971a). "Relevance feedback and the optimization of retrieval effectiveness." Bkz. Salton (1971b: 324-336).
- Salton, Gerard. (ed.) (1971b). *The SMART retrieval system: experiments in automatic document processing*. Englewood Cliffs, NJ: Prentice-Hall.
- Salton, Gerard ve M.J. McGill. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.
- Salton, Gerard ve Chris Buckley. (1990). "Improving retrieval performance by relevance feedback," *Journal of the American Society for Information Science* 41(4): 288-297.
- Shepherd, Michael A. (1981). "Text passage retrieval based on Colon Classification: retrieval performance," *Journal of Documentation* 37(1): 25-35.
- _____. (1983). "Text retrieval based on Colon Classification: failure analysis," *Canadian Journal of Information Science* 8: 75-82.
- Shneiderman, B. (1986). *Designing the user interface: strategies for effective human-computer interaction*. Reading, MA: Addison-Wesley.
- Svenonius, Elaine. (1983). "Use of classification in online retrieval," *Library Resources & Technical Services* 27(1): 76-80.
- Swanson, Don R. (1977). "Information retrieval as a trial-and-error process," *Library Quarterly* 47, 128-148.
- Van Rijsbergen, C.J. (1979). *Information retrieval*. 2nd ed. London: Butterworths.
- Walker, Stephen ve R. de Vere. (1990). *Improving subject retrieval in online catalogues. 2: Relevance feedback and query expansion*. British Library Research Paper, no. 72. London: The British Library.
- Walker, Stephen ve Micheline Hancock-Beaulieu. (1991). *Okapi at City: an evaluation facility for interactive information retrieval*. British Library Research Report 6056. London: The British Library.