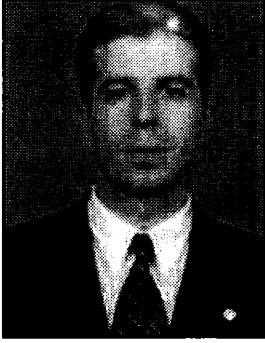


## Hakemsiz Yazılar

### Opinion Papers

#### Veri Madenciliği ve Bilgi Merkezleri



\* Sacit Arslantekin \*

#### Öz

*Günümüzde bir çok kurum ve kuruluş işlem ve hizmetlerini otomasyona geçirmiştir. Bu kurum ve kuruluşlar kullanıcı odaklı olarak faaliyetlerini Internet üzerinden de sürdürmeye çalışırken, daha büyük kazanımlar elde etme amacındadır. Elde etmek istedikleri kazanımlar kalitelerini arttırmalarının yanında, kullanıcı memnuniyetine de dayanmaktadır. Kullanıcılarının ürün ve hizmetlerinden ne kadar yararlandıkları, memnun kalıp kalmadıkları ancak, onların davranışları izlenerek ortaya konabilir. Bilgi merkezleri de kullanıcı davranışlarını izleyen ve elde ettiği bilgilere göre işlem ve hizmetlerini geliştiren kurumlardandır. Bilgisayara dayalı işler sürecinde otomatik olarak oluşan verinin işlenip analiz edilmesi, bilgi merkezlerine çalışanları ve kullanıcıları hakkında anlamlı bilgiler sunacaktır. İşlemler ve kullanım sırasında oluşan veri büyük bir yığın halinde olup, bu yığından anlamlı veri elde edilmesi bazı özel yöntem ve tekniklerin kullanılmasını gerektirmektedir. Veri madenciliği oluşan verilerin toplanması, bir yapay zeka aracılığıyla analiz edilmesi ve yorumlanarak amaç ve hedeflere ulaşmada gerekli adımları önermesi sürecidir.*

\* Yrd. Doç. Dr. Sacit Arslantekin, Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Bilgi ve Belge Yönetimi Bölümü öğretim üyesidir. E-posta: atekin@humanity.ankara.edu.tr

## Giriş

1960'lardan bu yana iş yaşamında bir moda gibi hızla değişen planlar, analizler ve stratejiler söz konusudur.. Bu değişime ayak uydurabilmek için ise üst seviye yöneticilerin, analistlerin ve bilgi profesyonellerinin gereksinim duydukları şey daha fazla bilgidir. Masalardaki çok güçlü bilgisayarlara ve iletişim sistemlerine rağmen uzmanlar, karar mekanizmasını oluşturan yöneticiler ve danışmanlar, örgütlerinde zaten mevcut olan kritik bilgilere ya da enformasyonlara ulaşamama sıkıntısını yaşamaktadırlar.

Örgütler az veya çok milyarlarca byte uzunluğunda, kullanıcıları, süre giden işleri, hizmet ve ürünleri, çalışanları vb. hakkında veri üretirler. Ancak bu veriler bilgisayar sistemlerinin içinde her geçen gün ulaşılması daha da zor bir hal alarak gömülür gider. Uzmanlar ise, mevcut verilerden elde edilmiş, işlenmiş ve depolanmış sadece küçük bir parça veriye dayanarak değerlendirme yapabilmektedirler. Örgüt bünyesinde toplanan verilerin yalnız raporlama amaçlı kullanılması, yöneticilere yeterince karar desteği sağlanamamasına neden olmaktadır (Karakas:2002).

Dünyadaki ani değişmelerin beklenmedik bir şekilde kurum ve kuruluşlara olan etkisi karşısında bilgi teknolojileri kritik bir misyonu da üstlenmektedirler. Bilgi teknolojileri ile ilgilenenlerin stratejik misyonları, teknolojiyi kesintisiz ve mekandan bağımsız bir halde sunarak organizasyonların performansını arttırmaktır. Örgütün iş stratejisinin merkezindeki yeriniz, yüksek kalitedeki veriyi doğru insana doğru zamanda sunabilmenize bağlıdır.

Bilgisayar sistemleri ve veri tabanları sayesinde hızla toplanan veriler her geçen gün artmaktadır. Gerek özel sektör, gerekse kamu yöneticileri, bu toplanan veriyi karar destek amaçlı bilgiye dönüştürmek, bilgiye dayalı yönetime geçmek durumundadırlar. Sonuçta bu veriler, insanları, onların koşullarını, isteklerini, gereksinimlerini, sorularını içermekte ve eğilimlerine karşı da ipuçları vermektedir. İstatistik; bu verilerden yararlanmanın yolunu açmakta ve kurumların, uyguladıkları programların, yönedikleri kesimin gereksinimlerini karşılayıp karşılamadıklarının farkına varmalarını sağlamaktadır.

İstatistik, kendisinden yararlananların veriyi analiz etmesi ve verideki gizli bilgileri karar destek amaçlı kullanılabilmesini sağlar. Özellikle günümüzün ekonomik koşulları ve yaşanan hızlı değişim ortamında, iş deneyimi ve ön sezilere dayanılarak alınan kararlarda yanlış karar alma riski çok yüksektir. Riski azaltmanın tek yolu bizleri "BİLGİYE" dayalı yönetime taşıyan karar destek çözümlerdir. İstatistik ve veri madenciliği gereçleri gerçek anlamda bir karar destek sistemi oluşturmada olamazsa olmaz ikilidir. Bu noktada bilgi teknolojilerinden yararlanmak elbette kaçınılmazdır.

## Veri Ambarları

Veri ambarlarının amacı, operasyonel veri tabanlarındaki verileri kullanılabilir kılmak ve bunları diğer dışardan alınan bilgi ve verilerle harmanlamaktır.

Kurum genelindeki bilgileri veri ambarına atmak, onları birleştirmek, özet halinde raporlamak bilinen veri ambarı teknikleridir. Veri ambarlarına genelde alınmayan bilgiler, site klik bilgileridir. Yani, “site ziyaretçisi 2 no’lu sayfaya baktı, üç ürüne/bibliyografik kimliğe baktı, birisini bibliyografyasına ekledi ya da birinin detayına gitti” gibi bilgiler site klik bilgileridir. Ziyaretçiyi “gözetler” gibi toplanan bu bilgiler sitenizi geliştirmek için çok isinize yarayacaktır. Bu bilgileri toplamanın çeşitli yolları bulunmaktadır. Apache, IIS gibi programlar üzerinde birkaç değişiklik gerçekleştirilerek bu tür bazı programlar kullanıcıyı takip eder hale getirilebilir. Takip bilgisini almanın yollarının başında, ‘servis programı’ kullanmak gelir. ‘Sayfa servis’ programları sadece sayfa ile uğraşırlar. Sayfa servis eklentileri ile de kullanıcı takip ettirilebilir. “Servis programı” ve onların üzerine kurulan ticaret uzantıları, ziyaretçi hakkında her türlü e-ticaret bilgisini içerecektir. Mesela ziyaretçi sepeti hakkında bilgiler oradan alınabilir (Bayramlı:2002).

Bildiğimiz gibi birçok kurum ya da kuruluş Internet üzerinde kendilerini, ürün ve hizmetlerini tanıtan bir site yapmaktadır. Eski ekonomi şirketleri denilen ve gerçek mağazası olanlardan tutun, sadece sanal ortamda vücut bulan şirketler bile bu yeni ortamdaki kazanç elde etmeye başladılar. Bir süre sonra, bu sanal mağazalar kullanıcılarına daha iyi hizmet verebilmek için sitelerine giren kullanıcının her hareketini kaydeder oldular. Kullanıcılarının yoğunlukla kullandığı sayfaları ve ürün ya da hizmetleri bu yolla anlayacaklar ve buna göre site tasarımını düzeltme şansları olacaktı. Bu amaç için, sayfa ve sayfa üzerinde gösterilen malların kayıtlarını içeren, kullanıcının hareketlerine göre yazılan bir günlük kütüğü yarattılar. Her kullanıcı hareketi, bir şekilde kütüğe gecti. Örneğin “x.html” sayfasına dört kez bakan A kullanıcısı, bu kütükte dört kayıt yaratmış olacaktı. Bu sayede hangi sayfaların rağbet gördüğü anlaşıldı, fakat esas önemli olan kullanıcı istihbaratı denen önemli ve stratejik bilgi sayfalardan alınamadı. Ziyaret verisinin kişiselleştirilmesi gerekiyordu, bu da ziyaret verisini zaten kayıtlı olan kullanıcı bilgisi ile birleştirmek suretiyle başarıldı.

Veri ambarları bu seviyede devreye girmiştir ve ziyaret kütüğü kullanıcı kütüğü ile birleştirilerek, raporların kullanıcı bazında çıkartılması başarılmıştır..

Bu gibi bir uygulama, tahmin edileceği gibi muazzam denecek kadar büyük miktarda veri ile oynamaya mecbur bırakmaktadır. Büyükçe bir kurum ya da kuruluşun mesela 100.000 sayfa ziyareti ve 5.000 kayıtlı kullanıcısı olduğunu farz etsek, bu veri içinden yararlanılabilecek bilgi çıkarmanın zorluklarını gözümüzün önüne getirebiliriz. Sonuç olarak kurum ve kuruluşlar, kendileri için yararlı olabilecek bu türde bilgileri çıkarmak için iki türlü yolu seçmişlerdir :

Bunlardan ilki, bir varsayımda bulunarak üzerinde çalışma yapmaktır. Örneğin, kullanıcılarımız arasında “x” veritabanına bakmasına rağmen bu veri tabanından yararlanmayan bir “a” grubu vardır. Bu kullanıcıları belirleyip onların istedikleri veri tabanına üye olmak, onları yeniye yönlendirmek gerekebilir. Bu varsayım sonrasında kullanıcı bilgilerimizi veri ambarındaki verilerle eşleştirip çeşitli yazılım ve sorgulama dilleri kullanarak bazı sonuçlara ulaşmamız mümkündür. Elde ettiğimiz sonuç bizim yeni bir tavır içine girmemizi gerektirebilir ya da gerektirmeyebilir. Bunu ancak bu sorgulamalar sonucu elde ettiğimiz verileri kendi ölçütlerimizle kıyaslayarak ortaya koyabiliriz.

Diğer bir yol ise, bilgisayar ve kullanılan sistemlerin otomatik olarak verileri analiz etmesi ve ilginç görülen durumların rapor edilmesi yöntemidir.

### Veri Madenciliği

Karar aşamalarında çok kritik bazı bilgiler vardır ki, sonuçların etkileri bu bilgilerin doğruluğuyla orantılıdır. Bir çok durumda cevabını tam olarak veremediğimiz sorular doğrultusunda karar verebiliriz. Kullanıcılarımızın ilgi alanları, bize karşı olan bakış açıları, diğer bilgi merkezleri ile olan ilgileri, kaynaklarımızdan yararlanma ve kullanım sıklıkları, eğitim ve bilgi düzeyleri gibi bilgiler onlara hizmetlerin kalitesi üzerinde çok net etkiler yapacaktır. Bu tür bilgiler teorik olarak her ne kadar sistemlerimizde kayıt altında olsa da, kullanılabilir bir şekilde açık ve net yanıtlara ulaşabilmemiz mevcut kullanımdaki sistemlerle neredeyse imkansız denecek kadar zordur.

Çok büyük veri yığınları altında saklı olan bu bilgilere ulaşmak için, uzun yıllar boyu yapılagelen çalışmaların neticesinde bir dizi metodoloji geliştirilmiştir. Veri madenciliği uzun yıllardır özellikle Batı ülkelerinde üzerinde çalışılan bir konu olmasına rağmen, gerçek hayatta, yazılım endüstrisinin son yıllarda üretmiş olduğu ileri teknoloji ürünü yazılımlar ile kullanılmaya başlamıştır.

Veri madenciliği, büyük miktarda veriden anlamlı bilgi çıkarma sanatıdır. Bindiği gibi, ‘veri’ ile ‘bilgi’ aynı şey değildir. Bilgi bir anlamıyla, üzerinde irdeleme yapılabilen ve eylem planı oluşturabilen bir ileti, gözlem, ve sayısal/görsel raporlar toplamıdır. Veri madenciliği, kullanılan veri madenciliği programı yardımıyla, veri ambarı üzerine çeşitli yollar ve formatlarda toplanan verileri işleyerek bize kritik ve çalışmalarımıza kullanıcı memnuniyetini sağlayacak yönde bilgi sağlamaktadır. Veri madenciliğini, örgütlerin karar aşamaları için yeni bilgiler üreten ya da gelecekle ilgili tahminler ve planlar yapmamızı sağlayan bir dizi teknikler ve anlayışlar bütünü olarak tanımlayabiliriz.

Harrod’s sözlüğünde veri madenciliği “ilk bakıldığında pek fazla anlam taşımayan veri setlerinden anlamlı bilgi çıkartma sürecidir” şeklinde tanımlanmıştır (Harrod’s, 2000 : 208).

Veri madenciliği toplanan büyük yığın halindeki veriler arasında örnek kalıpların tanımlanması, eğilimlerin belirlenmesi ve gerekli ilişkilerin kurulması işlemlerine ait bir süreçtir.

Veri madenciliği, veri tabanları, istatistik ve yapay öğrenme gibi bilgi yönetimi alt bileşenlerinin kavramlarına dayalı teknikler kullanır ve bir bilgi yönetim işinin diğer modülleri ile uyumlu bir fikirsel yapı oluşturur.

Zaman zaman veri madenciliği istatistiksel programlar ya da müşteri takip programları olarak algılandığı görülür. Dikkatten kaçmaması gereken ayrıntı veri madenciliğinin bir sorgu işleme mekanizması, yapay öğrenme sistemi olmadığıdır.

Ülkemizde son dönemlerde sık kullanılmaya başlayan veri madenciliği kavramının, Avrupa ve Kuzey Amerika ülkelerinde birbirinden çok farklı alanlarda aynı amaçla, 10 yılı aşkın süredir kullanıldığını biliyoruz. Pazarlama ve satış alanında hedef pazarların tespitinde, müşteri ilişkilerinin yönetiminde, sepet analizinde, çapraz satışlarda, pazar segmentasyonlarında ve müşteri hatırlamada sık sık veri madenciliğinden yararlanılmaktadır. Veri kaynaklarını işlemek için müşteri kartı bilgilerinin kaydedilmesinde, müşteri şikayetlerinin incelenmesinde, yaşam biçimi çalışmalarında, web ve/veya e-ticarete veri madenciliği oldukça büyük işleve sahiptir. Öte yandan birliktelik bilgilerine dayanan tahminler ve analizler de veri madenciliğinin kullanıldığı alanlar arasında sayılmaktadır.

Veri madenciliği temelde şu yararları sağlar:

- 1) Satış sırasında ve satış sonrasında kazancın artmasını sağlar,
- 2) Kayıpları azaltarak kazancın artmasını sağlar,
- 3) Daha fazla ölçülebilir değerler ortaya koyar (Data, 2003:4).

Yapay öğrenme kavramı da son yıllarda pek çok sektörde kullanılmaya başlanmış olup, zaman zaman veri madenciliği olarak da anılmaktadır. Yapay öğrenme bir makine ve yazılımının “önceki sonuçlara dayalı olarak performans sağlayabilme yeteneği” olarak tanımlanabilir (Foldoc, 2003). Gerçekte bu iki kavramın pek çok ortak noktaları bulunmaktadır. Bunun için örneğin konuşma tanımasını gerçekleştiren bir yazılımı ele alalım. Aynı sözcük insanlar tarafından farklı biçimde söylenebilmektedir. Ağızdan çıkan ses kişinin cinsiyetine, yaşına, aksanına, hatta sağlık durumuna göre farklılık gösterebilmektedir. Yapay öğrenme kendisine verilen örneklerden model çıkartarak daha sonra yapılması gerekeni kendisi belirleyebilmektedir. Bunun gibi görüntü tanıma, el yazısı tanıma vb. yazılımların da kullanıldığını bilmekteyiz.

Veri madenciliği istatistik ve bilgisayar bilimlerinin bulunduğu bir alandır. Geçmiş örnekler bir örnek kümesi oluşturuyor. İstatistik bize örnek kümeden çıkarım yapmak için gereken teoriyi sağlar. Bilgisayar bilimleri de bu çıkarımı ya-

zılım veya donanım açısından yüksek başarımlı gerçeklemede devreye girmektedir.

### **Web Madenciliği:**

Anlaşılacağı gibi veri madenciliği kurum içi ağlarda kullanılabileceği gibi, İnternet üzerinde yapılandırılmış sitelerde de büyük ölçüde kullanım imkanına sahiptir. Özellikle www yapısı üzerinde oldukça yoğun biçimde kullanılmaya başlanmıştır. Bu durum web madenciliği teriminin veri madenciliği bünyesinde geliştiğini göstermektedir. Web madenciliği sayesinde, www ile ilişkili aktiviteler ya da insan yapımı nesnelere kesin bilgi ve yararlı olabilecek örnekler ile ilginç çıkarımlarda bulunulabilmektedir. Temel olarak web madenciliği; web dokümanlarına erişimi, web üzerindeki bilginin seçim ve işlenmesini, web sitelerinden ve bunlar arasındaki geçişlerden örneklerin meydana çıkartılmasını ve bulunan örneklerin analizini içermektedir.

Web madenciliği temel olarak üç şekilde kategorize edilmektedir: Web-içerik madenciliği, web-yapı madenciliği ve web-kullanım madenciliği (Hsu, 2003 : 438).

Web-içerik madenciliği, dokümanın içeriği ya da tanımlamalarından elde edilen verilerin işlenmesidir. Bu yöntem kavramların dizinlenmesi temeline dayanan kaynak keşfi biçimindeki web metinlerinin madenciliğini içerir.

Web-yapı madenciliği, ise, www üzerindeki örgütsel yapılanma ve bağlantılardan çıkarılan bilgilerin işlenmesi anlamındadır. Web-yapı madenciliği, sayfalar üzerindeki metin ve verilere bakmak yerine, web sitelerinin yapısından bilgi çıkartmaya yönelmektedir. Bu yöntem akıllı metinler ve diğer bağlantılar gibi, web sitesi üzerindeki dokümanlar arasında var olan bağlantılar üzerinde inceleme yapar. Bu yöntemde, geçerli web sayfalarının bulunacağı linkler aracılığıyla nakledilen bilgiden yararlanılır ( Hsu, 2003 : 438).

Web-kullanım madenciliği, ki aynı zamanda web-kütükleri madenciliği olarak da bilinir, web erişim kütüklerinden ve diğer web kullanım bilgilerinden çıkarılan ilginç örneklerin işlenmesi anlamındadır. Web-kullanım madenciliği, diğer yöntemlerin izlediği yol yerine, web kullanıcısının davranışlarını ya da bir kullanıcının web'i nasıl kullandığı ve web ile nasıl karşılıklı etkileşim içine girdiğine odaklanmıştır. Bu yöntem ikincil veri ya da kullanıcı etkileşiminden üretilmiş veri ile ilgilenir. İki yolla bunu sağlar: Genel ulaşım örneğinin izlenmesi (ulaşım örnek ve eğilimlerini daha iyi anlamak için web kütüklerinin analizini yapar), isteğe uyarlanmış kullanım izlemesi (bireysel eğilimleri analiz eder).

Bunlardan web-içerik madenciliği, web tabanlı veri, doküman ve sayfalardan yeni veri ve bilgi keşfetmeyle ilgilidir. Kosala ve Blockeel'e göre, web-içerik madenciliğinde temel iki yaklaşım bulunmaktadır: bilgi erişim bakışı ve verita-

banı bakışı. Bilgi erişim bakışı; hem serbest metin gibi yapılandırılmamış hem de HTML ve hyperlink şeklindeki verilerde olduğu gibi yarı yapılandırılmış sayfalarda çalışmak için düzenlenmiştir. Bu bakış kümeleme, sınıflama, metin örnekleri bulma ve kurallar çıkarma gibi tekniklerle doküman analizine temellendirilmiş örnekler ve modeller tanımlamaya çalışır. İçerik madenciliğinde bir başka yaklaşım da benzer yapılandırılmamış doküman madenciliğinde kullanılan tekniklerin yanında, yarı yapılandırılmış dokümanlar için ortam elemanlarının bir çeşidini içeren doküman analizini gerçekleştirecek karmaşık yöntemleri de beraberinde kullanılmasıdır (Hsu, 2003 : 438-439).

Ayrıca web tabanlı veri içeren veritabanlarından daha iyi sorgulama yapılacağı düşünülen dil odaklı uygulamalar da bulunmaktadır. Araştırmacılar web kütükleri ve web SQL gibi www'den veri toplayan veri tabanı sorgulama standartlarını yaygınlaştırmak için web-yönlü sorgulama dilleri geliştirmişlerdir ( Hsu, 2003 : 438). Böylelikle farklı uygulama ve ortamlardan gelen farklı biçimlerdeki bilgiler arasında ilişki kurarak işlem yapılabilmektedir.

## **Bilgi Merkezleri ve Veri Madenciliği**

Veri madenciliği, anlamlı bilgi çıkarmak için otomasyon sürecini kullanan bir yapay zekadır. Veri madenciliğinin kütüphanelerde kullanımı her ne kadar sınırlı olsa da, bilimsel ve iş çevrelerinde birey ve gurupların davranışlarını izlemek ve daha bir çok uygulamada yıllardır kullanılmaktadır (Banerjee,1998 : 29 ).

Çevremizde terabyte'larla ölçülen bu kadar veri bulunmasına rağmen dünyada birkaç bilgi merkezi dışında gerçek veri madenciliği ve onunla beraber bulunan teknolojiyi kütüphane veri kümelerinin analizinde kullanmamaktadır.

Bugünkü endüstri kolları gibi bilgi merkezleri de ellerindeki deliller doğrultusunda karar vermek için topladıkları veriyi gruplandırmak (dilimleme ve kümeleme) için arayış içindedir. Her ne kadar terabyte'larca veri arasından tarama yapmak sorun olmasa da; bilgi merkezleri fiziksel ve sanal iş çevrelerinden farklı hareketler sonucu oluşarak bir araya gelen verinin toplanması, hazırlanması, analizi ve yorumlanması zorluklarıyla karşı karşıyadır. Bilgi merkezlerinde toplamadan analize kadar veri madenciliğinin uygulanmasına ilişkin bazı anahtar kavramlar bize rehberlik edebilir (Guenther, 2000 : 60).

Bu süreç üç aşamada tanımlanabilir: Veri seçimi ve sağlanması, veri hazırlanması ve işlenmesi, yorumlama ve bütünleştirme (Guenther, 2000: 61-63).

## Seçim ve Sağlama

Veri toplamanın ilk adımı, iki soruya karşılık arayarak en iyi şekilde yerine getirilebilir. “Toplamak istediğimiz veri nedir?” ve “Veri toplamada kullandığımız yöntem ya da işlem nedir?”. Veri seçme ve sağlamadaki yolumuz bizim birbirinden farklı hizmet hatlarımızda çakışır. Pek çok bilgi merkezinde veri; ödünç verme, derme ve danışma gibi çeşitli işlevlerin yer aldığı alanlarda üretilir.

Ödünç verme ve danışma, kullanıcı ile bilgi merkezi arasında oluşan hareketlere dayalı hizmetlerdir. Örneğin ödünç verme işlemlerinde, ödünç verilen materyalin işlemleri sırasında oluşan çok sayıdaki hareket ve ödünç verme, süre uzatma, kütüphaneler arası ödünç verme istekleri ya da bilgi kaynağı dağıtımındaki hareketleri ölçümleriz. Danışma hizmetlerinde, verinin kullanıcı gereksinimleri için yararlı olup olmadığının karşılaştırılması ile veri toplarız. Bu noktada genellikle verinin gereksinime doğru yanıt verip vermediği şeklinde veriler toplanmaktadır.

Derme geliştirmede, veriyi, kullanıcı gereksinimlerine yönelik kaynakların dermede bulunması, bunların geliştirilmesi ve devamının sağlanması sırasında bir araya getiririz. Derme geliştirme verisi genellikle ödünç verme ve danışma bölümlerinin işlemlerinin sonucunda oluşturulan verilerin kümelenmesi şeklindedir.

Yakalamak istediğimiz detayı kolaylıkla tanımlayabiliriz. Fakat yakalanan ya da elde edilen bu veri, karmaşık bütünde yeni bir katman olacaktır. Yaratılan verinin çok sayıda değişik platform ve uygulamadan gelmesi, bunun sonucunda değişik formatlarda olması zorlukları katlamaktadır.

Capital One’s Information Technology Futures’un veri tabanı uzmanı Troy Borough “Veri kullanıcılarının hepsinin aynı şeyi istediğini bunun da tüm veri kaynaklarına tek bir yer yada tarama mekanizmasından erişebilmek olduğunu söylemiştir.

Ne yazık ki bu durum yapılandırılmış yada yapılandırılmamış veri kaynaklarından taramanın oldukça zorluğunu ortaya koymaktadır. Yapılandırılmış data tipik olarak ilişkisel veri tabanlarında ya da benzer şekildeki sabit yapıdaki veri ambarlarında bulunmaktadır. Yapılandırılmamış veri ise sunu dokümanları (Power Point gibi), hesap tabloları, kelime işlem dokümanları, web sayfaları gibi başka her ortamı kapsamaktadır. Yapılandırılmış ya da yapılandırılmamış verinin her birini karmaşık uzman dizinleme mekanizmaları ile taramak oldukça kolaydır. Bununla birlikte bu iki veri şeklinin ikisini de, hepsinden önemli olan üstveri (metadata) ile bağlamadığımız takdirde, elimizdeki tüm bilgi birikimini kapsayan bir tarama gerçekleştirmek zordur.”

Bu durum farklı dosya biçimlerinden, değişik metodolojiler kullanarak elde edilen, toplanan verilerden oluşan bilgi merkezleri veri kümeleri için de aynıdır. Veri gerçek zamanda meydana gelen kontrolsüz hareketlerin döküm sayfaları yo-



luyla toplanır, örneğin danışma hizmetinde bir isteğe verilen yanıt gibi. Bunun yanında veri bize içerik sağlayıcılarımız tarafından sağlanabilir. OPAC sunucumuzdaki kullanımlar da bize diğer verileri yaratmaktadır. Bilgi merkezleri ayrıca, web siteleri ve kendi içinde kullanmakta olduğu veri tabanından veri sağlayan diğer sistemleri içeren bazı kütüphane otomasyon programları kullanmaktadır. Birbiriyle etkileşim sağlayan arayüz olmadıkça veri sağlayan çoklu sistemler çok geçerli değildir (Guenther, 2000 : 61).

### **Hazırlama ve İşleme:**

İyi düzenlenmiş bir ortamda tümleşik sistemlerden veri sağlamamız ve seçmemiz mümkündür ve ayrıca anlamlı bilgi çıkartmak için mevcut sistemler arasında sorgulama yapmak için herhangi bir ekstra adıma gereksinim duyulmaz. Fakat gerçekte pek azımız bu mükemmel dünyada yaşıyoruz. Gerçekte basılı dökümlerden, sürücü kütük dosyalarından ve bunların analizlenmesinden ya da elimizdeki veri tabanına yazdığımız küçük programlar vb. aracılığıyla topladığımız gibi, pek çok metodolojiyi kullanarak birbiriyle uyumsuz sistemlerden veri toplanmaktadır. Şirketler, yapılandırılmış ve isimlendirilmiş veri tabanlarını standart bir veri haline dönüştürmek için formal planlar hazırlamada ve veri mimarilerini uygulamada milyonlarını harcamaktadır. Şirketlerin elindeki standart veri yapısı ve değişkenler içeren yaygın veri mimarisi, tartışmaya açık veri tabanı sorgulamalarını ortadan kaldırır. Bir sorgulama birden fazla veri tabanının birleşiminden oluşan verilerin tümünü yönetmektedir. Farklı formattaki (ilişkisel veri tabanı yönetim sistemi dosya formatı) ya da farklı alan yapılandırmalarındaki (alan adı ve değişkenler) verileri bir araya getirmek için birkaç basamak daha işleme ya da gelişmiş araçlara sahip olmak gerekir.

Bilgi merkezleri için uygulamadan tarafsız veri üretmek için iki strateji vardır. Birinci strateji, farklı fiziksel platformdaki, veri işleyişinden periyodik olarak bilgiler almak ve analiz için aldıkları bu bilgileri tek bir veri tabanına yüklemeyi gerektirir. Bu tip standardizasyon ASCII tabanlı, düz dosya formatında (sıklıkla virgül yada kesme ile ayrılmış), bir program yardımıyla bilinen bir alan kümesi haline getirerek ve bilinen bir ilişkisel veri tabanı yönetim sistemine aktararak sağlanabilir. Düz dosya formatı, analiz yapılmasında bir tek veri tabanına veri aktarma sürecini en azından ortak paydada birleştirme imkanı verir. Bütün kaynak veri tabanları tarafından paylaşılamayan verilerin uygun olmaması ya da aktarımda kaybolan veriler ile zaten bozuk olan veriler aktarım sürecinde kesinlikle rol oynar.

İkinci strateji, bütün bilgi merkezi sisteminden bilinen bir veri mimarisi gerçekleştirerek ve veriyi bütünden seçip çıkartarak uygulanıdır. Bu uyumlu sistem platformlarını seçmek ve bir "İlişkisel Veri Tabanı Yönetim Sistemi"nin ben-

zer veri alan ve karakteristiklerini, üstveriyi, dizinleme şemalarını (benzeri olmayan tanıtaçlar) uygulamak anlamına gelir. Kısaca kaynak veri tabanlarından veri bütünlüğü içinde besleme yapmak anlamındadır. Fakat bu durum çok basit görülmemelidir. Bilgi merkezleri için veri yönetimine yönelik bir mimari yaratmak önemli bir durumdur. Bunun yanında söz konusu içerik sağlayıcılar olduğunda, sağlayıcıların bize verdiği içerikten (çevrimiçi dergiler, e-kitaplar ve veri tabanları gibi) ne hakkında veriye ihtiyaç duyuyoruz? Bunların her biri kendi veri mimarisini sahip olup, pek çok durumda bizim kendi dilimizde değildir. Sağlayıcı tarafından paketlenmiş ve önceden belirlenmiş veri toplanır ve PDF, HTML, Excel yoluyla sayısallaştırılmış ya da doğrudan Word yada Excel formatında belirli dönemlerde çok çeşitli formatlarda sunulabilir. Pek çok örnekte format, veri tipi ve özgüllük için fazla seçeneğimiz yoktur. Genellikle sağlayıcının gönderdiği veriyi diğer bilgi merkezi verisiyle bütünleştirmek için önemli işlemler yapmak gerekir, sürücüler evrensel boyutta kabul edilmiş veri değişim formatlarına gereksinim duyarlar. Değişim formatına dönüştürülmüş veri, toplama ve analizi işlemi kolaylaştırmak amacıyla XML olarak alınabilir. XML verinin hem dağıtım hem de değişimini kolaylaştırır.

Borough “Sağlayıcılar kendi alanlarında depoladıkları tüm verileri XML tabanlı depolamak için bazı çalışmalar yapmakta” olduğunu ileri sürmektedir. XML, verilerin farklı yerlerde depolanmalarının yönetimi için bir çözüm sağlar. Bir XML veri deposu XML formatında gelen tüm doküman tiplerini içerir. Nitekim bir çalışma tablosu verilerin arasına eklenebilir ve sonrasında bir ilişkisel veri tabanı tablosu formatında ya da bir kelime işlemci dokümanı olarak erişilebilir. XML hem yapılandırılmış hem de yapılandırılmamış verilerin bir araya getirilmesine izin veren ortak bir dildir.

Gelecekte XML, veri toplama işlemlerini doğru bir yola sokmak isteyen kütüphaneler için bir cevap olabilecektir. Uzun vadeli çözümlerde kütüphane konsorsiyumları, içerik sağlayıcılar ve yayıncıların veri değişim için üzerinde birleştikleri bir formattır. Tamamen kütüphane çevresine odaklanmış olan hem MARC kayıtları ve hem de özel kütüphane tabanlı XML taglarıyla ilişkiyi nasıl yöneteceğini ortaya koyan Document Type Definition (DTD) konusunda, Kongre Kütüphanesi MARC DTD'sini kullanarak, MARC formatındaki kataloglama verisini SGML (Standard Generalized Markup Language) formatına dönüştüren bir çalışma gerçekleştirmektedir (Guenther, 2000 : 61-62).

## **Yorumlamak ve Bütünleştirmek**

Veri toplamanın son safhası, topladığımız anlamsız veriden anlam çıkartmaktır. Verinin farklı veri havuzlarından gelmesi, farklı uygulamalarda kullanılabilmek için birtakım işlemleri gerektirecektir. Örneğin, diğer bir analiz için bir en-

vanter sayfasını bir çalışma tablosuna kopyalayabiliriz. Çalışma tabloları kolaylıkla hazırlanabilen, geçici kullanımlara kolaylıkla uyarlanabilen ve veri analizi için bir "Tak ve Kullan" aracı olarak kabul edilir.

İlişkisel veri tabanı yönetim sistemleri pek çok gücü bünyelerinde barındırır. Bunları geliştirmek çokfazla düşünmeyi gerektirir ve zaman almasına rağmen bir veri tabanının yapısal çevrede veri barındırması bakımından önemlidir. Bu durum raporlama için çok fazla seçeneğin ve karmaşık istatistiki analizler ile taramada çok büyük bir güç oluşturur. Kütüphanenin işlevsel bilgi alanları arasında ilişkiyi tanımlamak ve veri elementleri için örnek tipteki alanları ifade etmede bir veri tabanı geliştirmek büyük bir güç olacaktır (Guenther, 2000 : 62-63).

## Sonuç

Veri madenciliği ve beraberinde gelen uygulamalar (özellikle web madenciliği) uzun yıllardır Avrupa ve Kuzey Amerika'da kullanılmaktadır. Bu yöntemler öncelikle iş ve endüstri alanlarında kullanılmaya başlamıştır. Temel amaç ürün pazarlamada kullanıcı davranışlarını belirlemek, üretim ve pazarlamalarını bu yönde geliştirerek daha fazla kazanç sağlamaktır.

Belirtilen yöntemlerin bilgi merkezlerinde kullanılması ise oldukça yenidir. Ancak unutmamak gerekir ki, tıpkı iş ve endüstri alanlarında olduğu gibi, bilgi merkezleri de kullanıcı üzerine odaklanmış kurumlardır. Bilgi merkezleri veri madenciliği yöntemlerini kullanarak işlem ve hizmetleri sırasında üretilen ve kendileri için çok anlam ifade etmeyen verilerini anlamlı hale getirebilirler.

Bilgi merkezleri materyal sağlama, kataloglama ve sınıflama, ödünç verme, danışma hizmetleri vb. sırasında küçümsenmeyecek boyutlarda işlem yapmaktadırlar. Veri madenciliği yöntemleri ile otomasyona geçmiş bir bilgi merkezinde işlemler sırasında yapılan hareketler izlenerek, bu işlemler sırasında uğranılan kayıplar önlenebilir.

Uluslararası veri tabanları yine bilgi merkezlerinin büyük yatırım yaptıkları kaynaklardır. Bu yatırım, kullanıcının veri tabanlarını ne kadar, nasıl kullandığı, gerçek anlamda yararlanıp yararlanmadığı gibi sorulara yanıt almayı zorunlu kılmaktadır. Bu zorunluluk kullanıcının veri tabanları üzerindeki işlemlerini incelemeyi ve bu inceleme sonucunda analiz yapmayı gerektirmektedir.

Bilgi merkezleri hizmetlerini küresel ortamda yaymak ve kullanıcılarına ulaşabilmek için hizmetlerini İnternet ortamına taşımaktadır. İnternet üzerinde www aracılığıyla verilen hizmetlerde yalnızca verilen hizmetin hangi detayda olduğu değil, bu hizmetten yararlanacak kullanıcının web sayfalarını kullanımı ve beklentileri de önemlidir. Web madenciliği yöntemleri, kullanıcıların davranışlarını inceleyerek bilgi uzmanlarının web sayfalarını daha verimli bir şekilde, yeniden ve gerekirse kullanıcıya uyarlanmış sayfalar düzenleyebilmelerini sağlayacaktır.

Veri madenciliği ile gerek bilgi merkezi içinden, gerekse Internet üzerinden kullanıcı davranışları incelenip analiz edilerek belirlenecek kullanıcılara ait profil yaratılması mümkün olabilir. Bu işlem bireysel ya da grup belirlenerek de gerçekleştirilebilecek, hizmetler otomatik olarak detaylandırılacaktır.

Bu örnekleri çoğaltmak mümkündür. Bununla beraber bütün çalışmalar göstermektedir ki, adı ve türü ne olursa olsun, özellikle otomasyona geçmiş ve Internet üzerinden hizmet veren bilgi merkezleri yakın gelecekte, veri madenciliği uygulamaları ile veri ambarlarında biriken verileri kullanarak kendileri için anlamlı bilgiye ulaşmak bazı adımlar atmak zorunluluğu hissedeceklerdir.

## Kaynakça

- Harrod's librarians' glossary and reference book (2000). 9th ed. Aldershot: Gower.
- Data Mining: Tools, Techniques and ROI (2003). London, Canada.: Info-Tech Research Group.
- Machine learning (2003). FOLDOC: Free On-line dictionary of computing içinde Elektronik adres: <http://foldoc.doc.ic.ac.uk/foldoc/foldoc.cgi?query=machine+learning&action=Search>
- Banerjee, Kyle(1998). "Is data mining right for your library?" *Computers in Libraries*, (November/December 1998) 28-31.
- Bayramlı, Burak (2002). Veri ambarları - site satış analizi nasıl yapılır? Bilgi Yönetimi. [Çevrimiçi]. Elektronik adres: [http://www.bilgiyonetimi.org/cm/pages/mkl\\_gos.php?nt=393](http://www.bilgiyonetimi.org/cm/pages/mkl_gos.php?nt=393)
- Guenther, Kim (2000). "Building digital libraries: Applying data mining principles to library data collection" *Computers in Libraries*, (April 2000) 60-63.
- Hsu, Jeffrey (2003). "Critical and future trends in data mining: a review of key data mining technologies/ applications" *Data Mining: Opportunities and Challenges* içinde Hershey: Idea Group Inc.
- Karakaş, Melikşah (2002). Veri ambarları genel yapısı. Bilgi Yönetimi [Çevrimiçi]. Elektronik adres: [http://www.bilgiyonetimi.org/cm/pages/mkl\\_gos.php?nt=131](http://www.bilgiyonetimi.org/cm/pages/mkl_gos.php?nt=131)
- Karakaş, Melikşah. (2002) Veri madenciliği üzerine. Bilgi Yönetimi [Çevrimiçi]. Elektronik adres: [http://www.bilgiyonetimi.org/cm/pages/mkl\\_gos.php?nt=132](http://www.bilgiyonetimi.org/cm/pages/mkl_gos.php?nt=132)