

## Metadata Kavramı

### The Concept of Metadata

Mehmet Emin Küçük\* ve Umut Al\*\*

#### Öz

İnternet, kullanıcılarına iletişimde hız ve kolaylıklar sunmakla birlikte, bilgi erişim açısından bir dizi sorunu da beraberinde getirmiştir. Bu sorunları çevrim içi ortamda yer alan belgelerin tamamının kapsanması, dizinlenmesi, seçimi, erişimi ve bu karmaşık çevrim içi ortamda farkedilebilmek olarak özetlemek olanaklıdır. Çalışmamızda, İnternet bilgi kaynaklarının düzenlenmesi ve erişim sorunlarına çözüm olarak ortaya çıkan metadata kavramı açıklanarak, metadata oluşturma araçları ve metadata standartlarından bazıları tanıtılmıştır.

**Anahtar sözcükler:** İnternet, Metadata, Metadata standartları, Elektronik kaynaklar-kataloglama, Metadata araçları.

#### Abstract

The Internet provides fast and effective communication. However, it has also introduced several problems which can be categorised as coverage, indexing, selection and retrieval of documents. In this descriptive study, the concept of metadata, which emerged as a solution for the problems of organising and retrieving Internet resources, and some of the metadata creation tools and the standards are described.

**Keywords:** Internet, Metadata, Metadata standards, Electronic resources-cataloging, Metadata tools.

### Giriş: İnternet ve Erişim Sorunları

Son yıllarda hepimizin gözlediği gibi İnternet büyük bir hızla büyümekte, uzun yıllardır literatürde "bilgi patlaması" adı verilen olguya bu ortamda daha yakından tanık olmaktadır. İnternet'in büyüklüğü bir başka deyişle içerdiği kaynak sayısını ifade etmek amacıyla bir dizi kestirimde bulunulmuş olmakla bir-

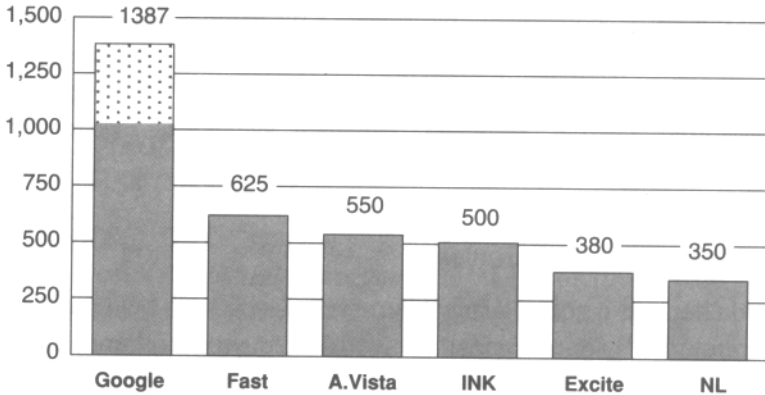
\* Yrd. Doç. Dr.; Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü (mkucuk@hacettepe.edu.tr).

\*\* Arş. Gör.; Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü (umutal@hacettepe.edu.tr).

likte kesin bir rakam verebilmek olanaksızdır. Gray (1996) tarafından Web sitesi sayılarına ilişkin yapılan çalışmada şu veriler elde edilmiştir:

Ay/Yıl	Web sitesi sayısı
6/93	130
12/93	623
6/94	2 738
12/94	10 022
6/95	23 500
1/96	100 000

Diğer bir kestirimde ise, 1997 başında 100 milyon, 1999 yılı ortasında ise 800 milyon Web sayfası tespit edilmiştir (Lowe, 2000). Netcraft (url<sup>1</sup>)\* tarafından yapılan çalışmada da Kasım 2001 tarihinde sunucu adına sahip 36.458.394 sitenin varlığı belirlenmiştir (Netcraft, 2001). Bu sitelerde yer alan dizinlenebilir belge sayısı ise bir varsayım olmaktan öteye geçemeyecektir. Arama motorlarınınca belirtilen, dizinlenen Web sayfalarının karşılaştırmalı tablosu aşağıda verilmiştir. Tablo incelendiğinde bir milyardan üzerinde Web belgesinin Google (url<sup>2</sup>) tarafından dizinlendiği görülmektedir.



**Kaynak:** Sullivan, 2001 The Search Engine Report.\*\*

\* Okumayı kolaylaştırmak için Web adresleri numaralandırılarak kaynakçadan sonra URL Listesi altında verilmiştir.

\*\* Grafikteki rakamlar milyonla çarpılmalıdır.

Web belgelerinin artışının yanı sıra İnternet kullanıcılarının sayısının da hızlı bir şekilde arttığı gözlemlenmektedir. Nua (url<sup>3</sup>) (2001) tarafından yapılan araştırmada, Ağustos 2001 tarihinde 513.41 milyon İnternet kullanıcısı olduğu kestiriminde bulunulmuştur. Dolayısıyla, büyük hacimli bilgi yığınının, çok fazla kullanıcı grubu tarafından kullanıldığı karmaşık bir çevrim içi ortam karşımıza çıkmaktadır.

İnternet'e yönelik sunulan veriler, İnternet'in karmaşık doğası nedeniyle birer kestirim olmaktan öteye gidememektedir. Ancak, belirgin olan ve herkesin üzerinde anlaşabileceği nokta İnternet'in giderek büyüdüğü, sayısal açıdan basılı bilgi kaynaklarıyla karşılaştırılamayacak hacime ulaştığıdır.

Kuşkusuz, bu hacimdeki bilginin düzenlenmesi ve istenilen bilgiye ulaşılması hiç de kolay değildir. İnternet ve onun en sık kullanılan aracı World Wide Web (Web), bilgi erişimde önemli gelişmeler göstermesine karşın, İnternet üzerinde yer alan bilgiye erişim ve seçme konusunda bir dizi sorunu da beraberinde getirmiştir. Bilindiği gibi arama motorları İnternet bilgi kaynaklarının dizinlenmesinde ve onlara erişimin sağlanmasında kullanılan araçlardır. Arama motorlarının İnternet'in tamamını dizinleyebilmeleri (bazılarının bunu iddia etmelerine karşın) olanaklı olmayıp, bunun doğal sonucu olarak İnternet üzerindeki tüm bilgi kaynakları ne yazık ki kullanıcılar için erişilebilir değildir. Arama motorlarının performanslarına yönelik yapılmış bir araştırmaya göre Northern Light (url<sup>4</sup>) İnternet'in %16'sını, Alta Vista (url<sup>5</sup>) %15,5'ini, HotBot (url<sup>6</sup>) %11.3'ünü, Infoseek (url<sup>7</sup>) %8'ini, Google %7.8'ini, Yahoo (url<sup>8</sup>) %7.4'ünü, Excite (url<sup>9</sup>) % 5.6'sını, Lycos (url<sup>10</sup>) % 2.5'ini, EuroSeek (url<sup>11</sup>) %2.2'sini dizinleyebilmektedir (Browser News, 1999). Kuşkusuz bu araştırmadan sonra, arama motorları ve dizinleme servislerinin performanslarındaki gelişmelere paralel olarak yukarıdaki verilerin değişmiş olması da olasıdır. Ancak, İnternet'in tamamının dizinlendiğini ve üzerindeki tüm bilginin erişilebilir olduğunu da iddia etmek mümkün değildir.

İnternet'in tamamının dizinleme servisleri tarafından kapsanamamasının yanı sıra erişim isabeti ve kesin isabet oranları (recall and precision) bir diğer sorunlu alandır. Yüzde 10-20'lik erişim isabeti ve kesin isabet oranı genelde kabul edilebilir olmakla birlikte, Web arama motorlarının, kesin isabet oranının %1'den düşük olduğu iddia edilmektedir (Cathro, 1997). Soydal (2000,

s. 86) tarafından yapılmış olan çalışmada, kesin isabet oranı Northern Light %64, Alta Vista ve Excite % 50, Infoseek % 49 ve HotBot %48 olarak belirlenmiştir. Bir başka çalışmada ise, arama motorlarının ortalama kesin isabet oranının %20 ila %40 arasında değiştiği ifade edilmiştir (Gordon ve Pathak, 1999, s. 154). Sonuçlardaki bu farklılığın önemli nedenlerinden biri, Web'in büyüklüğü nedeniyle tamamının uygulama kapsamına alınamaması ve arama motorlarının performanslarında görülen değişimdir.

Sorunu örnekleme amacıyla, literatürde yer alan bir tarama, aynı arama motoru ve anahtar kelime ile son 3 yılda tarafımızdan tekrarlanmıştır. ANZ-WERS (url<sup>12</sup>) isimli arama motorunu kullanarak, "IETF" (Internet Engineering Task Force) anahtar kelimesi Ağustos 1997'de tarandığında 896.354 belgenin varlığı saptanmıştır (Cathro, 1997). Aynı terim, 5 Ağustos 1999'da tarandığında 91.017, 12 Nisan 2000'de 100.888, 22 Mart 2001'de 80.952, 7 Aralık 2001'de ise 93.300 belgenin varlığı belirlenmiştir. Kuşkusuz aradan geçen süre içerisinde adı geçen arama motorunun performansında olumlu bir gelişme olmasına karşın varlığı tespit edilen belgelerin çoğuna kazara (içinde "ietf" harflerinin yanyana geldiği) erişildiği söylenebilir. Bu örnek, arama motorlarının bir İnternet belgesi içindeki önemli ve rastgele kelimeler arasındaki ayırımı yapamadığını ve dolayısıyla taramayla ilişkisiz bir dizi bilgiyi de karşımıza çıkarabileceğini göstermektedir.\*

Kapsam, erişim ve seçim sorunlarına ek olarak, Web üzerinde yaratılmış olan bir sayfa ya da siteden başkalarının haberdar edilmesi de bir başka sorun olarak karşımıza çıkmaktadır. Bilindiği gibi, dinleme servisleri, dinleme robotlarının siteleri ziyaret etmesiyle veya kendini kaydetme ile dinleme işlemini gerçekleştirmektedir. Dinleme robotları tarafından ziyaret edilmiş (dolayısıyla dinlenmemiş) bir İnternet belgesinin kullanıcılar tarafından farkına varılabilmesi olanaklı değildir. Bunun yanı sıra, bazı dinleme servisleri İnternet belgelerini dinlerken sadece meta bilgilerini (metadata) kullanmaktadır. Eğer bir belgenin metadatası yerleştirilmemişse, bu belge dinlenemeyecek ve dolayısıyla potansiyel kullanıcılar tarafından farkedilmeyecektir.

\* Arama motorlarının performans değerlendirme çalışma örnekleri için bkz. Oppenheim, Morris ve McKnight, 2000; Soydal, 2000.

## Metadeta Nedir ?

İnternet, sınıflaması olmayan büyük hacimli bir bilgi kaynağıdır. Bilindiği gibi, sınıflama yapmadan büyük hacimli bilgi yığınlarıyla başa çıkmak kolay değildir. Kütüphaneciler tarafından uzun süredir bilinen ve kullanılan bu olgu, İnternet üzerinde bulunan bilginin dizinlenmesinde, İnternet belgelerine metadeta sınıflama bilgisinin eklenmesi gerekliliğiyle tekrar gündeme gelmiş ve metadeta kavramı olarak karşımıza çıkmıştır.

Türkçe'de üst veri (Olgun ve Sever, 2000) olarak da kullanılan metadeta\*\* kavramını açıklamak üzere bir dizi tanım yapılmış olup, en basit ve sıkça kullanılan tanım "veri hakkında veri/bilgi" (data about data) şeklinde karşımıza çıkmaktadır. Literatürde yer alan tanımlardan bazıları şöyledir:

Metadeta bir kaynağın öğelerini tanımlayan veridir. Bu nedenle bibliyografik veri olabildiği gibi içerik, kullanım koşulları, kapsam ve teknik ya da erişim özelliklerine ilişkin diğer tanımlamayı da içerebilir (Dempsey, 1998).

Web kaynakları ya da diğer şeyler hakkında makinece anlaşılabilir bilgi (W3 Consortium (url<sup>13</sup>), 2001).

Metadeta veri hakkında bilgidir, ve bu nedenle bir eserin yazarı, oluşturulduğu tarih, ilişkili çalışmalara bağlantılar vb. temel bilgileri sağlar (Miller, 1996).

Tanımlar incelendiğinde metadeta kavramının bibliyografik tanımlamayı içerdiği ve ifade ettiği açıkça görülecektir. Bu açıdan yaklaşıldığında bir kütüphanedeki kart katalog veya bir veri tabanındaki bibliyografik verileri de metadeta olarak isimlendirmek olanaklıdır. Kart üzerindeki ya da elektronik ortamdaki bibliyografik veri, bir bilgi kaynağının metadetasıdır. Kuşkusuz bu noktada akla gelen ilk soru, kütüphanecilerin "bibliyografik veri" ya da "katalog bilgisi" diye adlandırdığı veriye niçin metadeta denildiği? Aslında bu sorunun yanıtı oldukça basittir. Kütüphanecilerin uzun yıllardır araştırdığı ve kullandığı, büyük hacimli bilginin düzenlenmesi ve erişimi tekniklerinin, kütüphanecilik mesleği dışındaki meslek gruplarının konuya daha yoğun ilgi göstermesi ve geleneksel kütüphane teknik ve yöntemlerini kendi meslek jargo-

\*\* Kavram kargaşasına yol açmamak için bu makalede metadeta terimi tercih edilmiştir.

nu içinde isimlendirmesinin yanı sıra, metadada kavramının bibliyografik tanımlamayla birlikte erişim, haklar, kullanım, format, orijinalliğini belirleme (authentication) ve güvenlik gibi dijital dünyaya ilişkin unsurları da içermesidir.

### Metadada Türleri ve Metadada ile İlgili Yaklaşımlar

Metadada kavramının bibliyografik tanımlamadan, erişim yönetimine kadar geniş bir yelpazeyi içermesi nedeniyle, bu kavram Swetland (2000) tarafından 5 ayrı (idari, tanımlayıcı, koruma, teknik ve kullanım) kategoriye ayrılmıştır. Bu kategoriler ve özellikleri özetlenerek aşağıda verilmiştir:

Tür	Tanım	Örnek
İdari (administrative)	Bilgi kaynaklarının yönetiminde kullanılan metadada	Sağlama bilgileri
Tanımlayıcı (descriptive)	Bilgi kaynaklarını tanımlama ve kimliklemeye kullanılan metadada	Katalog kayıtları
Koruma (preservation)	Bilgi kaynaklarının korunmasına ilişkin metadada	Kaynakların fiziksel ve dijital korunmasına yönelik eylemlerin dokümantasyonu
Teknik (technical)	Sistemin nasıl çalışacağı veya metadadanın nasıl kullanılacağına ilişkin metadada	Yazılım ve donanım dokümantasyonu, sayısallaştırma bilgisi (formatı, sıkıştırma oranı, güvenlik verisi gibi).
Kullanım (use)	Bilgi kaynaklarının kullanım türü ve düzeyine ilişkin metadada	Kullanım ve kullanıcıyı izleme, bilginin birden fazla formatta tutulması (html, pdf gibi).

Metadada türlerinin ortaya çıkışındaki en önemli etken, farklı yapıya sahip bilgi türlerinin (html metin, coğrafik görüntü, hareketli görüntü, tıbbi görüntü, arşiv belgesi gibi) farklı işlenmesi (sağlama, saklama, paketlenme gibi) ve hizmete sunulması olduğunu söylemek olanaklıdır. Örnek vermek gerekirse, Directory Interchange Format (DIF) uydu görüntüleri için metadada sağlamaktadır (Miller, 1996). Text Encoding Initiative (TEI) (url<sup>14</sup>), bilimsel anlamda

yapılan arařtırmalar için elektronik metinlerin yorumlanmasında bir rehber geliřtirmeyi amaçlayan uluslararası bir proje olup, bir anlamda National Spatial Data Infrastructure'ın (NSDI) (url<sup>15</sup>) devam eden bir gözden geçirmesi şeklindedir. ABD'ndeki bu uygulamaya karşılık, Birleşik Krallık'ta geçici olarak National Geospatial Database (url<sup>16</sup>) olarak adlandırılan proje, uzayla ilgili verilerin tanımlanmasını içeren karmaşık konular için benzer bir şema oluşturmayı amaçlamaktadır (Miller, 1996; TEI, 2001).

Var olan metadatalar projeleri son derece karmaşık ve uzmanlar tarafından yaratılıp bilgisayarlar tarafından anlaşılması (yorumlanması) gereken yapılarıdır. Örneğin Resource Organisation and Discovery'in Subject-based Services (ROADS) (url<sup>17</sup>), Joint Information Systems Committee (JISC) (url<sup>18</sup>) tarafından finanse edilen, farklı konularda ve coğrafik alanlardaki İnternet kaynaklarına erişim sağlamak ve bu kaynakları tanımlayarak değerlendirmek amacıyla geliştirilen bir projedir. Geliştirilen yazılım, konu girişlerinin oluşturulmasına olanak sağlamakla birlikte, söz konusu kaynaklara ait kayıtların Web erişimli bir veri tabanıdır. Her bir kayıt, URL adresi, başlık, tanım gibi alanları içerir ve tüm girişler taranabilmektedir (Kirriemuir, 1998).

Bu tür uzmanlaşmış metadatalar oluşturma/geliştirme yaklaşımlarının yanı sıra arama motoru yaklaşımı diye adlandırabileceğimiz, arama motorlarını Web belgeleri içindeki HTML meta tagları kullanmaya doğru götüren ikinci bir yaklaşımdan söz edilebilir. Örneğin Alta Vista arama motoru, sayfaların tanımlama (description) ve anahtar kelime (keywords) öğelerini kullanarak sayfaları dizinlemektedir (Miller, 1996).

Metadataların işlenmesine yönelik karşımıza çıkan modellerden birisi olan Resource Description Framework (RDF), World Wide Web Konsorsiyum (W3C) tarafından önerilmektedir (Weibel, 1999). RDF, Web üzerindeki maki-nece anlaşılabilir bilginin uygulamalar arasında işlenebilirliğini ve Web kaynaklarının otomatik işlenmesini sağlayan olanakları içermektedir. RDF'in kullanıldığı çeşitli uygulamaları örneklemek gerekirse;

- Arama motorlarının kaynak keşfi yeteneğini artırmaktadır.
- Bir Web sitesinin içerik ve içerik ilişkilerinin tanımlanmasında (kataloglamada) kullanılmaktadır.

- Akıllı yazılım araçları tarafından bilgi paylaşımı ve değişiminde kullanılmaktadır.
- Web sayfalarına kullanıcı erişim haklarının belirlenmesi, içerik erişim değerlerinin belirlenmesi ve sayfaların özel kullanım ve kopyalama haklarının belirlenmesinde kullanılmaktadır (Olgun, 1999).

### Metadeta Standartları

Elektronik ortamdaki bir bilgi kaynağının standart bir biçimde tanımlanması, bibliyografik verilere erişimi sağlayacak yazılımların kullanacağı standartları belirler ve verileri sistemden sisteme transfer edebilme olanağını yaratır. Aksi takdirde bibliyografik tanımlama kaosunun oluşması ve aynı kaynağın birden fazla kişi/kurum tarafından benzer emek, zaman ve maliyetin harcanarak tanımlanması gibi bir durum ortaya çıkacaktır. Kütüphanecilik mesleğinde bu durum yıllar öncesinde çözülmüş ve standart tanımlama kodları geliştirilmiştir. Kütüphanecilik dışında gelişen metadeta uygulamalarında ise, Content Standards for Digital Geospatial Metadata (url<sup>19</sup>), Encoded Archival Description (url<sup>20</sup>), Text Encoding Initiative, Categories for the Description of Works of Art (url<sup>21</sup>), Dublin Core vb. bir dizi metadeta standardı ortaya çıkmış bulunmaktadır. Bu standartlardan bazıları daha önce belirtildiği gibi (uydu görüntüleri vb.) belirli tip bilgi kaynaklarını tanımlamak ve onlara erişim sağlamak amacıyla geliştirilmişlerdir. Dublin Core (DC) gibi bazıları da farklı tipteki bilgi kaynaklarını tanımlamak ve tek bir tarama işlemiyle bunlara erişim sağlamak amacıyla geliştirilmişlerdir.

Kuşkusuz, İnternet gibi son derece karmaşık bir çevrim içi ortamda, bir anda üzerinde görüş birliğine varılmış bir standardın (hatta sonsuza dek) olmasını beklemek oldukça iyimser bir yaklaşım olacaktır. Ancak, bu standartlar arasından bir ya da birkaçının diğerlerine oranla daha fazla kullanılacağını en azından akademik veya profesyonel yayıncılar gibi belirli çevrelerde daha fazla uygulama alanı bulacağını kestirmek olanaklıdır. Nitekim bu standartlar içinde Dublin Core, konu ve kaynak tipi ayırımı gözetmemesinden dolayı en fazla tanıtılan ve kullanılan standarttır. Dublin Core metadeta element seti, 1995 ve 1996 yıllarında aralarında kütüphanecilerin de bulunduğu bir



grup tarafından geliştirilmiş olup, W3 Consortium tarafından da desteklenmektedir. Aşağıda DC element setinin 15 data ögesi listelenmiştir (Dublin Core Metadata Initiative, 1999):

- Title (eser adı)
- Author or Creator (yazar ya da yaratan)
- Subject and Keywords (konu ve anahtar kelimeler)
- Description (tanım)
- Publisher (yayıncı)
- Other Contributors (diğer katkıda bulunanlar)
- Date (tarih)
- Resource Type (kaynak tipi)
- Format (biçim)
- Resource Identifier (kaynak tanımlayıcı)
- Source (kaynak)
- Language (dil)
- Relation (ilişki)
- Coverage (kapsam)
- Rights Management (hakların yönetimi).

### **Metadata Oluşturma Araçları**

Gerek geleneksel olarak gerekse İnternet ortamında olsun, metadatanın temel amacı, bilgi erişimi geliştirmek ve olanaklı kılmaktır. Fonksiyonlarının tamamının mesleki eğitim almış kütüphaneciler tarafından gerçekleştirildiği geleneksel kütüphane ortamında, makinece okunabilir kataloglama (MARC) gibi karmaşık metadata sistemleri kaynak tanımlamada kabul edilebilir araçlardır. Ancak, daha karmaşık olan çevrim içi ortamda, sürekli yeni kaynaklar or-

taya çıkmakta ve bu kaynaklar genellikle bireyler tarafından oluşturularak yapılmaktadır. Bu kaotik yapı nedeniyle bilgileri bulmak ve değerlendirmek kolay değildir. Tüm olumsuzluklarına karşın yeni bilgilere erişebilmekte arama motorlarını kullanmaktan başka seçeneğimiz de yoktur.

Böyle bir ortamda, metadatanın kullanılması kaçınılmazdır. Fakat, oluşturulacak metadatanın arama motorları ve insanlar tarafından yorumlanabilme-yeye uygun olması, herhangi bir Web yaratıcısının hazırladığı sayfayı ve içeriğini kolayca tanımlayabilmesi gerekir ki, en uygun tanımlama bilgilerini basit bir teknikle oluşturulabilsin. Metadata yaratmanın en uygun zamanı, tek bir işlemin parçası olarak, kaynak yaratılırken oluşturulmasıdır. Ancak, İnternet üzerinde daha önceden üretilmiş bir dizi yararlı kaynağın var olduğu ve kişilerin geriye dönük olarak bunlara metadata eklemek isteyecekleri göz önüne alındığında bu tür uygulamalara yönelik olarak İnternet üzerinde ücretsiz elde edilebilen metadata yaratıcılarının (metadata generators) da bulunduğu unutulmamalıdır.

Metadata oluşturma araçlarını editörler ve yaratıcılar olmak üzere iki grupta toplamak olanaklıdır. Editör olarak adlandırılan araçlar, ekrana boş bir form getirirler ve boşluklar uygun ifadelerin girilmesiyle doldurulur. Destek yazılımı, girilen içeriği HTML tag haline getirir ve oluşturulan HTML tag bir belgeye kesilip yapıştırılabilir. Yaratıcı olarak adlandırılan araçlar ise mevcut HTML kodlu belgelerden metadatayı alarak içeriği HTML tag haline getirir. Yaratıcılar, HTML V3.2, HTML V4.0 ve XML gibi bir dizi çıktı seçeneği sunabilmektedir (National Library of Australia, 1999). Mevcut metadata oluşturma araçlarından bazıları aşağıda sunulmuştur:

### Editörler

*Nordic Web Index DC Metadata Template* (url<sup>22</sup>):

Kuzey Avrupa ülkeleri için Dublin Core metadata yaratılmasına destek sağlamak amacıyla oluşturulmuş bir proje olarak karşımıza çıkmaktadır. Nordic Metadata Project'in sahip olduğu editör detaylı yardım hizmeti ile dikkat çekmektedir (Koch ve Borell, 1997).

*Gateway to Educational Materials (GEM) (url<sup>23</sup>):*

Amerikan Eğitim Bakanlığının (url<sup>24</sup>) girişimiyle eğitimcilerin İnternet tabanlı ders planlarına ve diğer eğitim kaynaklarına erişimini artırmak amacıyla oluşturulan bir proje olup, bu kaynaklara erişimi sağlayan metadata standartları ve teknik mekanizmalar sunar. Bünyesinde binlerce ders planı, müfredat ve eğitimle ilgili kaynak bulunduran GEM, özellikle öğretmenler için yararlı olmaktadır. GEM'in amacı, eğitimle ilgili kaliteli kaynak koleksiyonu sağlayarak, kaynak keşif problemini çözmektir. Eğitimciler GEM'e bağlandıklarında; eğitim düzeyine, anahtar kelimeye veya konuya göre düzenlenmiş listeleri tarayabilmektedirler. GEM'de Şubat 2001 itibarıyla 7800'ün üzerinde kayıt yer almaktadır (GEM, 2001).

*Reggie (url<sup>25</sup>):*

Metadata editörü bir program aracılığıyla RDF formatı, HTML 3.2 ve HTML 4.0 standartlarını kullanarak farklı element setlerine ve farklı dillere dayalı olarak metadata yaratabilmektedir. Söz konusu diller İngilizce, İtalyanca, Fince, Fransızca, Almanca, İspanyolca, Norveççe ve Portekizce'dir. Ayrıca birbirinden farklı şemaları seçebilme olanağı da verebilmektedir. Bu şemalar Dublin Core, Global Information Locator Service (GILS) (url<sup>26</sup>), Australia New Zealand Land Information Council (ANZLIC) (url<sup>27</sup>), Australian Government Locator Service (AGLS) (url<sup>28</sup>), Education Network Australia (EdNA) (url<sup>29</sup>), IMS, GEM ve vCard şeklinde sıralanmaktadır (Resource Discovery Unit, 1998).

*State Library of Tasmania's Generic Editor (url<sup>30</sup>):*

Tasmania Devlet Kütüphanesinin (url<sup>31</sup>) bünyesinde bulunan Tasmania Online (url<sup>32</sup>) tarafından sağlanmaktadır. SuperNoteTab isimli metin editörü kullanılarak var olan HTML dosyalarına metadata koymak ve üzerinde değişiklikler yapmak olanaklıdır. Metadata Dublin Core (url<sup>33</sup>), AGLS ve ADMIN Core'u (url<sup>34</sup>) içermektedir (Campbell, 1998).

### *H-DCEdit (Hacettepe-Dublin Core Üstveri Elemanları Editörü):*

Hacettepe Üniversitesi Bilgisayar Mühendisliği Bölümünde Türkçe elektronik kaynak içeriklerinin tanımlanmasında kullanılacak bir yazılım geliştirmek amacıyla yapılan yüksek mühendislik tezinin ürünüdür. 1999 yılında geliştirilen yazılım ilk Türkçe Dublin Core editörü olarak karşımıza çıkmaktadır (Küçük, Olgun ve Sever, 2000).

### **Yaratıcılar**

#### *Tag Gen (url<sup>35</sup>):*

Hiawatha Island Software Şirketinin programı olan TagGen, metadata yaratan bir ürün olarak 1998 yılında piyasaya çıkmıştır. "Web üzerinde sitenizi görünür yapın" sloganıyla yola çıkan yazılım, Web sitesi tanıtımında ve bilgi yönetiminde söz sahibi olmak amacını gütmektedir (Hiawatha Island Software, 2001).

#### *DC-dot (url<sup>36</sup>):*

United Kingdom Office for Library and Information Networking (UKOLN)'de (url<sup>37</sup>) çalışan Andy Powell tarafından geliştirilmiştir. DC-dot, tanımlanması istenilen Web sayfasına erişerek otomatik olarak metadatasını yaratmakta, ayrıca metadata üzerinde değişiklik yapabilme olanağı\* vermekte, farklı formatlara dönüştürme işlemini de gerçekleştirebilmektedir. Bu formatlar, US-MARC, Summary Object Interchange Format (SOIF), Internet Anonymous Ftp Archives - IAFA/ROADS, TEI headers, GILS ve RDF'tir (Powell, 2000).

#### *Medical Metadata Creator (url<sup>38</sup>):*

University of Newcastle Upon Tyne'in geliştirdiği ürün, tıbbi konulardaki Web sayfaları için metadata yaratmakla birlikte tıbbi konulara ilişkin kapsamlı anahtar kelimeler listesini de bünyesinde bulundurmaktadır (DSTC, 1998).

---

\* Bu özelliği Dc-dot'ı aynı zamanda metadata editörü yapmaktadır.

### Metadatalar Harvesting Protokolü (MHP)

Web, elektronik kaynakların tanımlanması ve dizinlenmesinin yanı sıra bunların kullanımına ilişkin bir dizi sorunu (lisans anlaşmaları, telif hakları, erişim hakları gibi) beraberinde getirmiştir. Web'in sunduğu hızlı erişim, aktarma olanakları sayesinde akademik dergiler hızlı bir biçimde tam metin olarak Web üzerinde yer almaya başlamışlardır. Ancak, yayıncıların ücret ödemediği yazarından aldıkları bu yayınları, ücret karşılığında (oldukça yüksek ücretler) kullanıma sunmaları, bu yazıları üreten kişi ve kurumları yeni arayışlara itmiştir (binlerce makale üreten bir üniversitenin, kendi yazarlarının yazılarının da yer aldığı veri tabanlarına ve ücretsiz yazısı alınan bir yazarın başka bir yazarın makalesini okuyabilmek için ücret ödemesini gözönüne alınız). Bu arayışlar sonucunda belge (metin, fotoğraf, hareketli görüntü vb.) üreticiler, bu belgeleri herkese açık arşivler denilebilecek sunuculara yerleştirmeye başladılar. Böylelikle, üretilen belgeler daha fazla kişi tarafından ve ücretsiz olarak erişilmeye başlandı. Ancak bu arşivlere ve arşivlerde yer alan belgelere daha etkin erişimin sağlanabilmesi için uygun yöntem arayışları da başlamıştır.

Bunlardan biri The Open Archives Metadatalar Harvesting Protokolü'dür. Bu protokol, dağıtık ağa dayalı bilgi hizmetlerini desteklemede (*distributed networked information service*) önemli bir alt yapı bileşenidir. Belge üreticilerinin belgelerinin metadatalarını oluşturmalarını kolaylaştıran bir mekanizma olarak Metadatalar Harvesting Protokolü, hızlı bir şekilde kullanım bulmuş ve farklı topluluklara yönelik sistem mimarilerini ve yeni hizmetleri olanaklı kılmaya başlamıştır. Daha yalın biçimde ifade edilecek olursa, MHP, sunucular üzerinde yer alan nesnelerin tanımlanmasına (belirlenmesi) yönelik dışarıdan gelen istekleri yanıtlayan bir ara yüzdür (Lynch, 2001). Bu arayüz sayesinde bir sunucu üzerindeki dijital nesnelerin neler olduğu tanımlama bilgileri yani metadatalar aracılığıyla öğrenilir ve/veya başkalarının bu metadataları elde etmelerine imkan tanınır (bu yüzden metadatalar hasadı yapmak anlamına gelen, "harvesting" ifadesi kullanılmıştır).

## Sonuç

İnsanoğlunun iletişimde dördüncü aşama olarak tanımlayabileceğimiz (dil, yazı, matbaa ve www) Web, kullanıcılarına iletişimde hız ve kolaylıklar sunmakla birlikte, bilgi erişim açısından bir dizi sorunu da beraberinde getirmiştir. Bu sorunları, çevrim içi ortamda yer alan belgelerin tamamının kapsanması, dizinlenmesi, seçimi, erişimi ve bu karmaşık çevrim içi ortamda farkedilebilmek olarak özetlenebilir.

Mevcut Web belgelerinin, belgenin tamamının dizinleme robotlarınca okunarak dizinlenmesi Web hacmi gözönüne alındığında olanaklı olmadığı gibi erişim isabeti ve kesin isabet açısından da en uygun çözüm yolu değildir. Ancak Web belgesinin tamamı yerine, standart bir biçimde oluşturulmuş bibliyografik verilerini okuyarak yapılan bir dizinleme, dizinleme robotlarının olduğu gibi erişim isabeti açısından arama motorlarının da performansını artıracaktır. Bu da, her bir Web belgesinin standart bir biçimde kimliklenmesi yani metadatasının oluşturulmasını zorunlu kılmaktadır.

Nitekim, kütüphaneciler büyük hacimli bilgi yığınlarıyla başa çıkmada bir dizi standart tanımlama kuralları, yöntemleri geliştirmişler ve uygulamışlardır. Kütüphanelerde meslek eğitimi almış profesyonellerce uygulanan bu yöntem ve kuralların, ilgili eğitimi almamış kişilerce de uygulanabilmesini beklemek, hele İnternet gibi her düzey ve yaştan belge yaratıcısı ve kullanıcısının olduğu karmaşık bir ortamda, iyimserliğin ötesinde bir yaklaşım olacaktır.

Web belgelerinin metadatasını oluşturabilmenin bir dizi yolu bulunmaktadır. Bunlardan biri belgeleri oluşturan kişi tarafından html taglarda tanımlama bilgilerinin verilmesi, diğeri ise bir araç aracılığıyla (editör ve yaratıcılar) metadatası oluşturmaktır. Kuşkusuz bu bilgilerin standart bir biçimde ve içerikte olması farklı uygulama yazılımlarının bu verileri kullanabilme olanağını sağlayacaktır. Bu gereksinimin doğal bir sonucu olarak standart metadata oluşturma yaklaşımları bir dizi farklı standardın oluşmasına neden olmuştur. Farklılığın en temel nedeni Web belgelerinin türü, işlenme biçimi ve hizmet amacıdır. Geliştirilmiş olan metadata standartları arasında özellikle Web belgeleri açısından Dublin Core Metadata Standardı'nın yaygınlaştırılmaya çalışıldığını söylemek olanaklıdır (aynı şekilde, arşivler için de Encoded Archival Description Metadata Standardı'nın önerildiğini söyleyebiliriz). Bir başka yak-

laşım, Metadata Harvesting Protocol adı altında başlatılan, sunucularda yer alan belgelerin metadatalarının hangi standartta olursa olsun bir ara yüz aracılığıyla (Metadata Harvesting Protocol) istemci/sunucu yazılımlar tarafından alınabilmesi ve verilebilmesini hedefleyen (Z39.50'ye benzer bir yapı) metadata oluşturan/toplayan yeni bir uygulamadır.

Batıda İnternet içeriğinin düzenlenmesi ve İnternet'te kaynak keşfi yeteneğinin artırılması konusunda bir dizi çalışma yapılırken, Türkçe içeriğin düzenlenmesi ve erişiminin etkinleştirilmesi konusuna yeterince yönelebildiğimizi söylemek olanaklı değildir. Kütüphaneciler olarak, büyük hacimli bilgi yığınlarının düzenlenmesinde ve hizmete sunulmasındaki deneyimlerimizi diğer meslek gruplarının üyeleriyle paylaşmak, geliştirmiş olduğumuz yöntem ve teknikleri aktarmak durumundayız.

### Kaynakça

- Browser News. (1999). *Statistics- search engines*. [Çevrim içi]. Elektronik adres: [http://www.upsdell.com/BrowserNews/stat\\_search.htm](http://www.upsdell.com/BrowserNews/stat_search.htm) [10 Aralık 2001].
- Campbell, D. (1998). *Generic edit tool*. [Çevrim içi]. Elektronik adres: [http://www.dstc.edu.au/Research/Projects/metaweb/generic\\_tool.html](http://www.dstc.edu.au/Research/Projects/metaweb/generic_tool.html) [12 Aralık 2001].
- Cathro, W. (1997). *Metadata: On overview*. [Çevrim içi]. Elektronik adres: <http://www.nla.gov.au/nla/staffpaper/cathro3.html> [11 Aralık 2001].
- Dempsey, L. (1998). *Metadata: A UK HE perspective*. [Çevrim içi]. Elektronik adres: <http://www.ukoln.ac.uk/services/papers/bl/blri078/content/report~27.htm> [10 Aralık 2001].
- DSTC. (1998). *Analysis of current metadata creation tools*. [Çevrim içi]. Elektronik adres: <http://www.dstc.edu.au/Research/Projects/metaweb/tool-post.html> [10 Aralık 2001].
- Dublin Core Metadata Initiative. (1999). *Dublin Core Metadata Element Set, Version 1.1: Reference Description*. [Çevrim içi]. Elektronik adres: <http://dublincore.org/documents/1999/07/02/dces/> [10 Aralık 2001].

- GEM. (2001). *The Gateway to Educational Materials*. [Çevrim içi]. Elektronik adres: <http://www.geminfo.org/networker.html> [11 Aralık 2001].
- Gordon, M. ve Pathak, P. (1999). Finding information on the World Wide Web: the retrieval effectiveness of search engines. *Information Processing and Management*, 35(2): 141-180.
- Gray, M. (1996). *Web growth summary*. [Çevrim içi]. Elektronik adres: <http://www.mit.edu/people/mkgray/net/web-growth-summary.html> [11 Aralık 2001].
- Hiawatha Island Software. (2001). *HiSoftware-TagGen*. [Çevrim içi]. Elektronik adres: <http://www.hisoftware.com/taggen.htm> [12 Aralık 2001].
- Kirriemuir, J. (1998). *What is ROADS?* [Çevrim içi]. Elektronik adres: <http://www.ukoln.ac.uk/metadata/roads/what/> [05 Eylül 2000].
- Koch, T. ve Borell, M. (1997). *Dublin Core Metadata Template*. [Çevrim içi]. Elektronik adres: <http://www.lub.lu.se/cgi-bin/nmdc.pl> [12 Aralık 2001].
- Küçük, M. E., Olgun, B. ve Sever, H. (2000). Application of metadata concepts to discovery of Internet resources. Tatyana Yakhno (ed.) *Advances in information systems: first international conference; proceedings/ ADVIS 2000, İzmir, Turkey, October 25-27, 2000* içinde (s.304-313). Berlin: Springer.
- Lowe, D. (2000). *Improving Web search relevance: using navigational structures to provide a search context*. [Çevrim içi]. Elektronik adres: <http://ausweb.scu.edu.au/aw2k/papers/lowe/paper.html> [11 Aralık 2001].
- Lynch, C. (2001). Metadata Harvesting and the Open Archives Initiative. *ARL Bimonthly Report* [Çevrim içi], 217: 1-9. Elektronik adres: <http://www.arl.org/newsltr/217/mhp.html> 217: 1-9. [14 Aralık 2001].
- Miller, P. (1996). *Metadata for the masses*. [Çevrim içi]. Elektronik adres: <http://www.ariadne.ac.uk/issue5/metadata-masses/> [10 Aralık 2001].
- National Library of Australia. (1999). *Meta matters: Tools*. [Çevrim içi]. Elektronik adres: <http://www.nla.gov.au/meta/tools.html>



- Netcraft. (2001). *Netcraft Web Server Survey*. [Çevrim içi]. Elektronik adres: <http://www.netcraft.com/survey/> [11 Aralık 2001].
- Nua. (2001). *How many online*. [Çevrim içi]. Elektronik adres: [http://www.nua.ie/surveys/how\\_many\\_online/](http://www.nua.ie/surveys/how_many_online/) [11 Aralık 2001].
- Olgun, B. (1999). *Dublin Core üstveri elemanları editörü*. [Çevrim içi]. Yayımlanmamış yüksek mühendislik tezi. Hacettepe Üniversitesi, Ankara. Elektronik adres: <http://ata.cs.hun.edu.tr/~km/baha/rdf.html> [10 Aralık 2001].
- Olgun, B. ve Sever H. (2000). Kaynak keşif yeteneğinin artırılması için İnternet kaynaklarının içeriklerinin standart biçimde tanımlanması. *Bilgi Dünyası*, 1(1): 56-88.
- Oppenheim, C., Morris, A. ve McKnight, C. (2000). The evaluation of WWW search engines. *Journal of Documentation*, 56 (2): 190-211.
- Powell A. (2000). *DC-dot*. [Çevrim içi]. Elektronik adres: <http://www.ukoln.ac.uk/metadata/dcdot/> [12 Aralık 2001].
- Resource Discovery Unit. (1998) *Reggie-metadata editor*. [Çevrim içi]. Elektronik adres: <http://metadata.net/dstc/> [12 Aralık 2001].
- Soydal, İ. (2000). *Web arama motorlarında performans değerlendirmesi*. Yayımlanmamış bilim uzmanlığı tezi. Hacettepe Üniversitesi, Ankara.
- Sullivan D. (2001) *Search engine sizes*. *The Search Engine Report, August 15*. [Çevrim içi]. Elektronik adres: <http://searchenginewatch.com/reports/sizes.html> [11 Aralık 2001].
- Swetland A. J. (2000) *Introduction to metadata: Setting the stage*. [Çevrim içi]. Elektronik adres: [http://www.getty.edu/research/institute/standards/intrometadata/2\\_articles/index.html](http://www.getty.edu/research/institute/standards/intrometadata/2_articles/index.html) [11 Aralık 2001].
- TEI (2001). *What is the TEI Consortium*. [Çevrim içi]. Elektronik adres: <http://www.tei-c.org/Consortium/index.html> [10 Aralık 2001].
- W3 Consortium. (2001). *Metadata and resource description*. [Çevrim içi]. Elektronik adres: <http://www.w3.org/Metadata/> [10 Aralık 2001].
- Weibel, S. (1999). The state of the Dublin Core Metadata Initiative. *D-Lib Magazine* [Çevrim içi], 5(4). Elektronik adres: <http://dlib.org/dlib/april99/04weibel.html> [11 Aralık 2001].

### **URL Listesi**

- 1 Netcraft - <http://www.netcraft.com/>
- 2 Google - <http://www.google.com/>
- 3 Nua - <http://www.nua.ie/>
- 4 Northernlight - <http://www.northernlight.com/>
- 5 Alta Vista - <http://www.altavista.com/>
- 6 HotBot - <http://www.hotbot.com/>
- 7 Infoseek - <http://www.infoseek.com/>
- 8 Yahoo - <http://www.yahoo.com/>
- 9 Excite - <http://www.excite.com/>
- 10 Lycos - <http://www.lycos.com/>
- 11 EuroSeek - <http://www.euroseek.com/>
- 12 ANZWERS - <http://www.answers.com.au/>
- 13 W3 Consortium - <http://www.w3.org/>
- 14 Text Encoding Initiative - <http://www.tei-c.org/>
- 15 National Spatial Data Infrastructure - <http://www.fgdc.gov/nsdi/nsdi.html>
- 16 National Geospatial Database - <http://www.ngdf.org.uk/>
- 17 Resource Organisation and Discovery in Subject-based Services - <http://www.ilrt.bris.ac.uk/roads/>
- 18 Joint Information Systems Committee - <http://www.jisc.ac.uk/>
- 19 Content Standards for Digital Geospatial Metadata - <http://www.fgdc.gov/metadata/contstan.html>
- 20 Encoded Archival Description - <http://lcweb.loc.gov/ead/>
- 21 Categories for the Description of Works of Art - <http://is.gseis.ucla.edu/impact/f95/Cdwa/HOME PAGE.HTML>

- 22 Nordic Web Index DC Metadata Template - <http://www.lub.lu.se/cgi-bin/nmdc.pl>
- 23 Gateway to Educational Materials - <http://www.geminfo.org/>
- 24 US Department of Education - <http://www.ed.gov/>
- 25 Reggie - <http://metadata.net/dstc/>
- 26 Global Information Locator Service - <http://www.gils.net/>
- 27 Australia New Zealand Land Information Council - <http://www.anzlic.org.au/>
- 28 Australian Government Locator Service - [http://www.naa.gov.au/recordkeeping/gov\\_online/agls/summary.html](http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html)
- 29 Education Network Australia - <http://www.edna.edu.au/>
- 30 State Library of Tasmania's Generic Editor - [http://www.dstc.edu.au/Research/Projects/metaweb/generic\\_tool.html](http://www.dstc.edu.au/Research/Projects/metaweb/generic_tool.html)
- 31 State Library of Tasmania - <http://www.statelibrary.tas.gov.au/>
- 32 Tasmania Online - <http://www.tas.gov.au/>
- 33 Dublin Core - <http://dublincore.org/>
- 34 ADMIN Core - <http://metadata.net/admin/>
- 35 TagGen - <http://www.hisoftware.com/taggen.htm>
- 36 DC-Dot - <http://www.ukoln.ac.uk/metadata/dcdot/>
- 37 UKOLN - <http://www.ukoln.ac.uk/>
- 38 Medical Metadata Creator - <http://medir.ohsu.edu/bicc-informatics/ebm/latest.htm>