

Bilgi Taramalarının Nitelik Açısından Değerlendirilmesinde "Kesin İsbet" (KI-Precision) ve "Erişim İsbeti" (EI-Recall) Oranları

Precision and Recall Ratios in Evaluating The Quality of Literature Searches

Nazlı Alkan

Öz

Genellikle her hizmetin değerlendirilmesinde bazı ölçütler vardır. Bilgi tarama ya da bilgi erişim hizmetlerinin değerlendirilmesinde ve başarısının ölçülmesinde de bazı ölçütler geliştirilmiştir. Bunlar nitelik, süre ve maliyet ölçütleridir. Bu ölçütler kullanılmaksızın varılan değerlendirme yargıları güvenilir olmaktan uzaktır. Bu makalede nitelik ölçütlerinden anlamlılık ve onun ölçülmesini sağlayan KI ve EI oranları, değerlendirme ölçütü olarak ele alınmış ve bu oranların ne oldukları, nasıl elde edildikleri ve tarama başarısını belirlemedeki önemleri tartışılmıştır. Bu iki oranın tarama etkinliğinin özünü oluşturan anlamlılığı ölçebilmeleri, onların diğer nitelik ölçütlerinden, tüm süre ve maliyet ölçütlerinden daha öncelikli bir konumda olduklarını düşündürmektedir. Bununla beraber, tüm değerlendirme ölçütlerine kısaca değinilmiş, KI ve EI oranlarının bu bütün içindeki konumları belirlenmeye çalışılmıştır. Bu oranları nitelik ölçütü olarak kullanılmaksızın varılan değerlendirme yargularının yüzeysel olmaktan öteye geçemeyeceği görüşü ile, sunulan bilgi sonuçlandırılmıştır.

Abstract

Criteria by which literature search services may be evaluated, are quality, time and cost criteria. Evaluation judgements are far from being reliable, if not based on these criteria. Relevance, which is the most important quality criteria, is discussed briefly. Precision and Recall ratios, which reflect the numerical value and measure of Relevance, are discussed in some detail. These two criteria frequently used to judge the performance of searches. Significance of Precision and Recall is described and definitions are given. It is considered that these two ratios are the most important performance criteria among all, for they measure Relevance, and Relevance is the most significant component in search effectiveness. For this reason, the present study is limited mainly to the topics of Precision and Recall. However, the text includes some brief description of other criteria as well. In the present study, it is stated as a conclusion that Precision and Recall ratios should be used as Relevancy criteria in judging the performance of literature searches, otherwise all the quality judgements on searches will be false and unreliable.

Giriş

Bilgi tarama hizmetlerinin temel amacının, bilgi ihtiyaçlarını gidermek üzere hizmet isteyen kullanıcıları, istek konularıyla ilgili literatüre en kısa zamanda yönlendirmek olduğu bilinmektedir. Teknolojik gelişmeler,

1960'ların ortalarından itibaren taramaların bilgisayarla yapılmasını sağlamış, hizmetleri başarılı kılan öğelerden hızlı yanıt süresini garantilemiştir. Hızlı yanıt süresinden daha önemli bulunan öğe ise, taramaların ne oranda niteliği yakalayabildiğidir. Nitelik, istek konusu açısından anlamlı olan, yani konuyla ilgili bilgi ihtiyacını karşılayan künyelere ne oranda başarı ile erişilebildiği, taranan veri tabanlarının kapsamının ne olduğu, güncel bilgilerin ne derecede karşılanabildiği, tarama çıktısının yeterlilik derecesi ve harcanan kullanıcı çabasının miktarı ile ilgilidir. Kullanıcıların yakından ilgilendiği bu öğeler, taramaların başarısını ölçen "nitelik ölçütleri"ne, yani anlamlılık, bütünlük (kapsam), yenilik, tarama çıktısı ve kullanıcı çabası ölçütlerine dönüşmüştür. "Nitelik ölçütleri", taramaların hızını ölçen "süre ölçütleri" ve hizmetin ekonomik boyutunu ölçen "maliyet ölçütleri" ile birlikte, hem geleneksel, hem bilgisayar taramanın değerlendirildiği ölçütler olarak kabul edilmişlerdir. Kuşkusuz hizmetin kullanıcıları onu, sadece niteliğine göre değil, aldığı süre ve maliyetine göre de değerlendirme eğiliminde olmuşlardır. Bu eğilim zamanla tarama değerlendirme ölçütlerini yaratmıştır. Böylelikle, hizmetlerde neyin değerlendirileceği sorusunun yanıtı bulunmuştur. Değerlendirilmesi gereken unsurlar, bu üç ölçütün çevresinde yer alırlar.

Değerlendirmenin nasıl yapılacağına gelince; ölçütleri elde etmek üzere toplanan verilerden ve erişim etkinliğini ölçme tekniklerinden yararlanılır⁽¹⁾. Bu veri ölçme teknikleri bazı ifade ve sayısal değerlerin elde edilmesini ve değerlendirme sonucuna ulaşılmasını sağlarlar.

Tarama hizmetlerinin kullanıcı ve sistemler açısından en yüksek düzeyde değerlendirilmesi ve başarısının ölçülmesi ancak bu ölçütler kanalıyla sağlanabilir. Bu ölçütler ele alınmaksızın yürütülen değerlendirme çalışmaları, düşük bir bilimsel değer taşımaktan ileri gidemeyeceği gibi, güvenilir yargılara da ulaştırılmazlar. Bunlar, bir taraftan istek sahiplerinin tarama sistemine ilişkin beklentilerini belirleyen, diğer taraftan onların bilgi ihtiyaçlarını tatmin etmek üzere sistemin yeteneğini ortaya koyan ölçütlerdir. Sadece kullanıcılar açısından değil, hizmeti sunan kurumların yöneticileri, değerlendiriciler, hatta veri tabanı tasarımcı ve üreticileri açısından da önem taşıyan değerlerdir.

Nitelik ölçütleri ilk kez 1964'de Cleverdon⁽²⁾ tarafından tarama değerlendirmelerine kullanılmıştır. Bütünlük (kapsam), anlamlılık (relevancy) ve onun ölçülmesini sağlayan KI ve EI oranları, yanıt süresi, tarama çıktısı ve kullanıcı çabası ölçütleri nitelik ile ilgili bulunmuştur. Bu ölçütler o zamandan beri değerlendirmenin vazgeçilmez unsurları olarak kullanılmaktadır.

lar. Daha sonraki yıllarda yanıt süresi, doğal olarak süre ölçütleri içinde yer almış, yenilik oranı nitelik ölçütlerine eklenmiştir⁽³⁾. Kullanıcı ihtiyaçları ile taramanın yürütüldüğü sistemin başarısını ilişkilendiren bu ölçütler, gelecekteki olsun, bilgisayara dayalı olsun, tarama tekniklerinin değerlendirilmesinde de kullanılarak, güvenilir sonuçlara ulaştırmada güçlü bir rol oynarlar. Nitelik ölçütlerini belirleyerek, bunlardan güvenilir sonuçlar çıkarmak, süre ölçütleri sonuçlarını belirlemekten çok daha zordur. Diğer taraftan tüm nitelik ölçütleri arasında sürekli bir etkileşim vardır. Hepsisi birbirleriyle yakından ilgilidir.

Nitelik yönünden başarının iki temel değerinden ilki, kullanıcının aradığını elde edip etmediği, ikincisi ise aradığını ne derecede tam ve doğru olarak elde ettiğidir. Bu bakımdan "anlamlılık", yani kullanıcının bilgi ihtiyacını karşılayan, onun aradığı konu ile ilgisi olan bilgi parçacıklarını elde etme durumu önem kazanır.

Bu makalenin konusu, bilgi erişim ve tarama hizmetlerinin başarısını değerlendirmenin özünü oluşturan anlamlılık ölçütü ve anlamlılığı esas alarak onun sayısal olarak ifade edilmesini sağlayan KI ve EI oranları ile sınırlandırılmıştır. Bu iki oran, anlamlılığın ölçülmesini sağlayan, tarama başarısını yansıtan temel, standart ve değerli nitelik ölçütleridir⁽⁴⁾.

Anlamlılık

Bu iki oran hakkında ayrıntıya girmeden önce, tarama hizmeti açısından anlamlılık kavramı üzerinde durmanın yerinde olacağı düşünülmüştür. Bilgi erişim sistemleri ve tarama hizmetlerinin temel amacı, kullanıcılarının anlamlı bilgiye erişmelerini sağlamak olduğuna göre, anlamlılık kavramını erişim sistem ve hizmetlerinin başarısını değerlendirmede anahtar unsur olmakta, enformasyon bilimi ve bilgi iletişiminin en önemli kavramlarından biri olarak ele alınmaktadır. Erişilen künyeler istek sahipleri tarafından anlamlı bulunduğu zaman, erişim başarılı olmuş demektir. Bundan dolayı anlamlılık, erişim başarısını değerlendirmede temel nitelik ölçütü olmaktadır⁽⁵⁾. Anlamlılık, kaynağın içeriğindeki bilginin kullanıcının bilgi ihtiyacına ne derecede yanıt verdiğini ortaya koyan bir ölçüdür. Başka bir deyişle, bilgi ihtiyacını belirleyen istekle, kaynağın içerdiği bilgi arasındaki ilişkinin saptanması söz konusudur.

Diğer taraftan anlamlılık, soyut bir kavramdır. Aldatıcı, kaygan ve görel bir yapısı vardır. Kaynak ile kullanıcının ihtiyacı arasındaki ilişkiyi tanımlamak güçtür. Bu tanımlanması güç yapıya rağmen, bilgi taramasının

anlamlılık açısından ne derecede başarılı olduğunu saptamak üzere anlamlılık değerlendirilmesinin yapılması gerekir. Bu değerlendirme, anlamlı olan künyeleri, anlamlı olmayanlardan ayırma işlemine dayanır. Tarama sonuçlarını veren çıktıdan, anlamlılık değerlendirmelerinin isteği yapan kullanıcı tarafından yürütülmesi gerektiği görüşü yaygın olarak benimsenmiştir.

Saracevic⁽⁶⁾ gibi bazı araştırmacılar, değerlendirmede kararların, çeşitli anlamlılık dereceleri göz önüne alınarak verilmesi gerektiğini savunmuşlardır. Buna karşılık, iki dereceli değerlendirmelere göre bir künye ya anlamlıdır, ya da değildir. İki anlamlılık derecesinin kullanılması birçok koşulda geleneksel olmuştur. İki'den fazla seçenekli değerlendirmelerde, kullanıcı çeşitli anlamlılık dereceleri (tamamen, muhtemelen, kısmen veya sınırdan anlamlı gibi) arasında ince ayrımlara zorlandığı için belirsizlikler ortaya çıkmaktadır⁽⁷⁾. Üstelik iki dereceli değerlendirmeler, anlamlılığın ölçülmesini sağlayan Kİ ve Eİ oranlarının elde edilmesinde kolaylık getirmektedir.

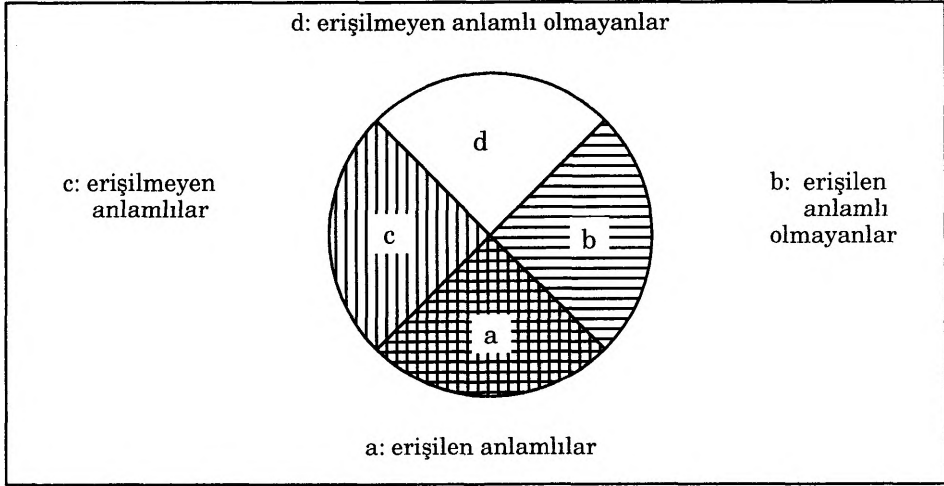
Kİ ve Eİ Oranları

Bilgi tarama hizmetinin başarısını değerlendirmede yaygın olarak kullanılan bu iki önemli nitelik ölçütü, ilk kez Kent⁽⁸⁾ tarafından önerilmiştir. Cleverdon ise bu oranları tarama değerlendirmesinde kullanılan ve yaygınlaştıran ilk kişidir. Bir taramanın başarısını sergilemede birlikte kullanılmaları gereken bu iki oranı, yüzde olarak elde etmek mümkündür.⁽⁹⁾

$$Kİ: \frac{\text{Erişilen anlamlı künye sayısı}}{\text{Erişilen künye sayısı}} \times 100$$

$$Eİ: \frac{\text{Erişilen anlamlı künye sayısı}}{\text{Veri tabanındaki tüm anlamlı künye sayısı}} \times 100$$

Kİ, taramanın eriştiği künyelerden ne kadarının istek yapılan konu açısından anlamlı olduğunu, Eİ ise veri tabanındaki tüm anlamlı künyelerden ne kadarına erişildiğini ifade eden oranlardır. Bu oranları başka bir yaklaşımla ifade edebilmek için koleksiyonu ya da veri tabanını şekil 1'de görüldüğü gibi dört bölüme ayırmak gerekir. Şekil 1'deki ilk ayırım (tek çizgi), tarama sonucunda erişilen ve erişilemeyen künyeleri, ikinci ayırım (çift çizgi) ise anlamlı künyeleri ve anlamlı olmayanları birbirinden ayırmaktadır.

Şekil 1: Koleksiyon/Veri Tabanının Kısımlara Ayrılması⁽¹⁰⁾

Bu şekle göre a+b erişilenler (yatay çizgili alan), c+d erişilemeyenler, a+c anlamlılar (dikey çizgili alan), b+d ise anlamlı olmayanlardır.

Kİ ve Eİ oranlarını yüzde olarak elde etmek üzere kullanılan formüllerin şekil 1'e göre ifadesi şu şekildedir:

$$Kİ: \frac{a}{a+b}$$

$$Eİ: \frac{a}{a+c}$$

Bu ifadelerdeki a, erişilen anlamlı künye sayısını, a+b anlamlı olsun veya olmasın tüm erişilenleri, a+c ise erişilsin veya erişilmesin tüm anlamlıların göstermektedir.

Bilgi taramasının yürütüldüğü sistem, bir yandan konu dışı künyeleri reddetme, bir yandan da konu ile ilgili künyelerin büyük bir kısmına erişme yeteneğine sahip olmalıdır. Süzgece benzetilen tarama sistemleri, bütün süzgeçlerde olduğu gibi, istenmeyenleri süzgeçten geçirmeme, istenenleri geçirme yeteneğini sergileyebilmelidir. Eİ, sisteminin konu açısından anlamlı künyelere erişme yeteneğini, Kİ ise, tersine, anlamlı olmayan künyeleri reddetme yeteneğini ölçer⁽¹¹⁾.

Oranlardan sadece biri, tarama etkinliğini veremez. İkisi birlikte bu etkinliği ortaya koyar. Bir taramanın %70 oranında Eİ, %50 oranında Kİ ile sonuçlandığını düşünelim. Başka bir tarama ise yine %70 oranında Eİ, %14 oranında Kİ ile sonuçlanmıştır. İlk tarama daha başarılıdır. Çünkü ikinci

tarama anlamlı olmayan künyeleri yeterince reddedememiştir. %50 oranında Kİ sonucu veren ilk tarama, çok daha fazla bir süzme yeteneği ortaya koymaktadır⁽¹²⁾.

Kİ ve Eİ oranları, kullanıcının anlamlılık değerlendirmeleriyle, sistemin anlamlılık tahmininin ne derecede uyum içinde olduğunu verirler. Mükemmel bir taramada iki anlamlılık değeri tam olarak birbirlerine denk düşerler. Bu durumda hem Kİ, hem de Eİ oranı yüksektir. Her iki oranın %100 olduğu durumda en mükemmel tarama gerçekleştirilmiş olur. Ne yazık ki, bu tip taramalar çok enderdir. En yaygın olarak rastlanan sonuçlar, belli derecelerde Kİ ve Eİ oranlarını veren sonuçlardır. Anlamlı künyelerin çoğuna erişildiğini, ancak hepsine erişilemediğini, diğer taraftan anlamlı olmayan künyelerin çoğunun atıldığını, fakat tümünün reddedilemediğini gösteren sonuçlara yaygın olarak rastlanır⁽¹³⁾.

Diğer taraftan Kİ ve Eİ oranları ters bir ilişki içindedirler. Genellikle daha yüksek Eİ oranı elde edebilmek üzere tarama konusu genişletildiğinde Kİ oranı düşer. Tersine, bir taramanın Kİ oranını yükseltmek amacıyla, o taramanın kapsamı daraltıldığında Eİ oranı bozulur. İki oran arasındaki bu çelişki, bu karşıt ilişki kaçınılmazdır⁽¹⁴⁾.

Değişik kullanıcıların farklı Kİ ve Eİ oranlarına ihtiyaçları olabilir. Uzun süreli bir araştırma projesine başlayan veya kitap yazan bir kullanıcı, ilgilendiği konuyla ilgili hiçbir künyeyi kaçırmadığını garantilemek ister. Böyle bir kullanıcı birkaç künye ile yetinmeyecek ve yüksek Eİ oranını, yüksek Kİ oranına tercih edecektir. Diğer taraftan, konusuyla ilgili birkaç anlamlı künyeyi hemen isteyen bir kullanıcı için yüksek Kİ oranı daha önemlidir. Aslında tarama isteğinde bulunanların büyük çoğunluğu bu iki uçta değildir. Kabul edilebilir düzeyde Kİ ve Eİ oranları çoğunu memnun eder. Zaten taramaların çoğu da belli derecede Kİ ve Eİ ile sonuçlanmaktadır.

Taramaların Kİ ve Eİ oranlarını bulmanın ilk adımı, tarama sonucu erişilen künyeleri, isteği yapan kullanıcıya ileterek, ondan bir anlamlılık değerlendirmesi yapmasını istemektir. Tarama sonuçlarındaki tüm künyeler için anlamlılık yargıları verildikten sonra Kİ ve Eİ hesaplamaları açık ve doğru sonuçlar vermektedir. Kullanıcının, konusu açısından anlamlı bulunduğu künyeler toplamı, erişilen anlamlılardır. Bu değer, taramanın eriştiği tüm künye sayısına bölünerek Kİ oranı kolaylıkla elde edilir.

Asıl sorun Eİ oranının hesaplanmasıdır. Bu oran, erişilen anlamlı künye sayısı yanında, veri tabanında bulunan anlamlı fakat erişilmemiş

künye sayısının elde edilmesini gerektirmektedir. Bu sayının saptanabilmesi için ise, değerlendiricinin veri tabanındaki tüm künyeleri inceleyerek anlamlı olanları ayıklaması koşulu vardır. Normal büyüklükteki bir veri tabanında bu işlemi gerçekleştirmek olanaksızdır. Bu nedenle taramanın tam ve gerçek Eİ oranı elde edilmez. Değerlendirmelerde tam ve gerçek Eİ oranından vazgeçme durumunda kalınmakta, bunun yerine mümkün olan en sağlam Eİ oranını elde etme yoluna gidilmektedir. Bunun için bazı teknikler geliştirilmiştir. Lancaster⁽¹⁵⁾ bu teknikler hakkında ayrıntılı bilgi vermektedir.

En geçerli tekniklerden biri paralel taramalarla yürütülen "Görelî Eİ" (Relative Recall) tekniğidir. Bu teknikte tamamlanmış bir taramanın Eİ oranını elde etmek üzere, aynı konuda başka taramalar yapılır. Kullanıcı tarafından anlamlılık değerlendirilmesi yapılmış A taramasının Eİ oranı saptanmak istendiğinde, kullanıcının bilgi ihtiyacını belirleyen yazılı ifade, A taramasını sonuçlandırmış olan tarama uzmanının dışında kalan tarama uzmanlarına verilir. Onların B ve C taramalarını yapmaları sağlanır. Bu kişilerin A tarama uzmanının kullandığı stratejiyi görmesine izin verilmez. Bu durumda B ve C tarama uzmanları farklı künyelere erişebileceklerdir. Sonuçta, B ve C taramalarındaki, A taramasında yer almayan, yeni künyeler belirlenerek bunlar anlamlılık değerlendirmesi yapması için kullanıcıya sunulur ve anlamlı künyeler elde edilir. Bu durumda A taramasının Eİ oranı şu yaklaşımla elde edilir:

$$Eİ: \frac{\text{A taramasında yer alan anlamlı künye sayısı}}{\text{A, B ve C taramalarında, birbirlerinden farklı olarak, yer alan anlamlı künyeler toplamı}}$$

Dikkat edilmesi gereken husus, B ve C taramalarının, A taramasının Eİ oranını elde etmek üzere yürütüldükleri ve bunların değerlendirilmesinin söz konusu olmadığıdır.

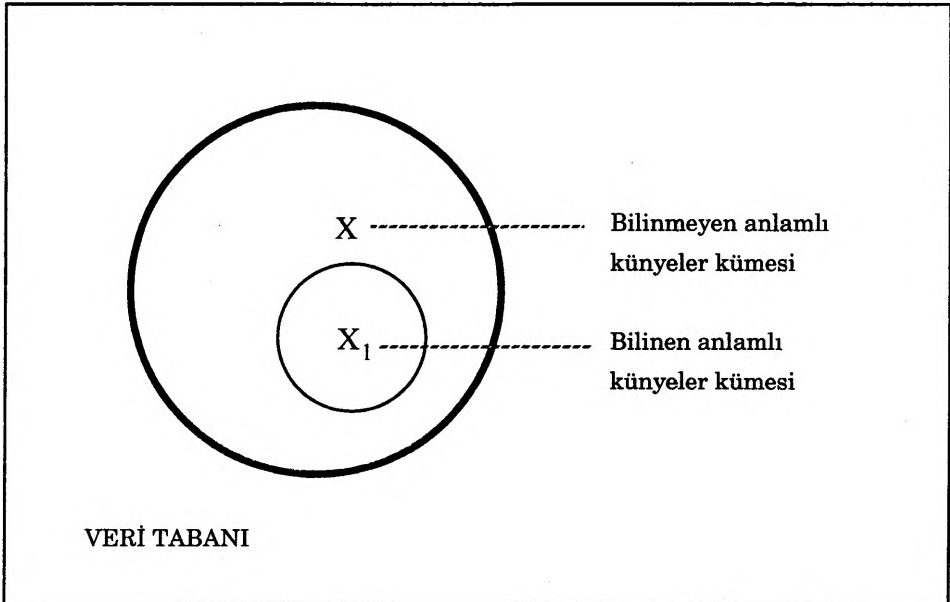
Bilgisayar taramalarının Eİ oranını bulmak için, veri tabanının karşılığı olan basılı indeksten paralel tarama yürütülebilmektedir. Aynı şekilde geleneksel taramanın Eİ oranı da bilgisayarla paralel tarama yapılarak bulunabilir. Anlamlılar belirlendikten sonra, örneğin bilgisayarla yapılan taramanın Eİ oranı aşağıdaki yaklaşımla elde edilebilir:

$$Eİ: \frac{\text{Bilgisayarla yapılan taramanın eriştiği anlamlı künye sayısı}}{\text{Bilgisayarla yapılan taramanın eriştiği anlamlı künye sayısı + Geleneksel taramanın eriştiği farklı anlamlı künye sayısı}}$$

Yukarıda örnekleri verilen bu tekniklerle mümkün olan en sağlam Eİ oranı elde edilebilmektedir. "Görelî Eİ" tekniğinde, asıl tarama ile paralel taramaların birbirlerinden farklı olarak eriştikleri anlamlı künyeler toplamının, veri tabanında bulunabilecek anlamlı künyeler toplamı olduğu var sayılmaktadır. Yani, çeşitli taramalarla anlamlı künyelerin çoğuna erişilebilecektir. Bu var sayım tamamen değerli olmayabilir. Çünkü paralel taramaların veri tabanındaki diğer anlamlı künyelere ulaşamaması olasıdır. Ancak tam ve gerçek Eİ oranını saptamak olanaksız, en azından aşırı derecede güçtür.

"Görelî Eİ" tekniği ile Eİ oranını elde etmek birçok değerlendirme amaçları için yeterli bulunmaktadır. Şekil 2'de gösterildiği gibi, "Görelî Eİ" tekniği, veri tabanındaki X bilinmeyen anlamlı künyeler kümesinin içinde yer alan X_1 bilinen anlamlı künyeler alt kümesini bulmamıza yaramakta ve bu alt küme için Eİ elde edilebilmektedir. X_1 , X'in temsilcisi olduğuna ve onun bütün özelliklerini taşıdığına göre, X_1 'in Eİ oranının, X'in Eİ oranına yaklaştığı savunulmakta ve bu var sayımla hareket edilmektedir. Birçok koşul altında, "Görelî Eİ" tekniği ile elde edilen Eİ, mümkün olan en iyi sonucu vermektedir.

Şekil 2: Koleksiyondaki/Veri Tabanındaki Bilinen ve Bilinmeyen Anlamlı Künyeler⁽¹⁶⁾.



Lancaster⁽¹⁷⁾ ve Pao⁽¹⁸⁾ tarafından belirtildiğine göre, bir tarama sonucu hakkında bilmek isteyebileceğimiz tüm verileri "2X2 tablosu" olarak adlandırılan tabloda göstermek, yaygın olarak geçerlidir (Tablo 1). Bu araştırmacılar, Kİ ve Eİ sonuçlarının bu tablodan yararlanarak kolaylıkla elde edilebileceğini vurgulamaktadırlar.

Tablo 1: Bir Bilgi Taraması Sonuçlarının Gösterildiği "2X2 Tablosu"

		KULLANICININ ANLAMLILIK KARARLARI		
		ANLAMLILIK	ANLAMLILIK OLMAYAN	TOPLAM
SİSTEMİN ANLAMLILIK TAHMİNİ	ERİŞİLEN	a İSABETLER, ERİŞİLEN ANLAMLILAR	b İŞE YARAMAYANLAR, ERİŞİLEN ANLAMLILIK OLMAYANLAR	a+b TÜM ERİŞİLENLER
	ERİŞİLEMİYEN	c KAÇIRILANLAR, ERİŞİLEMİYEN ANLAMLILAR	d DOĞRU OLARAK REDDEDİLENLER, ERİŞİLEMİYEN ANLAMLILIK OLMAYANLAR	c+d TÜM ERİŞİLEMİYENLER
	TOPLAM	a+c TÜM ANLAMLILAR	b+d TÜM ANLAMLILIK OLMAYANLAR	a+b+c+d TÜM DERME

Bilgi taramalarında, veri tabanını sorgulamak üzere kullanılan tarama stratejisi ile denkleşen künyelere erişilir (a+b), bu strateji ile çakışmada başarısız olanlara ise erişilmez (c+d). Erişilen anlamlılar (a) ve erişilen anlamlı olmayanlar (b), kullanıcının anlamlılık değerlendirmesiyle belirlenmektedir. Bir taramanın erişemediği anlamlı künyeler toplamı, paralel taramalarda bulunan anlamlı künyelerden yararlanılarak elde edilebilmektedir (c). Hem erişilemeyen, hem de anlamlı olmayanları ise, tarama süreci doğru olarak reddetmektedir (d). Kİ ve Eİ oranlarının Tablo 1'e göre ifadesi, şekil 1'den çıkarılan ifade ile aynıdır:

$$Kİ: \frac{a}{a+b}$$

$$Eİ: \frac{a}{a+c}$$

Yapılan tüm taramaları temsil eden örnekleme yer alan her tarama için Kİ ve Eİ değerleri saptandıktan sonra, sıra örnekleme taramalarının or-

talama Kİ ve Eİ oranlarını elde etmeye gelir. Salton ve Mc Gill'e⁽¹⁹⁾ göre ortalama oranlar şu şekilde bulunur:

$$\text{Ortalama Kİ: } \frac{\text{Tüm taramalarda erişilen anlamlı künyeler toplamı}}{\text{Tüm taramalarda erişilen künyeler toplamı}}$$

$$\text{Ortalama Eİ: } \frac{\text{Tüm taramalarda erişilen anlamlı künyeler toplamı}}{\text{Tüm taramalarda erişilen ve erişilemeyen anlamlı künyeler toplamı}}$$

"Görelî Eİ" tekniğinin veri tabanındaki tüm anlamlı künyeleri elde etmeyi sağladığı hatırlanmalıdır.

Tarama hizmetlerini değerlendirmede temel ağırlık, ortalama hesaplarına verilmelidir ki, taramanın yürütüldüğü sistemin ortalama Kİ ve Eİ sonuçları belirlenebilsin ve bu yolla o sistemin başarılı olup olmadığı ortaya konabilsin. Örneğin, Lancaster⁽²⁰⁾, MEDLARS'ı değerlendiren araştırmasında sistemin ortalama olarak %50.4 Kİ, %57.7 Eİ oranına ulaştığı sonucuna varmıştır.

Diğer taraftan ortalama Kİ ve Eİ sonuçları iki ya da daha fazla tarama sistemi veya tarama tekniğinin karşılaştırılmasına olanak sağlarlar. Geleneksel ve bilgisayara dayalı tarama tekniklerinin nitelik açısından karşılaştırmalı değerlendirilmesinde Kİ ve Eİ çok değerli ölçütler olmaktadır.

Literatürde, geleneksel ve bilgisayara dayalı (daha çok çevrim-içi) taramaları karşılaştırarak ortalama Kİ sonuçlarına ulaşan bazı araştırmacıların bulguları, geleneksel tarama Kİ oranı ortalamalarının, bilgisayar taramalarının Kİ ortalamalarından daha yüksek olarak elde edildiğini ortaya koymaktadır. Olive, Terry ve Datta⁽²¹⁾, Smith⁽²²⁾, Elchesen⁽²³⁾, Cleverdon⁽²⁴⁾, Crawley ve Adams⁽²⁵⁾ gibi araştırmacılar, basılı kaynaklardan yapılan taramaların, bilgisayara dayalı taramalara göre daha yüksek Kİ oranı sunduklarını belirtmişler ve araştırma bulgularıyla bunu kanıtlamışlardır.

Bazıları ise, bilgisayara dayalı tekniğin daha çok sayıda anlamlı künyeye eriştiğini, bu durumun da Eİ oranını arttırdığını belirtmektedirler. Elchesen⁽²⁶⁾ ve East⁽²⁷⁾ bu tarama tekniğinde anahtar sözcüklerle yaklaşım olanaklarının, yüksek Eİ oranına ulaşılmasına neden olduğunu vurgulamaktadırlar. Tıp taramalarını değerlendiren Ohta⁽²⁸⁾, özellikle geniş konulu taramalarda çevrim-içi tekniğin çok daha fazla anlamlı künyeye eriştiğini belirtmiştir. Bilgisayara dayalı tekniğin daha yüksek Eİ sonuçları ver-

diğini belirten ve bunu kanıtlayan araştırmacılar arasında, yine bilgisayara dayalı teknik için daha düşük Kİ ortalamaları elde etmiş olan Elchesen⁽²⁹⁾, Cleverdon⁽³⁰⁾, Crawley ve Adams⁽³¹⁾ gibi araştırmacılar da vardır.

Sonuç

Çeşitli yönleriyle irdelenen Kİ ve Eİ oranları, tarama değerlendirme ölçütleri olarak ülkemizde pek bilinmemekte, dolayısıyla kullanılmamaktadır.

1964'den beri anlamlılığın ölçülmesini ve sayısal olarak ifade edilmesini sağlayan Kİ ve Eİ oranlarını elde etmeksizin ve onları taramaların nitelik açısından değerlendirilmesinde ölçüt olarak kullanmaksızın varılan değerlendirme yargılarının yüzeysel olmaktan öteye geçemeyeceği gerçeğinin meslektaşlarımız tarafından benimsenmesi sevindirici olacaktır.

Notlar

- 1- C.J. Van Rijsbergen. *Information Retrieval*. London: Butterworths, 1979, 145-146.ss.
- 2- Cyril W. Cleverdon. *Indentification of Criteria for Evaluation of Operational Information Retrieval Systems*. Cranfield: College of Aeronautics, 1964.
- 3- F.W. Lancaster. *Information Retrieval Systems: Characteristics, Testing and Evaluation*. 2nd ed. New York: John Wiley and Sons. 1979, 109.s.
- 4- D. Soergel. *Organizing Information: Principles of Data Base and Retrieval Systems*. Orlando: Academic Press, 1985, 127.s.
- 5- Miranda Lee Pao. *Concepts of Information Retrieval*. Englewood: Libraries Unlimited, Inc., 1989, 54.s.
- 6- Tefko Saracevic. "Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science", *Journal of the American Society for Information Science*, XXVI, 6, 1975, 327.s.
- 7- G. Salton ve M.J.Mc Gill. *Introduction to Modern Information Retrieval*. New York: Mc Graw Hill, 1983, 174.s.
- 8- A. Kent. "Machine Literature Searching. VIII. Operational Criteria for Designing Information Retrieval Systems", *American Documentation*, VI, 1955, 93.s.
- 9- F.W. Lancaster. y.a.g.e. 112.s.
- 10- G. Salton ve M.J. Mc Gill. y.a.g.e. 164.s.
- 11- F.W. Lancaster. y.a.g.e. 112.s.
- 12- F.W. Lancaster ve E.G. Fayen. *Information Retrieval Online*. Los Angeles: Melville, 1975, 126.s.
- 13- F.W. Lancaster. *Information Retrieval Systems*. y.a.g.e. 113.s.
- 14- Aydın Köksal. "Bilgi Erişim Sorunu ve Bir Belge Dizinleme ve Erişim Dizgesi Tasarım ve Gerçekleştirimi." Yayınlanmamış Doçentlik Tezi. Ankara: Hacettepe üniversitesi, 1979, 64.s.
- 15- F.W. Lancaster. *Information Retrieval Systems*. y.a.g.e. 130-131.ss.
- 16- F.W. Lancaster. *Information Retrieval Systems*. y.a.g.e. 130.s.

- 17- F.W. Lancaster. *Information Retrieval Systems*. y.a.g.e. 113.s.
- 18- Miranda Lee Pao. y.a.g.e. 226.s.
- 19- G. Salton ve M.J. Mc Gill. y.a.g.e. 169.s.
- 20- F.W. Lancaster. *Evaluation of the MEDLARS Demand Search Service*. Bethesda: National Library of Medicine, 1968, 41.s.
- 21- G. Olive, J.E. Terry ve S. Datta. "Studies to Compare Retrieval Using Titles With That Using Index Terms. SDI from Nuclear Science Abstracts", *Journal of Documentation*, XXIX, 2, 1973, 169.s.
- 22- R.G. Smith. "Before You Scrap the Old Ways, Compare Retrieval Systems. Manual vs. Online", *Online*, I, 2, 1977, 26.s.
- 23- D.R. Elchesen. "Cost-effectiveness Comparison of Manual and Online Retrospective Bibliographic Searching", *Journal of the American Society for Information Science*, XXIX, 2, 1978, 62.s.
- 24- Cyril W. Cleverdon. "Letter to the Editor", *Journal of Information Science*, I, 4, 1979, 237.s.
- 25- J. Crawley ve C. Adams. "Info-Access Project: Comparing Print, CD-ROM and In-house Indexes", *The Canadian Journal of Information Science*, XVI, 1, 1991, 34-35.ss.
- 26- D.R. Elchesen. y.a.g.e. 62.s.
- 27- H. East. "Comparative Costs of Manual and Online Bibliographic Searching: A Review of the Literature", *Journal of Information Science*, II, 2, 1980, 106.s.
- 28- M. Ohta. "A Comparison of Some Demand Subject Searches. Machine vs. Human", *Bulletin of the Medical Library Association*, LV, 4, 1967, 408.s.
- 29- D.R. Elchesen. y.a.g.e. 62.s.
- 30- Cyril W. Cleverdon, "Letter...". y.a.g.e. 273.s.
- 31- J. Crawley ve C. Adams. y.a.g.e. 34-35.ss.