

**Prof. Dr. K. Glbn Baydur'a Armađan**

Yayına Hazırlayanlar:  
zgr Klc, Tolga akmak ve Nevzat zel

Hacettepe niversitesi  
Bilgi ve Belge Ynetimi Blm  
Ankara, 2012

## Bilgi Sınıflama, Bilgi Düzenleme ve Bilgi Erişim

Yaşar Tonta\*

*Öz: İnsan, duyu yoluyla çevresindeki değişimleri algılar, edindiği deneyim ve bilgileri sınıflar, düzenler, belleğine kaydeder ve gerektiğinde bu bilgilere yeniden erişir. İnsan belleği dışındaki ortamlara kaydedilen bilgilerin de erişim amacıyla sınıflanması ve düzenlenmesi gerekir. Ama varlıkları, soyut kavramları ve bilgiyi sınıflamak ve düzenlemek kolay değildir. Yüzyıllardır entellektüel çaba harcanmasına ve teknolojik gelişmeler bu amaçla kullanılmasına karşın, veri ve bilgileri sınıflama, düzenleme ve bu bilgilere gerektiğinde etkin bir biçimde erişim sağlama sorunu henüz çözülebilmemiş değildir. Bu çalışmada bilgi sınıflama, bilgi düzenleme ve bilgi erişim kavramları tanımlanmakta ve bu kavramlar birbirleriyle ilişkilendirilmektedir. Geleneksel bilgi düzenleme sistemlerinin dijital ortamda kayıtlı bilgileri düzenlemede yetersiz kalmasının nedenleri ile bilgi erişim sorununun çözümü için yapılması gerekenler tartışılmaktadır.*

**Anahtar Sözcükler:** Sınıflama, kategorizasyon, bilginin düzenlenmesi, ilgililik, soru yanıtlama sistemleri.

### Giriş

İnsanlar da dahil olmak üzere canlılar çevrelerini algılama yetenekleri sayesinde yaşamlarını sürdürebilir ve türlerini devam ettirebilirler. Çiçekler havadaki bir derecelik sıcaklık değişimini fark edebilmektedir. Arılar yaklaşık 60 kilometre uzaklıktaki çiçeklerin kokusunu algılayabilmektedir. Anne karnındaki bir bebek annesinin sesini diğer seslerden ayırt edebilmekte, doğumdan sonra ise annesinin yüzünü ve kokusunu kolayca tanıyabilmektedir. Başka bir deyişle, canlılar genetik miras ve deneyimle algılama yeteneğine sahip olmaktadır.

İnsanların edindikleri bilgilerin çoğu algılara dayanır. Ama insanları diğer canlılardan ayıran en önemli özellik sınıflama yetenekleridir. Sınıflamak insana ilişkindir (Bowker ve Star, 2000, s. 1). İnsanlar somut nesnelere ek olarak soyut kavramları da sınıflayarak öğrenmekte, nesnelere ve kavramlar hakkında bilgi sahibi olmakta, doğayı ve toplumu anlamakta, bireysel ve toplumsal yaşamlarını buna göre düzenlemektedirler. Standartlar ve sınıflamalar yaşamımızda merkezi bir yer tutmasına karşın genellikle bunların farkına varmayız. Bu sınıflamaları görmezden gelir ve bunun sonucu olarak yaşamımız zorlaşırsa, ancak o zaman sınıfların varlıklarını fark ederiz. Kategorilerin *ne* oldukları, kimler tarafından yaratıldıkları ve değiştirildikleri, nasıl ve ne zaman görünür hale geldikleri, nasıl yayıldıkları, bu görünmez ve güçlü varlıklar (entities) tarafından yaratılan toplumsal ve ahlaki (moral) düzen konularında pek bilgi sahibi değilizdir. Oysa sınıflama hem günlük

---

\* Prof. Dr., Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü, tonta@hacettepe.edu.tr

yaşamımızı hem de kamu politikalarını derinden etkilemektedir. Cinsiyet, ırk, din, hastalık, eğitim, meslek sınıflamaları bunlardan birkaçıdır. Sınıflama sistemlerini tasarlayan ve seçen bilginin çok azı bu sınıflamaların toplumbilimsel, ahlaki, kültürel ve estetik seçimler içerdiğinin farkındadırlar (Bowker ve Star, 2000, s. 2-4). Bilgi düzenleme ve bilgi erişim ise bu seçimlerle çok yakından ilişkilidir. Bu çalışmada sınıflama, düzenleme ve erişim kavramları ve bu kavramların birbirleriyle ilişkileri kısaca tartışılmaktadır.

## Sınıflama

Sınıf, sınıflama, tasnif, kategori, kategorize etme, kategorizasyon sözcükleri eş anlamlıdır<sup>1</sup>. Aristoteles'ten günümüze kadar birçok felsefeci, matematikçi, mantıkçı, dilbilimci ve bilişsel bilimci (örneğin, Bacon, Russell, Gödel, Wittgenstein, Lakoff ve Rosch) sınıflama/kategorizasyon terimini tanımlamaya çalışmışlardır. Klasik görüşe göre "şeyler ancak ve ancak belli ortak özelliklere sahipse aynı kategoridedirler". Kategoriler birbirini dışlayacak şekilde ve eksiksiz olarak açıkça tanımlanabilir. Sınıflama evrenindeki her varlık önerilen kategorilerden sadece ve sadece biri altında yer alabilir.<sup>2</sup> Kategoriler insan yeteneklerinden bağımsız olarak varlıklarını sürdürürler. Bu yaklaşım cinsiyet, ırk,<sup>3</sup> din gibi özelliklere göre insanları sınıflamak için kullanılmaktadır. Aynı yaklaşım botanik ve zoolojide bitkileri (flora) ve hayvanları (fauna) sınıflamak için geliştirilen taksonomiler<sup>4</sup> ya da hastalıkların sınıflanması için de geçerlidir.

Kategorizasyonla ilgili Aristoteles'in görüşleri uzun süre geçerliliğini korumuştur. Ancak 20. yüzyılın başında Aristoteles'in tanımının kısıtlayıcı olduğu görülmüştür. Matematiksel mantık alanında klasik sınıf kuramına dayanan Bertrand Russell'ın görüşleri Kurt Gödel ve Ludwig Wittgenstein tarafından eleştirilerek yeni kuramlar geliştirilmiştir.<sup>5</sup> Bir sınıfta (class, category) yer alan bir varlık başka sınıf ya da sınıfların özelliklerine de sahip olabileceği gibi, bu varlığın bir sınıfın bütün özelliklerini taşıması da gerekmez.<sup>6</sup> Lotfi Zadeh (1965) bu esasa dayanan bulanık kümeler (fuzzy sets) ve bulanık mantık (fuzzy logic) kavramlarını geliştirmiştir. Günümüzde bulanık mantık bilginin de dahil olmak üzere hemen hemen her alanda kullanılmaktadır.

Wittgenstein'a göre bütün oyunları kapsayan ortak özellikler yoktur. Sınıf/kategori kavramı ailevi benzerlikler (family resemblances) metaforuyla daha iyi açıklanabilir. Bir

<sup>1</sup> Ama "sınıflama" kütüphanecilik ve bilginin literatüründe, "kategorizasyon" ise dilbilim, bilişsel bilimler ve bilgisayar mühendisliğinde daha yaygın olarak kullanılmaktadır. Örneğin, Online Dictionary of Library and Information Science (ODLIS) adlı sözlükte sadece "classification" terimi listelenmiştir ([http://www.abc-clio.com/ODLIS/odlis\\_c.aspx](http://www.abc-clio.com/ODLIS/odlis_c.aspx)). Bu çalışmada yararlanılan literatüre göre bazen sınıflama, bazen kategorizasyon terimi birbiriyle eş anlamlı olarak kullanılmıştır.

<sup>2</sup> Bkz. <http://en.wikipedia.org/wiki/Categorization>

<sup>3</sup> Güney Afrika Cumhuriyeti'nde 1994 öncesi ırk ayrımı (apartheid) döneminde uygulanan ırk sınıflaması ve yeniden sınıflamanın sonuçları için bkz. Bowker ve Star (2000, Bölüm 6).

<sup>4</sup> Taksonomi: "Canlıların sınıflandırılması ve bu sınıflandırmada kullanılan kural ve prensipler. Taksonomi terimi Yunanca *taxis* (düzenleme) ve *nomos* (yasa) sözcüklerinden türetilmiştir." Bkz. <http://tr.wikipedia.org/wiki/Taksonomi>

<sup>5</sup> Bkz. [http://en.wikipedia.org/wiki/Principia\\_mathematica](http://en.wikipedia.org/wiki/Principia_mathematica)

<sup>6</sup> Nitekim 2005'te yayımlanan *Türkçe Sözlük*'te de "kategori" sözcüğü "[a]ralarında herhangi bir bakımdan ilgi veya benzerlik bulunan şeylerin tamamı, grup, ulam" olarak tanımlanmaktadır (abç).

aileye mensup bireyler bazı yönlerden birbirine benzerler, ama her birey aynı zamanda diğerlerinden farklı özelliklere sahiptir. Oyunlar da böyledir. Eleanor Rosch klasik sınıflama kuramına göre her üyenin bir kategorinin tüm özelliklerini taşımak zorunda olduğunu, bu nedenle bir kategoriye temsil eden diğerlerinden daha iyi bir örnek olamayacağını ve tüm üyelerin her yönden birbirine benzemeleri gerektiğini, oysa gerçek yaşamda bazı üyelerin bir kategoriye daha iyi temsil ettiğini öne sürmektedir. “Prototip kuramı”nı geliştiren Rosch’a göre, prototipler bilişsel referans noktalarıdır. Nesnelere ve şeyleri prototipe benzerliklerine göre sınıflarız. Örneğin, bülbül “kuş” prototipine pengüenden daha çok benzer.<sup>7</sup>

Kategorilerin insan yeteneklerinden bağımsız olarak varlıklarını sürdürdükleri tezi de günümüzde geçerliliğini yitirmiş, sınıflamanın kültür ve deneyimle yakından ilişkili olduğu görülmüştür. Avustralya yerlileri kadın, ateş ve tehlikeli şeyleri birlikte kategorize etmektedirler (Lakoff, 1990). Genelde meyve olarak sınıflanan domatesi Amerikan Anayasa Mahkemesi 1893 yılında sebze olarak sınıflamıştır (Rosenfeld ve Morville, 1998, s. 24). Hazar Denizinin deniz ya da göl olarak sınıflanması sınır ülkelerin ekonomik çıkarlarına göre değişebilmektedir (Tonta, 2005, s. 5). Nesnelere ve kavramların nasıl sınıflandıkları kültürden kültüre de değişmekte ve sınıflama bize insanların farklı zihin yapıları ve düşünce biçimleri hakkında fikir vermektedir (Nisbett, 2005).

## Bilgi Düzenleme

Devekuşu veya penguenler uçamadıkları halde “kuş”, balinalar ise yüzebildikleri halde “memeli” olarak sınıflanmaktadır. Bitkileri ve hayvanları sınıflamak zordur.<sup>8</sup> Yeni türler, alt türler ve cinsler keşfedildikçe ya da DNA dizilim analizleri sonucu familyaların sınırları daha iyi anlaşıldıkça ilgili taksonomiler güncelleştirilmektedir.<sup>9</sup> Somut nesnelere ve soyut kavramları sınıflamak daha da zordur. Bilgiyi düzenlemek ise en zordur. Çünkü bilgi düzenleme iki farklı varlık (entity) olan belgeler ile belgelerin içerdiği eserlerin birlikte ve birbirleriyle ilişkilendirilerek düzenlenmesini gerektirir (Svenonius, 2000, s. 10). Bu bağlamda bilginin temel kavramları arasında yer alan “bilgi”, “belge” ve “eser” terimlerini kısaca tanımlamakta yarar vardır. “Bilgi”, bilgilenme sonucu edinilen şey, bir mesajın ya da iletişimin içeriğidir. “Belge”, bilgi kırıntısı, bilgi taşıyan bir yazı ya da üzerinde insan düşüncesini temsil eden işaret ya da simgeler bulunan bir materyaldir.<sup>10</sup> “Eser” ise bilgi, bir mesajdan ayrılmış içerik veya ifade edilen düşüncedir (Svenonius, 2000, s. 7-9). Kolayca görülebileceği gibi, bu tanımlarda somut bir varlık olan belge ile soyut varlıklar olan bilgi ve eser iç içe geçmiştir. Svenonius (2000, s. 10) bilgi düzenleme ile başka şeyleri düzenleme arasındaki farkı şöyle açıklamaktadır:

<sup>7</sup> Bkz. [http://en.wikipedia.org/wiki/Prototype\\_theory](http://en.wikipedia.org/wiki/Prototype_theory)

<sup>8</sup> Örneğin, kış sonunda açan ve kısa ömürlü bir çiğdem türü olan Ankara çiğdemini (*Crocus ancyrensis*) APG III (Angiosperm Phylogeny Group) sınıflama sistemindeki yeri genelden özele şöyle belirlenmiştir: Alem: Bitkiler → Bölüm: Kapalı tohumlular (Magnoliophyta) → Sınıf: Bir çenekliler (Liliopsida) → Takım: Asparagales → Familya: Süsengiller (Iridaceae) → Cins: Çiğdem (*Crocus*) → Tür: Ankara çiğdemi (*C. ancyrensis*). Bkz. [http://tr.wikipedia.org/wiki/Ankara\\_çiğdemi](http://tr.wikipedia.org/wiki/Ankara_çiğdemi)

<sup>9</sup> Bkz. <http://en.wikipedia.org/wiki/Asparagales>

<sup>10</sup> “Bilgi” ve “belge” terimleriyle ilgili daha geniş bir tartışma için bkz. Buckland (1991; 1997).

Düzenlemenin çeşitli türleri olabilir. Düzenlemenin prototipi sınıflamadır. Sınıflama birbirine benzeyen şeyleri bir araya getirir. Geleneksel sınıflamalarda birbirine benzeyen şeyler bir ya da daha fazla tanımlanmış özelliğe göre bir araya getirilir. Aynı boy veya renk, aynı konu ya da aynı yazar gibi bazı özellikler bilgi içeren belge sınıfları yaratmak için kullanılabilir. Ama amacı bilgi düzenlemek olan bir sistem için en önemli özellik “aynı eseri içermek” özelliğidir. Düzenleme gücü açısından başka hiçbir özellik bu özelliğin yerini tutamaz çünkü bu özelliği paylaşan belgeler temelde aynı bilgiyi içerir. Bilgi düzenleme başka hiçbir anlama gelmese bile aynı bilgiyi bir araya getirmek demektir.

Svenonius (2000, s. 11-12) belgelerdeki bilgileri düzenleyen bibliyografik sistemlerin hem *tamamen* aynı bilgileri hem de *hemen hemen* aynı bilgileri bir araya getirdiğini, “hemen hemen aynı bilgi” ifadesinin ise işleri zorlaştırdığını vurgulamaktadır. Sezgisel olarak bu ifadeyi anlamak kolay olsa bile işletimselleştirmek zordur. Çünkü “eser nedir?” “bilgi nedir?” gibi tanımsal engellerle karşılaşmaktadır. “Eser”i tanımlamak zordur, çünkü eseri tanımlamak aynı zamanda “bilgi”yi tanımlamak anlamına gelir. Svenonius “eser” ile “bilgi” arasındaki farkı tanımlamanın zorluğunu şu örneklerle açıklamaktadır:

*İlyada*’nın özgün Yunancası İngilizce çevirisiyle aynı bilgiyi mi içerir (aynı eseri mi temsil eder)? İki farklı İngilizce çeviri aynı eseri mi temsil eder? (Bu iki sorunun yanıtı genellikle evettir.) Bir başka ortama (medium) çevirmek eser olma niteliğini (workhood) geçersiz kılar mı? *Hamlet*’in film sürümü metin sürümüyle aynı bilgiyi mi içerir? (Bu tür bir sorunun yanıtı genellikle hayırdır.) Bir senfoninin biri CD, diğeri video üzerindeki iki kaydı aynı eser midir? (Burada yanıt belirsiz gözüküyor.) . . . . Bilginin nasıl tanımlandığı neyin düzenleneceğini ve nasıl düzenleneceğini belirler. (Svenonius, 2000, s. 12)

Sorun sadece eseri içeren belgenin somut özelliklerini (yazarı, başlığı, boyutları, sayfa sayısı, vd.) tanımlamak değildir. Örneğin, *Hamlet*’in yazarının kim olduğu ya da eserin hangi yıl kaleme alındığı nispeten kolayca saptanabilir. Esas sorun *Hamlet*’in içerdiği bilginin, konusunun tanımlanmasıdır. Bir belgenin (örneğin, *Hamlet*) somut özelliklerinin nitelenmesi “tanımlayıcı kataloglama”, içerdiği bilginin, yani konusunun nitelenmesi ise “konu kataloglaması” olarak bilinir. Bir eserin somut özellikleri, konusu ya da içerdiği bilgi bir ya da birkaç konu başlığı veya anahtar sözcükle tanımlanır ve o eseri temsil eden üst veriler (metadata, “bilgi hakkında bilgi”) oluşturulur.

Ama üst veriler asla asıl eserin yerini tutamaz, eserin entellektüel içeriğini ya da bağlamını eksiksiz temsil edemez. Eserlerin temsilinde belgelerin tam metinleri kullanılsa bile belirsizlik (indeterminacy) vardır (Blair, 1990, s. 22). Bunun çeşitli nedenleri vardır. İlki, işin doğası gereği, bir şeyi tanımlamak onu sınırlamak, kısıtlamak anlamına gelir. Bu sınırlama konu kataloglamasının gücünü azaltmaktadır. Örneğin, Amerikan Kongre Kütüphanesi tarafından eserlere atanan konu başlığı sayısı ortalama iki civarındadır. Bunun başlıca nedeni ekonomiktir (katalog kartlarının basımı ve çoğaltılması). Ama çevrimiçi (online) katalogların ortaya çıkışıyla birlikte ortalama konu başlığı sayısında büyük bir artış gözlenmemiştir. Doğal olarak, bir eseri temsil eden konu başlığı sayısı azaldıkça üst verilerin eserin aslını temsil etme gücü de azalmaktadır. İkincisi, tanımlama sırasında yapılan hatalardır. Eserin içerdiği bilgiler veya bu bilgilerin bir kısmını temsil eden konular için konu başlıkları/anahtar terimler atanmayabilir (error of omission) veya yanlış atanabilir (error of commission). Üçüncüsü, aynı eseri tanımlayan farklı uzmanlar, aynı sistemi kullansalar bile, o esere farklı konu başlıkları atayabilmektedirler. Başka bir deyişle, uzmanlar arasındaki

tutarlılık (consistency) son derecede düşüktür (Tonta, 1991; Olson ve Wolfram, 2008; Wolfram, Olson ve Bloom, 2009).

Bir eseri temsil eden kaydı oluştururken hem eseri içeren belgeyi hem de eserin konusunu tanımlamak için yapay dillerden (konu başlıkları, kavramsal dizinler, vb. gibi) yararlanıldığı da unutulmamalıdır. Bu diller kültürden kültüre, hatta aynı kültür içinde bile, değişebilmektedir. Bütün ülkelerde yaygın olarak kullanılan uluslararası konu başlıkları ve sınıflama sistemleri olmasına karşın, bu sistemler yerel konuları yeterince (veya hiç) yansıtmamakta ve geliştirildiği ülkelerdeki dünya görüşünün izlerini taşımaktadır. Örneğin, Kongre Kütüphanesi Konu Başlıkları sisteminde kölelerin ayaklanmaları “isyan”, beyazların ayaklanmaları ise “devrim” olarak nitelendirilmektedir. “Indians of North America, Civilization of” (Kuzey Amerika Yerlileri, Uygarlığı)<sup>11</sup> başlığı altında yerlilerin uygarlığıyla ilgili literatür yerine, yerlileri “uygarlaştırmak” için yapılan çalışmalarını içeren literatürün yer alacağı not edilmiştir. Geçmişte ise “cinsel sapıklıklar” başlığı altında eşcinsellikle ilgili literatür yer almaktaydı (Berman, 1971’den aktaran: Buckland, 2012b, s. 159-160). Buckland, Berman’ın verdiği örneklerin çoğunun erkek egemen ve Hıristiyan dünya görüşünü, modası geçmiş toplumsal tutumları ve eski tıbbi ve psikolojik terminolojiyi yansıttığını vurgulamaktadır. Konuyla ilgili bir diğer örnek Dewey Onlu Sınıflama (DOS) sisteminden verilebilir. DOS’ta İslam diniyle ilgili şema (297) eksik ve hatalı olduğundan konuyla ilgili eserler büyük bir olasılıkla daha genel ya da başka konular altında sınıflandırılmış olmalıdır (Tonta ve Al, 2007, s. 47). İslam diniyle ilgili şema yaratıldıktan sonra eskiden sınıflanan eserler yeni şemaya göre yeniden sınıflandırılmamıştır. Robert Fairthorne’un “marking and parking” (kataloglama ve rafa yerleştirme) olarak adlandırdığı bu yaklaşım çok yaygındır (Fairthorne, 1961, s. 84-85’ten aktaran Buckland, 2012b, s. 155). Başka bir deyişle eserler belli bir zamanda geçerli olan sisteme göre tanımlanmakta, daha sonra meydana gelen politika değişiklikleri eskiden tanımlanan eserlere yansıtılmadığından aynı konudaki literatür bölünmektedir.

Bilgi düzenlemek için kullanılan bibliyografik sistemlerin aynı bilgileri ya da hemen hemen aynı bilgileri içeren eserleri bir araya getirmesi (collocation), aynı ya da hemen hemen aynı bilgileri içeren eserlerin de hangi yönlerden birbirinden farklı olduklarını ayırt etmesi (differentiation) beklenir. Ama yukarıda anılan sorunlar (tanımlama hataları, birkaç konu başlığına indirgeme, tutarsızlık, vd.) bibliyografik sistemlerin bu iki işlevi kusursuz bir biçimde gerçekleştirmesini olanaksız hale getirmektedir. Örneğin, aynı bilgileri içeren *İlyada*’nın çeşitli basımlarını ve çevirilerini bu tür bir sistem başarıyla bir araya getirebilir ve basım yılı, çevireni, dili vb. gibi özelliklerine dayanarak birbirlerinden farklarını ayırt edebilir. Ama aynı konu başlığı ya da anahtar sözcükler atanmış olan iki eseri aynı bilgileri içerip içermedikleri yönünden birbirinden ayırt edemez. Örneğin, *Çaya Çorbaya İnternet* ile *İnternet ve Hukuk* adlı eserlerin ikisi de “İnternet” konu başlığı altında listelenmiştir.<sup>12</sup> Sadece eser adlarına bakarak bile bu iki eserin İnternet hakkında tamamen aynı bilgileri içermediği söylenebilir. *Hamlet*’in metin sürümüyle bu metne

<sup>11</sup> Kart kataloglarında konu başlıkları temel giriş ögesi olarak büyük harflerle yazılırdı (KUZHEY AMERİKA YERLİLERİ, UYGARLIĞI gibi).

<sup>12</sup> Bkz. Coşkun Tülcü, *Çaya Çorbaya İnternet*. İstanbul: Beta Basım Yayın, 1996; Yeşim M. Atamer (der.), *İnternet ve Hukuk*. İstanbul: İstanbul Bilgi Üniversitesi, 2004. Kaynak: Boğaziçi Üniversitesi Kütüphanesi (<http://seyhan.library.boun.edu.tr/>)

dayanan film sürümünün bile aynı bilgileri içermediğini kabul ediyorsak, yukarıdaki iki eserin de İnternet’le ilgili birbirinin aynı ya da hemen hemen aynı bilgileri içermediğini kabul etmek zorundayız. Ama konu başlığı açısından ele alınacak olursa bu iki eserin İnternet’le ilgili içerdikleri bilgiler açısından birbirinden hiçbir farkı yoktur.

### **Bilgi Erişim**

Bilgi düzenlemenin temel amacı düzenlenen bilgilere daha sonra erişim sağlamaktır. Bilgi erişim bir eseri temsil eden üst veriler aracılığıyla gerçekleştirilir. Bilgi gereksinimi olan kullanıcılar<sup>13</sup> bu gereksinimlerini tanımlayarak bir arayüz aracılığıyla bibliyografik sisteme girerler. Sistem kullanıcının bilgi gereksinimini tanımlayan konu başlıkları veya anahtar terimlerle bilgi düzenleme aşamasında eserlere atanan anahtar sözcükleri ve konu başlıklarını karşılaştırır. Birbiriyle çakışanları sonuç listesine ekler. Ama kullanıcıların bilgi gereksinimlerini layıkıyla tanımlayabileceklerini kabul etsek bile, bilgi sınıflama ve düzenleme sürecindeki eksiklik ve belirsizlikler nedeniyle kullanıcılar tam olarak istedikleri bilgilere muhtemelen erişemeyeceklerdir.<sup>14</sup> Bu durum bilgi erişim sistemlerinin performansını olumsuz yönde etkilemektedir.

Bilgi sınıflama ve bilgi düzenlemede temel amaç nasıl ki aynı bilgileri içeren eserleri bir araya getirmek ve bu eserlerin birbirinden farklarını ayırt etmek ise, bilgi arayan kullanıcıların da amacı belli bir konuda gereksinim duydukları tüm bilgilere ve sadece aradıkları bilgilere erişmektir. Hatta bilgi erişim sistemlerinin performansları da bu ölçütlere göre değerlendirilir. Arama sonucunda belli bir konuda dermedeki (koleksiyon, derlem) tüm ilgili bilgilere erişilmesi anma (recall), sadece ilgili belgelere erişilmesi ise duyarlık (precision) olarak tanımlanır.<sup>15</sup> Anma, sistemin kullanıcının istediği bilgileri içeren ilgili belgeleri bir araya getirme başarısını, duyarlık ise sistemin ilgisiz belgeleri eleme başarısını ölçer.

Bilgi erişimde tanımlanması en zor kavramlardan biri de “ilgililik” (relevance) kavramıdır. İdeal olarak ilgililik, erişilen belgelerin içerdiği bilgilerle kullanıcının sorgu cümlesinde ifade ettiği bilgilerin aynı ya da hemen hemen aynı olmasıdır. Uygulamada ise ilgililik, eserleri sınıflamak ve düzenlemek için atanan konu başlıkları ya da anahtar sözcüklerle kullanıcının sorgu cümlesinde yer alanların birbiriyle çakışmasıdır (“nesnel ilgililik”). Çakışan belgeler potansiyel olarak kullanıcının istediği bilgileri içerdikleri varsayılarak arama sonucunda listelenir. Çakışmayanların ise kullanıcının istediği bilgileri içermediği varsayılır.

---

<sup>13</sup> Kullanıcıları sadece insanlarla sınırlamak gerekir. Yapılan aramaların giderek artan bir yüzdesi kullanıcıları temsilen “akıllı ajanlar” ya da arama motoru örümcekleri tarafından gerçekleştirilmektedir.

<sup>14</sup> Burada bilgi erişim süreci çok basit biçimde verilmektedir. Kullanıcıların bilgi gereksinimlerini ifade etmeleri, kullanıcı arayüzü, sonuçların yorumlanması vb. gibi hususlarla ilgili sorunlar göz ardı edilmiştir. Örnekler konu aramalarından seçilmiş, somut olarak bilinen öğelere göre (yazar adı, yayın yılı, vb. gibi) yapılan aramalara değinilmemiştir. Ama bu seçim “known-item search” olarak bilinen bu tür aramalarda sorun olmadığı anlamına gelmemektedir. Bilgi erişim ve arama başarısızlıklarıyla ilgili ayrıntılı bilgi için bkz. Tonta (1992, 1995) ve Tonta, Bitirim ve Sever (2002).

<sup>15</sup> İdeal erişim hem anma hem de duyarlık değerlerinin mükemmel olduğu erişim sonucu anlamına gelir. “Kendisinden olanları içeren, olmayanları dışlayan”, “ne eksik ne fazla, eksikliği artığı olmayan” anlamına gelen “efradını cami, ağyarını mâni” deyimini mükemmel erişimi tanımlamak için kullanılabilir.

Bu varsayımlara dayanan erişim sonuçları kuşkusuz her zaman kullanıcının aradığı bilgileri içeren belgelere erişimle sonuçlanmaz. Bunun temel nedeni sınıflama ve düzenleme aşamasında yapılan hatalar ile ilgililik kavramının öznel olmasıdır. *Çaya Çorbaya İnternet ve İnternet ve Hukuk* adlı eserleri örnek olarak verecek olursak; bu iki eserin İnternet hakkında aynı bilgileri içermeyebileceğini daha önce vurgulamıştık. Ama bu iki eser “İnternet” konu başlığıyla arama yapan bir kullanıcının karşısına çıktığında işin hukuk yönüyle ilgilenmeyen bir kullanıcı muhtemelen ikinci eseri ilgisiz olarak değerlendirecektir. Yani sistem ilgisiz bir belgeye erişmiş olacak, duyarlık oranı düşecektir. İkinci esere “İnternet – Hukuk ve Mevzuat” konu başlığı verilmiş olmasına karşın, “Bilişim Hukuku” başlığı ile arama yapan bir kullanıcı, bu başlık sınıflama ve düzenleme aşamasında tercih edilmediği için, bu esere bilişim hukukuyla ilgili bile olsa erişemeyecektir. Yani sistem ilgili bir belgeye erişememiş olacak, anma oranı düşecektir. Kaldı ki, sınıflama ve düzenleme aşamasında hata yapılmamış olsa bile, ilgililik değerlendirmesi kişiden kişiye değişebileceğinden bilişim hukuku konusundaki bu eseri bazı kullanıcılar ilgili, bazıları ise çeşitli nedenlerle (örneğin, konunun işleniş biçimi, düzeyi, vs.) ilgisiz bulabilirler. Ya da aynı kullanıcılar farklı durum ve zamanlarda bu eseri bazen ilgili, bazen ilgisiz bulabilirler (Wilson, 1973). Örneğin, bilişim hukukuyla ilgili ders notu hazırlayan bir öğretim üyesi bir panelde sunulan bildirimleri içeren bu eseri çok spesifik bulduğu için öğrenciler için hazırladığı okuma listesine koymak istemeyebilir. Ama aynı öğretim üyesi aynı eseri “Siber Terorizm ve Ulusal Güvenlik” ile ilgili bir bildiri içerdiği için bir başka zaman ilgili bulabilir.<sup>16</sup>

Dikkat edilecek olursa yukarıdaki ilgililik değerlendirmeleri “ilgili” ve “ilgisiz” olmak üzere iki değerlidir (bivalent, binary). Bilgi düzenleme sürecinde bir konuda aynı bilgileri içeren eserlerle hemen hemen aynı bilgileri içeren eserler arasında genellikle bir ayrım yapılmaz ve ilgili konu başlığı ya da anahtar sözcük bütün eserlere atanır. Benzeri bir biçimde kullanıcılar da sorgu cümlelerini oluştururken böyle bir ayrım yapmazlar; yani erişilecek eserlerde hangi bilgilerin daha ağırlıklı işlenmesi gerektiği yönünde bir istekte bulunmazlar. Çoğu bilgi erişim sistemleri (örneğin çevrimiçi kataloglar) iki değerli ilgililik değerlendirmelerine göre erişim sonuçlarını verir. Oysa kullanıcılar ilgililik değerlendirmelerini “hiç ilgisiz” ile “tamamen ilgili” skalasına göre yaparlar. Yani, örneğin, İnternet’le ilgili bilgi arayan bir kullanıcı ilk eseri (*Çaya Çorbaya İnternet*) diğerine oranla daha az ilgili bulabilir.

Günümüzde bazı bilgi erişim sistemlerinde iki değerli mantığa dayanan algoritmalar daha karmaşık sınıflama, düzenleme ve bilgi erişim algoritmaları kullanıldığı öne sürülebilir. Örneğin, MEDLINE veri tabanında yer alan makalelerde bir makalede ağırlıklı olarak işlenen konu ya da konulara verilen başlıklar yıldız (\*) işaretiyle ayırt edilmekte, kullanıcılar da arama sonuçlarını sadece yıldızlı konu başlığı verilen belgelerle sınırlayabilmektedirler.<sup>17</sup> Ya da vektör uzayı modeli, olasılıksal (probabilistik) model ve

<sup>16</sup> “Siber terorizm” ile ilgili bilgi arayan kullanıcılar da tıpkı “bilişim hukuku” ile ilgili bilgi arayanlar gibi bu esere “siber terorizm” konu başlığıyla arama yaparak ulaşamayacaklardır.

<sup>17</sup> Bu durumda ilgililik değerlendirmesi artık ikili (ilgili/ilgisiz) bir karar değil, üçlü bir karardır (ilgisiz/daha çok ilgili/daha az ilgili). İlgili belgeler kendi içinde “daha çok ilgili” (yıldızlı olanlar) ve daha az ilgili diye ayrıca sınıflanmaktadır.



dil modelini kullanan bilgi erişim sistemlerinde eserlerin konularını niteleyen anahtar sözcükler ağırlıklandırılabilen ve erişim sonuçları da buna göre sıralanabilmektedir.

MEDLINE örneğinde makalelerin belirli bir konuyla ilgili olup olmamasına entellektüel bir değerlendirme sonucunda karar verilmektedir. Diğer modellerde ise sözcük sayısı, bağlantı (link) sayısı gibi özellikler dikkate alınarak belgelerin belirli bir konuyla ilgilik derecesi saptanmaktadır. MEDLINE’da ve geleneksel bilgi erişim sistemlerinde (örneğin, kütüphane katalogları) “anlamsal ilgililik”; vektör uzayı, PageRank algoritması gibi modelleri kullanan (örneğin, Google) sistemlerde ise “istatistiksel ilgililik” değerlendirmesi yapıldığı söylenebilir (Zadeh, 2006, s. 164). 1990’lardan itibaren bilgileri temsil etmek için istatistiksel anahtar sözcük ve benzerlik yöntemlerinin yaygın bir biçimde kullanılması ağ ortamında bilgi temsil kalitesini karma hale getirmiştir. Bu durumu tanımlamak için “düzenli karmaşa” (organized chaos) terimi kullanılmaya başlanmıştır (Chu, 2003, s. 4).

MEDLINE benzeri sistemlerde bilgi düzenleme sürecinde yapılan hatalara daha önce değinilmişti. Diğer modellere göre işleyen bilgi erişim sistemlerinin performansları da mükemmel değildir. Örneğin, Gerard Salton tarafından geliştirilen vektör uzayı modelinde bir belgede ya da sorgu cümlesinde geçen sözcükler o belgenin ya da sorgu cümlesinin konusunu belirleyen içerik terimleri olarak kabul edilmekte ve birer vektör olarak gösterilmektedir. Bir belgeyle dermedeki diğer belgeler ya da bir sorgu cümlesiyle belgeler arasındaki benzerlik, belgeyi/sorguyu temsil eden terim vektörleri arasındaki açının kosinüsü alınarak hesaplanmakta ve erişim sonuçları buna göre sıralanmaktadır (Salton ve McGill, 1986). Bu modelde bir terimin bir belgede geçiş sıklığıyla (term frequency) bütün dermedeki belgelerde geçiş sıklığı (inverse document frequency), terimin belgenin neresinde geçtiği (örneğin, başlık, özet, vb. gibi), belge uzunluğu gibi çeşitli ölçütler dikkate alınarak benzerlik hesaplamaları normalleştirilse de sonuçta ilgililik kararı istatistiksel ilgililik dikkate alınarak verilmektedir. İstatistiksel ilgililik her zaman anlamsal ilgililik demek değildir. *Çaya Çorbaya İnternet* adlı eserin başlığında geçen çay ve çorba sözcükleri eserin konusu hakkında bize hemen hemen hiçbir ipucu vermemektedir.<sup>18</sup>

Sonuç olarak, bir eserin hangi konu ya da konularda bilgi içerdiğine ister entellektüel çaba harcanarak karar verilsin isterse (vektör uzayı modelinde olduğu gibi) otomatik dizinleme yapılarak ya da olasılık veya dil modellerine göre karar verilsin, sonuçta anlamsal ilgililik kararı büyük ölçüde o eseri temsil eden bir konu başlığı ya da terimin (token) varlığına ya da yokluğuna indirgenmektedir. Konu başlıkları ya da terimler ise bir (yapay) dilin temel öğeleridir ve belgeleri temsil eden tanımları (üst verileri) yaratmak için kullanılır. Belgelerin temsilinde ise her zaman bir belirsizlik (ambiguity) söz konusudur. Bu bakımdan bilgi erişimde temel sorun dil ve temsil sorunudur (Blair, 1990, 4. Bölüm). Kolayca görülebileceği gibi, bu sorunu daha hızlı ve işlem kapasitesi yüksek bilgisayarlar, ağlar ya da veri tabanı yönetim sistemi yaklaşımıyla çözmek mümkün değildir.

<sup>18</sup> Vektör uzayı modelinde belgelerde çok seyrek geçen terimlerin benzerlik sıralamasında listenin alt sıralarında yer alacağı, dolayısıyla erişim statüsü değerlerinin çok düşük olacağı haklı olarak öne sürülebilir. Ama belgelerde geçen terimler ya da bağlantılar erişim amacıyla optimize edilebilir ya da bunlarla oynanabilir. Böylece belgelerin erişim sıralamasındaki yerleri değiştirilebilir. Arama motoru optimizasyonunun (search engine optimization) temel amacı budur.

## Yeni Dijital Düzensizliğin Gücü<sup>19</sup>

Weinberger (2007) dijital çağda bilgi düzenleme ile ilgili konuları işlediği eserinin “Düzenin Yeni Düzeni” başlığını taşıyan ilk bölümünde bilgi düzenlemenin üç derecesini açıklamaktadır. Birinci derece düzenlemede bilgi taşıyan varlıkların, nesnelerin kendisi (örneğin, ofislerdeki yazışma dosyaları) düzenlenir ve her nesne sadece bir sınıf altında yer alabilir. İkinci derece düzenlemede nesnelerin üst verileri düzenlenir, bir nesne birden fazla sınıf altında listelenebilir (örneğin, kart kataloglar). İkinci derece düzenlemede üst veriler uzmanlar (örneğin, katalogcular) tarafından hazırlanır. Her iki düzenleme biçiminde de fizik yasaları geçerlidir. Atomlardan oluşan fiziksel nesnelere belli bir zamanda ancak bir yerde bulunabilir. Üçüncü derece düzenlemede ise atomlar yerine “bit”ler düzenlenir. Üst veriler herkes tarafından yaratılabilir, dijital nesnelere aynı anda birden fazla yerde bulunabilir (Weinberger, 2007, s. 16-23).

Birinci derece düzenlemede üst veri genellikle nesnenin üzerine kaydedilir. İkinci derece düzenlemede ise nesnenin kendisiyle nesneyi tanımlayan üst veriler birbirinden ayrılmıştır. Bir nesne birden fazla üst veriyle tanımlanabilse de genellikle üst veriler sınırlıdır. Sadece uzmanlar tarafından yaratılan üst veriler kullanılarak nesnelere (ve dolayısıyla bilgiye) erişim sağlanabilir. Üçüncü derece düzenlemede artık üst verileri tanımlamak için bir otoriteye gerek duyulmamaktadır. Kullanıcılar kendi üst verilerini yaratabilirler. Üst veriler ile bilgi içeren dijital nesne birbirinden ayrı olabileceği gibi birlikte de bulunabilir. Kullanıcılar kendileri, başkaları ya da otomatik sistemler tarafından yaratılan üst verileri kullanarak dijital ortamdaki bilgilere erişebilirler. Günümüzde LibraryThing ([www.librarything.com](http://www.librarything.com)), Flickr ([www.flickr.com](http://www.flickr.com)), Delicious ([delicious.com](http://delicious.com)) gibi sistemlerde bilgi erişim bu şekilde gerçekleşmektedir.

İkinci derece düzenleme sistemleri araştırma ürünü basılı literatür için geliştirilmiştir (Şekil 1). Şekil 1’deki piramit bir buzdağı olarak düşünülecek olursa, kütüphaneciler ve bilginimciler yüz yılı aşkın bir zaman diliminde buzdağının görünen kısmındaki 40-50 milyon kitap ile birkaç yüz milyon makaleyi düzenleyebilmişlerdir. Oysa günümüzde bilimsel bilgiler ve ham veriler giderek tamamen dijital ortamda (born digital) üretilmekte, yayınlanmakta ve saklanmaktadır. Dijital “belge” sayısı muhafazakâr bir tahminle birkaç yüz milyardır. Daha 2008 yılında Google’ın dizinlediği belgelerdeki bağlantı (URL) sayısı bir trilyona ulaşmıştı. Basılı ortamdaki bilgileri düzenlemek için geliştirilen bu sistemler dijital ortamdaki bilgileri, türetilmiş bilgileri ya da ham verileri düzenlemede yetersiz kalmaktadır. Başka bir deyişle, buzdağının görünmeyen kısmı henüz yeterince düzenlenememiştir.

Ayrıca ikinci derece düzenleme sistemlerinde “analiz birimi” genellikle kitap ya da makale iken, üçüncü derece düzenleme sistemlerinde analiz birimi çok daha küçülmüştür (web sayfaları, tartışma listelerine ve twitter gibi mikro-bloglara gönderilen mesajlar, türetilmiş bilgiler ve ham veriler gibi). Dijital çağda çevrimiçi ortamlarda üretilen ve yayınlanan bilgiler de en az basılı kitaplardakiler, makalelerdeki kadar değerli hale gelmiştir. 1990’larda geleneksel düzenleme araçlarıyla “İnternet’i ortaklaşa kataloglamak” için girişimlerde bulunulmuştur (örneğin, OCLC’nin Cooperative Online

<sup>19</sup> Bu ara başlık Weinberger’in (2007) kitabının adından alınmıştır.

Resource Catalog projesi). İkinci derece düzenleme sistemleriyle çok daha çözünür (granular) olan, örneğin 140 karakterden oluşan twitter belgelerini ya da türetilmiş bilgi ve ham verileri düzenlemek hemen hemen olanaksızdır.



Şekil 1. Tüm bilimsel veriler çevrimiçi (Gray, 2009, s. xxviii)

Öte yandan, dijital ortamda bilimsel iletişim süreci de değişime uğramaktadır. Şekil 1’deki piramidin ortasında ve tabanında yer alan bilgi ve ham verileri de bilimsel üretim sürecinin bir parçası haline getirebilmek için alt yapı geliştirilmektedir. Böylece etkileşimli, hakemlerin ve okuyucuların araştırmada kullanılan ham verileri yeniden analiz edebilecekleri ya da görselleştirebilecekleri “çalıştırılabilir bildirimler” (executable papers) yayınlamak mümkün hale gelmektedir (Nowakowski ve diğerleri, 2011). “Makine dostu”, yani makinece okunabilir ve üzerinde makine tarafından işlem yapılabilir (machine-actionable) bilimsel yayınlar üretmek ve veri setlerini bilimsel yayıncılık sistemiyle bütünleştirmek için çalışmalar yapılmaktadır (Van de Sompel ve Lagoze, 2009). Bilimsel iletişim süreci sadece fen bilimlerinde değil insani bilimlerde de değişmektedir. Ham veriler insani bilimlerde de bilimsel iletişim sürecinin bir parçası olarak düşünülmektedir. Örneğin, bir yazarın yayımlanmış biyografik bir eseriyle birlikte eseri yaratma sürecindeki düşüncelerini içeren el yazması notlar, gazete kesikleri (kupür), görüntüler, çalışma taslaqları ve detaylı notların da paylaşılması amaçlanmaktadır (Buckland, 2012a).<sup>20</sup>

Dijital nesnelere de dahil her tür eseri daha iyi tanımlamak amacıyla FRBR (Functional Requirements for Bibliographic Records), FRSAD (Functional Requirements for Subject Authority Data) gibi varlık-ilişki (entity-relationship) modeline dayanan ve çerçeve

<sup>20</sup> İlgili projenin web sayfası için bkz. Knowledge Unix: Sharing and enrichment of (digital) data. <http://ecai.org/KnowledgeUnix/>.

kurallar içeren ikinci derece düzenleme sistemleri geliştirilmiştir. Dijital nesnelere XML (Extensible Markup Language), RDF (Resource Description Format) gibi işaretleme (markup) dilleri ya da üst veri sistemleri kullanılarak da etiketlenilmekte ve bu etiketler erişim amacıyla kullanılmaktadır. Söz konusu işaretleme dilleri ile bir metindeki paragrafları, cümleleri, hatta sözcükleri atomik düzeyde etiketlemek ve bu verilere belgelerin tanımlayıcı özelliklerini kullanarak veri tabanı yaklaşımıyla erişmek son derecede kolaydır. Ama bu tür düzenleme sistemlerini kullanarak belgelere konularına göre erişmek zordur. Örneğin, belirli bir kişinin bir tartışma listesine hangi mesajları gönderdiğini bu etiketler aracılığıyla kolayca öğrenebiliriz ama devam eden bir tartışmaya bu kişinin kavramsal açıdan ne tür bir katkıda bulunduğunu öğrenemeyiz.

Öte yandan dijital nesnelere basılı nesnelere kadar sabit değildir. Örneğin, arXiv e-baskı arşivinde (arxiv.org) çıkan bilimsel bir makale yapılan tartışmalar sonucu daha sonra güncellenebilir. Ya da çevrimiçi bir dergide yayımlanarak sabitlenmiş bir makale hakkında bloglarda ya da twitter'da tartışmalar devam edebilir ve bu tartışmalar esas yayınla birlikte listelenebilir. İkinci derece sistemler bilgi içeren sabit nesnelere (kitap, bildiri, vs.) düzenlemek için tasarlandığından dinamik nesnelere düzenlemekte yetersiz kalmaktadır.

Önemli sorunlardan biri de ikinci derece düzenleme sistemleriyle üçüncü derece düzenleme sistemlerinin birlikte çalışabilirliği (interoperability). Bir MARC (Machine Readable Cataloging) kaydı Dublin Core ya da XML formatına kolayca dönüştürülebilir ama birlikte çalışabilirlik genellikle üst verilerle sınırlıdır. Literatürü içeren basılı nesnelere içeriği "bit"lere dönüştürülerek işaretlenmedikçe, sadece geleneksel üst veriler aracılığıyla erişilebilen bilgilerin kullanımı sınırlı kalacaktır. Hatta bazı nesnelere için yeniden işaretleme yapmak gerekecektir. (Örneğin, geçmişte "Yugoslavya" olarak etiketlenmiş olan ve coğrafi bilgi içeren bir nesneyi bugün "Hırvatistan", "Sırbistan", "Kosova" vb. gibi sekiz ayrı etiketle temsil etmek mümkündür.)<sup>21</sup> Öte yandan, türetilmiş bilgiler ve ham veriler için henüz yeterince standart geliştirilmediğinden birlikte çalışabilirlik bu alanda daha da büyük bir sorundur.

Weinberger'in (2007) ikinci derece bilgi düzenleme sistemlerinin dijital bilgileri düzenlemede yetersiz kalması, dijital ortamda bilgi erişim için mutlaka otoriteler tarafından atanan üst verilerin gerekmemesi ve isteyen herkesin özellikle dijital nesnelere üst veri atayabilmesi, bir dijital nesnenin birden fazla yerde depolanabilmesi vb. gibi hususlardaki saptamaları yerindedir. Ama dijital nesnelere daha çok sayıda üst veri atayarak bilgi erişim sistemlerinin performansının da buna bağlı olarak yükseleceği düşünülmemelidir. Daha önce değinilen dil ve belgelerin entelektüel içeriğinin ve bağlamının eksiksiz temsil edilmesiyle ilgili sorunları dijital nesnelere çok sayıda anahtar sözcük atayarak çözmek mümkün değildir.

---

<sup>21</sup> Aslında "Yugoslavya" ile ilgili sorun burada sunulandan daha da karmaşıktır. Coğrafi olarak eski Yugoslavya'nın günümüzde hangi ülkelere karşılık geldiği nispeten daha kolay etiketlenilebilir. Ama "Yugoslavya" olarak adlandırılan coğrafi bölgenin hangi dönemler arasında hangi ülkeleri içerdiğiyse ilgili zamansal (temporal) bilgi genellikle tatminkâr bir biçimde etiketlenmemektedir. Örnek için bkz. Madnick ve Zhu (2006, s. 465-466).

Son yıllarda ikinci derece bilgi düzenleme sistemleri tarafından atanan konu başlıklarının web ortamında kullanıcılar tarafından atanan etiketlerle de desteklenmesi yönünde çeşitli projeler yürütülmektedir. Örneğin, kullanıcılar Flickr'a konulan Kongre Kütüphanesi dermesinde bulunan bazı fotoğrafların tanımlanmasına katkıda bulunmuşlardır (Springer ve diğerleri, 2008). Kopenhag Kraliyet Kütüphanesi de kullanıcıların yarattığı etiketlerden yararlanmaktadır (Hjorland, 2012). Folksonomi ya da etiketleme (tagging) olarak bilinen işlemler sonucunda bilgi sadece kültürel açıdan değil, bireysel ve kurumsal açılardan da sınıflanmaktadır (Glushko, Maglio, Matlock ve Barsalou, 2008). Belgelerin içeriğini tanımlamak için kullanılan konu başlıklarında tutarsızlıklar ve başlık seçiminde kültürel ön yargıların devreye girmesi gibi sorunlara Kongre Kütüphanesi Konu Başlıkları sisteminden örnekler vererek kısaca değinmiştik. İçerik (kültür) ile ilgili konu başlıkları ya da sınıflama etiketleri eserin entellektüel içeriği konusunda genel bir fikir vermektedir. Ama bireysel ve kurumsal sınıflama sistemleri spesifik amaçlar için tasarlandığından bu sistemlerin herkes tarafından kullanımı daha problematik olabilmektedir. Örneğin, bireysel kullanım amacıyla "mutlaka oku", "süper" gibi etiketler atanmış bir belgenin içeriği hakkında bir fikir sahibi olmak mümkün değildir. Kültürel, bireysel ve kurumsal etiketleme sistemlerinin bilgi erişim amacıyla birlikte kullanılması birlikte çalışabilirlikle ilgili sorunları da beraberinde getirmektedir.

Son yıllarda Google gibi şirketler tarafından gerçekleştirilen büyük ölçekli dijitalleştirme projeleri mevcut içerik üzerinde tam metin arama yapılabilmesine olanak sağladığından, sınıflama ve dizinleme yapma gereksiniminin zamanla ortadan kalkacağını düşünenler olabilir. Hatta Google'dan sonra sınıflama gerekli mi sorusu sorulmaktadır (Hjorland, 2012). Sınıflama yapmanın amacı sadece kitapları rafa yerleştirmekle sınırlanırsa dijitalleştirmeye birlikte sınıflama ve düzenlemeye olan gereksinimin de ortadan kalkacağı düşüncesi ilk bakışta doğruymuş gibi görünebilir. Ama bu durum bilgi sınıflama gereksiniminin tamamen ortadan kalkacağı anlamına gelmemektedir. Tam tersine, dijital bilgi yönetiminde sınıflama ve düzenleme daha da önem kazanmaktadır.

Günümüzde çoğu bilgilerin dijital ortamda üretilmesi ve basılı ortamda üretilenlerin de giderek dijital ortama aktarılması bilgi erişim açısından yeni olanaklar sunmaktadır. Basılı bilgilerin hızla dijital ortama aktarılması ve dijitalleştirilen belgelerin tam metinlerinin arama motorlarının örümcekleri tarafından dizinlenmesi bilgi erişim performansını kuşkusuz artırmaktadır. Ama basılı bilgilerin tamamı dijitalleştirilse bile mevcut bibliyografik sistemler (örneğin, çevrimiçi kataloglar) kullanılarak bilgi erişim sorununu çözmek mümkün gözükmemektedir.

### **Arama Motorlarından Soru Yanıtlama Sistemlerine**

Zadeh (2006, s. 163) "Arama Motorlarından Soru Yanıtlama Sistemlerine" başlıklı makalesinde Google başta olmak üzere günümüzdeki arama motorlarını başarılı bulmasına ve performanslarının günden güne geliştiğini teslim etmesine karşın, arama motorlarının bilgi erişim sorununu çözmede yetersiz kaldığını söylemektedir. Zadeh'ye göre bunun temel nedeni arama motorlarının çıkarsama yani tümdengelim (deduction) yeteneklerinin olmamasıdır. Zadeh arama motorlarının iki değerli mantığa ve olasılık kuramına dayanan araçları kullanarak çıkarsama yeteneğine sahip olan soru yanıtlama (question answering) sistemlerine dönüştürülüp dönüştürülemeyeceği sorusunu sormakta

ve bu soruya olumsuz yanıt vermektedir. Zadeh'ye göre bu tür bir dönüştürümün önünde üç engel bulunmaktadır: (1) dünya bilgisi (world knowledge); (2) ilgililik kavramı; ve (3) algılama temelli bilgiden çıkarsama yapma. Aşağıda bu engeller sırasıyla incelenmektedir (Zadeh, 2006, s. 163).

Zadeh'ye göre ilk engel insanların deneyim, iletişim ve eğitimle elde ettikleri dünya bilgisidir. Dünya bilgisinin arama, ilgililik değerlendirmesi ve çıkarsama yapmada merkezi bir rolü vardır. Dünya bilgisi önermesel (Fransa'nın başkenti Paris'tir), kavramsal (iklim), ontolojik (yağış iklimle ilgilidir), varoluşçu (bir kimsenin iki babası olamaz) ve bağlamsaldır (uzun). Dünya hakkındaki bilgilerin çoğu algılara dayanır. Algılara dayalı bilgiler ise kesin değildir, dolayısıyla bu tür bilgiler için iki değerli mantık ve olasılık kuramı kullanılamaz (Zadeh, 2006, s. 163; 2005, slayt 22). Aslına bakılırsa kütüphane dermelerindeki yayınlarda yer alan dünya bilgisinin de kesin, tutarlı, iki değerli mantık ve olasılık kuramı kullanmaya elverişli olduğunu söylemek mümkün değildir. Birbirinin tam tersi kuramlar, görüşler ve araştırma sonuçlarını içeren literatür raflarda yan yana bulunabilmektedir. Bu literatürdeki tüm bilgiler dijitalleştirilse ve bu bilgilere dayanarak dünya bilgisi oluşturulsa bile, çıkarsama yeteneğine sahip olan soru yanıtlama sistemleri, literatüre dayanan mevcut bilgi tabanında (knowledge base) yer alan bilgilerden yararlanarak her soruya doğru yanıt veremez.

İkinci engel çok karmaşık ve henüz çözülmemiş olan ilgililik kavramıdır. İki tür ilgililik vardır: (a) soru ilgililiği ve (b) konu ilgililiği. "Doğal Dil Anlamaya Yeni Bir Yaklaşım" başlıklı bir bildiri makine çevirisiyle ilgili bir soruyla ne kadar ilgilidir? İlgililik değerlendirmesine "anlamsal" ve "istatistiksel" olmak üzere iki şekilde yaklaşılabilir. Arama motorlarında daha çok istatistiksel ilgililik (bağlantı ve sözcüklerin sayılması) kullanılmaktadır. Anlamsal ilgililik değerlendirmesi zordur ve iki değerli mantık ve olasılık kuramıyla çözülemez (Zadeh, 2006, s. 164). Örneğin:

İlgililik değerlendirmesindeki zorluğun temel kaynağı ilgililik fonksiyonunun tümlenememesidir (non-compositionality). Daha spesifik olarak, bir soru,  $q$ , ve iki önermemiz,  $p$  ve  $r$ , olduğunu varsayalım.  $R(q/p, r)$  değeri  $R(q/p)$  ve  $R(q/r)$  değerlerinden oluşturulabilir mi?<sup>22</sup> Yanıt genelde hayırdır. Web ile ilgisi olmayan basit bir örnek;  $q$ : Vera'nın yaşı kaçtır;  $p$ : Vera Irene ile yaşittir;  $r$  ise Irene 65 yaşındadır, olsun. Bu durumda  $R(q/p) = 0$ ;  $R(q/r) = 0$ ; ama  $R(q/p, r) = 1$ 'dir. Bunun anlamı, ilgililik tek başına, başkalarından ayrı olarak (in isolation) değerlendirilemez. Bu da ilgililik ile başkalarından ayrı ilgililik (i-relevance, a-İlgili) arasında bir ayrım yapma gereğine işaret eder. Başka bir deyişle,  $p$  önermesi kendi başına ilgiliyse a-İlgili, kendi başına ilgili değilse a-İlgisizdir, ama diğer önermelerle birlikte ilgili olabilir. (Zadeh, 2006, s. 167)

Konu ilgililiği değerlendirmesinde ise istatistiksel teknikler kullanılmaktadır. Mevcut arama motorlarının konu ilgililiğini değerlendirmede soru ilgililiğini değerlendirmeden daha başarılı olmasının nedeni budur (Zadeh, 2006, s. 164). Nitekim Zadeh'nin Google'da çalıştırdığı test sorularından konusal (topical) olanlar başarılı sonuçlar verirken çıkarsama yapmayı gerektiren sorular başarısız sonuçlar vermiştir.<sup>23</sup> Başka bir deyişle, Google arama

<sup>22</sup>  $R(q/p, r)$  ilgililik fonksiyonunu,  $q$  soruyu veya konuyu,  $p$  ve  $r$ 'nin her biri birer önermeyi, konuyu, belgeyi, web sayfasını veya bu tür nesnelerin bir koleksiyonunu,  $R$  ise  $p$  ve  $r$ 'nin  $q$  ile ilgililik derecesini temsil eder (Zadeh, 2006, s. 167).

<sup>23</sup> Test sorusu örnekleri için bkz. Zadeh, 2006, s. 168-172.

motorunda çıkarsama yapma yeteneği bulunmamaktadır. Aynı durum “hesaplamalı bilgi motoru” (computational knowledge engine) ya da anlamsal arama motoru olarak nitelendirilen Wolfram|Alpha ve Hakia gibi arama motorları için de geçerlidir.

Zadeh’ye göre üçüncü engel ise algılama temelli bilgiden çıkarsama yapılmasıdır. Örneğin:

Soru, *q*, “İsveçliler ortalama ne kadar uzundur?”, mevcut bilgi, *p*, “Çoğu yetişkin İsveçliler uzun boyludur” olsun. Bir başka örnek: Mevcut bilgi “Robert genellikle akşam saat 6’da işten eve döner”, soru “Robert’in 6:15’te evde olma olasılığı nedir?” olsun. Ne iki değerli mantık ne de olasılık kuramı bu tür sorunlara çözüm getirebilir. İşin zorluğu temelde kesin olmayan ve belirsiz önermelerden çıkarsama yapılmasıdır (Zadeh, 2006, s. 164).

Zadeh dünya bilgisi, ilgililik ve çıkarsama yapma sorunlarının altında yatan temel sorunun doğal dil anlama (natural language understanding) sorunu olduğunu vurgulamaktadır. “Dünya bilgisinin ve web bilgisinin çoğu bir doğal dille ifade edilir. Bir doğal dil temelde algıları tanımlama sistemidir. Algılar aslen belirsiz olduğundan doğal diller de, özellikle anlambilim (semantics) alanında, belirsizdir” (Zadeh, 2006, s. 164).

Zadeh’ye göre soru yanıtlamayı mekanikleştirmenin ön koşulu doğal dil anlamayı mekanikleştirmek, doğal dil anlamayı mekanikleştirmenin ön koşulu ise kavramların anlamını kesinleştirmek (precisiation) ve bir doğal dilden önermeler çıkarmaktır.<sup>24</sup> Zadeh dünya bilgisi, ilgililik, çıkarsama ve kesinleştirme sorunlarını çözebilmek için yeni araçlar gerekli olduğunu öne sürmektedir:

Bu araçların başlıcaları Kesinleştirilmiş Doğal Dil (Precisiated Natural Language, PNL), Protoform Kuramı ve Genelleştirilmiş Belirsizlik Kuramıdır (Generalized Theory of Uncertainty, GTU). Bu araçlar bulanık mantıktan -her şeyin bir derece sorunu olduğu ya da olmasına izin verildiği mantık- yararlanır. Bunların en önemlisi genelleştirilmiş sınırlılık kavramıdır. Genelleştirilmiş sınırlılık kavramının önemi . . . bilginin doğası gereği istatistiksel olduğu konusundaki evrensel olarak kabul edilmiş görüşü, genelleştirmenin temeli olarak almasından kaynaklanır. Daha spesifik olarak . . . genelde bilgi bir genelleştirilmiş sınırlılıklar sistemi olarak temsil edilebilir, istatistiksel bilgi özel bir vakayı oluşturur. Bu, çok daha genel, bilgi bakış açısı dünya bilgisi, ilgililik, çıkarsama ve kesinleştirme sorunlarının etkin bir biçimde üstesinden gelmek için gereklidir (Zadeh, 2006, s. 164-165).

Zadeh’ye (2006, s. 165) göre bulanık mantığın standart olmayan bu kavramları, fikirleri ve araçları arama motorlarına çıkarsama yapma yeteneği eklenmesine yol açacaktır. Ama kütüphane dermelerinde bulunan belgeler dijitalleştirildikçe bu belgelerde yer alan bilgiler daha büyük bir bilgi tabanının (knowledge base) parçası haline gelecek ve arama motoru örnekleri tarafından dizinlenecektir. O zamana kadar Zadeh’nin dünya bilgisi, ilgililik, çıkarsama ve kesinleştirme ile ilgili çözümleri olgunlaştırılabilir ve doğal dil anlama ve doğal dildeki belirsizlikleri giderme (language disambiguation) yöntem ve teknolojileri geliştirilebilirse, soru yanıtlama sistemleri kütüphanelerin içerdiği

<sup>24</sup> Çok bilinen bir örnektir: Doğal dil anlama sistemleri “Akraba ziyareti sıkıcı olabilir” (Visiting relatives can be a nuisance) cümlesinin anlamını bağlamdan ve diğer cümlelerden (önermelerden) bağımsız olarak doğru yorumlayamaz. Çünkü bu cümlede sıkıcı olanın akrabalarınızın sizi ziyareti mi yoksa sizin akrabalarınızı ziyaretiniz mi olduğu belirsizdir.

bilgilerden de yararlanabilecektir. Ancak o zaman Weinberger'in sözünü ettiği üçüncü derece sistemler başarıya ulaşmış olacaktır.

Ama bu geçiş sürecinin nispeten uzun zaman alacağı anlaşılmaktadır. Çünkü Swanson (1988) yaklaşık 25 yıl önce "Bilgi Erişim ve Bir Yanılsamanın Geleceği" başlıklı çalışmasında Zadeh'nin de incelediği bazı konulara değinmiştir. Swanson'ın bilgi gereksinimlerinin soru cümlesi olarak formüle edilmesi, ilgililik ve anlam konusundaki "yetersizlik ön kabulleri" (postulates of impotence, PI) aşağıda verilmektedir.

PI 1: Bir bilgi gereksinimi arama sorusu olarak bağlamla ilgili sayısız varsayımdan (presupposition) bağımsız biçimde tamamen ifade edilemez -bağlamın kendisini tam olarak tanımlamak olanaksızdır, çünkü başka şeyler yanında soruyu soranın kendi arka plan bilgisini de içerir. Gerçekten de, arama başlangıçta bilinmeyen bir şey için yapıldığından, yanıt bulunmadıkça soru kesin olarak oluşturulamaz.

PI 2: Bir makine ifade edilen arama isteğini bir dizi uygun arama terimine çevirmesi için programlanamaz. Arama terimleri hipotezler, buluşlar veya tahminlerdir; kuralları yoktur.

PI 3: Bir belge istek yapanın dikkate alabileceği diğer tüm belgelerden bağımsız olarak bir bilgi gereksinimi için ilgilidir denemez. İlgililik sabit değildir; [ilgililiğe] değişken bir çerçevede karar verilir. . . .

PI 4: Bir soru için ilgili bütün belgelerin bulunup bulunmadığını doğrulamak asla mümkün değildir, çünkü ilgililiğe ancak bir belgeyi inceleyerek karar verilir ve kimse uygulamada, hatta ilke olarak bile, tüm belgeleri asla inceleyemez. . . .

PI 5: Makineler anlamı tanıyamaz ve dolayısıyla insan muhakemesinin belgelerin dizinlenmesi ve sınıflanmasına katkısını ilke olarak taklit edemez....

PI 6: Sözcüklerin kullanım istatistikleri ne anlamı temsil eder, ne de anlam yerine geçer. Ama bu tür veriler . . . bir metnin [anlam ve ilgililik açısından] potansiyel olarak verimli alanlarını belirli bir başarıyla işaret edebilir.

PI 7: Bir bilgi erişim sisteminin yineleyici bir süreci destekleme yeteneği insanın bir kerede verdiği ilgililik değerlendirmesine bakılarak değerlendirilemez. Birden çok yinelemelerde sistemin soruyu yaratıcı bir biçimde değiştirmeyi özendirme gibi yeni ölçütler ortaya çıkar.

PI 8: Ya incelikli, ustaca yapılmış ilgililik değerlendirmelerine ya da çok etkili mekanik süreçlere sahip olabilirsiniz, ama ikisine birden değil.

PI 9: Özet olarak, ilk sekiz ön kabul tutarlı bir şekilde etkin ve tam otomatik dizinleme ve erişimin mümkün olmadığını göstermektedir. Bilgi erişimin kavramsal sorunları -anlam sorunları- düşünmekten ya da herhangi bir akıllı davranış türünden daha az derin değildir. Düşünmenin maddi bilgiler içeren bir veri tabanının kurallara göre işletilmesine indirgenebileceği yönünde henüz bir kanıt yoktur. İlgililik değerlendirmemiz ya da düşünmek, diğer şeylerin yanı sıra, hayal gücümüzün mantık, muhakeme ve tutarlılığın soğuk eli tarafından sınırlanmamış ustalıklı sıçramalar yapmasını gerektirir; daha önemlisi kim olduğumuzu, ne olduğumuzu, nasıl bir dünyada yaşadığımızı ve aradığımızı niçin istediğimizi bilmeyi gerektirir. . . . (Swanson, 1988, s. 95).

Swanson makinelerin bilgi erişime büyük yardımı dokunduğunu, ama daha çok bilgiye daha hızlı erişimin tek başına bu işi daha iyi yapmayı öğrendiğimiz anlamına gelmediğini vurgulamaktadır. Nitekim, Swanson'ın soruları ifade etme, ilgililik ve dil



konusundaki yetersizlik ön kabullerinin halen geçerli olduğu görülmektedir. Henüz veri tabanı yönetim sistemleri, veri madenleme teknikleri ve bulanık mantık bu yetersizlik ön kabullerini tamamen giderme ve bilgi erişim sistemlerindeki mevcut durumu değiştirme yönünden pek etkili olamamıştır.

## **Sonuç**

Kolayca görülebileceği gibi, bilgi sınıflama, bilgi düzenleme ve bilgi erişim kütüphanecilik, bilişsel bilimler, bilgisayar mühendisliği gibi birçok disiplini ilgilendiren son derece karmaşık araştırma alanlarıdır. Bilgi sınıflama, bilgi düzenleme ve bilgi erişim sorunları teknik sorunlar olarak görüldüğünde çözümleri nispeten kolaydır, ama kavramsal sorunlar olarak görüldüğünde çözümleri çok zordur. Örneğin, teknik yönden (hız, depolama kapasitesi, bilgi işleme gücü vb. gibi) günümüzde hayal gücümüzü zorlayan gelişmeler olmaktadır. Ama teknik gelişmeler ancak veri tabanlarında depolanan yapısal bilgilerle ilgili sorunların çözümünde bize yardımcı olmaktadır. Oysa tahminlere göre örgütlerdeki bilgilerin sadece %20'si yapısaldır, %80'i ise yapısal olmayan, metin türü bilgilerdir (Mavroudakıs ve Karanikas, 2007).

Metin türü bilgilerin tanımları yani üst verileri yetersizdir. Dijital ortama aktarılan metin türü bilgiler için RDF, XML, OWL (Web Ontology Language) gibi makinece işlenebilir tanımlar gereklidir. Dahası, bir metin dermesinden çeşitli yazılımlar (örneğin, OntoClassify, OntoGen) yardımıyla otomatik olarak ontolojiler inşa edilebilmeli ve bu ontolojiler otomatik olarak yönetilebilmelidir. Ontolojilerin amacı bilgi paylaşımı, işlevleri ise belgeleri sınıflamak ve düzenlemektir. Metin türü verilerden bilgi keşfetmekte kullanılan doğal dil anlama, makine öğrenme, veri madenlemesi ve gizli anlam dizinleme teknikleri geliştirilmelidir. Böylece dijital ortama aktarılan bilgiler yapısal hale getirilebilir, metin türü bilgilerden ve ham verilerden yeni bilgiler türetilir.

Şekil 1'deki piramidin ortasındaki tabandaki bilgi ve verileri düzenleme ve bu bilgilere erişme konusundaki deneyimlerimiz henüz sınırlıdır. Bu yönde geliştirilen çözümler genellikle tek tek alanlara özgüdür (domain-specific). Zadeh'nin sözünü ettiği dünya bilgisi, ilgililik, çıkarsama ve kesinleştirme sorunları tüm alanları kapsayacak şekilde çözümlenirse, ancak o zaman arama motorları soru yanıtlama sistemlerine dönüştürülebilir.

Bilgi sınıflama, bilgi düzenleme ve bilgi erişim eğitimi piramidin sadece üst kısmında yer alan literatür bilgilerinin sınıflanması, düzenlenmesi ve bu bilgilere erişimle sınırlanmamalıdır. Günümüzde özellikle e-bilim projelerinde piramidin orta ve alt kısmındaki bilgileri işleyebilen, sınıflayabilen, düzenleyebilen ve bu bilgilere erişebilen bilgi profesyonellerine, veri bilimcilere (data scientists) gereksinim duyulmaktadır. Mezunlarımızın gerek e-bilim projelerinde gerekse eğitim-öğretim, araştırma ve yönetim süreçlerinde yer alan diğer profesyonellerle birlikte çalışabilmeleri ve etkili olabilmeleri için ders programlarında doğal dil anlama ve işleme, veri madenlemesi, belge mühendisliği, ontoloji geliştirme, makine öğrenimi gibi konulara da yer verilmelidir.

## **Teşekkür**

Metni gözden geçiren ve önerilerde bulunan Prof. Dr. Serap Kurbanoglu, Güven Köse, Müge Akbulut, Gülela Düzyol Doğan ve İpek Şencan'a teşekkür ederim.

## Kaynakça

- Berman, S. (1971). *Prejudices and antipathies: A tract on the LC Subject Heads concerning people*. Metuchen, NJ: Scarecrow.
- Blair, D.C. (1990). *Language and representation in information retrieval*. Amsterdam: Elsevier.
- Bowker, G.C. ve Star, S.L. (2000). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.
- Buckland, M.K. (1991). *Information and information systems*. New York: Praeger.
- Buckland, M.K. (1997). What is a “document”? *Journal of the American Society for Information Science*, 48, 804-809.
- Buckland, M. (2012a). Benefitting from the work of others. Teldap International Conference, Taipei, Taiwan, Feb 21-24, 2012. 16 Nisan 2012 tarihinde <http://people.ischool.berkeley.edu/~buckland/teldapplenary12.ppt> adresinden erişildi.
- Buckland, M.K. (2012b). Obsolescence in subject description. *Journal of Documentation*, 69, 154-161.
- Chu, H. (2003). *Information representation and retrieval in the digital age*. Medford, NJ: American Society for Information Science & Technology.
- Fairthorne, R.A. (1961). *Towards information retrieval*. London: Butterworths.
- Glushko, R.J., Maglio, P.P., Matlock, T. ve Barsalou, L.W. (2008). Categorization in the wild. *Trends in Cognitive Sciences*, 12(4): 129-135. 25 Nisan 2012 tarihinde <http://faculty.ucmerced.edu/tmatlock/papers/paper-categorization-in-wild.pdf> adresinden erişildi.
- Gray, J. (2009). Jim Gray on eScience: A transformed scientific method. T. Hey, S. Tansley ve K. Tolle (Ed.). *The fourth paradigm: Data intensive scientific discovery* içinde (s. xix-xxxii). Redmond, WA: Microsoft Research. 24 Nisan 2012 tarihinde [http://research.microsoft.com/enus/collaboration/fourthparadigm/4th\\_paradigm\\_book\\_jim\\_gray\\_transcript.pdf](http://research.microsoft.com/enus/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf) adresinden erişildi.
- Hjørland, B. (2012). Is classification necessary after Google? *Journal of Documentation*, 68, 299-317.
- Lakoff, G. (1990). *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Madnick, S. ve Zhu, H. (2006). Improving data quality through effective use of data semantics. *Data & Knowledge Engineering*, 59, 460-475.
- Mavroudakos, T. ve Karanikas, H. (2007). Intelligent search via ontology driven metadata analysis (poster). 10th International Protégé Conference, July 15-18, 2007, Budapest, Hungary. 2 Mayıs 2012 tarihinde <http://protege.stanford.edu/conference/2007/posters/mavroudakos.pdf> adresinden erişildi.
- Nisbett, R.E. (2005). *The geography of thought: How Asians and Westerners think differently – and why*. New York: Free Press.
- Nowakowski, P. ve diğerleri (2011). The Collage Authoring Environment. *Procedia Computer Science*, 4, 608-617.
- Olson, H.A. ve Wolfram, D. (2008). Syntagmatic relationships and indexing consistency on a larger scale. *Journal of Documentation*, 64, 602-615.
- Rosenfeld, L. ve Morville, P. (1998). *Information architecture for the World Wide Web*. Sebastopol, CA: O'Reilly
- Salton, G. ve McGill, M.J. (1986). *Introduction to modern information retrieval*. New York: McGraw-Hill.

- Springer, M. ve diğeri. (2008). *For the common good: The Library of Congress Flickr Pilot Experiment*. 21 Nisan 2012 tarihinde [http://www.loc.gov/rr/print/flickr\\_report\\_final.pdf](http://www.loc.gov/rr/print/flickr_report_final.pdf) adresinden erişildi.
- Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge, MA: MIT Press.
- Swanson, D.R. (1988). Historical note: Information retrieval and the future of an illusion. *Journal of the American Society for Information Science*, 39, 92-98.
- Tonta, Y. (1991). A study of indexing consistency between Library of Congress and British Library catalogers. *Library Resources & Technical Services*, 35(2), 177-185.
- Tonta, Y. (1992). Analysis of search failures in document retrieval systems: a review. *The Public-Access Computer Systems Review*, 3(1), 4-53. 22 Nisan 2012 tarihinde <http://journals.tdl.org/pacsr/article/viewFile/6047/5679> adresinden erişildi.
- Tonta, Y. (1995). Bilgi erişim sistemleri. *Türk Kütüphaneciliği*, 9(3), 302-314. 22 Nisan 2012 tarihinde <http://tk.kutuphaneci.org.tr/index.php/tk/article/view/1377/2754> adresinden erişildi.
- Tonta, Y. (2005). Internet and electronic information management. *Information Services & Use*, 25(1), 3-12.
- Tonta, Y. ve Al, U. (2007). Türkiye'nin bilimsel yayın haritası: Türkiye'de dergi yayıncılığı üzerine bir araştırma. (Proje sonuç raporu) (TÜBİTAK Sosyal Bilimler Araştırma Grubu SOBAG Proje No. SOBAG 105K088). Ankara: Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü. 22 Nisan 2012 tarihinde <http://yunus.hacettepe.edu.tr/~tonta/yayinlar/cdmb-projesi-sonuc-raporu-24ocak.pdf> adresinden erişildi.
- Tonta, Y., Bitirim, Y. ve Sever, H. (2002). *Türkçe arama motorlarında performans değerlendirme*. Ankara: Total Bilişim Ltd. Şti.
- Van de Sompel, H. ve Lagoze, C. (2009). All aboard: Toward a machine-friendly scholarly communication system. T. Hey, S. Tansley ve K. Tolle (Ed.). *The fourth paradigm: Data intensive scientific discovery* içinde (s. 193-199). Redmond, WA: Microsoft Research. 24 Nisan 2012 tarihinde [http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th\\_paradigm\\_book\\_part4\\_sompel\\_lagoze.pdf](http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part4_sompel_lagoze.pdf) adresinden erişildi.
- Weinberger, D. (2007). *Everything is miscellaneous: The power of the new digital disorder*. New York: Holt.
- Wilson, P. (1973). Situational relevance. *Information Storage and Retrieval*, 9, 457-471.
- Wolfram, D., Olson, H.A. ve Bloom, R. (2009). Measuring consistency for multiple taggers using vector space modeling. *Journal of the American Society for Information Science and Technology*, 60, 1995-2003.
- Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, 8, 338-353. 21 Nisan 2012 tarihinde <http://www-bisc.cs.berkeley.edu/Zadeh-1965.pdf> adresinden erişildi.
- Zadeh, L.A. (2005). From search engines to question-answering systems – The problems of world knowledge, relevance and deduction (keynote lecture). *6th WSEAS International Conference on Fuzzy Systems (FS '05), June 16-18, 2005, Lisbon, Portugal*. 29 Nisan 2012 tarihinde <http://www.worldses.org/plenary/2005/lisbon/zadeh-2005.pdf> adresinden erişildi.
- Zadeh, L.A. (2006). From search engines to question answering systems – The problems of world knowledge, relevance, deduction and precisiation. Sanchez, E. (ed). *Fuzzy logic and the Semantic Web* içinde (s. 163-210). Amsterdam: Elsevier. 21 Nisan 2012 tarihinde <http://wwwbisc.cs.berkeley.edu/zadeh/papers/From%20Search%20Engines%20to%20Question-Answering%20Systems...2006.pdf> adresinden erişildi.