

Elektronik Kaynakların Bibliyografik Denetimi ve Metadata

Bibliographic Control of Electronic Resources and Metadata

Elif Aytek Kaynak*

Öz

İnternet'teki elektronik kaynaklarının nicelik ve nitelik olarak hızlı artışı dijital çağda bilgi erişim sorunlarının artmasına neden olmuştur. Mevcut küresel arama motorları Web'i "spider" ve "crawler" olarak adlandırılan programlar aracılığı ile tam metin, başlık ve benzerlerinde geçen kelimeleri kullanarak otomatik olarak indekslemektedir. Web dizinleri olarak adlandırılan diğer arama mekanizmaları ise insana dayalı sistemlerdir ve konuları ilgili başlıklar altında hiyerarşik olarak listelemektedir. Ancak, her iki sistem de İnternet'te etkin bilgi erişim için yetersiz kalmaktadır. Elektronik kaynakların doğası ve çeşitliliği, İnternet'te bibliyografik denetime olan ihtiyacın altını çizmektedir. Bu çalışmanın amacı; elektronik kaynakların bibliyografik denetimi ve metadata konuları için genel bir bakışı sağlamaktır.

Anahtar sözcükler: Elektronik kaynaklar, Bibliyografik denetim, Arama motorları, Metadata, Dublin Core, Birlikte işlerlik.

Abstract

The rapid growth of the electronic resources on the Internet, both in quantity and quality, had lead the information retrieval problem to increase in the digital age. Current global search engines indexing the Web automatically via the programs being called "spider" and "crawler" use the terms derived from the fulltext, header etc. Other global searching mechanisms, such as Web directories are human generated and enumerate the subjects hierarchically under the related topics. However, both systems are inadequate for efficient information retrieval on the Internet. The nature and variety of electronic resources under-

* Uzman Kütüphaneci, TÜBİTAK ULAKBİM Cahit Arf Bilgi Merkezi (elif@ulakbim.gov.tr).

score the need for bibliographical control on the Internet. The purpose of this paper is to provide an overview for the bibliographic control of the electronic resources and" the metadata.

Keywords: Electronic resources, Bibliographic control, Search engines, Metadata, Dublin Core, Interoperability.

Giriş

Temeli ilk kez 1960'lı yıllarda atılan İnternet, o yıllarda, bir nükleer savaş olması halinde kesintisiz iletişimi sağlamaya yönelik tasarlanan paket anahtarlama (packet switching) veri iletim tekniğine dayalı askeri bir ağ (ARPANET) niteliğindedir. Bugünkü İnternet'e çok benzer "Galactic Network" kavramından ise ilk kez 1962'de DARPA (Defence Advanced Research Project Agency)'nın başında olan J.C.R Licklider bahsetmiştir. Buna göre, küresel olarak birbirine bağlı ağlar ile isteyen herkes veri ya da herhangi bir programa hızlı ve kolay bir şekilde erişebilecekti (Leiner, Cerf ve Clark, 1997 s.17). Bugün gelinen noktada ise İnternet, özellikle 1990'lı yılların başında Web'in ortaya çıkışı ile birlikte tüm dünyada büyük bir hızla yaygınlaşmıştır. NUA Surveys (2002) tarafından yapılan bir araştırmaya göre Şubat 2002 itibarıyla 544.2 milyon kişi İnternet'e bağlı durumdadır. Ocak 1993'de 1,313,000 olan İnternet'e bağlı makine (host) sayısı ise, Ocak 2002 itibarıyla 147,345,000'lere yükselmiştir (Internet Domain Survey, 2001; Data, 2002).

1990'lı yılların ikinci yarısından itibaren İnternet, bilgi teknolojileri ve elektronik yayıncılıktaki gelişmelerin etkisiyle, elektronik ortamda yayınlanan bilgi kaynakları gerek nicelik, gerekse nitelik ve format olarak büyük bir hızla artmaya ve çeşitlenmeye başlamıştır. Ancak buna karşın, kaynak keşfi, indeksleme ve erişim ile ilgili uygulamaların büyük oranda ihmali, İnternet'te etkin ve hızlı bilgi erişim sorunlarını gündeme getirmiştir.

İnternet mevcut şekliyle, kaynakların yerlere rastgele saçıldığı hemen hemen her türde koleksiyonu içeren küresel bir kütüphaneye benzetilmektedir. Bu durumda, belirli bir organizasyona tabi tutulmamış küresel bir kütüphanede bilgi arayan kişi, İnternet karşısında saatlerini geçirebilmekte ve tüm bu emeğine karşın aradığı bilgiye ulaşamayabilmektedir.

Bu çalışmada, İnternet bilgi kaynaklarına erişim sorunları irdelenmekte ve elektronik kaynakların organizasyonu ve kaynak keşfine yönelik uygulamalara değinilmektedir.

İnternet'te Bilgi Arama Araçları: Arama Motorları

Bilindiği üzere, İnternet'te Web bilgi kaynaklarına erişim için arama motorları (search engines), dizinler (directories) ve meta arama motorları (meta search engines) gibi çeşitli küresel arama araçları kullanılmaktadır.

En popüler olanları arasında *Altavista* (url¹), *Yahoo* (url²), *HotBot* (url³), *Excite* (url⁴), *Lycos* (url⁵), *Google* (url⁶)'ı sayabileceğimiz genel arama motorları, anahtar kelime sorgulaması yoluyla tüm İnternet genelinde arama yapmak için kullanılmaktadır. Arama motorları, "örümcek (*spider*)" ya da "crawler" olarak bilinen yazılımlar aracılığı ile İnternet'i otomatik olarak dolaşarak, Web sayfalarını indeks/katalog olarak adlandırılan veri tabanlarında indekslemektedir. Bu şekilde, bir bilgi tarayıcısı, herhangi bir genel arama motoru ile tarama yaptığıında, aslında bu veri tabanını taramaktadır. *Spider*'lar indekslenen sayfaları düzenli aralıklarla tekrar ziyaret ederek, değişiklikleri tekrar indekslemek suretiyle bu veri tabanlarına aktarmaktadır (Sullivan, 2001).

Yahoo benzeri birçok arama motoru aynı zamanda Web dizinleri sınıfına girmektedir. Dizinlerde indeksleme işlemleri, diğerlerinin tersine otomatik olarak değil, bilgi avcıları (information hunters) ve indeksçiler olarak adlandırılan kişiler tarafından yapılmakta ve kaynaklar, belirli kategori ve konular altında, hiyerarşik düzende sınıflama suretiyle indekslenmektedir. Ancak, dizinlerde kategorizasyon ve konu bölümlenmelerinin güçlüğü, başlıklar altında uzun listeler oluşma durumu ve çapraz yönlendirmelerin yetersizliği gibi nedenler bilgi erişim açısından sorun teşkil etmektedir. Bunun yanı sıra, indekslemenin elle (manuel) yapılması, otomatik indekslemeye oranla çok daha emek gerektiren pahalı bir işlem olup, İnternet'in büyüme hızının gerisinde kalmaktadır.

İnternet'te bilgi arama araçlarından bir diğeri meta-arama (meta-search engines) motorlarıdır. Bunlar, sorgulamada kullanılan kelime/leri çeşitli ara-

ma motorlarından tarayarak elde ettikleri sonuçları "arama motoru adı" ve "arama sonucu erişim yüzdesi" ile birlikte tek bir ekranda listelemektedir. *Metacrawler* (url⁷), *Dogpile* (url⁸) *Mamma* (url⁹), *ixquick* (url¹⁰) ve *Search.com* (url¹¹) bu tür popüler meta-arama motorlarından birkaçıdır ("Metacrawlers", 2001).

Bir diğer gruba giren "All in one search" türü arama motorları ise, meta-arama motorlarının tersine, kişiye çeşitli arama motorlarının isimlerini listelemektedir. Bu şekilde, bilgi arayan kişi, sorgulamayı kendisinin seçeceği arama motorları üzerinden yapabilmektedir. *SearchBug.com* (url¹²) *Skworm* (url¹³), *Vivisimo* (url¹⁴), ve *Proteus* (url¹⁵) bu gruba giren arama motorlarıdır ("Metacrawlers", 2001).

Tüm İnternet'i tarayan küresel arama motorlarının yanı sıra, bölgesel arama motorları da bulunmaktadır. Bunlara örnek olarak, İngiltere için *Altavista.UK* (url¹⁶) *EverydayUK* (url¹⁷) ve Orta Doğu için *Arabist* (url¹⁸), *Middle East Internet Pages* (url¹⁹) verilebilir ("Regional search engines",2001). Türkçe kaynakları indeksleyen arama motorlarından bazıları ise, *List2000* (url²⁰), *Find-İt* (url²¹), *Arabul* (url²²), *IndexTurkey* (url²³), *Netbul* (url²⁴) ve *Arama!com* (url²⁵) 'dur.

İnternet'te bilgi aramak için çeşitli alternatiflerin varlığına karşın, NEC Araştırma Enstitüsünün yaptığı araştırma sonuçlarına göre, küresel arama motorlarının çoğu Web'in % 10'undan azını indeksleyebilmektedir. Bugün en kapsamlı arama motorlarının tümünün toplamı dahi Web'i ancak % 40'lar civarında indeksleyebilmektedir (Hwei ve Smeltzer, 2000). Bu veriler, bir yandan Web'in ne denli hızlı büyüdüğünü, bir yandan da Web'deki mevcut indekslemelerin ne denli yetersiz kaldığını göstermektedir.

Küresel arama motorları Web sayfalarını, HTML içindeki başlıklar (title, header) ve Meta TAG'lerde verilen kelimeleri kullanmak suretiyle ya da kelimelerin metin içinde kullanım sıklıklarına göre, otomatik yazılımlar aracılığı ile indekslemektedir. Bu şekilde, arama sonuçları, sorgulamada kullanılan anahtar kelimeler ile indekslemede kullanılan ve metin içerisinde geçen kelimelerin eşleştirilmesi yoluyla listelenmektedir. Bu tür bir eşleşmenin olmaması durumunda, kaynağa, aranan konuyla doğrudan ilgili de olsa, erişileme-

mektedir. Bilgi aramadaki bu önemli engeli aşabilmenin yolu, yapılandırılmış bir tanımlamayla, ortak bir terminolojiyi kullanmaktan geçmektedir. Standart konu başlıkları listeleri, yapılandırılmış kavramsal dizinler (thesaurus) aracılığı ile konu tanımlamada kontrollü terimler kullanma ve kaynakların alan tanımlamalarını yapma bu sorunu çözmeye yöneliktir (Milstead ve Feldman, 1999).

İyi bir indeks, bilgi arayan kişiyi ilgili kaynağa kolayca yönlendirebilmeli, diğer bir deyişle, aramada “erişim isabeti” (recall) ile “kesin isabet” (presicion) arasında denge kurmalıdır. Web için küresel indeks sağlayan arama motorlarının çoğu tam metin indekslemeye dayalı olduklarından, sorgulama sonucu kesin isabet oranı düşük olmaktadır (Cleveland, 1990, s.143). Tam metinden yapılan indekslemelerde, aranılan konuyla ilgili birçok sayfanın gözden kaçırılma oranı oldukça yüksektir. Bunun yanı sıra, sorgulama yapılan arama motorunun aranılan kaynağı indekslememiş olma olasılığı, sitelerin HTML metinlerine ek olarak resim (image), veri tabanı, PDF gibi farklı formatlarda kaynakları içerebilmesi, arama motorlarının 2-3 hiyerarşik seviyede indeksleme yapması, güncellemelerin yavaşlığı, indekslenen kaynaklarda bir konu bölümlenmesine gidilmemesi ve sayfalar arası içi içe yönlendirmelerin varlığı gibi nedenlerden ötürü ilgili kaynaklara erişim oranı düşük olmaktadır (Taylor, 1999).

İnternet'te bilgi erişimi etkileyen bir diğer sorun ise, elektronik bilgi kaynaklarının dinamik yapısı nedeniyle sürekli adres değiştirebilme ve silinebilme olasılıklarının bulunmasıdır. Bu tür sabit olmayan kaynakları indeksleme, bilgi erişim açısından sorun teşkil etmektedir. Bu konuda, Tekbiçim Kaynak Adı (URN), Tekbiçim Kaynak Belirteci (URI) ve Dijital Nesne Tanımlama (DOI) türü standart sistemler üzerinde çalışılmaktadır (Tonta, 1996).

Elektronik Bilgi Kaynaklarının Tanımlanması: Metadata

Metadata, kısaca “veri hakkında veri” ya da “bilgi kaynağının tanımlanması” olarak da ifade edilmektedir. “Meta” Yunanca kökenli bir kelime olup “değişim” anlamına gelmektedir. O nedenle metadata kelime anlamı itibarıyla “form değiştiren veri” şeklinde de tanımlanabilir (A User Guide, 1998).

Metadata, aslında kütüphanecilerin çok yakından bildiği katalog ve indekslere eş değer bir kavramdır (Miller, 1996). Bir kütüphane kataloğu da işlevi itibarıyla bir metadattır. Basılı bilgi kaynakları için indeks ve kataloglarda bibliyografik alan (başlık, yazar, kaynak vb.) tanımlamaları yapma, aramalarda kesin isabet oranını artırmak için kontrollü terimler kullanma türü uygulamalar, kütüphanecilikte uzun yıllardır kullanılmaktadır. Ancak, günümüzde bilginin kayıtlı olduğu ortam ve sunum şekillerinin değişmesi ile birlikte, ağ bilgi kaynaklarının belirli özel alanlarda tanımlanması, ya da bir başka deyişle makinece anlaşılabilir yapılandırılmış bilgiler “*metadata*” olarak ifade edilmektedir.

Yakın geçmişe kadar elektronik bilgi kaynakları için metadatalar oluşturma işlemi büyük oranda ihmal edilmiştir. Kaynağa kimlik vermek olarak da yorumlayabileceğimiz metadatalar kaynağın içeriğini etiketleme, bibliyografik tanımlama yapmak suretiyle etkin kaynak keşfi ve bilgi erişimi olanaklı kılmaktadır (Milstead ve Feldman, 1999).

Kütüphanecilerin uzun yıllardır basılı kaynakları kataloglamada kullandıkları AACR2 ve MARC türü standart sistemler, mevcut şekilleri ile elektronik kaynakları tanımlamada yetersiz kalmaktadır. Bu nedenle, çalışmalar bir yandan geleneksel sistemleri elektronik kaynakları tanımlamada kullanabilecek nitelikte yenileme, bir yandan da çeşitli disiplin ve kullanıcı gruplarının gereksinimlerine özel, farklı seviyelerde metadata sistemleri geliştirme ve sistemler arasında etkileşimli ortamlar oluşturma yönündedir (Vellucci, 1997).

Kütüphaneciler için “organizasyon” bibliyografik denetime eş değerdir. Bibliyografik denetim ise “kayıtlı bilgiye erişim için organizasyon” anlamında olup, kataloglama ve indeksleme şeklinde iki yaygın türü mevcuttur. Kataloglama mevcut koleksiyona (dergi, kitap vd.), indeksleme ise koleksiyonun içeriğine (makale vd.) erişime yöneliktir. Web ortamında “koleksiyon” kavramı, genel anlamda düşünüldüğünde, çok sayıda Web sayfalarından oluşan “Web site”lerine karşılık gelmektedir. Bu durumda, gerek Web sayfaları gerekse Web siteleri bilgiyi barındıran ve bibliyografik denetim gerektiren birer bibliyografik ünite durumundadır (Weibel, 1995).

Ağ bilgi kaynakları, yapı itibarıyla basılılardan farklı, dinamik ve değişkendir. İnternet’te bugün var olan bir kaynak ertesi gün adres değiştirmiş ya da

tamamen silinmiş olabileceği gibi, aynı kaynak farklı formatlarda (Postscript, ASCII text vb.) ve farklı adreslerde yer alabilmektedir. Tüm bunlar ağ bilgi kaynaklarının organizasyonunu güçleştiren unsurlardır (Heery, 1996). Günümüzde, ağ bilgi kaynaklarının nitelik ve nicelik olarak hızlı artışı ve beraberinde getirdiği bilgi erişim sorunlarının çözümüne yönelik, disiplinlere özel, farklı seviyelerde çeşitli metadata çalışmaları yürütülmektedir. Bunlardan birisi de, temel bibliyografik tanımlama ve kullanım kolaylığı açısından yaygın olarak kullanılan "Dublin Core Metadata" (url²⁶)'dir.

Dublin Core

Kısaca "Dublin Core" ya da "DC" olarak bilinen "Dublin Core Metadata Öge Set'i", 1-3 Mart 1995 tarihleri arasında, OCLC (Online Computer Library Center) ve NCSA (National Center for Supercomputing Applications)'in sponsorluğunda Dublin'de düzenlenen ve kütüphaneci, arşivci, yazar, yayıncı, yazılımcı, araştırmacı gibi farklı disiplinlerden çok sayıda uluslararası temsilcinin katılımıyla gerçekleşen çalıştay (workshop) sonucunda belirlenmiş bir metadata setidir. Metin, ses, görüntü gibi birçok farklı formatta olabilen, "Belge Benzeri Nesnelere" (Document Like Objects - DLOs) olarak adlandırılan, ağ bilgi kaynaklarının keşfini kolaylaştırmak için gerekli 15 metadata ögesinden oluşmaktadır (Weibel, 1995). Bunlar;

1. *Konu (Subject)*: Kaynağın konusu ile ilgili anahtar kelimeler vb.
2. *Başlık (Title)*: Kaynağın ismi.
3. *Yazar/Yaratıcı (Author/Creator)*: Kaynağın entellektüel içeriği ve oluşturmundan sorumlu kişi(ler) ya da kurum(lar).
4. *Tanımlama (Descriptions)*: Kaynağın içeriği ile ilgili kısa metinsel tanımlama.
5. *Yayıncı (Publisher)*: Kaynağı yayımlayan kurum ya da kuruluş.
6. *Diğer Aracılar/Katkıda Bulunanlar (Other Agents/Contributors)*: Kaynağa katkıda bulunanlar. (Örn. Editör(ler), Çevirmen(ler) vd.)
7. *(Tarih) Date*: Kaynağın yayımlandığı tarih.

8. *Nesne Türü (Object Type)*: Kaynağın türü. (Örn. Roman, şiir, resim, sözlük, Web sayfası vb.)
9. *Biçim (Form)*: Kaynağın formatı. (Örn. Postscript, text/html,ASCII vb.)
10. *Tanımlayıcı (Identifier)*: Kaynağı özgün olarak tanımlayan alanlar. (Örn. URLs, URNs, ISBN vd.)
11. *İlişki (Relation)*: Kaynağın diğer kaynaklarla ilişkisi. (Örn. Belge içinde bir resim, kitap'ın bir bölümü vb.)
12. *Kaynak (Source)*: Basılı ya da elektronik ortamda yayımlandığı kaynağın adı.
13. *Dil (Language)*: Entellektüel içeriğin dili.
14. *Kapsam (Coverage)*: Kapsamı.
15. *Haklar (Rights)*: Telif ve kullanım hakları ile ilgili bilgiler.

Dublin Core öğelerinden ilk 13'ünün anlamları açıkça tanımlanmasına karşın, "kapsam" ve "haklar" gibi bazı öğeler (14. ve 15.) henüz deneysel aşamadır.

Metadata bilgileri, HTML işaretleme dilinde başlık <HEAD>....</HEAD> alanında yer alan <META> ve <LINK> işaretleri (tags) ile verilmektedir. Tanımlamalar META alanında "İSİM" (NAME) ve "İÇERİK" (CONTENT) başlıkları altında yer almaktadır. Bir örnek vermek gerekirse, Husby (1997), tarafından yazılan "*Metadata: ELAG'97, Gdansk, June 18th 1997*" başlıklı makalenin Dublin Core kullanılarak oluşturulan metadatası aşağıdaki gibidir.

```
<META NAME="DC.date.current" CONTENT="(SCHEME=ISO31) 1997-06-11">
```

```
<META NAME="DC.title" CONTENT="Metadata">
```

```
<META NAME="DC.title.subtitle" CONTENT="ELAG'97, Gdansk, June 18th 1997">
```

```
<META NAME="DC.creator.name" CONTENT="Husby, Ole">
```

```
<META NAME="DC.creator.email" CONTENT="ole.husby@bibsys.no">
```

```
<META NAME="DC.creator.affiliation" CONTENT="BIBSYS">
```


<META NAME="DC.subject" CONTENT="Metadata, Network documents, f
Dublin Core, Nordic Metadata project">

<META NAME="DC.subject" CONTENT="(SCHEME=HUMORD) Katalogisering">

<META NAME="DC.subject" CONTENT="(SCHEME=HUMORD) Automatisk indekse-
ring">

<META NAME="DC.subject" CONTENT="(SCHEME=UDC) 681.3:02">

<META NAME="DC.subject" CONTENT="(SCHEME=DDC) 005.72">

<META NAME="DC.description" CONTENT="The article discusses cataloguing of
network documents and metadata systems. A special treatment is given to the Dublin Co-
re system, and the Nordic Metadata project.">

<META NAME="DC.publisher" CONTENT="BIBSYS">

<META NAME="DC.date.creation" CONTENT="(SCHEME=ISO31) 1997-06-10">

<META NAME="DC.type" CONTENT="Article">

<META NAME="DC.format" CONTENT="(SCHEME=imt) text/html">

<META NAME="DC.identifier.url" CONTENT="http://www.bibsys.no/elag97/metada-
ta.html">

<META NAME="DC.language" CONTENT="(SCHEME=NISOZ39.53) ENG">

<META NAME="DC.rights" CONTENT="Public domain">

DC'nin tüm öğeleri seçimli (optional) ve tekrar edilebilir özellikte olup, di-
siplin, dil ve kültür farkı gözetmeksizin, her türde kaynağın çekirdek tanımla-
masına uygun ve evrenselidir. DC'nin Fransızca, Almanca, İspanyolca, Arap-
ça ve Çince gibi hâlihazırda birçok dilde sürümü (version) bulunmaktadır. Te-
melde, kaynakların yazar ya da bilgi sağlayıcıları tarafından tanımlanmasına
yönelik tasarlanmasından ötürü, herkesçe kolay anlaşılır, basit öğelerden
(simplicity) oluşmaktadır. Kullanımı için MARC türü standartlarda olduğu gi-
bi, uzman kataloglama bilgisine gerek yoktur. Bunun yanı sıra, 15 çekirdek
ögeye ek olarak niteleyiciler (qualifiers) ile, tanımlanan kaynağa ilişkin daha

detaylı bilgiler de verilebilmektedir. Niteleyiciler, kodlama şemaları (encoding schemes) ve öge iyileştirme (element refinement) şeklinde iki türdedir (Dublin Core Qualifiers, 2000). Örneğin, konu (subject) alanında "scheme" olarak Kongre Kütüphanesi Konu Başlıkları (LCSH), Dewey Onlu Sınıflama Numarası (DDC), Tıp Konu Başlıkları (MeSH) gibi farklı sistemlerden tanımlama bilgileri verilebilmektedir.

Örn:

```
<META NAME="DC.Subject" SCHEME="LCSH" CONTENT="metadata">
```

Öge iyileştirme olarak adlandırılan tür (type) niteleyicileri ise, tanımlanan alana ilişkin daha spesifik tanımlamalar eklemeye kullanılmaktadır. Örneğin "Tarih" alanına, kaynağın oluşturulduğu, yayımlandığı, en son gözden geçirildiği ve değişiklik yapıldığı, hangi tarihler arasında kullanılır (available) olduğu gibi niteleyici bilgiler eklemek mümkündür.

Dublin Core Metadata girişimi, temelde kaynak keşfini kolaylaştırmak için, elektronik kaynakların içerik tanımlamasına yönelik olmasının yanı sıra, farklı metadata setleri arasında birlikte işlerlik (interoperability) sağlamayı da amaçlamaktadır. DC, kullanım kolaylığı, genişletilebilirliği, disiplin bağımsız olması ve birlikte işlerliğe uygun olması gibi birçok nedenden dolayı İnternet genelinde sayısız metadata projesinde yaygın olarak kullanılan bir *de facto* standart durumunda olup, resmi bir standart haline getirilmesi yönünde çalışmalar yürütülmektedir.

Günümüzde birçok ülkede yürütülen metadata projelerinde Dublin Core kullanılmaktadır. OCLC-CORC (Cooperative Online Resource Catalog) (url²⁷), Kaliforniya Üniversitesi – Berkeley, Dijital Kütüphane Kataloğu (Digital Library Katalog) (url²⁸), Avustralya Eğitim Ağı (EdNA-Education Network Australia) (url²⁹), Eğitim Materyalleri Geçiti (GEM-Gateway to Educational Materials) (url³⁰), Hollanda Milli Kütüphanesi, Nordic Metadata Projeleri ve Avrupa'daki konu tabanlı metadata projesi DESIRE, bu projelerden bazılarıdır (An Oveview of.Resource Description Issues, 1998).

Metadatalar Arası Birlikte İşlerlik

Ağ bilgi kaynaklarının makinelerce anlaşılabilir, yapılandırılmış veriler düzeyinde tanımlandığı metadatalar, elektronik bilginin keşfini ve erişim oranını arttırmaktadır. Bu nedenle, Web’de, etkin bilgi erişimi için elektronik kaynakların tanımlanmasına yönelik metadata uygulamaları hızla yaygınlaşmaktadır.

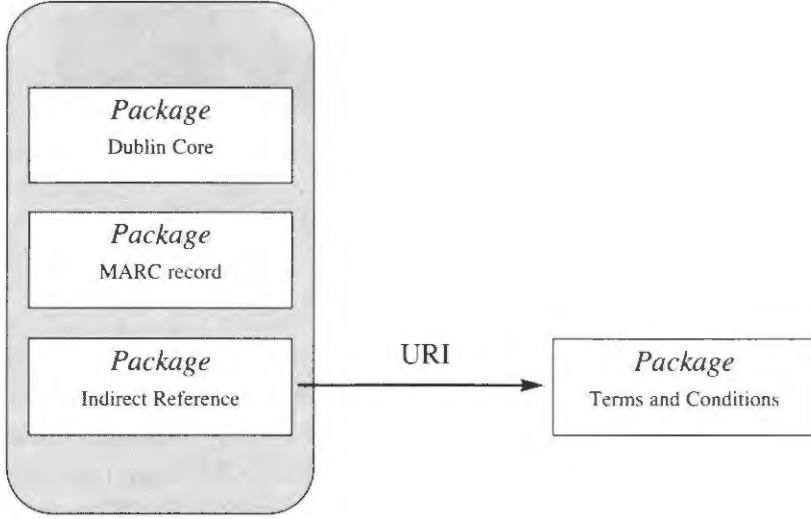
Elektronik bilgi kaynaklarının çeşitliliği ve disiplin özel tanımlamalara duyulan gereksinim, farklı tür ve seviyede metadata formatlarının (DC, IAFA, TEI Header, RFC 1807, MARC vd.) geliştirilmesine neden olmuştur. Ancak bu noktada, farklı formatlar arası etkileşimi mümkün kılacak uygulamalara gerek duyulmaktadır. Heterojen koleksiyonlar ve metadata formatları arasındaki etkileşim, uygulamalar arası birlikte işlerliğin (interoperability) sağlanması ile mümkün olmaktadır.

Whatis.com’da, birlikte işlerliğin sözlük karşılığı, “belirli bir sistem ya da ürünün, diğer sistem ya da ürünlerle, kullanıcı tarafından özel bir çaba gerektirmeksizin işlerliği” olarak verilmektedir (Miller, 2000). Metadatalar arası birlikte işlerlik, farklı setlerin etkileşimli çalışabilecekleri çatı (framework) yapıları ile sağlanmaktadır. Warwick Çatısı (Warwick Framework) ve Kaynak Tanımlama Çatısı (Resource Description Framework– RDF) türü uygulamalar, metadatalar arası birlikte işlerliği sağlamaya yönelik birleştirici yapılar sunmaktadır.

Warwick Çatısı (Warwick Framework)

Warwick Framework, UKOLN ve OCLC’nin organizasyonu ile Nisan 1996’da Warwick Üniversitesinde gerçekleşen, ikinci DC Metadata çalıştayının sonucu bir uygulamadır. Metadatalar arası değişim ve birlikte çalışılabilirliğe yönelik farklı metadata paketlerini birleştirici bir yapı sunmaktadır. Şekil-1’de görüleceği üzere basit bir Warwick Çatısı uygulaması, farklı formatlardan metadata bilgilerinin yer aldığı “paketler” (örneğin Dublin Core, MARC kaydı) ve bunları içeren bir taşıyıcı (container) yapıdan oluşmaktadır (Lagoze, 1996). Bu tür uygulamalarda farklı adreslerdeki metadatalara referans için güvenilir, kalıcı ve özgün adres tanımlamayı olanaklı kılan Tek Biçim Kaynak Numara-

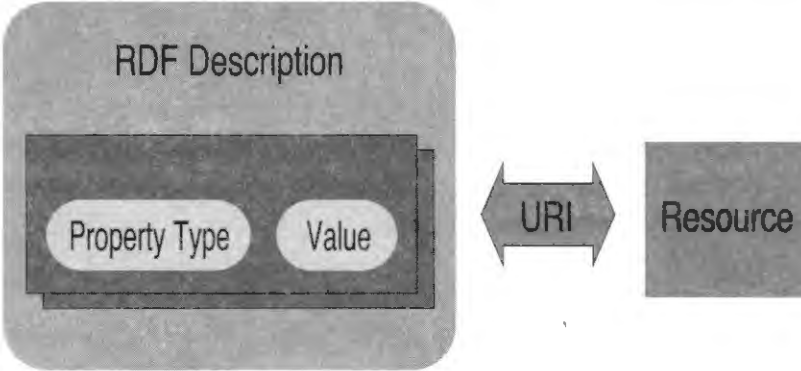
sı (URNs), PURL (Persistent URL), Handle ve Tekbiçim Kaynak Belirteci (URIs) türü standartlar kullanılmaktadır.



Şekil -1. Warwick Framework Modeli (Lagoze, 1996)

Kaynak: Tanımlama Çatısı (Resource Description Framework-RDF)

World Wide Web Konsorsiyumu (W3C) tarafından geliştirilen Kaynak Tanımlama Çatısı (RDF), farklı metadata uygulamalarını ortak semantik, söz dizimi ve yapı altında birleştirmek koşuluyla metadatadalar arası etkileşimi ve birlikte işlerliği olanaklı hale getirmektedir. RDF uygulamasında, metadata tanımlama ve değişiminde ortak söz dizimi için XML (eXtensible Markup Language) dili kullanılmaktadır (Miller, 1998). Bir XML uygulaması olan RDF, URI ile isimlendirilen herhangi bir elektronik belgenin (doküman) makinece anlaşılır biçimde tanımlanmasını sağlamaktadır. Temel bir RDF, Şekil-2'de görüleceği üzere, URI ile tanımlanan bir "kaynak", yazar, konu vb. tanımlama öğelerine karşılık olarak kullanılan "özellik türü" (property type) ve buna ilişkin "değer" (value) ilişkilerinden oluşmaktadır (Iannella, 1998).



Şekil -2. Basit RDF Modeli (Iannella, 1998)

RDF'de kaynağa ilişkin URI'lara işaret eden "XML Namespace"ler kaynağın özgün bir biçimde tanımlandığı özellikler setini belirler. Şekil-3'de görüleceği üzere "Namespace" olarak RDF ve DC gibi iki ayrı metadada sisteminden tanımlama bir arada verilebilmektedir. Şekil-3'de URI, ana bölümde belirtilen "http://dstc.com.au/report.html" adresli kaynağa ilişkin DC tanımlamalarını işaret etmektedir. Bunun gibi, aynı kaynağa ilişkin birçok diğer metadada formatında tanımlama bilgileri de yine RDF yapısı içerisinde belirtilebilmektedir.

```
<? xml version="1.0" ?>
<RDF xmlns = "http://w3.org/TR/1999/PR-rdf-syntax-19990105#"
  xmlns:DC = "http://purl.org/DC#" >
  <Description about = "http://dstc.com.au/report.html" >
    <DC:Title> The Future of Metadata </DC:Title>
    <DC:Creator> Jacky Crystal </DC:Creator>
    <DC:Date> 1998-01-01 </DC:Date>
    <DC:Subject> Metadata, RDF, Dublin Core </DC:Subject>
  </Description>
</RDF>
```

Şekil-3. Basit bir RDF modelinin XML'de gösterimi (Iannella, 1998).

Sonuç

İnsanoğlunun hayalinin dahi ötesinde bir hızla büyüyen küresel ağ, yeni teknolojileri, buna bağlı olarak yeni sorun ve çözümleri de beraberinde getirmektedir. Günümüz İnternet'inde mevcut en büyük sorunlardan birisi de, nicelik ve nitelik olarak büyük bir hızla artan ağ bilgi kaynaklarına etkin erişim sorunudur. Her gün bir yenisinin eklendiği küresel arama motorlarının, mevcut şekilleriyle İnternet'i indeksleme konusunda yetersiz kaldıkları artık bir gerçektir. Bu nedenle, kaynak keşfi ve bilgi erişim için ağ bilgi kaynaklarının tanımlanması yönünde çok sayıda metadata çalışması yürütülmektedir.

Metadataalarının oluşturulması ile kaynak keşfi ve bilgi erişim oranlarının artırılması yönündeki çalışmalar, arama motorlarının yakın gelecekte "akıllı araçlar" (intelligent agents) olarak çalışmalarını sağlayacaktır. Bu konuda Web'in yaratıcısı Tim Berners-Lee'nin öngörüsü, "Semantik Web" olarak tanımlanan ve bugünkünden çok farklı "makinece anlaşılabilir, işlenebilir metadataalar Web'inin, geleceğin Web'i olacaktır. Mevcut şekilleriyle tam metinlerde, başlıklarda vb. öğelerde geçen kelimeleri çakıştırma yoluyla çalışan arama motorları, yakın gelecekte, farklı metadataalar arası birlikte işlerliği mümkün kılan RDF türü uygulamalar ve güçlü kavramsal dizinler desteğinde akıllı sistemler şekilde çalışacaklardır.

Kaynakça

A User Guide for Simple Dublin Core. Draft Version 5. (1998, June). [Çevrim içi]. Elektronik Adres: <http://www.lib.helsinki.fi/meta/UserGuide5.html> [04.04.2001].

An Overview of Resource Description Issues. A review of metadata: a survey of current resource description formats. Work Package 3 of Telematics for Research project DESIRE (RE 1004). (1998, Jun). [Çevrim içi]. Elektronik Adres: http://www.ukoln.ac.uk/metadata/desire/overview/rev_01.htm [12.01.2001].

Cleveland, D.B. ve Cleveland, A.B (1990) *Introduction to indexing and abstracting.* (2nd ed.). Englewood, Colorado.Libraries Unlimited.

- Data on Internet activity worldwide.* (February, 2002). [Çevrim içi]. Elektronik Adres: <http://www.gandalf.it/data/data1.htm> [05.03.2002].
- Dublin Core Qualifiers* (2000, Jan). [Çevrim içi]. Elektronik Adres: <http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/> [07.02.2001].
- Heery, R.(1996). *Review of Metadata Formats. UKOLN.* [Çevrim içi]. Elektronik Adres: <http://www.ukoln.ac.uk/metadata/review.html> [01.01.2001].
- Hwei, J.,K. ve Smeltzer, L. (June, 2000). Finding the needle: Controlled vocabularies, resource discovery, and Dublin Core. *Library Collections, Acquisitions, and Technical Services*, 24(2): 205-215.
- Husby, O. (1997, June) *Metadata: ELAG'97, Gdansk, June 18th 1997.* [Çevrim içi]. Elektronik Adres: <http://www.bibsys.no/elag97/metadata.html> [14.04.2001].
- Iannella, R. (1998, September). An Idiot's Guide to the Resource Description Framework. *The New Review of Information Networking* [Çevrim içi], 4. Elektronik Adres: <http://archive.dstc.edu.au/RDU/reports/RDF-Idiot/> [05.03.2001].
- Internet Domain Survey, January 2001. *Number of Hosts advertised in the DNS Internet Software Consortium.* (2001). [Çevrim içi]. Elektronik Adres: <http://www.isc.org/ds/WWW-200101/index.html> [24.03.01].
- Jul, E. (1997, October) Cataloging Internet Resources: Survey and Prospectus. *ASIS Bulletin* [Çevrim içi], Elektronik Adres: <http://www.asis.org/Bulletin/Oct-97/jul.htm> [04.12.2000].
- Lagoze, C. (1996, July/August). The Warwick Framework: A Container Architecture for Diverse Sets of Metadata. *D-Lib Magazine* [Çevrim içi], Elektronik Adres: <http://www.dlib.org/dlib/july96/lagoze/07.lagoze.html> [05.10.1999].
- Leiner, B. M., Cerf, V.G. ve Clark, D. D (1997, July / August). A Brief History of the Internet Part 2, *On the Internet*,3(3):16-25.
- Metacrawlers. *Search Engine Watch.* (2001). [Çevrim içi]. Elektronik Adres: <http://searchenginewatch.com/links/Metacrawlers/> [24.03.2001].

- Miller, E. (1998, May). An Introduction to the Resource Description Framework. *D-Lib Magazine* [Çevrim içi], Elektronik Adres: <http://www.dlib.org/dlib/may98/miller/05miller.html> [20.05.2001].
- Miller, P. (1996, September). Metadata for the masses. *Ariadne* [Çevrim içi], Elektronik Adres: <http://www.ariadne.ac.uk/issue5/metadata-masses/> [06.12.2000].
- _____ (2000, June). Interoperability. What is it and why should I want it ?” *Ariadne* [Çevrim içi]. Elektronik Adres. <http://www.ariadne.ac.uk/issue24/interoperability/intro.html> [28 Haziran 2001].
- Milstead, J. ve Feldman, S. (1999, January). Metadata:Cataloging by any other name. *Online* [Çevrim içi], Elektronik Adres: <http://www.online-inc.com/onlinemag/OL1999/milstead1.html> [02.09.1999].
- NUA Surveys. *How Many Online?* (2002). [Çevrim içi]. Elektronik Adres: http://www.nua.ie/surveys/how_many_online/world.html [25.03.2002].
- Regional Search Engines. *Search Engine Watch*. (2001). [Çevrim içi]. Elektronik Adres: http://searchenginewatch.com/links/Regional_Search_Engines/ [25.03.2001].
- Sullivan, D. (2001, June 26). *How Search Engines Work*. *Search Engine Watch*. (2001). [Çevrim içi]. Elektronik Adres: <http://searchenginewatch.com/webmasters/work.html> [25.03.2001].
- Taylor, C. (1999, April). *An Introduction to Metadata*. [Çevrim içi]. Elektronik Adres: <http://www.library.uq.edu.au/iad/ctmeta4.html> [04.11.2000].
- Tonta, Y. (1996, Eylül). Internet, Elektronik Kütüphaneler ve Bilgi Erişim. *Türk Kütüphaneciliği* [Çevrim içi], 10(3): 215-230. Elektronik Adres: <http://yunus.hun.edu.tr/~tonta/inetekutbe.html> [02.05.2001].
- Vellucci, S. L. (1997, October) Options for Organizing Electronic Resources: The Coexistence of Metadata. *ASIS Bulletin* [Çevrim içi], Elektronik Adres: <http://www.asis.org/Bulletin/Oct-97/vellucci.htm> [04.03.2001].
- Weibel, S. (1995, July). Metadata:The Foundations of Resource Description. *D-Lib Magazine* [Çevrim içi], Elektronik Adres: <http://www.dlib.org/dlib/july95/0/weibel.html> [09.04.2001].

Elektronik Adresler (URLs):

1. Altavista arama motoru - <http://www.altavista.com>
2. Yahoo arama motoru - <http://www.yahoo.com>
3. HotBot arama motoru - <http://www.hotbot.com>
4. Excite arama motoru - <http://www.excite.com>
5. Lycos arama motoru - <http://www.lycos.com>
6. Google arama motoru - <http://www.google.com>
7. MetaCrawler meta arama motoru - <http://www.metacrawler.com>
8. Dogpile meta arama motoru - <http://www.dogpile.com/>
9. Mamma meta arama motoru - <http://www.mamma.com/>
10. Ixquick meta arama motoru - <http://www.ixquick.com/>
11. Search.com meta arama motoru - <http://www.search.com/>
12. SearchBug.com -<http://www.searchbug.com/>
13. Skworm - <http://www.skworm.com/>
14. Vivisimo - <http://vivisimo.com/>
15. Proteus - <http://www.thrall.org/proteus.html>
16. AltavistaUK - <http://uk.altavista.com>
17. EverydayUK - <http://www.everydayuk.co.uk>
18. Arabist - <http://www.arabist.com/>
19. Middle East Internet Pages - <http://www.middle-east-pages.com/>
20. List2000 - <http://www.list2000.com/>
21. Find-İt - <http://find.egenet.com.tr/indexe.html>
22. Arabul - <http://arabul.dominet.com.tr/>>
23. IndexTurkey - <http://www.indexturkiye.com/>

24. Netbul - <http://www.netbul.com/>
25. Arama!com - <http://www.arama.com/>
26. Dublin Core Metadata Initiative - <http://dublincore.org>
27. OCLC CORC - <http://www.oclc.org/corc/about/>
28. Digital Library Catalog - <http://sunsite.berkeley.edu/>
29. EdNA - Education Network Australia - <http://standards.edna.edu.au/metadata/>
30. GEM - Gateway to Educational Materials <http://gem.syr.edu/>