T.C.
ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
BİLGİ VE BELGE YÖNETİMİ
ANABİLİM DALI

BÜYÜK VERİNİN ÜNİVERSİTE KÜTÜPHANELERİ AÇISINDAN ÖNEMİ VE KÜTÜPHANE HİZMETLERİNDE KULLANIMI: ANKARA ÜNİVERSİTESİ KÜTÜPHANELERİ ÖRNEĞİ

Yüksek Lisans Tezi

Korcan DOĞAN

Ankara-2015

T.C.
ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
BİLGİ VE BELGE YÖNETİMİ
ANABİLİM DALI

BÜYÜK VERİNİN ÜNİVERSİTE KÜTÜPHANELERİ AÇISINDAN ÖNEMİ VE KÜTÜPHANE HİZMETLERİNDE KULLANIMI: ANKARA ÜNİVERSİTESİ KÜTÜPHANELERİ ÖRNEĞİ

Yüksek Lisans Tezi

Korcan DOĞAN

Tez Danışmanı:

Prof. Dr. Sacit Arslantekin

Ankara-2015

T.C.
ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
BİLGİ VE BELGE YÖNETİMİ
ANABİLİM DALI

BÜYÜK VERİNİN ÜNİVERSİTE KÜTÜPHANELERİ AÇISINDAN ÖNEMİ VE KÜTÜPHANE HİZMETLERİNDE KULLANIMI: ANKARA ÜNİVERSİTESİ KÜTÜPHANELERİ ÖRNEĞİ

Yüksek Lisans Tezi

Tez Danışmanı:

Prof. Dr. Sacit Arslantekin

Tez Jürisi Üyeleri:

Adı ve Soyadı	İmzası
Prof. Dr. Doğan Atılgan	……..
Prof. Dr. Sacit Arslantekin	……..
Yrd. Doç. Dr. Mustafa Bayter	……..

Tez Sınavı Tarihi: ……/……../ 2015

TÜRKİYE CUMHURİYETİ

ANKARA ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Bu belge ile, bu tezdeki bütün bilgilerin akademik kurallara ve etik davranış ilkelerine uygun olarak toplanıp sunulduğunu beyan ederim. Bu kural ve ilkelerin gereği olarak, çalışmada bana ait olmayan tüm veri, düşünce ve sonuçları andığımı ve kaynağını gösterdiğimi ayrıca beyan ederim.(……/……/200…)

Tezi Hazırlayan Öğrencinin

Adı ve Soyadı

Korcan Doğan

İmzası

……

ÖNSÖZ

Büyük veri günümüzde yeni bir devrin başlangıcı olarak yorumlanmaktadır. Büyük veri ile birlikte dünyada bu derece büyük bir dönüşüm gerçekleşirken, kütüphanelerinde bu dönüşümün dışında kalmaları mümkün değildir. Kütüphaneler açısından konu iki temel başlık altında incelenmiştir. Bunlar;

i. Kütüphanelerin dermesinde büyük verinin yer alabilmesi,
ii. Kütüphane işlem ve hizmetlerinde büyük verinin, bu işlem ve hizmetleri geliştirmek için kullanılabilirliğidir.

Bu dönüşüm sürecinde ise üniversite kütüphanelerine de yapılarından dolayı büyük görevler düşmektedir ve bu dönüşümde öncü olmaları gerekmektedir.

Bu çalışma Ankara Üniversitesi Kütüphanelerinde çalışan, Bilgi ve Belge Yönetimi/Kütüphanecilik Bölümünden mezun olan kütüphanecilerin Büyük Verinin kütüphanelerde kullanımı ile ilgili, büyük veri konusunda kavramsal bilgileri yokken veya sınırlı bilgileri varken ki bakış açıları ile kavramsal düzeyde bilgi verildikten sonraki bakış açılarını değerlendirmek için ve gelecekte diğer üniversite kütüphanelerinde de bu konuda çalışmalar yapılırken, katkı sağlayacağı ve fikir vereceği düşünülerek hazırlanmıştır.

Öncelikle bu çalışmada neredeyse zaman ve ortam gözetmeksizin günün her saatinde ve ortamda çalışmanın her aşamasında bana çok büyük katkı sağlayan, benden yardımlarını ve desteğini hiç esirgemeyen, değerli fikirleriyle ve yönlendirmeleriyle, ilgili çalışmama katkıda bulunan, uzun saatler boyunca meslekle ilgili pek çok konudaki sorularıma sabırla ve hoşgörü ile yanıt veren, bu konuyla ilgilenmemde ve

anlamamda çok önemli katkıları olan, büyük veri ile katıldığım pek çok organizasyona katılmam için beni sürekli destekleyen ve bu organizasyonlara beninle birlikte gelen çok değerli hocalarım ve tez danışmanlarım Prof. Dr. Sacit Arslantekin ve Prof. Dr. Tülay Fenerci'ye sonsuz teşekkürlerimi sunarım.

Anket uygulamasının gerçekleştirilmesi ve bilgilendirici sunumunun uygulanmasında destek veren, çok değerli hocam Prof. Dr. Doğan Atılgan'a ve beni tez süresince sürekli destekleyen Prof. Dr. Fatoş Subaşıoğlu'na ve Bölümümüzdeki diğer Hocalarım ve çalışma arkadaşlarıma çok teşekkür ederim.

Son olarak çalışma süresince benden desteğini hiç esirgemeyen canım eşim Derya Doğan'a ve aileme çok teşekkür ederim.

İÇİNDEKİLER TABLOSU

ÖNSÖZ i

İÇİNDEKİLER TABLOSU iii

KISALTMALAR viii

TABLOLAR LİSTESİ ix

I . BÖLÜM 1

GİRİŞ 1

I.1 Konunun Önemi 1

I.2 Araştırmanın Amaçları 2

I.3 Araştırma Problemi, Soruları ve Hipotezleri 3

I.4 Araştırmanın Kapsamı, Yöntemi, Veri Toplama ve Değerlendirme Teknikleri 4

I.5 Kaynaklar 7

II . BÖLÜM 9

BÜYÜK VERİ' YE KAVRAMSAL YAKLAŞIM 9

II.1 Veri nedir? 9

II.1.1 Veri Tabanı 12

II.1.2 Veri Tabanı Yönetim Sistemi 13

II.1.3 Veri Ambarı 14

II.1.4 Yapılandırılmış Veri (Structured Data) 15

II.1.5 Yapılandırılmamış Veri (Unstructured Data) 16

II.1.6 Üstveri (Metadata) 17

II.2 Büyük Veri Kavramı 18

II.2.1 Büyük Veriyi Niteleyen Unsurlar 24

II.3 Nesnelerin İnterneti 29

II.4 Veri Madenciliği 30

II.5 Metin Madenciliği 32

II.5.1 Doğal Dil İşleme (Natural Language Processing - NLP) 33

II.5.2 Sayısallaştırma ve Verileştirme Arasındaki Fark 37

II.5.3 Veri Dışatımı 40

II.6 Yapay Zeka 41

II.7 NoSQL 43

II.8 Google MapReduce 44

II.9 Hadoop 45

II.10 Verinin Değeri 46

II.10.1 Verinin Gerçek Değeri ve Opsiyon Değeri Arasındaki İlişki 46

II.10.2 Verinin Amortisman Değeri 48

II.10.3 Veri Simsarlığı 49

II.11 Açık Veri 50

II.12 Büyük Veri ile ilgili Farkındalık Yaratan ve Yatırım Yapan Kuruluşlar 51

II.13 Medyada Büyük Veri 52

II.14 Dünyada Büyük Verinin Kullanım Alanları ile ilgili Örnekler 53

II.14.1 PriceStats - Billion Prices Project 53

II.14.2 Walmart 54

II.14.3 Barnes & Noble 55

II.14.4 World Cup 2014 Brazil 56

II.15 Büyük Veri ve Etik 57

III . BÖLÜM 63

KÜTÜPHANE İŞLEM VE HİZMETLERİNDE BÜYÜK VERİ 63

III.1 Kütüphane İşlem ve Hizmetlerinde Büyük Verinin Kullanımı 66

III.2 Kütüphanelerde Büyük Verinin Derme İçinde Yer Alması 69

III.2.1 RDA, Linked Data, VIAF ile Büyük Veri İlişkisi 69

III.2.2 Açık Veri ve Kütüphanecilik 72

III.2.2.1 LOC , Twitter ve Büyük Veri 72

III.2.2.2 CiteSeer^{x} Dijital Kütüphanesinde Büyük Verinin Entegrasyonu 74

III.2.2.3 OCLC ve Büyük Veri 74

III.2.3 Araştırma Verilerinin Kütüphane Dermesinde Yer Alması 76

IV . BÖLÜM 80

ANKARA ÜNİVERSİTESİ KÜTÜPHANELERİNDE ÇALIŞAN BİLGİ YÖNETİCİLERİNİN BÜYÜK VERİ FARKINDALIĞI 80

IV.1 Anket Çalışması, Kapsam ve Yöntem 81

IV.1.1 Ankara Üniversitesi Kütüphanecilerinin Büyük Veri Konusunda Genel Bulguları ve Değerlendirmesi 85

IV.1.2 Kütüphanecilerin, Kendileri ile İlgili Veri Toplayan Araçları Kullanma Düzeylerine İlişkin Değerlendirme ve Bulgular 89

IV.1.3 Kütüphanecilerin, Meslekleri ile İlgili Yeni Teknolojileri Takip Etme Düzeylerine İlişkin Değerlendirme ve Bulgular 94

IV.1.4 Kütüphanecilerin Büyük Veri Konusunda Temel Seviyede Bilgilerinin Ölçülmesine İlişkin Değerlendirme ve Bulgular 97

IV.1.5 Kütüphanecilerin Veri Toplayan Nesneler Hakkındaki Farkındalıklarının Ölçülmesine İlişkin Değerlendirme ve Bulgular 99

IV.1.6 Anket Bazında Genel Değerlendirme 103

IV.1.6.1 Birinci anket 103

IV.1.6.2 İkinci anket 105

V BÖLÜM 107

SONUÇ VE ÖNERİLER 107

KAYNAKÇA 111

EKLER 122

Ek-1: Anket Öncesi Bilgilendirme Metni 122

Ek-2: Birinci Anket 123

Ek-3: İkinci Anket 125

Ek-4: Sunum İçeriği 126

Ek-5: Sunum 127

ÖZET 138

ABSTRACT 140

KISALTMALAR

AB : Avrupa Birliği

ABD : Amerika Birleşik Devletleri

BİT : Bilgi ve İletişim Teknolojileri

ETL : Alınma (extraction) , dönüştürülme (transforming) ve yükleme (load)

NoSQL : SQL ve Daha Fazlası (not only SQL)

MARC : Machine Readable Catologing (Makinece okunabilir kataloglama)

MIT : Massachusetts Institute of Technology

MLA : Modern Language Association

NLP : Doğal Dil İşleme (Natural Language Processing)

OCLC : Online Computer Library Center

RDA : Kaynak Tanımlama ve Erişim (Resource Description and Access)

SQL : Yapılandırılmış Sorgu Dili (Structured Query Language)

URL : Uniform Resource Locator

WWW : World Wide Web

YÖK :Yükseköğretim Kurulu

TABLOLAR LİSTESİ

Tablo 1. Bazı Medya Kuruluşlarının Büyük Veri ile İlgili Yaptığı Haberlerin Başlıkları ve Yayın Tarihleri ….…….…….……52

Tablo 2. Kütüphanecilerin Mezuniyetlerinden Günümüze Kadar Geçen Yıllara Göre Sıklığı ……...……...…….……..….…….86

Tablo 3. Kütüphanecilerin Eğitim Seviyelerine Göre Sıklığı…...…….……...…..87

Tablo 4. Kütüphanecilerin Büyük Veri Kavramını Daha Önce Duyup/Duymama Durumlarına Göre Sıklığı …..……...…..……..87

Tablo 5. Kütüphanecilerin Büyük Veri Konusunda Herhangi Bir Yazı Okuyup, Okumamalarına Göre Sıklığı ……..……..88

Tablo 6. Kütüphanecilerin Sosyal Medya Uygulamalarını Kullanım Sıklığı ……...89

Tablo 7. Kütüphanecilerin Kullandıkları Birbirlerinden Farklı Sosyal Medya Uygulamaları Sıklığı ……..……..90

Tablo 8. Kütüphanecilerin e-Posta Kullanım Durumlarına Göre Sıklığı ……..…….91

Tablo 9. Kütüphanecilerin Bulut Bilişim Hizmetlerini Kullanım Sıklığı …….…..91

Tablo 10. Kütüphanecilerin Kredi Kartı veya Alışveriş Kartı Kullanım Durumlarına Göre Sıklığı ……..92

Tablo 11. Kütüphanecilerin Kendileri Hakkında Veri Toplayan Cihazların Farkındalık Durumlarına Göre Sıklığı ……...……..93

Tablo 12. Kütüphanecilerin Meslekleri İle İlgili Yeni Teknolojileri Takip Edip Etmemelerine Göre Sıklığı ……..…..94

Tablo 13. Kütüphanecilerin Meslekleri ile İlgili Yeni Teknolojileri Takip Ettikleri Kaynaklar ve Araçların Durumu ……..95

Tablo 14. Kütüphanecilerin Meslekleri ile İlgili Yeni Teknolojileri Birbirlerinden Farklı Ortamlarda Takip Etme Sıklığı ……...…….…….…….96

Tablo 15. İşlem Öncesi ve Sonrası Büyük Veri Bilgi Düzeyinin Ölçülmesinde Kullanılacak Sorular ……..…….…….…….97

Tablo 16. İşlem Öncesi ve Sonrası Bilgi Düzeyi Puanlarının Farklarına İlişkin İki Bağımlı Örnek için Wilcoxon İşaretli Sıra Sayıları Testi Sonucu …….98

Tablo 17. Kütüphanecilerin Veri Toplayan Nesneler Hakkındaki Farkındalıklarının Ölçülmesinde Yanıtların Sıklığı ……...….…….100

Tablo 18. Kütüphanecilerin Birinci Anketin İkinci Bölümüne Verdikleri Yanıtların Sıklıkları…….……..…….…….……103

Tablo 19. Kütüphanecilerin İkinci Ankete Verdikleri Yanıtların Sıklıkları ..……..105

I . BÖLÜM

GİRİŞ

I.1 Konunun Önemi

Gelişen bilgi ve iletişim teknolojilerinin (BİT) parçalarından bazıları olan internet teknolojileri; web sayfaları, bloglar, sosyal medya uygulamaları, sensörler ve daha pek çok veri toplayan cihaz ve uygulamalar sayesinde her an bilimsel olan veya olmayan verileri toplar hale gelmiştir. Günümüzde bu durum nesnelerin interneti bağlamında ele alınmaktadır.

Toplanan bu veriler, pazarlama, halkla ilişkiler, bankacılık, güvenlik vb. pek çok alanın yanında araştırmacıların yaptıkları araştırmalarda da kullanılabilir nitelik taşıyabilmektedir.

Bu verilerin günümüzde hız, çeşitlilik, kapasite (hacim) açısından büyük artış göstermesi ve bu artışa teknolojinin de destek vererek, yeni çözümler üretmesi ile birlikte "Büyük Veri" kavramı ortaya çıkmıştır.

Bu kadar büyük ve değerli niteliklere sahip veri yığınının pek çok araştırmada da kullanılması kaçınılmazdır. Nitekim bu veri yığının değerinin anlaşılması sonucunda, bu veriyi toplama, işleme, kullanıcılara hazır hale getirme, erişime sunma, saklama, analiz etme gibi aşamalarda pek çok farklı teknik de kullanılabilir hale gelmiştir.

Bu gelişmelerle birlikte kütüphaneler için büyük veri kavramı, gerek dermesinde bulundurulması bakımından, gerekse işlem ve hizmetlerinde kullanılması bakımından son derece önemli hale gelmiştir.

Bilimsel araştırma niteliği açısından, ülkemizde ve dünyadaki en önemli kütüphane türleri arasında yer alan üniversite kütüphaneleri bu gelişmeler karşısında hazırlıksız yakalanmıştır.

Araştırmamızın ortaya çıkışı, ülkemizde üniversite kütüphanecilerinin büyük veri karşısında ne gibi bir tutum içine girdikleri ve bu durumun farkında olup olmadıkları noktasından oluşmuştur.

I.2 Araştırmanın Amaçları

Ankara Üniversitesi Kütüphanelerinde çalışan ve Bilgi ve Belge Yönetimi\Kütüphanecilik Bölümlerinden mezun olmuş kütüphanecilerin örneklem olarak ele alındığı araştırmamızın temel amaçları şunlardır:

- Kütüphanecilerin büyük veri konusundaki kavramsal bilgi düzeylerini belirlemek,
- Kütüphanecilerin büyük veri ile ilgili teknolojilere yatkınlıklarını ve kullanım durumlarını saptamak,
- Büyük veri kavramı hakkında yeterli bilgiye sahip olanların veya olmayanların, büyük veri teknolojilerinin araçlarını kullanmaları nedeniyle, bu konuyla ilgili farkında olup olmama durumlarını belirlemek,

- Büyük veri ile ilgili verilen temel bilgi sonrasında bilgi ve farkındalık düzeylerindeki artışı saptamak,
- Büyük verinin kütüphane hizmetlerinde kullanımıyla ilgili ve büyük verinin kütüphanecilik mesleğinde kullanılmasıyla ilgili bilgi düzeylerini ölçmektir.

I.3 Araştırma Problemi, Soruları ve Hipotezleri

Çalışmamızda büyük verinin günümüzde gelmiş olduğu noktayı ortaya koyduktan sonra kütüphanelerin bu olgu karşısında yapması gerekenleri saptamaya çalışacağız. Bu nedenle şu sorulara yanıt arayacağız:

- Kütüphaneciler ya da bilgi yöneticileri büyük veriyi kullanarak yeni eğilimlerle kütüphane hizmet ve işlemlerinde gelişme sağlayabilecekler mi?
- Kendi yeteneklerinin farkına varıp büyük veri üzerinde fikir geliştirebilecekler mi?
- Kurumlarında büyük veriyi toplayıp işleyebileceleri ve gerek öngörülerde bulunabilecekleri, gerekse büyük veriyi hizmete sunabilecekleri alt yapıya sahipler mi?

Bütün bu sorular aklımıza öncelikle şu problemi getirmektedir. “Büyük veri ile yapılacak her türlü işlem ve hizmet konusunda bilgi yöneticileri bir farkındalığa sahipler midir?”

Çalışmamızın hipotezi yukarıdaki soru bağlamında aşağıdaki şekilde oluşturulmuştur:

“Ankara Üniversitesi Kütüphanelinde çalışan Kütüphanecilik ya da Bilgi ve Belge Yönetimi Bölümü mezunları büyük veriye kaynak teşkil edecek birçok uygulamaları kullanmalarına rağmen bunların veri ürettiklerinin ve daha başka pek çok kaynaktan sağlanarak oluşturabilecekleri büyük veri dermesini hizmete sunabilecekleri ve kendi işlemlerinde kullanabileceklerinin farkında değillerdir.”

Bu temel hipoteze bağımlı olarak aldıkları temel eğitimin doğal sonucu olarak farkındalığın yaratılması durumunda büyük veri kavramı üzerinde çalışmak ve bu konuyu uygulamaya koymak için çaba sarf etmeyi isteyecekleri alt hipotezini de ortaya koymak mümkündür.

I.4 Araştırmanın Kapsamı, Yöntemi, Veri Toplama ve Değerlendirme Teknikleri

Araştırmamız, Ankara Üniversitesi Kütüphanelerinde, bütün Bilgi ve Belge Yönetimi\Kütüphanecilik Bölümü mezunu kütüphanecilerini kapsayacak şekilde planlanmıştır.

Araştırma evreni Ankara Üniversitesi kütüphanelerinde çalışan Bilgi ve Belge Yönetimi Bölümü mezunu 35 kütüphaneciden oluşmaktadır. Bilgi ve Belge Yönetimi Bilim dalında eğitim görmüş olmaları ve mesleklerinin kütüphanecilik olması nedeniyle, araştırma evrenini Bilgi ve Belge Yönetimi Bölümü mezunu kütüphaneciler oluşturmuştur.

Araştırmanın aşamaları aşağıdaki gibidir:

- Araştırmanın ilk aşamasında kuramsal temeli oluşturmak amacıyla literatür taraması yapılmıştır. Ulusal ve uluslararası literatür taranarak büyük verinin kavramsal olarak nasıl ifade edildiği, çalışma kapsamında nasıl değerlendirileceği, büyük veri ile birlikte kullanılan kavramlar, büyük verinin tarihsel gelişimi, büyük veri ile birlikte kullanılan teknolojiler, büyük veri ile birlikte yeniden şekillenen kavramlar, dünyada büyük verinin kullanım alanları, büyük verinin kütüphanelerde kullanımı ve büyük veri ile birlikte bilim dalında yaşanan değişimler incelenmiştir.
- Araştırmanın ikinci aşamasında bulgular toplanırken Ankara Üniversitesi Kütüphanelerinde çalışan Bilgi ve Belge Yönetimi Bölümü mezunu 35 kütüphaneciye bir anket uygulanması ardından büyük veri konusunda temel bilgiler veren kısa bir sunum yapılması; bunun ardından da ikinci bir anketin tam sayım şeklinde uygulanması hedeflendirilmiştir. İlk ankete 28 kütüphaneci katılmış, sunum ve ikinci ankete ise 22 kütüphaneci katılmıştır. Ankete verilen yanıt oranı ilk anket için %80, sunum ve ikinci anket için ise %63'dür.
- Birinci anketin birinci bölümde büyük veri ile ilgili büyük veri konusunda genel bilgileri elde etmek amacıyla çoktan seçmeli sorular sorulmuştur.
- Birinci anketin ikinci bölümünde ve ikinci anketin birinci bölümünde büyük veri konusunda bilgi düzeylerini ve nesnelerin interneti konusunda tutumlarını, büyük verinin kütüphane hizmet ve işlemlerinde

kullanılması ile ilgili tutumlarını ölçen Likert Ölçeği düzeninde bir anket uygulaması yapılmıştır. Bu bölümler de sorulan sorulara, kütüphanecilerin sırasıyla verilen "Kesinlikle Katılmıyorum", "Katılmıyorum", "Kararsızım/Yeterli Bilgim Yok", "Katılıyorum", "Kesinlikle Katılıyorum" şeklinde yanıtlardan birini seçmesi istenmiştir. Likert ölçeği günümüzde kullanılan en yaygın tutum ölçme tekniğidir (Tezbaşaran 5). Likert tipi ölçekler bireyin kendisi hakkında bilgi vermesi esasına dayanmaktadır. Bireyin kendisi hakkında bilgi vermesi çok çeşitli biçimlerde ortaya çıkabilmektedir. Burada birey, çeşitli özellikleri bakımından kendisini gözleyerek, kendisi hakkındaki gözlem sonuçlarını bildirir. Tekniğin uygulanışında genellikle bireye bir soru listesi (anket, envanter, ölçek, test) verilir ve bireyden listedeki ölçek maddelerine tepkide bulunması (soruları yanıtlaması) istenir. Bu sorularda, bireyden hipotetik olarak ortaya konan durumlarda takınacağı tavrın veya göstereceği davranışın ne olacağını belirtmesi istenerek, tutumlarının ölçülmesi amaçlanır. (Tezbaşaran 6, 7).

- İkinci anketin ikinci bölümünde ise açık uçlu soru şeklinde büyük verinin kütüphanelerinde hangi hizmet ve işlemlerde kullanabileceği ile ilgili görüş ve öneriler toplanmıştır.

I.5 Kaynaklar

Araştırma kapsamında yapılan literatür taramasında kullanılan kaynaklar aşağıda görülmektedir.

- ACM Digital Library
- Ankara Üniversitesi Kütüphane Kataloğu
- Bilgi Dünyası
- Bilkent Üniversitesi Kütüphane Kataloğu
- CiteSeerX
- Ebrary
- Google Books (https://books.google.com)
- Google Scholar (https://scholar.google.com.tr)
- Milli Kütüphane Katalogları
- Orta Doğu Teknik Üniversitesi Kütüphane Kataloğu
- Türk Kütüphaneciliği
- YÖK Tez Kataloğu

Bu kaynaklar taranırken aşağıdaki anahtar kelimeler ve bunların çeşitli şekillerde birleşimleri kullanılmıştır.

- Büyük Veri (Big Data)
- Kütüphane Hizmetleri (Library Services)
- Veri Tabanı (Database)
- Veri Tabanı Yönetim Sistemi (Database Management System)
- Veri Ambarı (Data Warehouse)
- Yapılandırılmış Veri (Structured Data)

- Yapılandırılmamış Veri (Unstructured Data)
- Kısmen Yapılandırılmış Veri (Semi-Structured Data)
- Üstveri (Metadata)
- Büyük Verinin Tarihçesi (The History of Big Data)
- Nesnelerin İnterneti (Internet of Things)
- Veri Madenciliği (Data Mining)
- Metin Madenciliği (Text Mining)
- Doğal Dil İşleme (Natural Language Processing)
- Kültüromi (Culturomics)
- Gerçek Madenciliği (Reality Mining)
- Veri Dışatımı (Data Exhausting)
- Yapay Zeka (Artificial Intelligence)
- Verinin Değeri (Value of Data)
- Veri Simsarlığı (Data Broker, Information Broker)
- Büyük Veri ve Etik (Big Data and Ethics)

Araştırma raporunun yazım aşamasında, tez düzeninin oluşturulması için, Ankara Üniversitesi Sosyal Bilimler Enstitüsü'nün "Tez Yazım Yönergesi" kullanılmıştır.

Tez metni içerisinde yapılan göndermeler ve hazırlanan kaynakça, Modern Language Association (MLA) sistemine göre düzenlenmiştir.

II . BÖLÜM

BÜYÜK VERİ' YE KAVRAMSAL YAKLAŞIM

II.1 Veri nedir?

Çalışmamızın tamamının dayandığı kavram "veri" dir. Bu nedenle burada verinin tanımları üzerinde durma gereği olduğunu düşündük.

Türk Dil Kurumu Güncel Türkçe Sözlükte veri şu şekilde tanımlanmaktadır (Türk Dil Kurumu):

1. *"Bir araştırmanın, bir tartışmanın, bir muhakemenin temeli olan ana öge, muta, done*
2. *Bir sanat eserine veya bir edebî esere temel olan ana ilkeler*
3. *Gözlem ve deneye dayalı araştırmanın sonuçları*
4. *Bilgi, data*
5. *Bir problemde bilinen, belirtilmiş anlatımlardan bilinmeyeni bulmaya yarayan şey*
6. *Olgu, kavram veya komutların, iletişim, yorum ve işlem için elverişli biçimli gösterimi."*

Diğer bir kaynakta veri için *"Veri, tanım itibariyle, herhangi bir işleme tabi tutulmadan, gözlem veya ölçüm yöntemleri ile ortamdan elde edilen her türlü değerdir"* (Şeker 22) şeklinde ifade edilmektedir.

Bir başka kaynak ise veriyi "Veri, veri tabanında bulunan enformasyon için kullanılan genel bir terimdir" (Prytherch 195) şeklinde tanımlamaktadır.

Yılmaz (Yılmaz 98) veriyi, *"tek başına anlam ifade etmeyen veya kullanılamayan, bununla birlikte enformasyona ve bilgiye temel oluşturan ilişkilendirilmeye, gruplandırılmaya, yorumlanmaya, anlamlandırılmaya ve analiz edilmeye gereksinim duyulan ham bilgi"* şeklinde tanımlamıştır.

Veri araştırmalardan, gözlemlerden, internetten, sosyal medyadan, sensörlerden vb. çok farklı ortamlardan elde edilip, herhangi bir düzen, ilişkilendirme gibi belirli bir yapıda düzenlenmeyen genel bir terimi ifade etmektedir.

Veriler aşağıdaki şekilde gruplara ayrılarak nitelendirilebilir (Jeffery 6):

- Yapılandırılmış, Yapılandırılmamış, Yarı yapılandırılmış
- Statik, dinamik, akan
- Güveli / açık, özel / halka açık
- Ücretli / ücretsiz
- Açık hükümet verisi
- Açık veri
- Büyük veri

Verinin bu kadar değişik şekillerde gruplandırılması ve eskiden bu grupların çok azının yönetilebiliyor olması, çalışmada büyük veri konusuna gelindiğinde bir bakıma konunun temelini oluşturacaktır.

Veri ve büyük veri konusunda bu kavramların bilgiden ve özellikle enformasyondan farkının ortaya konulması gerekmektedir.

İletişim süreci bağlamında enformasyon için ise Gürdal (Gürdal 3), *"alıcıda şüpheyi indirgeyen ve onun üzerinde etki bırakan ve bu nitelikleri taşıyan bilginin uygun bir kanal aracılığı ile iletimi eylemi"* tanımını yapmıştır.

Enformasyonda ilgili ve ilişkili verilerin, belli bir amaç doğrultusunda bir araya getirilen veriler olduğu görülmekte ve veriler anlam kazanarak enformasyona temel oluşturmakta olduğundan bu bağlamda enformasyona kısaca anlam kazandırılmış veya yüklenmiş veri de denilebilir (Yılmaz 98).

Gürdal (Gürdal 2) enformasyondan farklı olarak bilgiyi, *"bireyin zihninde tutulan ve yalnız bireyin sahip olduğu ve/veya bilgi kayıt ortamları aracılığıyla toplumun bütün bireylerince elde edilebilen organize edilmiş, anlamlı ve ilişkili veriler bütünüdür"* şeklinde tanımlamıştır.

Kurumların ve kuruluşların sahip oldukları büyük verinin, hangi durumlarda veri, hangi durumlarda enformasyon ve hangi durumlarda da bilgi olarak tanımlanacağı, büyük veri ile gelen yeni bir tartışma konusu olabilecektir. Çünkü ileride görülebileceği gibi büyük verinin tanımı kapsamında değerlendirildiğinde saklanan büyük verinin erişim uçlarının tam olarak belirlenemediği ve tam olarak anlam kazanamadığı durumlar da meydana gelecektir. Bu da sahip olunmasına rağmen organize edilmediği, ilişkilendirilmediği hatta hangi veriye sahip olunduğunun bilinmediği durumlarda hala başkaları için enformasyon olarak nitelendirilebilecek bir yığının bir başkası için veri olarak da yorumlanabileceğini gösterebilir.

Veri kavramından bu kadar bahsettikten sonra çalışmamız kapsamında önem taşıdığını düşündüğümüz, Veri Tabanı, Veri Tabanı Yönetim Sistemi, Veri ambarları,

Yapılandırılmış ve Yapılandırılmamış Veri, Üstveri kavramlarına da değinmemiz gerekir.

II.1.1 Veri Tabanı

Yukarıda belirtilenler doğrultusunda verilerin derlenmesi, gerektiğinde ilişkilendirilip bir enformasyon haline dönüştürülebilmesi için belirli bir düzen ve sistematik doğrultusunda kayıt altına alınmaları gerekir. Bu kayıtlara gerektiğinde erişilebilmesi de temel unsurlardan birini oluşturur. Bu noktada "veri tabanı" kavramından bahsetmek gerekir.

Uysal (Uysal 31), veri tabanının tanımını, *"birbiri ile ilişkili veriler topluluğudur; veri tabanı sadece veriler yığınını değil, bunlar arasındaki ilişkileri de saklar"* şeklinde yapmıştır.

Veri tabanı, belirli bir amaç için ya da son kullanıcıların belirli bir kümesi için verilerin bilgi alanları ile organize edilerek gruplanmasını ve verinin saklama, gruplama, erişime sunulması ve raporlama gibi manipülasyonlarına olanak veren araçları sağlar. Veri tabanı, bibliyografik veri veya sayısal, istatistiksel materyal içerebilir ve ticari olarak veya bir organizasyon, kütüphane veya bireysel olarak pazarlanması amacıyla toplanabilir. Online veri tabanına erişim yönetici aracılığıyla sağlanıyor olabilmektedir. Veri tabanı yasal olarak şu şekilde tanımlanır: "iş, veri veya diğer materyal dermesinin : (a) sistematik veya metodolojik yollarla düzenlenmesi ve (b) elektronik ve diğer yollarla bireysel olarak erişilebilmesidir.". Bu genel kabul görmüş tanıma göre daha kapsamlı bir tanımlamadır. (Prytherch 196)

II.1.2 Veri Tabanı Yönetim Sistemi

Büyük veri konusuna gelmeden önce verilerin nasıl saklandığı konusu çok önemlidir ve bu kapsamda veri tabanı yönetim sistemi kavramı değerlendirilecektir.

Veri tabanını oluşturmak, yönetmek, verileri saklanmak, gerektiğinde sorgulama yapmak ve daha pek çok başka işlemi yapılabilmek için bir yazılıma gereksinim duyulur. Bu yazılım "veri tabanı yönetim sistemi" olarak karşımıza çıkar.

Veri tabanı yönetim sistemi, veri tabanın oluşturulması ve idame ettirilmesi için gerekli olan yazılım paketleridir (Prytherch 196).

Bir başka tanımda ise *"veri tabanı yönetim sistemi terimi, tam olarak bir veri tabanını ve bu veri tabanı üzerindeki yönetimle ilgili bütün yazılımları kapsamaktadır"* şeklinde tanımlanmaktadır. (Şeker 23). *"*

"Bir veri tabanını oluşturmak ve üzerinde çeşitli işlemler gerçekleştirme imkânı sağlayan bir programlar koleksiyonudur. Bu nedenle veri tabanı yönetim sistemi, çeşitli uygulamalar için veri tabanlarını tanımlama, oluşturma ve üzerinde işlem yapma yeteneği olan genel maksatlı bir yazılımdır (general purpose software system). Veri tabanının tanımlanması ise, veri tabanını oluşturan verilerin tip ve uzunluklarının belirlenmesi olarak anlaşılmaktadır. Veri tabanının oluşturulması kavramı ise veri tabanı yönetim sistemi kontrolü altında, bir depolama ortamına verilerin yüklenmesini anlatmaktadır. Bir veri tabanı üzerinde işlem yapma (manipulation) kavramının anlattığı şey ise belirli bir verinin sorgulanması, veriler üzerinde meydana gelen değişiklikleri yansıtmak üzere veri tabanının güncellenmesi ve verilerden rapor üretilmesi gibi eylemlerin gerçekleştirilmesidir. En popüler veri tabanı yönetim

sistemleri Microsoft Access, Oracle, Progress, Sybase, Db2, SQL Server, MySQL gibi yazılımlardır (Uysal 31). *“*

“Veri Tabanı Sistemi; veri tabanı, bu veri tabanı üzerinde işlem yapan yazılımlar ve veri tabanı yöneticisinden oluşan bir sistemdir. (Uysal 32). *“*

Fenerci (Fenerci 124), veri tabanı yönetim sistemini *“tüm kuruluş için gereksinilen verileri işleyen ve kullanıcıların işlemleri yürütmelerinin yanı sıra sorunlarına yanıt arayan bir sistemdir.”* şeklinde tanımlamıştır.

II.1.3 Veri Ambarı

Kimi zaman iş süreçlerimizde kullanacak veriler çeşitli karar alma mekanizmalarında yeterli olmayabilir ve dışarıdan da büyük miktarda veri girişine olanak veren sistemlere ihtiyaç duyulabilir. Bu kapsamda veri ambarı konusunun kavramsal yapısının, büyük veri bağlamında açıklanması uygun olacaktır.

Çalışmaları ile ilgili hacim olarak çok büyük miktarlarda verileri bulunan fakat tek bir noktadan yönetilmesini sağlayacak yeterli araçları bulunmayan, özellikle büyük organizasyonlara sahip kurum ve kuruluşlar için planlama stratejilerinde problemler ortaya çıkmaktadır. Satış, envanter, maliyet profilleri, satış ekibi geri bildirimleri vb. dışarıdan elde edilen verilerle, kurum ve kuruluş içi bilgiler entegre edilebilir ve veri ambarında birleştirilebilir. Bütün veriler tek bir ana kaynağa yerleştirilir ve bunlar şirkete bütün faaliyetlerinde tutarlı ve güvenilir bir görüş sağlanmasında kullanılabilir. Müşteri tabanı, alım davranışları, iş fırsatları, yeni üretim alanları gibi faktörler veri ambarında bütün elverişli veriye karşı olası muhtemel sonuçların ölçülmesi için test

edilebilir. Günümüzde de ticari veri ambarı sağlayıcıları hızla çoğalmaktadır (Prytherch 196).

Arslantekin (Arslantekin 371), veri ambarının amacını "operasyonel veri tabanlarındaki verileri kullanılabilir kılmak ve bunları dışarıdan alınan bilgi ve verilerle harmanlamaktır" şeklinde belirtmiştir.

Yukarıda anlatılanlar göstermektedir ki çeşitli kanallardan toplanan veriler rasgele bazı depolarda tutulabilmektedir. Ancak bu depolarda herhangi bir düzen söz konusu değildir. Hatta bu durum bazı kaynaklarda "Too much data not enough information" (Çok fazla veri ama yetersiz bilgi) mottosuyla da vurgulanmaktadır.

II.1.4 Yapılandırılmış Veri (Structured Data)

Yapılandırılmış ve yapılandırılmamış veri ve bazen de kısmi-yapılandırılmış veri konusu büyük veri konusunda önemli bir yer tutmaktadır. Hatta ilerideki bölümlerde de görüleceği üzere, hangi verinin büyük veri olup olmadığı konusunda da bu kavramlar önemli rol oynayacaklardır. Bu doğrultuda bu terimlerin açıklamaları önem kazanmaktadır.

Yapılandırılmış veri tipi, belirli kurallar ve sistemler doğrultusunda depolandıkları için kolay erişilebilir, düzenlenebilir, kategorize edilebilir vb. yapıdadır.

Bu veri tipi tablolar üzerinde satır ve sütunlar halinde düzenlenmiş verilerdir. Bu tür veriler kağıt üzerinde olabileceği gibi, bilgisayarlarda istatistik paket programlarının satır ve sütunlarından oluşan matris şeklindeki yapılara da kayıt

edilmiş olabilir. İstatistik ve dolayısı ile veri madenciliği yapılandırılmış verilerle ilgilenmektedir (Oğuzlar 1).

"Yapılandırılmış veri tipik olarak ilişkisel veri tabanlarında ya da benzer şekildeki sabit yapıdaki veri ambarlarında bulunmaktadır" (Arslantekin 376).

Veri tabanı yönetim sistemlerinde yapısı gereği belirli bir düzen ve ilişkiler dahilinde depolanan veriler, yapılandırılmış verilerdir. Yaygın olarak yapılandırılmış verilerden bahsedilirken, veri tabanı yönetim sistemlerinde depolanan veriler anlaşılmaktadır.

II.1.5 Yapılandırılmamış Veri (Unstructured Data)

Bu veriler mektup, doküman, kitap gibi kağıt üzerindeki veya e-mail, web sayfası gibi bilgisayardaki metinlerden; fotoğraf gibi durağan ya da film gibi hareketli görüntülerden ve seslerden oluşur. Metin madenciliği, istatistiğin ve dolayısıyla veri madenciliğinin yapılandırılmış veriler için yaptığı işi yapılandırılmamış veriler üzerinde yapmaktadır (Oğuzlar 1,2).

Yapılandırılmamış veri sunu dokümanları, hesap tabloları, kelime işlem dokümanları, web sayfaları gibi her ortamı kapsamaktadır (Arslantekin 376).

Yukarıdaki açıklamalar ışığında yapılandırılmış veri ile yapılandırılmamış veri arasındaki temel farkın, yapılandırılmış verinin üzerinde her türlü işlem ve sorgulamanın yapılabileceği, ilişkilerinin kolaylıkla kurulabileceği bir veri tabanı yönetim sistemi üzerinde bulundurulması olarak görülebilir.

II.1.6 Üstveri (Metadata)

Verilerin kayıt altına alınırken bir standart kullanılması çok çeşitli faydalar sağlamaktadır. Üstveri ise bu kapsamda büyük veri ile ilgili işlemler yapılmasında son derece önemli bir kavram olarak karşımıza çıkmaktadır.

Günümüzde bilgi ve belge kaynaklarının sayısal ortamlarda üretilmesi, kullanılması ve saklanmasından, ayrıca nesnelerin interneti ile birlikte birbirleri ile iletişim kurabilen uygulamaların hızlı bir biçimde artmasından dolayı bilgi kaynaklarını tanımlama ve bunlara erişim konularında yeni araçların varlığına gereksinim duyulmaktadır. Geleneksel bilgi kayıt ortamlarına oranla elektronik ortamlardaki kaynakların metin, ses, resim, video, vb. içeriğinden oluşan daha farklı ve fazla türe sahip olması ve yaşanan veri patlaması ile birlikte, bu verilerin daha etkin bir biçimde kimliklendirilmesi sorunu ortaya çıkmıştır. Bu verileri kimliklendirme, düzenleme, yönetme, tarama ve erişim işlevlerini sürdürmek için kullanılan en önemli araç üst veridir (Bayter, Türkçe Web Belgelerinin Kataloglanması: Bir İşbirliği Modeli Önerisi 4).

"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 ile bağlanmadığı takdirde, eldeki tüm bilgi birikimini kapsayan bir tarama gerçekleştirmek zor olacaktır. (Arslantekin 376)*"*

Üstveri, kaynağı tanımlayan veya veri hakkındaki veri olarak tanımlanabilir. MARC[1] (Machine readable catologing) kayıtları ise üst veriye örnek olarak verilebilir (Prytherch 453).

Üstveri gerekli teknik enformasyonu elektronik arşivin içeriğine koymaya ve alınabilmesine, taşındıktan sonraki sistemlerde de kullanılabilmesine olanak tanımaktadır (Prytherch 453).

II.2 Büyük Veri Kavramı

Veri, günümüzde organizasyonlar için çok büyük rekabet avantajları ve fırsatlar sunmaktadır. Bir tarafta verinin etkin şekilde toplanması, yönetilmesi ve analiz edilmesi, daha iyi iş kararları alınmasını sağlamaktayken, bir yandan da rekabet üstünlüğü sağlamaktadır. 2012 yılında Davos'ta ki Dünya Ekonomik Forumu'nda tıpkı para, altın gibi varlıklara ek olarak, yeni bir ekonomik değer olarak "veri" den bahsedilmiştir. Yeni bir değer olmasına karşın, verinin ekonomik değerini bulmak oldukça zordur. Bir başka deyişle verinin kişi, kurum kuruluş vb.lerine ekonomik katkısını rakamlarla ifade edebilmek oldukça güçtür. 2011 yılında Amerika Birleşik Devletleri'nde 17 endüstri sektöründen 15' indeki şirket başına düşen veri miktarı Birleşik Devletler Kongre Kütüphanesinin sakladığı 235 Terabayt veriden daha çoktur. Wal-Mart Mağazaları Şirketi her saat bir milyondan fazla müşterinin veri işlemini veri tabanına saklamak zorunda kalmaktadır ve sakladığı veri miktarı 2,5 petabayta ulaşmıştır. Bu rakam Kongre Kütüphanesinin elinde bulundurduğu veri miktarının yaklaşık olarak 167 katıdır. Yine bunlara benzer olarak 2010 yılında 5

[1] MARC: Makinece okunabilir kataloglama

milyar cep telefonu kullanılmıştır ve 30 milyar adet Facebook içeriği paylaşılmıştır (Johnson 51,52).”

Büyük verinin, günümüzde kullandığımız anlamıyla ilk defa kimler tarafından kullanıldığı konusu da ayrı bir önem taşımaktadır.

Necmi Gürsakal’ın büyük veri kavramı üzerine yaptığı araştırmada, Büyük Veri kavramının ilk kez Ağustos 2000’de Francis X. Diebold tarafından “Makroekonomik ölçümler ve kestirim için büyük veri dinamik faktör modelleri” (Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting) isimli bildiri ile Seattle’da 8. Dünya Ekonometri Kongresinde ortaya atıldığını belirtilmektedir (Gürsakal 20). Francis X. Diebold ise büyük veri kavramının ilk defa Silicon Graphics (SGI)’den John Mashey tarafından 1998 de “Big Data and the Next Wave of InfraStress” isimli sunumunda kullanıldığını belirtmektedir (Diebold 3). Mashey’in bu sunumuna https://www.usenix.org/legacy/event/usenix99/invited_talks/mashey.pdf adresinden ulaşılabilmektedir (Mashey).

Şu anda Gartner’ın bir parçası olan Meta Group isimli şirket ise 2001 yılında Büyük Veriyi niteleyen hacim, hız ve çeşitliliği konularından bahsederek literatüre bu konuda önemli katkı yapmıştır (Laney) ve günümüzde neredeyse her ortamda büyük veri, 3V ile anılmaktadır. Büyük verinin bu terimler ile nitelendirilmesi Meta Group’un literatüre önemli bir katkısıdır.

2012 yılında Dünya Ekonomik Forum Davos toplantılarında “Büyük veri, büyük etki: Uluslar arası kalkınma için yeni imkanlar” (Big Data, Big Impact New Possibilities for International Development) başlıklı yayınlanan raporun ((World

Economic Forum)' dan aktaran (Gürsakal 21)) büyük veri ile ilgili yaşanan gelişmelerde önemli bir yeri vardır.

Schönberger ve Cukier (Schönberger ve Cukier, Büyük Veri 14) büyük verinin kesin bir tanımının bulunmadığını ifade etmektedirler. Büyük Veri' den önce bilgi hacmi çok büyüdüğünde incelenen veri miktarının, bilgisayarların işlem için kullandıkları belleğe uygun olmadığı düşünülmekteydi. Mühendislerin, verinin tümünü analiz etmek için kullandıkları araçları yenilemeleri gerekmekteydi. Google'ın MapReduce' u ve bunun Yahoo' da ortaya çıkan açık yazılımdaki karşılığı Hadoop gibi yeni işlem teknolojilerinin kaynağı olan bu araçlar, günümüzde eskisinden çok daha fazla verinin yönetilmesine olanak tanımaktadır ve bunlarla verinin önemli bir şekilde düzenli sıralara ya da klasik veri tabanı tablolarına yerleştirilmesi gerekmemektedir (Schönberger ve Cukier, Büyük Veri 14).

Schönberger ve Cukier (Schönberger ve Cukier, Büyük 14) büyük veriyi, *"yeni içgörüler çıkarmak ya da yeni değer biçimleri yaratmak amacıyla, piyasaları, organizasyonları, vatandaşlar ile organizasyonlar arasındaki ilişkileri ve daha fazlasını değiştiren biçimlerde, insanın daha küçük ölçekte yapılamayacak ama büyük bir ölçekte yapabildiği şeyleri ifade eder"* şeklinde tanımlamıştır. Ancak yazarlar bunun sadece bir başlangıç olduğunu, büyük veri devrinin yaşama ve dünyayla etkileşim kurma biçimimize meydan okuduğunu, en çarpıcı olaraksa toplumun basit korelasyonlar karşılığında nedensellik takıntısının bir kısmından sıyrılması gerekeceğini belirtmektedirler. Yine buna ek olarak 'Neden' i bilmemek ve 'Ne'yi bilerek yüzyılların yerleşmiş uygulamalarını alaşağı ettiğini ve nasıl karar alınacağına ve gerçekliğin nasıl kavranılacağına ilişkin en temel anlayışımıza meydan okuduğu da belirtilenler arasındadır. (Schönberger ve Cukier, Büyük Veri 14,15). Büyük Veri

‘Neden’ değil ‘Ne’ hakkındadır. Bir olayın nedenini bilmeden, olay hakkında bir araştırma yapmadan, sadece çeşitli algoritmalar ve uygulamalar çalıştırılarak veriler içinde var olduğunu hiç bilmediğimiz değişkenler arası bağıntılar kurulabilir (Schönberger ve Cukier, Büyük Veri 22). Büyük veri bilgisayar biliminin yapay zeka adı verilen kolunun ve daha özel olarak otomatik öğrenme denilen alanının bir parçası olarak tarif edilmesine rağmen, bu nitelemenin yanlış yönlendirmeye neden olabileceği, onun yerine daha doğru bir ifadeyle olasılıklar çıkarmak amacıyla çok miktarda veriye matematik uygulamakla ilgili olduğu yine bu yazarlar tarafından belirtilmektedir (Schönberger ve Cukier, Büyük Veri 19). Büyük veri, insanoğlunun dünyayı ölçme ve anlama arayışında önemli bir adıma işaret etmektedir. Eskiden asla ölçülemeyen, saklanamayan, analiz edilemeyen ve paylaşılamayan şeylerin büyük çoğunluğu verileştirilmeye başlanmıştır. Toplumun gelenekselleşmiş nedensellik tercihini terk etmeye ve bazı durumlarda da korelasyonlardan istifade etmeye yönlendirmektedir (Schönberger ve Cukier, Büyük Veri 25). Son cümle Schönberger ve Cukier’in, verinin tamamına sahip olunduğunda artık olayların nedenlerini aramak için kısıtlı örneklerle araştırma evrenini tahmin etmeye gerek olmayacağını belirtmesi ve bunun yerine verinin tamamına sahip olunduğundan değişkenler arası doğru ilişkileri bularak sadece bu ilişkilerin ‘Ne’ olduklarının saptanabileceğini belirtmesi büyük veri konusuna literatürde başka bir bakış açısını sunması bakımından son derece önemlidir.

Büyük veri ile birlikte dünyayı anlamak için bir olgu hakkında geçerli ve anlamlı bir hipotezimizin olması şart değildir. Hipotezin yönlendirdiği bir yaklaşım yerine, verinin yönlendirdiği bir yaklaşım kullanılabilmektedir (Schönberger ve Cukier, Büyük Veri 63).

Büyük miktarda veri kontrol edilebildiğinde, örnekleme kavramı da artık çok anlam ifade etmez. 'N=Hepsi' ni kullanmak, veri içinde derine inilebileceği anlamına gelmektedir ve örneklemler bu kadar iyi performans gösteremezler. Veriyi toplamak ve üzerinde çalışmak için kolay ve düşük maliyetli yolların ortaya çıkması ile birlikte gereksinim duyanlar için verinin tamamı işlenebilir hale gelmektedir (Schönberger ve Cukier, Büyük Veri 34, 35).

Büyük veri diyerek ne kastedilmektedir ve bu bizlere nasıl daha fazla içgörü veya gelir getirici fırsatlar sunabilecektir? Büyük veri genel olarak kullanılan programların, saklama yönetme ve işleme kapasitesinin ötesindeki veri kümelerini anlatmak için kullanılan bir terimdir. Büyük verinin devasa boyutları ile bundan ticari anlamda fayda sağlamak için gereken analizlerin karmaşıklığının birleşmesi yeni sınıf teknolojilerin ve bunları yönetecek araçların gelişmesine neden olmuştur. Aslında büyük veri farklı şekillerde kullanılabilmektedir. Genelde hem yönetilen verinin türünü hem de onu depolamak ve işlemek için kullanılan teknolojiyi anlatmaktadır. Aslında bu teknolojilerin büyük bir kısmı Google, Amazon, Facebook ve LinkedIn gibi şirketlerin inanılmaz büyük sosyal medya verisi ile uğraşırlarken kendileri için geliştirdikleri teknolojiden doğmuştur. Bu şirketler doğası gereği, düşük maliyetli hazırda bulunan donanım ve açık kaynaklı yazılımlara önem vermektedirler (Cackett 14).

Günümüzde pek çok kuruma yalnızca kendilerine ait operasyonel veri tabanları yetmemektedir. Dış kaynaklardan alınan verilerle çeşitli analizler yapılarak yeni bilgilerin üretilmesi ve bu bilgilerin kurum için süreçlerde kullanılması ihtiyacı doğmuştur. Yaygın olan ve alışılmış ilişkisel veri tabanı sistemleri ise dış kaynaklardan gelen bu verilerin kurum içi enformasyonun yönetiminde kullanılması

konusunda yeterli desteği verememektedirler ve çoğu zamanda faydasız kalmaktadırlar. Bu yüzden "Büyük Veri" kavramı ortaya çıkmış ve günümüzde pek çok büyük teknoloji şirketi büyük veri konusunda çok büyük yatırımlar yapmışlardır. Kavram, bu yatırımların sonucunda da günümüzde son derece popüler hale gelmiştir. Literatürde büyük veri yeni bir çağın başlangıcı olarak yorumlanmış; bu gelişmeler ışında verinin bu yeniçağda petrolün yerini aldığı belirtmiştir. Bazı uzmanlar ise bilinen bütün araştırma yöntemlerini ve istatistiksel teknikleri değiştireceğinden söz etmektedir.

McKinsey Global Institute'nün 2011 yılında tanımladığı (Manyika, Chui ve Brown 1) büyük veri, tipik ve geleneksel veri tabanı yazılımlarının yapamayacağı şekilde, bunların kabiliyetlerinin ötesinde, veri kümelerini alan, saklayan, yöneten, erişime sunan ve analiz eden araçları tanımlamak için kullanılmaktadır. Bu tanım, konusu büyük veri olan Kord Davis ve Doug Patterson'ın "Ethics of Big Data" (Davis ve Patterson 4) gibi başka yayınlarda da kullanılmış olduğundan konu açısından son derece önemlidir. Biz de çalışmamızda "Büyük Veriyi, McKinsey'in yaptığı bu tanım kapsamında ele alacağız.

Bu tanım kasıtlı olarak sübjektiftir ve bir veri kümesi için büyük veri tanımı nasıl yapılacağı konusunda kapsayıcıdır. Örnek olarak büyük veri yalnızca belli bir dosya boyutundan fazla olan terabaytlar gibi veriyi ifade etmemektedir (Manyika, Chui ve Brown 1). Zaman içinde geleneksel yöntemlerin işleyebileceği veri miktarı artacak ve teknolojiler zamanla daha fazla gelişecektir. Örnek olarak McKinsey Global Institute'nün bu raporu yayınladığı yıl verdiği örnekte geleneksel veri tabanlarının terabaytlar seviyesinde veri yönetebildiğini vurgulamaktadır (Manyika, Chui ve Brown 1). Ancak günümüzde en önemli geleneksel veri tabanı yönetim sistemlerinden

biri olan Microsoft's SQL Server 2012 Parallel Data Warehouse (PDW) tek bir cihazla verileri analiz etmek ve ölçeklendirerek saklamak için 6 petabayta kadar veriyi işleyebilmektedir (Microsoft). Demek ki yalnızca 4 senede geleneksel veri tabanı yönetim sistemlerinin yönetebileceği veri miktarı çok ciddi bir oranda artmıştır. Fakat ölçü birimleri ve miktarları büyük veri tanımında ve felsefesinde bir değişikliğe yol açmamaktadır.

Burada önemli olan bir noktada günümüzde geleneksel veri tabanı yönetim sistemlerinin önemli markaları büyük veri uygulamalarını kendi geleneksel veri tabanı yönetim sistemlerinden bağımsız olarak kullanıcıların hizmetine sunmaktadırlar (Microsoft) .

Gülle (Gülle 581), "Büyük Veri ya da İçgörü" isimli yazısında büyük veriyi, içgörü bağlamında değerlendirmiş ve Kurum ve Kuruluşların stratejilerinde öncelikli olarak yer alan "öngörü" modelinin yanı sıra "içgörü" ye de bir model olarak odaklanılmasının gerekli olduğu düşüncesinde olduğunu belirtmiştir.

II.2.1 Büyük Veriyi Niteleyen Unsurlar

Önceki bölümlerde görülebileceği gibi büyük veri tanımları çok büyük kesinlik taşımamaktadır. Belki büyük veri zamanla değişecek ve bugünün büyük verisi gelecekte aynı anlama gelmeyebilecektir. Bu yüzden büyük veri kavramının tanımlamasında yardımcı olması için genellikle verinin hacmi, hızı ve çeşitliliğini ifade eden "3 V"(volume, velocity, variety) (Cackett) notasyonu, yaygın olarak onu diğer veri türlerinden ayıran kavramlar olarak kullanılmaktadır. Bu kavramlar önceki

bölümde de belirtildiği gibi ilk defa Meta Group tarafından 2001 yılında kullanılmıştır (Laney).

Nitekim büyük veri kavramının tanımlanmasındaki değişim günümüzde bile kendini göstermektedir. Literatürde 3V'ye "verinin değerini" (value) ekleyerek 4V ile tanımlayanlarda olduğu gibi (Cackett 14), henüz büyük bir kabul görmemekle birlikte, yukarıdaki 3V'ye *"veracity, variabilty, visibilty"* yani doğruluğu, değişkenliği, görünürlüğü kavramları da eklenerek "6 V" şeklinde ifade edildiğine de rastlanmaktadır (Zikopoulos, deRoos ve Parasuraman 9) ve aynı zamanda her bir v büyük verinin yeni analiz sahasına uygulanmak için doğru olup olmadığını gösteren bir test haline gelmiştir (Cackett 14).

Büyük veri kavramı aşağıdaki başlıklarda görülebileceği gibi yalnızca verinin hacminin büyüklüğünü ifade etmemektedir. Hacmi yanında hızı, çeşitliliği, değeri, doğruluğu, değişkenliği, görünürlüğü de günümüz dünyasında hızla büyümektedir. Dolayısıyla veriyi ifade eden bütün bu özelliklerin hızla büyümelerine rağmen ihtiyaçlar doğrultunda geliştirilen büyük veri teknolojileri sayesinde bu büyüklük yönetilebilir hale gelmiştir.

Biz çalışmamızda konuyu çok fazla dağıtmamak için literatürde en yaygın olarak kullanılan 3V kavramları üzerinde duracağız.

Verinin hacmi, verinin büyüklüğü ve boyutunu ifade etmektedir. Verinin boyutunu rakamsal bir şekilde belirtmek genelde çok kısıtlayıcı olmaktadır. Çünkü teknoloji ilerledikçe, rakamlar hızlı bir şekilde değişmektedir ve kısıtlayıcı rakamlar artık geçerliliğini yitirmektedir. Bu yüzden verinin göreceli miktarını belirtmek daha faydalı olmaktadır. Eğer ilgilenilen verinin miktarı daha önce endüstride kullanılan

verinin üstündeyse muhtemelen büyük veri ile uğraşılmaktadır. Bu bazı şirketler için onlarca terabayt olurken, bazıları için onlarca petabayt olabilmektedir (Cackett 14).

Günümüzde artık pek çok cihaz veri üretebilir hale gelmiştir. Gerek bireylerin ve kurumların sakladığı verilerin, gerekse de internet dünyasında saklanan verinin büyüklüğü her geçen gün artmaktadır. Veri depolama birimlerinin fiyatlarındaki hızlı düşüş, saklanan verilerin oranının geometrik şekilde hızlanarak artmasında önemli bir etken olmaktadır. Eskiden yalnızca operasyonel veri tabanlarının kullanılması yeterli olurken şimdi ise, bilgi ve iletişim teknolojilerinde yaşanan gelişmeler doğrultusunda, veri ambarlarında toplanan bütün veri işlenip analiz edilebilir hale gelmiştir.

2009 yılında dünyada 0,8 zettabayt veri bulunmaktaydı. 2011 yılı sonunda bu rakam 1,8 zettabayt oldu. 2017 veya 2018 yıllarında bu rakamın 35 zettabayt'a erişeceği tahmin edilmektedir ki bu rakam yaklaşık olarak her biri 8 GB saklama alanına sahip 4 trilyon Ipod'a eşittir. Daha iyi anlaşılması için aşağıda 1 zettabayta eşit olan ölçü birimleri verilmiştir.

1 zettabayt,

1024 eksabayt'a

1.048.576 petabayt'a

1.073.741.824 terabayt'a

1.099.511.627.776 gigabayt

1.125.899.906.842.620 megabayt

1.152.921.504.606.850.000 kilobayt

1.180.591.620.717.410.000.000 bayt

1.208.925.819.614.630.000.000.000 bit'e eşittir.

Verinin hızı, elde edilen veri ile ilgili gerçek zamanlı (anlık) olarak harekete geçilebilmesini ifade etmektedir. Her ne kadar gerçek veri analizinin verinin geldiği dönemle ayni anda tamamlanması mümkün olmasa da; uygulamaya geçmedeki gecikmeler kaçınılmaz olarak yapılması istenen ve beklenen çalışmaların verimliğini kısıtlamakta, müdahale etmeyi zorlaştırmakta ve optimal olmayan süreçlere yol açmaktadır. Örnek olarak müşterinin nerede olduğuna dayanarak yapılan bir indirim/promosyon teklifi; eğer müşteri o noktadan geçtikten sonra müşteriye ulaşırsa başarılı olma şansı çok düşmektedir (Cackett 14).

Günümüzde BİT' de yaşanan gelişmeler, verinin üretildiği anda kullanılmasına olanak vermektedir. Hızla akan veriye en hızlı tepkiyi verip, daha veri akarken müdahale etmeyi, işlemeyi ve analiz etmeyi olanaklı hale getirmiştir. Verinin bu hızına yetişebilen firmalar daha veri yeni yaratıldığı anda yanlış yapılan bir işleme müdahale edebilmekte; bu veriler ortaya çıktığı anda kurumlar kendi analiz süreçlerine katabilmekte; karar destek sistemlerindeki analiz süreçlerine aynı anda bu veriler eklenip kullanılabilmektedir.

Verinin çeşitliliğinin söz dizimi (syntax) ve semantik (anlamsal) olmak üzere 2 boyutu vardır. Geçmiş dönemlerde bu iki boyut, hangi verinin güvenilir bir biçimde veri tabanlarında yapılandırıldığının ve analizin içeriği için ne kadar güvenilir olduğunun derecesini belirlemekteydi. Modern ETL[2] araçları görsel olarak gelen sanal

[2] ETL: Ex*traction , transforming ve load işlemlerinin kısaltması . Şirketlerin networklerinde farklı yerlerde / veri tabanlarında olan bilgilerin oradan alınması (extraction) , temizlenip belli bir*

sözdizimi verilerini çok başarılı bir şekilde işleyebilirken, serbest metin gibi semantik olarak zengin verilerin çözümünde daha başarısızlardır. Bu yüzden birçok organizasyon, enformasyon yönetim sisteminin veri kapsamını daha dar bir veri düzeni ile sınırlamışlardır. Bu sınırlamayı organizasyonların daha kapsayıcı, ek değer yaratması takip etmiştir ve bu muhtemelen Büyük Veri yaklaşımının en çekici olan özeliklerinden biridir (Cackett 15).

Günümüzde veri, önceden olduğu gibi sadece yapılandırılmış veriden değil, aynı zamanda yapılandırılmamış verilerden de oluşmaktadır. Bu veriler yalnızca metin formatında değil resimler, pdf ler, sosyal medya iletileri, Excel çalışma sayfaları, e-postalar (elektronik postalar), videolar vb. pek çok çeşitten oluşmaktadırlar. Hatta bazı büyük web siteleri kullanıcının imlecini nerelerde gezdirdiğinin ve web kullanım bilgilerini bile veri olarak saklamaktadırlar. Bu çeşitlilik hem kendi arasında, hem de alt dallarıyla her geçen gün hızla büyümektedir. Büyük veri ile birlikte günümüzde bu çeşitlik verinin biçiminden bağımsız olarak iş süreçlerinde de kullanabilir hale gelmiştir. Eskiden olduğu gibi yalnızca yapılandırılmış veriler değil aynı zamanda yapılandırılmamış olarak tüm bu çeşitler de iş süreçlerinde kullanılabilir hale gelmiştir.

Çalışmanın başlarında büyük verinin oluşumunda pek çok noktadan veri toplandığından bahsedilmişti. Bu doğrultuda günümüzde veri sağlayan pek çok araçtan söz edilebilir. Bu bağlamda da karşımıza bir başka yeni kavram olan "Nesnelerin İnterneti" çıkmaktadır.

---

formata dönüştürülmesi (transforming) ve veri madenciliği yapılacak veri tabanına yüklenmesini (load) belirtir.

II.3 Nesnelerin İnterneti

Nesnelerin interneti, her gün kullanılan nesnelerin içine çipler, sensörler ve iletişim modülleri yerleştirilerek, kısmen çevrimiçi ağ oluşturmakla, ama bundan da daha çok insanları çevreleyen her şeyi verileştirmekle ilgili bir kavramdır. Dünya bir kere verileştirildiğinde, bilginin potansiyel kullanımları temel olarak sadece kişinin marifetleri ile sınırlı olacaktır. Verileştirme, insanın kavrayışında temel bir zenginleşmeyi temsil etmektedir. Büyük veri ile birlikte, bundan böyle dünya temel olarak bilgiden oluşan bir evren olarak görülebilecektir (Schönberger ve Cukier, Büyük Veri 103,104).

Nesnelerin interneti kavramı, sensörlerinin kablolu ya da kablosuz bağlantıları yoluyla, iletişim kurabilen aygıtların birbirlerine bağlanabilme kabiliyetlerini tanımlamak için kullanılmaktadır. Bu aygıtlar termostat, arabanız, doktorunuzun sağlık durumuzu kontrol edebilmesi için yuttuğunuz hap şeklinde bir ilaç vb. olabilmektedir. Bu birbirine bağlı aygıtlar verileri iletmek, derlemek ve analiz etmek için interneti kullanmaktadırlar (The White House, Executive Office of the President).

Birbirlerine bağlanan ve birbirleriyle haberleşen bu cihazların artması, veri patlamasına neden olmakta, veri patlaması sonucu büyük veri oluşmakta ve bununla birlikte günümüzde verinin analiz edilmesi ile ilgilide önemli sorunlar doğurmaktadır. Büyük veriyi depolamak ve erişime sunmak kadar, analizi içinde yeni yaklaşımlar ve yöntemler geliştirilmektedir. Bu yöntemlerin en başında veri madenciliği ve metin madenciliği gelmektedir.

II.4 Veri Madenciliği

Veri madenciliği elde edilen büyük verinin analizi için en önemli yöntemlerden biridir.

Terim olarak ilk defa 1960'larda istatistikçilerin kullandığı veri avlama (data fishing) veya veri tarama (data dredging) terimleri, veri madenciliği kavramının atalarıdır denilebilir. Bu iki terim de denizcilik literatüründen türetilmiş, veri tarama fiili balık avlanması için denizi ağ ile tarama veya deniz tabanını tarama manasına gelen özel bir terimdir. İstatistikçiler bu terimleri detaylı analizlere başlamadan önce kötü verilerin bulunup ayıklanması için kullanmışlardır (Pektaş 99).

2000'li yılların başında veri tabanı madenciliği (database mining) kavramı San Diego' da bir veritabanı firması (o dönemki adı HNC) tarafından patentlenmiş ve veri madenciliği üzerine geliştirdikleri bir program grubunu piyasaya sürmek için kullanılmıştır. Zamanla veritabanı madenciliği kavramı veri madenciliği olarak kullanılmaya başlanmıştır (Pektaş 99).

Veri madenciliği, büyük ölçekli veriler arasından bilgiye ulaşma, bilgiyi bir madenden değerli bir cevheri çıkarıyormuşçasına gün yüzüne çıkarma işidir şeklinde ya da büyük veri yığınları içerisinden gelecekle ilgili tahminde bulunabilmemizi sağlayabilecek bağıntıların bilgisayar programı kullanarak aranmasıdır şeklinde tanımlanabilir (Pektaş 100).

Şeker ise (Şeker 25) veri madenciliğini *"çeşitli şekillerde ve çeşitli kaynaklardan toplanan verilerin üzerinde işlem yapılarak anlamlı bilgilerin çıkarılmasıdır"* şeklinde tanımlamıştır.

Veri Madenciliği İngilizce-Türkçe Ansiklopedik Bilişim sözlüğünde (Sankur 831) *"Doğal Dillerin Semantik yapısına dayanarak elektronik metin belgeleri içinde saklı kalmış ilintileri, örüntüleri, stratejik bilgileri, modelleri vb. bulup ortaya çıkarmayı amaçlayan araştırma tekniği"* şeklinde tanımlanmaktadır.

Veri madenciliği, veri ambarlarında saklanan yararlı olabilecek, aralarında bilinmeyen ilişkilerin olduğu verilerin keşfedilerek, bu verilerin hem anlaşılır hem de kullanılabilir bir şekle dönüştürülmesine yönelik geliştirilmiş yöntemler topluluğudur (Oğuzlar 5,6).

Arslantekin (Arslantekin 372, 373), veri madenciliğini *"büyük miktarda veriden anlamlı bilgi çıkarma sanatıdır"* şeklinde tanımlayarak *"toplanan büyük yığın halindeki veriler arasında örnek kalıpların tanımlanması, eğilimlerin belirlenmesi ve gerekli ilişkilerin kurulması işlemlerine ait bir süreçtir"* şeklinde devam etmiştir. Büyük veriye 12 sene önce yaptığı vurgu açısından tanımlara dikkat çekmek gerekir.

Cackett ise veri madenciliğini büyük miktarlı veriden, otomatik ve yarı otomatik yöntemler kullanmak suretiyle bilinmeyen enformasyonun çıkarılması şeklinde tanımlamıştır. Bazı yayınlarda geniş veri kümelerindeki bilgi keşfi olarak anılmaktadır ki bu tanım veri madenciliğinin geçmişi şeklinde düşünülebilmektedir (Cackett 10).

İlişkileri ve modelleri analiz etme amacı ile, gizli olan enformasyonu keşfetmek için ayrıntılı teknikler kullanarak, enformasyonun veri tabanından ve veri kümelerinden çıkarılması işlemidir. (Prytherch 195)

Veri madenciliğini büyük veri bağlamında değerlendirecek olursak, elde edilen büyük verinin içindeki gizli olan enformasyonun önceden güvenilirliği kanıtlanmış istatistiksel tekniklerle ortaya çıkarılmasıdır şeklinde tanımlayabiliriz. Büyük verinin

analizinde en temel yöntemlerden birisi olması bakımından veri madenciliği bu konuda son derece önemlidir.

II.5 Metin Madenciliği

Özellikle web ortamından alınan metin halindeki büyük verinin analiz edilmesinde metin madenciliği çok önemli bir istatistiksel tekniktir. Kimi zaman metin halindeki veriler sosyal medyadan elde edildiğinde buna sosyal medya madenciliği, web üzerinden elde edildiğinde buna web madenciliği gibi tanımlar yapılsa da ve her birinin kendine özgü yöntemleri olsa da, bunların geneli temel olarak veri madenciliği ve metin madenciliğine dayanmaktadır.

Metin Madenciliği İngilizce-Türkçe Ansiklopedik Bilişim sözlüğünde "İstatistik Yöntemlere dayanarak büyük veri hacimleri, veri tabanları, örün sunucuları içinde gizli kalmış yönsemeleri, stratejik bilgileri ilişkileri keşfetmeye yönelik açınsama çalışması" şeklinde tanımlanmaktadır (Sankur 208,209).

Metin madenciliği, metin içindeki kalıpları tanımlayarak, bilinmeyen bilgiyi ortaya çıkaran ve yapısı nedeniyle de metinleri bilgiye dönüştüren anahtar bir süreçtir (Oğuzlar 8).

Metin madenciliğini oluşturan temel alanlar istatistik, veri madenciliği, doğal dil işleme, web madenciliği ve bilgi erişimidir (Oğuzlar 20).

Doğal dil metinlerindeki örüntülerle ilgilenen ve yeni eski veya bilinmeyen enformasyonu keşfetmek amacıyla doğal dil işleme, veri madenciliği ve bilgi erişim tekniklerinin uygulayan teknolojidir. Metin madenciliği birbirleriyle ilişkili

metinlerden oluşan kaynakları bir araya getirmek için, bunları analiz etmek ve tanımlamak için, anahtar varlıkları ve bunların özelliklerini içeren nitelikleri çıkarmak için ve çıkarılmış nitelikleri birleştirmek için, yeni nitelikleri şekillendirmek için veya değerli içgörü kazanmak için kullanılabilir (Prytherch 688).

Yukarıdaki tanımlarda göstermektedir ki her tanım farklı farklı noktalara değinebilmektedir. Hatta bazı tanımları herkesin anlamasında güçlük çekilmesi de söz konusudur. Metin madenciliği pek çok farklı ortamdan olabileceği gibi özellikle web ortamından elde edilen yapısal olmayan metin yapısındaki verilerin içindeki gizli enformasyonun çıkarılmasını sağlayan, bu enformasyonun çıkarılması sırasında metinlerin içindeki ilişkileri bulan ve bunları çeşitli kalıplarda ifade edilmesine olanak sağlayan, bu kalıplara dayanarak geleceğe yönelik tahminlerde bulunulmasına olanak veren istatiksel analiz yöntemidir şeklinde de tanımlanabilir.

II.5.1 Doğal Dil İşleme (Natural Language Processing - NLP)

Büyük veri teknolojilerine yatırım yapan kurum ve kuruluşlar için elde edilen metinlerin analiz edilmesinde en önemli aşamalardan biri de doğa dil işlemedir. Bu kurum ve kuruluşların, çok büyük metin verilerine sahip olmaları ve bu metinleri analiz etme gereği duymaları nedeniyle zamanla doğal dili en iyi işleyen şirketler oldukları da söylenebilir.

NLP (doğal dil işleme) doğal dillerin kurallı yapısının çözümlenerek anlaşılması veya yeniden üretilmesi amacını taşımaktadır. Bu çözümlemenin insana getirdiği kolaylıklar ise yazılı dokümanların otomatik çevrilmesi, soru-yanıt makineleri, otomatik konuşma ve komut anlama, konuşma sentezi, konuşma üretme, otomatik

metin özetleme, bilgi sağlama gibi birçok başlıkla özetlenebilmektedir. Bilgi iletişim teknolojilerinin yaygın kullanımı, bu başlıklardan üretilen uzman yazılımların hayatımızın her alanına girmesini sağlamıştır. Örneğin tüm kelime düzeltme yazılımları bir imla düzeltme aracı taşır. Bu araçlar aslında yazılan metni çözümleyerek dil kurallarını denetleyen doğal dil işleme yazılımlarıdır (Oğuzlar 13).

NLP, yapay zeka (bilgi gösterimi, planlama, akıl yürütme vb.) biçimsel diller kuramı (dil çözümleme), kuramsal dilbilim ve bilgisayar destekli dilbilim, psikoloji gibi pek çok farklı alanlarda geliştirilmiş kuram, yöntem ve teknolojileri bir araya getirmektedir (Oğuzlar 12).

Dilin insan tarafından işlenmesinin algoritmaları kesin olarak bilinmemekle birlikte, her ne kadar çeşitli sistemler bir araya getirmeye çalışsa da, tam tatmin edici bir sürece henüz ulaşılamamıştır. Bu yüzden metin anlama sistemlerinin çoğunda problemi birkaç alt göreve ayırdıktan sonra, onları ayrı olarak çözen geleneksel böl ve yönet stratejisi kullanılmaktadır (Oğuzlar 14).

1954 yılında bir IBM bilgisayarı altmış Rusça cümleyi bilgisayarın kelime dağarcığında bulunan 250 kelime çifti ve altı dilbilgisi kuralıyla İngilizce' ye çevirmiştir. Sonuçlar çok umut verici olmasına karşı 1966 yılında ilgili komite başarısızlığını itiraf etmek durumda kalmıştır. 1980'lerin sonlarında IBM araştırmacıları, bir bilgisayarı bir sözlükle birlikte kesin dil kuralları ile beslemek yerine, bir dildeki kelime ya da cümleye diğer bir dilde en yakın olanı hesaplaması için bilgisayarın istatistiksel olasılığı kullanmasına karar vermişlerdir. 1990'larda, IBM'in Candide projesi ile istatistiksel makine çevirisi olarak bilinen teknik, çeviri zorluğunu akıllı bir şekilde büyük bir matematik problemine dönüştürmüştür. Ancak bu

kavramsal sıçramanın başarısının ardından, IBM şirketi, çok para harcamasına rağmen az bir ilerleme kaydedebilmiştir ve projeyi durdurmuştur (Schönberger ve Cukier, Büyük Veri 45).

Daha fazla verinin daha iyi algoritmaları gölgede bırakma derecesi doğal dil işleme alanında (bilgisayarların kelimeleri günlük konuşmalarda kullandığımız şekilde çözümlemeyi öğrenme yolu) güçlü bir şekilde kanıtlanmaktadır. 2000 yılı civarında, Microsoft şirketinin araştırmacıları Michele Banko ve Eric Bill, şirketin Word programının bir parçası olan dilbilgisi denetleyişini iyileştirmek için bir metot aramamışlardır. Var olan algoritmaları iyileştirmeye mi, yeni teknikler bulmaya mı, yoksa daha gelişmiş özellikler eklemeye mi çaba göstermeleri gerektiğinden emin olamamışlardır. Bu yollardan herhangi birini denemeden önce var olan metotları çok daha fazla veriyle beslediklerinde ne olduğunu görmeye karar vermişlerdir. Elde ettikleri pek çok yapay öğrenme algoritmaları, toplamda bir milyon ya da daha az kelimeden oluşan metin kitaplıklarına dayanmaktaydı. Banko ve Brill yaygın olan dört algoritmayı alarak üç basamağa kadar daha fazla veri yani 10 milyon kelime, sonra 100 milyon kelime ve son olarak da bir milyar kelime ile beslediğinde karşılaştığı sonuçlar şaşırtıcı olmuştur. İçeri daha fazla veri girdikçe dört algoritma tipinin de performansı oldukça iyileşmiştir. Aslında yarım milyon kelime ile en kötü performans gösteren algoritma, bir milyar kelimeyi öğüttüğünde diğer üçünden daha iyi performans göstermiştir. Doğruluk oranı yüzde 75'ten yüzde 95'in üzerine çıkmıştır. Tersine az veriyle en iyi işleyen algoritma daha büyük miktarlarla veriyle kötü performansı göstermekle birlikte yine de yüzde 86 civarı doğruluk oranından yaklaşık yüzde 94' e ulaşarak o da iyi bir gelişme göstermiştir. Banko ve Brill, konuyla ilgili araştırma raporlarının birinde, "Bu sonuçlar, zaman ve parayı algoritma geliştirmeye

harcamak ile kitaplık geliştirmeye harcamak arasındaki dengeyi yeniden düşünmek isteyebileceğimizi akla getirmektedir" şeklinde yazmışlardır (Schönberger ve Cukier, Büyük Veri 43,44).

2006 yılında Google, "dünyanın bilgisini düzenleme ve onu erişilebilir ve faydalı hale getirme" misyonunun bir parçası olarak çeviri işine girmiştir. Google, iki dilde tercüme edilmiş sayfalarca metin yerine, daha büyük ama aynı zamanda daha dağınık olan global internetin tümü ve daha fazlası şeklindeki veri setinden yararlanmıştır. Google'ın sistemi, bilgisayarı eğitmek için bulabildiği her çeviriye başvurmuştur. Çok dilli kurumsal web siteleri, resmi belgelerin özdeş çevirileri, Birleşmiş Milletler ve Avrupa Birliği gibi hükümetler arası kuruluşlardan gelen raporları, Google'ın kitap tarama projesindeki kitapların çevirileri ve daha pek çok içerik sisteme dahil edilmiştir. Google Translate' in Başkanı Franz Josef Och' a göre Candide projesi dikkatlice tercüme edilmiş üç milyon kelimeyi kullanmışken, Google'ın sistemi çok farklı kalitede milyarlarca sayfa metinden yararlanmıştır. Bir trilyon kelimeden oluşan kitaplığı, kalitesi her ne kadar güvenilmez olsa da 95 milyar İngilizce kelimeye ulaşmıştır. Google'ın çeviri sisteminin iyi işlemesinin nedeni daha akıllı bir algoritması olması değil daha fazla veri yüklenmesidir. Google IBM'in Candide projesinden on binlerce kat daha büyük bir veri seti kullanabilmiştir. 1960'lı yıllardaki Brown Corpus'dan farklı olarak Google'da Peter Norvig ve meslektaşları *The Unreasonable Effectiveness of Data* başlıklı tebliğlerinde "Basit modeller ve çok sayıda veri, daha az veriye dayanan daha ayrıntılı modelleri gölgede bırakmıştır" diye yazmışlardır. Norvig ve ortak yazarların açıkladıkları gibi "Bu kitaplık, bazı açılardan Brown Corpus' tan bir adım geridir: Filtrelenmemiş web siteleri web sayfalarından alındığı için yarım cümleler, yazım hataları, dilbilgisi hataları içermektedirler. Dikkatli

bir şekilde elle düzeltilmiş sözcük türü etiketlerle ek açıklamalar içermemektedir. Ancak Brown Corpus'tan milyon kere daha geniş olduğu gerçeği bu kusurlarından ağır basmaktadır." (Schönberger ve Cukier, Büyük Veri 45 - 47).

Görüldüğü üzere büyük veri ve doğal dilin işlenmesi iki başlık altında incelenebilir:

i. Büyük veri ile birlikte, çok büyük miktarda veri ile çalışıldığında bilinen pek çok doğal dilin işlenmesi yöntemi değişmiş ve büyük veri, doğal dilin işlenmesi yöntemleri için kaynak oluşturmuştur. Doğal dilin işlenmesi konusunda günümüzde yaşanan önemli gelişmelerde büyük veri konusunun önemli bir etkisi olduğu söylenebilir.
ii. Büyük veri ile ilgili çalışmalarda, büyük verinin işlenerek anlamlı yeni bilgi ve öngörüler oluşturulabilmesinde doğal dil işleme yöntemleri bir araç olarak kullanılmaktadır.

Bütün bu çalışmalar web üzerindeki milyarlarca verinin işlenmesini, akıllı algoritmalar ile işlenerek anlamlı yeni bilgi ve öngörüler oluşturabilmesini kolaylaştırmaktadır

II.5.2 Sayısallaştırma ve Verileştirme Arasındaki Fark

Son yıllarda yaşanan veri patlaması ile birlikte daha önceden ilgilenilmeyen pek çok şey, veri olarak saklamaya ve bunlardan yararlanılmaya başlanmıştır. Bu doğrultuda daha önce pek kullanmadığımız verileştirme (datafication) kavramı daha çok kullanılmaya başlanmıştır. Günümüzde bilgi merkezlerinde sayısallaştırma

çalışmaları tüm hızıyla devam ederken; sayısallaştırma ve verileştirmenin farkının da bu doğrultuda incelenmesinde fayda olacaktır.

Sayısallaştırma, geleneksel formatta olan enformasyon içeriğini sayısal-okunabilir bir hale dönüştürme sürecidir. Genel olarak bir tarayıcı bu amaç ile örnek olarak basılı bir kitabı sayısallaştırırken kullanılabilir. Yazma eserlerin, resimlerin, basılı çıktıların sayısal olarak görüntülenebilir hale getirilmesi bu terimin kapsamı altındadır. Sayısallaştırma, genellikle koruma tekniklerini geliştirmek için tercih edilir, böylece özel durumlar haricinde hassas orijinal nüshaları fiziki olarak incelemek gerekmez. Orijinallerine erişimi zor olan sayısallaştırılmış metinlerin kolayca aranabilmesiyle kullanıcı erişimi büyük ölçüde geliştirilmiştir. Soluk ya da yıpranmış orijinal materyaller sayısallaştırma sayesinde bilgisayarlarca düzeltilebilmektedir. Bazı ulusal kütüphaneler hassas veya nadir bulunan kitaplar ya da malzemeler için sayısallaştırma programları yürütmektedirler (Prytherch 215).

Schönberger ve Cukier (Schönberger ve Cukier, Büyük Veri 90,91) sayısallaştırmayı eserin bir sayısal resim hale getirilmesi olarak, verileştirmeyi ise bu sayısal resimin üzerindeki harfleri, kelimleri, cümleleri, sonuç ve paragrafları tanıyabilecek optik karakter tanıma yazılımı kullanılarak verileştirilmiş bir metin hale getirilmesi olarak tanımlamaktadır. Bu tanımla birlikte verdiği örnek çok önemlidir. Google 2004 yılında, telif hakları yasalarının izin verdiği ölçüde, eline geçirebildiği her kitabın aranmasına ve bunların erişilmesine izin verecek bir projeye başladığını duyurmuştur. Bu kitapların ise sadece sayısallaştırılıp bir resim haline getirilmesi ile değil daha fazlasını isteyerek, tamamının optik karakter yazılımından geçirilerek verileştirilmiş metin hale getirilmesini sağlamıştır. Bu sayede sayfanın üzerindeki bilgi sadece insan okurlar tarafından kullanılabilir olarak kalmamış, bilgisayarlar tarafından

da işlenip algoritmalar tarafından analiz edilebilir hale getirilmiştir. Verileştirme ile birlikte sayfalar endekslenebilir ve aranabilir hale gelmiştir (Schönberger ve Cukier, Büyük Veri 90,91).

Bilgisayarların sayısal ölçümleme ve depolama aygıtlarının gelişimi, verileştirmenin çok daha etkili yapılmasını sağlamaktadır. Ayrıca verileştirme, verinin görünmeyen değerini ortaya çıkarmak için matematiksel analizi etkinleştirmektedir (Schönberger ve Cukier, Big Data).

Büyük veri ile birlikte verileştirme önem kazanmış, bu sayede otomatik dizinleme kullanan bilgi merkezlerin sayısında ve bu bilgi merkezlerinin kullanıcılarının sayısında son derece büyük artışlar meydana gelmiştir.

Bütün bunların yanında büyük verinin işlenmesinde, insan davranışını ve kültürel eğilimleri, metinlerin niceliksel analizi yoluyla anlamaya çalışan bilgisayımsal sözcük bilimi olarak tanımlanan kültüromi (culturomics) (Schönberger ve Cukier, Büyük Veri 91) ve insan davranışı hakkında çıkarımlarda ve tahminlerde bulunmak için, cep telefonlarından gelen çok fazla miktarda verinin işlenmesine atıfta bulunan gerçek madenciliği (Schönberger ve Cukier, Büyük Veri 97,98) gibi bir çok kavram üzererinde de durulabilir.

MIT'nin (Massachusetts Institute of Technology) İnsan Dinamikleri (Human Dynamics) Laboratuvarı'nın direktörü Sandy Pentland ile Nathan Eagle "gerçek madenciliği" konusunun öncüleridir. Bir çalışmalarında hareketleri ve arama modellerini analiz ederek, grip olan insanların hasta olduklarını daha kendileri fark etmeden önce başarı ile tespit etmişlerdir. Ölümcül bir grip salgınında, bu beceri, kamu sağlığı çalışanlarının herhangi bir anda en fazla etkilenen bölgeleri bilmelerine imkan

tanıyarak milyonlarca insanın hayatını kurtarabilecek önemli bilgilerin edinilmesini sağlayabilecektir (Schönberger ve Cukier, Büyük Veri 97,98). Gerçek madenciliği bilgi merkezleri için kullanıcıları ile etkileşimlerinde büyük fırsatlar sunmaktadır.

Bilgi merkezlerinin kullanıcılarının eğilimlerini anlamakta, metinlerin analiz edilmesi yöntemini kullanması gelecekte kültürominin, önemini daha da artıracaktır, ve bu konunun ayrıca incelenmesinde fayda görülmektedir.

Bütün bunların yanında düşünce madenciliği (opinion mining), ileride çok kısa değindiğimiz gerçek madenciliği (realty mining) ve benzeri madencilikle ilgili daha pek çok yeni kavrama da literatürde rastlanmaktadır. Ancak bu yeni kavramların da, bizim çalışmamızın da dayandığı temel noktalar yukarıda açıkladığımız türlerdir.

II.5.3 Veri Dışatımı

Nesnelerin interneti ile birlikte büyük veriyi üreten cihazlar yalnızca bu cihazlardan istenen ve beklenen verileri toplamamaktadır. Bu cihazları kullanan kullanıcılar farkında olmadan çok daha fazla veri üretmektedirler ve bu fazladan üretilen veriler bazen bu cihazlardan veri bekleyenler tarafından kullanılarak katma değere dönüştürülmekte bazen de kullanılamamaktadır.

Verinin yeniden kullanılması değişik ya da gizli şekillerde olabilmektedir. Örnek olarak web firmaları kullanıcıların yaptığı her şey hakkında veri ele geçirebilmektedirler. Daha sonra her bireysel etkileşimi siteyi kişiselleştirmek, bir hizmeti iyileştirmek, tamamen yeni dijital bir ürün yaratmak vb. nedenler için geri bildirim olarak kullanmanın bir işareti olarak ele alabilmektedirler. İnsanların arkalarında bıraktıkları dijital izi tanımlamak için veri dışatımı terimi kullanılmaktadır.

Terim, insanların dünyadaki eylemlerinin ve hareketlerinin yan ürünü olarak saçtıkları veriye atıfta bulunmaktadır. İnternet için, kullanıcıların, nereye tıkladıkları, bir sayfaya ne kadar süre baktıkları, fare imlecini nerede beklettikleri, ne yazdıkları vb. çevrimiçi etkileşimlerini tarif etmektedir. Veri dışatımı, ses tanıma, istenmeyen e-posta filtreleri, dil çevirisi ve çok daha fazla hizmetin arkasındaki mekanizmadır. Örneğin kullanıcılar bir ses tanıma programına söyledikleri şeyi yanlış anladığını belirttiklerinde, aslında daha iyi olması için sitemi eğitebilmektedirler (Schönberger ve Cukier, Büyük Veri 119,120).

Bu çalışmalar da büyük verinin toplanmasından sonra işlenmesinde büyük önem taşıyan gelişmelerdir.

II.6 Yapay Zeka

Nabiyev (Nabiyev 33) yapay zekayı, *"bir bilgisayarın ya da bilgisayar denetimli bir makinenin, genellikle insana özgü nitelikler olduğu varsayılan akıl yürütme, anlam çıkartma, genelleme ve geçmiş deneyimlerden öğrenme gibi yüksek zihinsel süreçlere ilişkin görevleri yerine getirme yeteneği"* olarak tanımlamaktadır.

Genel olarak yapay zeka problemleri aşağıdaki gibi sınıflara ayrılabilir (Nabiyev 96) :

- Problem Çözümleme
- Oyunların Modellenmesi
- Bilgilerin Modellenmesi
- Otomatik Teorem İspatı

- Uzman Sistemler
- Doğal Dilin İşlenmesi
- Örüntü Tanıma
- Robotik

Elmas, günümüzde bilgisayarların çok karmaşık işlemleri anında çözümleyebilmelerine rağmen, idrak etme ve deneyimlerle kazanılmış bilgileri kullanabilme konusunda çok yetersiz olduklarını (Elmas 21) ifade etse de IBM, günümüzde Watson isimli süper bilgisayarına, insan beyninin öğrenme sürecindeki gözlemleme (inceleme), yorumlama, değerlendirme ve karar verme (hükme varma), özelliklerini kazandırmıştır (IBM). Watson'ın bu konudaki, özellikle de Jeopardy isimli bilgi yarışmasında gösterdiği başarısı (IBM) Elmas'ın o dönem söylediklerinin geçmişte kaldığını ve yapay zekalı makinelerin günümüzde ne duruma geldiklerini göstermektedir.

Yapay zekalarla birlikte günümüzde çok kullanılan bir kavram olan iş zekasının da incelenmesinde fayda görülmektedir.

İş zekası, yapay zekanın iş dünyasındaki karşılığı olarak kısaca tanımlanabilir.

"İş zekası kavramının farklı açılardan farklı tanımları olmasına karşılık, genel anlamda, bir şirketteki veya organizasyondaki çeşitli seviyelerde ve çeşitli kaynaklarda bulunan verilerin işlenmesi ve karar süreçlerine etkili ve başarılı bir şekilde taşınmasıdır. Ayrıca, yönetim bilişim sistemleri açısından, en kısa tanımıyla, veri tabanından bilgiye (knowledge) ulaşmayı amaçlayan süreçtir. Literatürde KDD ismi verilen "Knowledge Discovery in Databases" kelimelerinin baş harflerinden oluşan ve Türkçeye Veri Tabanlarında Bilgi Keşfi olarak çevirebileceğimiz ve kısaca

VTBK olarak ifade edebileceğimiz bir işlem olarak tanımlamak mümkündür (Şeker 26).*"*

İş zekası ile ilgili önemli olan nokta günümüzde pek çok büyük kurum ve kuruluş büyük veriyi faydaya dönüştürmekte, veriye katma değer katmakta iş süreçlerinde iş zekası çözümleri kullanmaktadırlar. Bu yüzden büyük veri ile ilgili teknoloji ve çözüm üreten firmalarda iş zekası konusuna çok büyük yatırımlar yapmaktadırlar.

II.7 NoSQL

Son on yılda, genellikle NoSQL (not only SQL) (SQL ve daha fazlası) sistemleri olarak adlandırılan, veri yönetim sistemlerinin yeni bir sınıfı ortaya çıkmıştır ve günümüzde de oldukça hızlı bir gelişme göstermektedir.

Veri tabanlarına erişmenin en yaygın dili uzun süre SQL (Structured Query Language) ya da Türkçe karşılığı ile yapısal sorgulama dili olmuştur. Son yıllarda, önceden belirlenmiş kayıt yapısı gerektirmeyen, NoSQL' e büyük bir kayma yaşanmaktadır. NoSQL farklı tip ve büyüklüklerde veriyi kabul etmekte ve başarı ile aranmasına imkan tanımaktadır. Bu veritabanı tasarımları, yapısal dağınıklığa izin vermesinin karşılığında daha fazla işlem ve depolama alanı gerektirmektedir. Yine de bu durum, depolama ve işleme maliyetleri düştüğünden gücümüzün yetebileceği bir değişim olarak düşünülebilir (Schönberger ve Cukier, Büyük Veri 52,53).

Veritabanı tasarımında dünyanın önde gelen otoritelerinden biri olan Pat Helland, bu kökten değişimi "Eğer çok fazla veriniz varsa, o zaman 'yeteri kadar iyi' yeteri kadar iyidir" başlıklı raporunda açıklamaktadır. Geleneksel veritabanı tasarımı,

zaman içinde tutarlı sonuçlar vermeyi vaat etmektedir. Ancak toplanan verinin miktarı büyüdükçe ve sisteme erişimi olan kullanıcı sayısı arttıkça, bu tutarlılığı korumak zorlaşmaktadır (Schönberger ve Cukier, Büyük Veri 53). Eğer kuruma ait operasyonel bir veritabanı yönetilmiyorsa ve tutarlığın biraz da olsa bozulması çok önemli değilse, bu dönüşümle elde edilecek faydalar geleneksel sistemlerden çok daha fazladır.

Büyük veri ile birlikte günümüzde geleneksel veri tabanı çözümlerinin yeterli olamaması nedeniyle büyük veri uygulamalarının hemen hemen hepsinde, NoSQL sistemler kullanılmaktadır.

II.8 Google MapReduce

MapReduce bir programlama modeli ve aynı zamanda büyük veri setlerinin işlenmesi ve oluşturulmasıyla ilişkili uygulamadır (Dean ve Ghemawat 1).

Google MapReduce programlama modelini birçok farklı amaç için kullanmaktadır. Google MapReduce başarısını çeşitli sebeplere bağlamaktadır. Birincisi bu modelin kullanımı paralel ve dağıtılmış sistem tecrübesi olmayan programcıların için bile kolaydır. İkincisi büyük çeşitlilikteki problemler, MapReduce hesaplamaları gibi kolayca ifade edilebilir. Örnek olarak MapReduce veri üretmek için, Google'ın ürünü olan arama motoru servisi için, sıralama için, veri madenciliği için, makine öğrenimi için ve daha pek çok diğer sistem için kullanılabilir. Üçüncü olarak Google MapReduce'un binlerce makineyi içeren büyük makine kümelerine ölçeklenebilir bir uygulamasını geliştirmiştir. Bu uygulama makine kaynaklarını efektif kullanılmasını sağlar ve bu nedenle Google'da karşılaşılan geniş sayısal problemler için kullanımı uygundur (Dean ve Ghemawat 12).

II.9 Hadoop

Hadoop, Google MapReduce' un en büyük rakibidir ve büyük veriyi işleme konusunda günümüzde çok popülerdir ve büyük veri ile gelen değişimin simgesi olmuştur. Büyük veriyi daha küçük kümelere bölmekte ve bunları başka makinelere paylaştırmaktadır. Verinin temiz ve düzenli olmadığını yani verinin işlenmeden önce temizlenemeyecek kadar büyük olduğunu varsaymaktadır. Temel veri analizi, veriyi analiz edileceği yere taşımak için ETL denen bir işlem gerektirirken, Hadoop bu işlem yerine veri miktarının çok büyük boyutlarda olduğunu, bu nedenle taşınamayacağını ve olduğu yerde analiz edilmesi gerektiğini kabul etmektedir. Hadoop' un çıktıları ilişkisel veri tabanları kadar kesin değildir. Yani örneğin banka hesaplarının ayrıntılarını sorgulamak için pek güvenilir değildir. Ama kesin yanıtların gerekmediği birçok alanda ilişkisel veri tabanlarından çok daha hızlıdır. Buna da bir örnek vermek gerekirse Visa, Hadoop kullanarak yaklaşık 73 milyar işlemi içeren 2 yıllık kayıtları için gereken işlem süresini yaklaşık bir aydan 13 dakikaya indirebilmiştir (Schönberger ve Cukier, Büyük Veri 53,54).

Schönberger ve Cukier (Schönberger ve Cukier, Büyük Veri 55) dağınıklıkla yaşamanın karşılığında, geleneksel metotlar ve araçlarla kendi kapsamlarında çok değerli hizmetler alındığını, bütün dijital verinin sadece %5' inin yapısal olduğunu (yani geleneksel veritabanlarına tam uyan biçimde) ; dağınıklık kabul edilmezse, web sayfaları ve videolar gibi geriye kalan % 95 yapısal olmayan verinin belirsiz kalacağını; kesin olmama durumuna izin vererek, içgörülerin yararlanılamayan evrenine bir pencere açılacağından bahsetmektedir.

Hadoop günümüzde büyük veri konusunda teknoloji geliştiren ve büyük veriyi kullanan pek çok kurum ve kuruluş için vaz geçilmez olmuştur.

II.10 Verinin Değeri

Verinin önemini anlatmak için ortaya konulabilecek en belirgin kriter, verinin yapılan çalışmalara kattığı değerdir. Bu nedenle verinin değerine ilişkin bazı kavramlara yer vermek gerekir. Bunlar; verinin gerçek değeri ve opsiyon değeri arasındaki ilişki, verinin amortisman değeri ve veri simsarlığı gibi kavramlardır.

II.10.1 Verinin Gerçek Değeri ve Opsiyon Değeri Arasındaki İlişki

Ann Winblad'ın söylediği gibi "Veri yeni petroldür" (Rotella) ya da Michael Palmer'ın söylediği gibi "Veri sadece ham petroldür, rafine edilmezse, değeri vardır ama kullanılabilir değildir" (Palmer) ifadeleri gelecekte veriye sahip olanlar ve veriyi iş süreçlerinde daha iyi kullananların çok daha değerli kurumlar/firmalar/kişiler olacaklarını ve bu şekilde rakiplerine karşı ciddi fark yaratacaklarını kastetmektedir. Bu cümleler günümüzde veri ile ilgili en popüler cümlelerdendir ve büyük veri ve veri yönetimi ile ilgili (Huang), (Jeffery) vb. pek çok kaynakta geçmektedir.

Andres Weigend ise "veri yeni petroldür" sözüne aşağıdaki eklemeyi yapmıştır.

> *"Şirketler artık veriyi optimizasyon için kullanır hale geldi. Veri aslıda petrol gibidir. Şirketlerin veriyi üretmesi lazım, saklaması lazım ve analiz ederek rafine etmesi gerekiyor Bu süreçleri yerine getiren şirketler, sadece ürettikleri bu veriyi ürün olarak kullanılabilir. Bu yeni bir oyun anlamına geliyor"* (Hürriyet)

Verinin gerçek değeri, okyanusta yüzen buzdağının görünen parçası olarak tanımlanabilir. İlk bakışta sadece ufak bir kısmı görünürken, büyük bölümü deniz

seviyesinin altında gizlidir. İnovatif şirketler bu değeri ortaya çıkarabilirler (Schönberger ve Cukier, Büyük Veri 110).

Eğer veri geleceğin çok büyük değerlerinden biri ise, kurumlar bu doğrultuda yeni iş modellerini düşünmek durumundadırlar ve yeni modellerin değerli kaynağı istifade edilebilir hale getirmesi gerekmektedir (Rotella).

Veri fizikçiler gibi tasavvur edecek olunursa, verinin içinde bir potansiyel enerji mevcuttur. Bu enerji veri kullanılmaya başlandığında kinetik enerjiye dönüşecektir. Veri ilk kullanımından sonra tekrar kullanılmak üzere potansiyel enerjisini korumaktadır ve uyku haline geçmiştir (Schönberger ve Cukier, Büyük Veri 111).

Verinin opsiyon değeri, verinin kullanılabileceği bütün olası biçimlerin toplamını ifade etmektedir. Görünüşte sonsuz olan bu potansiyel kullanımlar, opsiyonlar gibidir. Verinin değeri, bu seçeneklerin toplamıdır. Büyük Veri Çağında, veri, ana değeri kullanılmaya başlandıktan sonra uzun süre değer vermeye devam eden sihirli elmas madenine benzetilmektedir. Verinin değeri, yeniden kullanarak, veri setlerini birleştirerek, vb. yöntemlerle tekrar ortaya çıkarılabilir (Schönberger ve Cukier, Büyük Veri 111). Verinin değeri için en önemli nokta görünüşte sınırsız olan yeniden kullanım potansiyeli yani verinin opsiyon değeridir. Bilgiyi toplamak önemli ama yeterli değildir, çünkü verinin sahip olduğu potansiyel yalnızca sahip olunmasında değil kullanımında yatmaktadır (Schönberger ve Cukier, Büyük Veri 129).

2011 yılında Facebook' un muhasebe kayıtlarına göre ona biçilen değer olan 6,3 milyar Dolar ile piyasanın biçtiği değer olan 104 milyar Dolar arasındaki farkın açılanmasında verinin değeri ve bu değerin hesaplanmasındaki zorluk dikkati

çekmektedir. Bu durum o yıllarda Facebook'un her sahip olduğu kullanıcısının, Facebook'un verisinin kaynağı oldukları için 100 USD değerinde olduklarını göstermektedir. Kurumsal değeri doğru belirleyememek işletme risklerinin ortaya çıkmasına ve piyasalarda dalgalanmalara neden olabilmektedir. Bu yüzden Amerika Birleşik Devletleri Senatosu 2000 yılında finansal raporlama kurallarını modernleştirmek için, bu kurallar ile ilgili toplantılar yapmışlardır. Bir şirketin defter değeri ile piyasa değeri arasındaki fark "maddi olmayan varlıklar" kısmında muhasebeleştirilmektedir. Bu maddi olmayan varlıklar eskiden marka değeri, yetenek, stratejiler gibi fiziksel ve muhasebe sisteminin parçası olmayan kalemleri içerirken, günümüzde şirketlerin sahip oldukları verileri de içermektedir. Esas sorun ise verinin değerinin nasıl belirleneceği, hatta varsa bu verinin o an bilinmeyen opsiyon değerinin nasıl belirleneceğidir. Büyük Veri çağında kurumlar sahip oldukları veriyi üçüncü kişilere kullanım hakkını satmaktan çok bu verinin opsiyon değerini kullanabilen şirketlerden ödeme alabileceklerdir. Bu fiyat üzerinden de verinin opsiyon değeri ile ilgili de bir değer saptanabilmiş olacaktır. (Schönberger ve Cukier, Büyük Veri 125, 126).

II.10.2 Verinin Amortisman Değeri

Dijital veri depolamanın maliyeti düştüğü için, kurumların veriyi aynı ya da benzer amaçlarla verinin opsiyon değerinden de yararlanarak yeniden kullanmak üzere saklamak yönünde güçlü ekonomik nedenleri bulunabilmektedir. Fakat bu faydanın da bir sınırı olması doğaldır. Çoğu veri, zamanla faydasının bir kısmını kaybetmektedir. Bu nedenle kurumlar, veriyi sadece verimli olmaya devam ettiği sürece kullanma yönünde bir dürtüye sahiptirler ve bunun sonucu olarak da sürekli verilerini

incelemeleri ve değerini kaybeden bilgiyi ayıklamaları gerekmektedir. Buradaki zorluk hangi verinin artık faydalı olmadığını bilebilmektir. Bu kararı sadece zamana dayandırmak nadiren uygun olmakla beraber, genellikle kurumlar ilgisiz verileri ayıklamak için komplike modeller geliştirmektedirler. Bu şekilde, eski verinin faydası daha iyi değerlendirilebilinir ve dolaysı ile veri için daha doğru amortisman oranları modellenebilir (Schönberger ve Cukier, Büyük Veri 117,118).

II.10.3 Veri Simsarlığı

Kitaplar, makaleler, müzikler, filmler vb pek çok eser yıllardan beri piyasalarda bilgi kaynağı olarak işlem görmektedir. Son yıllarda ise Amerika Birleşik Devletleri'nde ki Acxiom, Experian, Equifax gibi şirketler, yüz milyonlarca tüketici hakkında kapsamlı kişisel bilgi dosyaları için yüklüce para alarak veri simsarlığı yapmaktadırlar. Facebook, Twitter, LinkedIn, ve diğer sosyal medya platformları ile kişisel bağlantılarımız, fikirlerimiz, tercihlerimiz, gündelik yaşam modellerimiz, zaten yıllardır kullanıma açık olan bizimle ilgili kişisel bilgi havuzuna eklenmektedirler. Kısacası veri uzun zamandır değerli olmasına rağmen, ya bir işi yönetmenin temel operasyonlarına yardımcı olarak ya da fikri mülkiyet veya kişisel bilgi gibi dar kategorilerle sınırlı olarak görülmüştür. Büyük veri çağında ise bütün veri özünde değerli görülecektir (Schönberger ve Cukier, Büyük Veri 107).

2008 yılında İzlanda' da kurulan Data Market isimli kuruluş Birleşmiş Milletler, Dünya Bankası ve EuroStat gibi başka kaynakların veri setlerine ücretsiz erişim sağlayıp ayrıca piyasa araştırma firmaları gibi tedarikçilerden aldıkları verileri yeniden satarak gelir elde etmeye başlamıştır (Schönberger ve Cukier, Büyük Veri 128).

Yukarıda anlatılanlardan da anlaşıldığı üzere veri alınır satılır bir meta halindedir. Üretilen verinin çeşitliliği, hacmi büyük teknoloji firmalarını bu meta ile uğraşma ve diğer firmalara bu yığından bir katma değer yaratma uğraşına itmiştir. Şüphesiz günümüzde yaratılan bu katma değer de büyük ücretler karşılığı sağlanır durumdadır.

II.11 Açık Veri

Arama motorları büyük verinin öncüleri olmalarına rağmen, onlardan çok daha öncelerde kamu kuruluşları bilginin kitlesel ölçekte ilk toplayıcıları olmuşlardır. Hükümetler bilgi toplama konusundaki özel konumlarına rağmen, günümüzde bu bilgileri paylaşma konusunda bu derece istekli değillerdir. Kamunun topladığı verilerden değer çıkarmak için bunları özel sektörle paylaşma fikri günümüzde öne çıkmaktadır. Bu fikrin arkasındaki ilke ise kamu, eğer vatandaş adına veri topluyorsa ve bu verileri vatandaşlar için kullanıyorsa, bu verileri özel hayatın ve önemli sırların gizliliğini koruyarak etik değerleri gözeterek paylaşmalıdır. Açık verinin savunucuları, kamunun topladıkları verilerin sadece emanetçisi olduklarını, bu verilere toplumun ve özel sektörün ulaşması halinde daha inovatif olunabileceğini belirtmektedirler. Açık hükümet verisi fikri 21 Ocak 2009'da Amerika Birleşik Devletleri Başkanı Barrack Obama' nın devlet kurumlarının yöneticilerinden olabileceği kadar çok veriyi toplumla paylaşmaları isteği ile gelişme göstermiştir. Amerika'yı, İngiltere, Avrupa Birliği, Avustralya, Brezilya, Şili gibi ülkeler takip etmişlerdir. Bu gelişmelere paralel olarak ABD' de (Amerika Birleşik Devletleri) Code for America ve Sunlight Foundation, İngiltere'de Open Knowledge Foundation gibi veriden çok daha fazla

fayda elde etmenin yollarını arayan vizyoner düşünce toplulukları oluşmaktadır (Schönberger ve Cukier, Büyük Veri 122 -124).

II.12 Büyük Veri ile ilgili Farkındalık Yaratan ve Yatırım Yapan Kuruluşlar

Günümüzde büyük verinin bu derece önemli bir konu haline gelmesini sağlayan unsur, başta Amerika Birleşik Devletleri ve Avrupa Birliği olmak üzere pek çok devletin bu konuya çok önemli yatırımlar yapması, ayrıca kurum ve kuruluşları teşvik etmesidir. Teknoloji firmaları da bu destekle birlikte büyük veri konusuna aşağıdaki maddelerde de bir kısmının görüleceği üzere çok ciddi yatırımlar yapmakta ve bu konuda çok ciddi projeler geliştirmektedirler.

- Avrupa Komisyonu
- Beyaz Saray
- Microsoft
- IBM
- SAP
- EMC vb.

Yukarıda isimleri verilen ve bunlardan ayrı pek çok devlet, topluluk, kurum ve kuruluşun web siteleri içinde, "büyük veri" başlıklı ayrı bölümleri vardır ve bu kısımlarda büyük veri ile ilgili, kendi alanlarına yönelik önemli bilgiler vermektedirler.

II.13 Medyada Büyük Veri

Büyük verinin günümüzde çok popüler bir kavram olmasında medyanın da rolü büyüktür. Özellikle 2011 yılından sonra büyük veri konusu medyada çokça yer almıştır. Aşağıdaki tabloda bu konuya örnek olarak medyada yer alan bazı başlıkları ve medyanın ticari isimleri verilmiştir.

Tablo 1. Bazı Medya Kuruluşlarının Büyük Veri ile İlgili Yaptığı Haberlerin Başlıkları ve Yayın Tarihleri

Yayın Tarihi	Medya Adı	Metnin Başlığı
27 Mayıs 2015	Hürriyet	Büyük veri çağı (Hürriyet)
09 Haziran 2013	Hürriyet	'Büyük veri'yi anında okuyup fırsat yaratmak (Hürriyet)
05 Mayıs 2013	Hürriyet	Big Data Çağı (Tanış, Big data çağı)
20 Mayıs 2015	Hürriyet	Veriyi petrol gibi işleyin (Can)
28 Mart 2014	Hürriyet	Büyük veriyle ilgili doğru bilinen yanlışlar (Hürriyet)
27 Mayıs 2013	Cumhuriyet	'Büyük Veri', Büyük Tehlike (Yıldızoğlu)
11 Mart 2014	Harvard Business Review, Türkiye	Büyük Veri, Büyük Fırsatlar, Büyük Kararlar (Demirkan ve Dal)
Ekim 2012	Harvard Business Review	Big Data: The Management Revolution (McAfee ve Brynjolfsson)
11 Şubat 2012	The New York Times	The Age of Big Data (Lohr, The Age of Big Data)
30 Mayıs 2015	Yeni Şafak	Büyük Veri'yle ilgili 5 yanlış (Dede)
12 Haziran 2015	Sabah	Büyük veri mi, bilgi çöplüğü mü? (Oğuz)

II.14 Dünyada Büyük Verinin Kullanım Alanları ile ilgili Örnekler

Bilgi merkezlerinde ki kullanım alanlarına geçmeden önce, bilgi merkezleri içinde fikir oluşturması maksadıyla, bu bölümde dünyada büyük verinin farklı alanlardaki bazı başarılı uygulamalarına örnekler verilecektir.

II.14.1 PriceStats - Billion Prices Project

2007 yılının Ekim ayında Alberto Cavallo, çevrimiçi fiyatlardan faydalanarak, Arjantin, Şili, Brezilya ve Kolombiya için günlük enflasyon rakamları hesaplayan "Scraped Data and Sticky Prices: Frequency, Hazards, and Synchronization" isimli doktora tezini yayınlamıştır. 2008 yılının Mart ayında ise Alberto Cavallo ve Roberto Rigobon veri dermelerini 50 ülkeyi kapsayacak şekilde genişleterek, çevrimiçi verileri kullanan akademik araştırmayı yürütmek için Massachusetts Institute of Technolgy (MIT) akademik girişimi ile "Billion Prices Project" (Milyar Fiyat Projesi)'i başlatmışlardır. 2014 yılında Türkiye'nin de enflasyon rakamları bu siteye eklenmiştir (PriceStats).

Resmi istatistik kurumlarının bu rakamlara ulaşmak için ciddi miktarda zaman, para ve işgücü kullanmaları gereklidir. Ancak zamanla ilgili gecikmelerden dolayı rakamlar genellikle eskimiş olmaktadır. Örnek olarak 2008 yılındaki finansal krizin de gösterdiği gibi resmi kurumların enflasyon rakamlarını hesaplaması için geçen birkaç hafta bile uzun bir gecikme olabilmektedir. Karar alıcıların enflasyon rakamlarına daha iyi tepki verebilmeleri için onlara daha hızlı erişmeleri gerekir fakat örnekleme ve değer biçmede kesinliğe odaklanan geleneksel metotlarla bunu elde etmeleri mümkün değildir (Schönberger ve Cukier, Büyük Veri 49).

2012 yılında "The Economist" dergisi Arjantin'in resmi enflasyon rakamları yerine bu proje kapsamında hesaplanan enflasyon verilerini kullanmaya başlamışlardır (PriceStats). Bu olayla birlikte PriceStats'ın büyük veriyi kullanarak hesapladığı rakamların, kimi otoritelerce ülkelerin resmi istatistikleri yerine daha güvenilir bulunarak kullanılması ise üzerinde düşünülmesi gereken önemli bir noktadır.

II.14.2 Walmart

Walmart'ın iki milyondan fazla çalışanı vardır. Yıllık 450 milyon dolarlık cirosu, dünyadaki ülkelerin beşte dördünün gayri safi yurtiçi hasılasından daha büyüktür ve dünyadaki en büyük perakende satış zinciridir. 1990'larda Retail Link adı verilen sistem aracılığı ile perakende sektöründe devrim yapmıştır. Bu sistemle envanter hacimlerinin mal tedarikçileri tarafından görünmesine izin vererek ve tedarikçileri stoklara dikkat etmeye zorlayarak dünyanın en büyük konsinye mal mağazası haline gelmiştir. Tedarikçiler Walmart mağazalarındaki stoklarını görebildikleri için, mağazaya koydukları malların ederlerini ürünler satıldıktan sonra aldıkları için stoklarını kendileri yönetmek zorunda kalmışlardır. Şirket verilerini paylaşarak envanter riskini azaltmıştır ve maliyetlerini düşürmüştür (Schönberger ve Cukier, Büyük Veri 60,61). Walmart daha o yıllarda verinin değerini kavramış ve bu verileri iş süreçlerinde etkin bir biçimde kullanmayı başarmıştır.

Walmart 2004 yılında Teradata' nın sayısal çözümleme uzmanları ile hangi müşterinin hangi ürünü aldığı, toplam maliyetleri, alışveriş sepetlerinde başka neler olduğu, günün saatleri ve hatta durumları gibi verileri içeren devasa veri tabanlarını incelemişlerdir. Bu incelemeyi yaparken şirket, bir kasırga öncesinde sadece el feneri

satışlarının değil aynı zamanda şekerli bir Amerikan gevreği olan Pop-Tarts satışlarının da arttığını fark etmiştir. Sonrasında hızla girip çıkan müşterileri için mağazanın ön tarafındaki kasırga malzemelerinin yanına Pop-Tarts' ları depolayarak satışları önemli ölçüde artırmıştır. Geçmişte, verinin toplanması ve fikirlerin test edilmesi için merkezdeki bir çalışanın önceden içine doğması gerekirken, günümüzde Walmart bu kadar büyüklükte veriye ve daha iyi araçlara sahip olduğu için, korelasyonları çok daha hızlı ve ucuz şekilde ortaya çıkarıp bunları şirket operasyonlarında kullanarak büyük faydalar sağlayabilmiştir ve günümüzde de fayda sağlamaktadır (Schönberger ve Cukier, Büyük Veri 61).

II.14.3 Barnes & Noble

E-kitap okuma cihazları, onları okuyan insanların yazınsal tercihleri ve alışkanlıkları üzerine çok miktarda veriyi ele geçirebilmektedirler. Okuyucuların bir sayfayı ya da bir bölümü okumalarının ne kadar sürmekte oldukları, nerede okudukları, göz ucu ile mi bakıp geçtikleri ya da gerçekten mi okudukları, bir pasajın altını çizdiklerinde, bir not aldıklarında bunların kaydedilmesi, vb. olaylar örnek olarak verilebilir. Bu tür veri toplama becerisi, uzun süre yalnız bir eylem olan okumayı bir çeşit müşterek deneyime dönüştürmektedir. Bir kere toplandığında veri dışatımı yayıncılara ve yazarlara önceden hiç bilmedikleri ve beğenmedikleri nicel şeyleri söyleyebilmektedir. Bu bilgiyi ticari olarak değerlendiren e-kitap firmalarının kitapların içerik ve yapısını iyileştirmek için yayıncılara sattıkları düşünülebilir. Barnes&Noble 'ın Nook e-kitap okuyucusundan aldığı veri üzerinde yaptığı analiz, okuyucuların kurgu dışı uzun kitapları yarısında bırakmaya meyilli olduklarını ortaya koymuştur. Bu keşif şirketin "Noon Snaps" isimli bir seri yaratmasına ilham kaynağı

olmuştur. Bunlar sağlık ve güncel gelişmeler gibi gündemdeki konular üzerine kısa çalışmalar şeklindedir (Schönberger ve Cukier, Büyük Veri 121).

Bu örnekte önemli olan konulardan biri de Barnes and Nobles'dan hizmeti aldıktan sonra bile son kullanıcının, eseri okumasının her aşamasında veri üretmeye devam ettiğinin ortaya koymasıdır.

II.14.4 World Cup 2014 Brazil

2014 yılında FIFA Dünya Kupası şampiyonu Almanya olmuştur (Brazilian Federal Government). Turnuvada SAP ve Almanya Futbol Federasyonu (DFB), kupadaki oyuncu performanslarını artırmak için büyük veriyi akıllı kararlara dönüştürecek bir sistem için inovatif bir işbirliği yapmışlardır. SAP HANA platformunda çalışan bu çözüm, antrenmanların, hazırlıkların ve turnuvaların analizi kolaylaştırmak için ve oyuncu - takım performanslarını artırmak için tasarlanmıştır. Oliver Bierhoff bu konuda 10 dakikada 10 oyuncunun, 7 milyon fazla veri noktasından veri ürettiğini belirtmektedir. İlgili çözüm ile bir sonraki maçın antrenmanlarının ve hazırlığının yapılmasında bu inanılmaz büyüklükteki verinin analizi gerçekleştirilebilmiştir (SAP SE).

Futbol dünyasında başarının gelmesinde payı olan bu çözüm hem büyük verinin kullanım alanlarının çeşitliliği ve geliştirilmesi konusunda hem de spor dünyası açısından çok önemli bir gelişmedir. Maç içinde toplanan verinin, enformasyona dönüşmesinde ve bu enformasyonun karar süreçlerinde kullanılarak önemli bilgiler elde edilmesinde büyük veri uygulamalarında ne kadar önemli olduğu bu örnekte değerlendirilebilir. Şüphesiz bu verilerin toplanması ve karar süreçlerinde kullanıma

hazır hale gelmesinde veri ve veri kaynakları üzerinde analitik düşünebilen bilgi yöneticilerine olan ihtiyaç da göz ardı edilmemelidir.

II.15 Büyük Veri ve Etik

Günümüzde büyük veri üzerinde durulan bir önemli nokta da etik yaklaşımdır. Kişiler/kurumlar hakkında değişik kaynaklarca toplanan verilerin üçüncü şahıslarla paylaşımının etik sorunları gündeme getirmesi doğaldır.

Büyük veri ve etik; kişi kimlik bilgileri (identify), mahremiyet (privacy) (özel yaşamın gizliliği), mülkiyet (ownership) ve itibardan (reputation) oluşan dört temel çerçevede incelenebilir (Davis ve Patterson 2,3). Bu 4 temel çerçeve aşağıda ayrıntıları ile açıklanmıştır.

- **Kimlik:** Bu bölümde çevrimdışı (offline) ve çevrimiçi (online) kimliklerimizin arasındaki ilişki üzerinde durulacaktır (Davis ve Patterson 3). Kim olduğumuzun yani kimliğimizin birçok boyutu bulunmaktadır ve normalde bütün bu boyutların tüketim için sadece bir kurum veya kişi tarafından toplanması çok mümkün değildir. Kimliğimizin birçok boyutu olduğu gibi değerlerimizin ve etik ilişkilerimizin de birden fazla boyutu olabilmektedir. Artık büyük veri sayesinde kurumlar bireylerin izinleri olmadan veya katılımları olmadan kolayca tüm değerlerimizi ve çevrimdışı kimliklerimiz ile çevrimiçi kimliklerimiz arasındaki bağlantıları kurabilecek duruma gelmişlerdir. Buradan da uygun olan veya uygun olmayan davranışlarımızı çıkarabileceklerdir. Özetle büyük veri bizim iznimiz veya onayımız

olmadan üçüncü kişilere kimliğimizi ve çevrimdışı/çevrimiçi davranışlarımızı daha kolay özetleme, toplama ve ilişkilendirme şansı vermektedir. Bu da aslında kimliklerimizin, değerlerimizin ve davranışlarımızın bir bütün olarak değerlendirilmesini beraberinde getirmektedir. Bunların hepsi daha açık bir şekilde değerlendirildikçe dünyadaki etik ilişkileri de değiştirmektedir. (Davis ve Patterson 16). Örnek vermek gerekirse bir kişi vergi dairesi ile kimlik bilgilerini paylaştığında ya da vergi dairesi bireyin izni olmadan bu verileri elde ettiğinde, o kişi sosyal medyayı farklı bir kimlikle kullanmasına rağmen vergi dairesi, büyük veri ile birlikte bu iki kimliği birbiriyle, bireyin izini olmadan ilişkilendirebilme yeteneğine kavuşabilecektir. Hatta vergi dairesi kimlik bilgilerini gizli tutarak yani mahremiyete dikkat ederek vergi bilgilerini toplumla paylaşsa bile; üçüncü şahıslar bu verilerle sosyal medyadaki verileri birleştirerek kişi hakkında gizli olması gereken bilgileri elde edebileceklerdir. Bu da büyük veri ve etik konusunun önemli bir parçasıdır.

- **Mahremiyet:** Bu bölümde verilere erişimin kim ve kimler tarafından kontrol edileceği üzerinde durulacaktır (Davis ve Patterson 2). Büyük verinin eğer mahremiyet sözcüğünün anlamını değiştirdiği düşünülmekteyse, o zaman mahremiyet kavramı ile ilgili nelerin değerli olduğu konusundaki tartışmalarının anlaşılması gerekir. Büyük verinin getirmiş olduğu değişim ile birlikte kişisel mahremiyeti korumaya çalışırken mahremiyetle ile ilgili nelerin değerli olup olmadığı da

sorgulanmaya başlıyor. Birçok insan teknolojinin gelişmesi ile birlikte dünyanın bizi algılanmasındaki kontrolü nasıl ele geçirdiklerini düşünürken (sosyal medyadaki profiller ve paylaşımlar) aslında asıl soru büyük veri ile birlikte dünyanın bizi algılaması üzerindeki kontrolümüzü nasıl kaybettiğimizdir. Örneğin bir birey bilinçli olarak özel hayatı ile ilgili bir bilgiyi saklı tutmak isterken, başka insanların bunu kamuya açıklamaya ne kadar hakkı vardır? Sizin görünmek istemediğiniz bir anda arkadaşınız Facebookta sizi etiketlemesi veya izniniz olmadan resminizi yayınlaması etik midir? Burada dolayısıyla iki sorun bulunmaktadır. Öncelikle mahremiyet çevrimiçi ve çevrimdışı (gerçek hayatta) ortamlarda aynı anlamımı ifade etmektedir? İkincisi bireylerin kendileri hakkında olan bilgileri tutmaya hakları olmalı mıdır ve hangi aşamaya kadar olmalıdır? Genellikle bu tartışmalar çevrimdışı davranışlar ve çevrimiçi beklentiler arasında gidip gelmektedir. Başkalarına bizim gizli kalmasını tercih ettiğimiz bilgileri açıklamalarının sebeplerini sorabileceğimiz gibi kendimize de şu soruyu sorabiliriz. “Neden çevrimdışı dünyada sahip olduğumuz, neyi ne kadar paylaşacağımızın kontrol ve seçme kabiliyetini çevrimiçi (online) dünyada aynı şekilde bekliyoruz (Davis ve Patterson 17,18). Büyük veri ve mahremiyetin bu tartışmalar ile birlikte detaylı bir şekilde değerlendirilmesi gerekmektedir.

- **İtibar:** Verinin güvenilirliğini nasıl belirleyeceğiz? Büyük verinin toplamış olduğu bilgi miktarı ve bizim onunla olan ilişkimiz çok hızlı şekilde artmaktadır. Bu da nasıl algılanacağımızın ve değerlendirileceğimizin zorluk derecesini arttırmaktadır (Davis ve Patterson 2). Eskiden itibar, özellikle kişiyi tanıyan ve onunla zaman geçiren insanlar için, insanların kişi hakkında bildikleri ve düşünceleri şeklinde tanımlanabilirdi. Eğer insanlar ünlü değillerse, başka sebepler için, genellikle itibarlarını insanlara iyi davranarak korurlardı. Bazı durumlarda, ikinci bir görüş açısı olarak, kendisini tanıyan insanların, kendisi hakkında söyledikleri, kişinin itibarını etkileyebilirdi. İtibar konusundaki en büyük değişikliklerden birisi, büyük veri sayesinde, kişinin nasıl bir insan olduğunu anlatacak ve bunu konu hakkında fikir verecek insan sayısının büyük ölçüde katlanarak artmasıdır. Hatta bu değişim en çok son yıllarda gerçekleşmiştir. Dahası, kişinin çevrim içi itibarını koruması, kişisel kontrolünün git gide dışına çıkmaktadır. Günümüzde birçok şirket için "itibar yönetimi" çok önemli bir hale gelmiştir. Geçmişteki itibar yönetimi anlayışının nasıl yönetileceği ve bunun dijital davranışa nasıl yansıyacağını yeteri kadar bilemediğimiz için bu konu üzerinde daha fazla tartışılma ve araştırma yapılmalıdır (Davis ve Patterson 18).

- **Mülkiyet:** Bu bölümde "Verinin sahibi kim? "Verinin hakları devredilebilir mi?" ve "Veriyi oluşturan ve kullanan insanların yükümlülükleri nelerdir?" gibi sorular üzerinde durulacaktır (Davis ve Patterson 2). Kişilerin özel bilgilerine ne kadar sahip olduğu mahremiyet hakları ve mahremiyet çıkarları arasında farklılığa göre değişmektedir. Çevrimdışı dünyada, kişiler boyları ve kiloları hakkındaki gerçeklere "sahip" midirler? Kendi varoluşları, üzerinde telif hakkı ya da yaratılışla ilgili hak sahibi oldukları, bir yaratıcı hareket oluşturur mu? Aile öyküleri, genetik yapıları ve fiziksel betimlemeleri hakkındaki gerçek, meşrubat tercihleri ya da basketbol sahasında serbest atış yapma becerileri sahip oldukları bir özelliği oluşturmakta mıdır? Bilginin özelliğine göre bilginin sahibi (mülkiyeti) değişmeli midir? Aslında çevrimiçi dünyada etik davranışların prensiplerini ve kurallarını oluşturmak için onlarca girişim ve programlarlar geliştirilmiş durumdadır. Açık veri kullanımları boyut ve karmaşıklık açısından geliştikçe, açık hükümet verileri çoğalmakta ve şirketler kişisel bilgileri kullanarak daha çok gelir elde etmektedirler. Kimin neye sahip olduğu sorusu ve veri izinde hangi noktada olduğu soruları gelecekte daha önemli bir tartışma konusu olacaktır (Davis ve Patterson 19).

Henkoğlu ve Yılmaz (Henkoğlu ve Yılmaz 467,468), yeterli seviyede olmamasına rağmen özellikle bireysel hak ve mahremiyetin korunması konusunda, AB'de diğer bölgelere oranla kişilere daha fazla hak tanındığını belirtmişlerdir. Ayrıca AB'nin bilgi güvenliği politikalarında öncelikli hedefinin; bireysel hakları koruyan yasal düzenlemelerle, yeni bilgi ve iletişim teknolojilerinin kullanımını

yaygınlaştırmak, çevrimiçi ekonomiyi canlandırarak rekabetçi bir ortak pazarın oluşumunu hızlandırmakta olduğu şeklinde görüşlerini belirtmişlerdir.

Hizmet sağlayıcıların kullanıcı ile yapmış oldukları sözleşmelerde; AB (Avrupa Birliği) üyesi ülkelerin, ABD'nin ve bazı ülkelerin (İsviçre ve Avusturalya gibi) kullanıcılarına ait verileri için yapılan anlaşmalara bağlı olarak gerekli güvenlik önlemlerinin alınacağı ifadesi yer almaktadır. Bilgi güvenliği politikası ve bu kapsamda hazırlanmış yasal düzenlemeleri olmayan ülkelerin kullanıcılarına ait verilerin güvenliğinin sağlanması sorumluluğu kullanıcıya bırakılmıştır (Henkoğlu ve Yılmaz 467).

Büyük veri ve etik konusunda çeşitli ülkelerin istihbarat servislerinin pek çok büyük veri teknolojisini kullanarak, kişi ve kurumların kişisel bilgilerini ele geçirmesi ve bu hak ihlalini de ulusal güvenlik çatısı altında yapmaları (Tanış, http://www.hurriyet.com.tr/yazarlar/23207665.asp), gerçekten büyük veri ve etik konusunun ne kadar önemli olduğunu ve bu konuda bütün dünyada daha sağlıklı politikalar ve yasalar ortaya konulması gerektiğinin göstergesidir.

Şubat 2012'de Beyaz Saray kişisel bilgilerin çevrimiçi nasıl kullanılacağı ve kişisel gizliliği korumayı güçlendirmek amacıyla tüketici haklarıyla ilgili bir plan ortaya koymuştur (The White House, Office of the Press Secretary).

III . BÖLÜM

KÜTÜPHANE İŞLEM VE HİZMETLERİNDE BÜYÜK VERİ

Tarihteki ilk örneklerinden günümüzdeki hallerine kadar olan süreçte kütüphaneler, her zaman bilginin toplandığı merkezler olmuşlardır. Günümüzde elektronik yayınların da etkisi ile yalnızca basılı materyallerin merkezi değil, aynı zamanda elektronik kaynaklarında merkezi haline gelmişlerdir. Elektronik kaynaklar ile kütüphanecilik mesleğinde de önemli bir dönüşüm yaşanmıştır. Bu dönüşüm ile birlikte kütüphanecilerin programlama, veritabanı yönetimi, istatistik, matematik vb. bilimlerde de uzmanlaşması gerekmiş ve bu uzmanlıkları bilgi merkezi yönetimi sürecinde kendileri için son derece önemli bir konu olmuştur.

Kütüphanelerde eski zamanlarda yalnızca kütüphane materyalleri ve bazı idari bilgiler tutulurken, günümüzde kullanıcılarının da bir takım bilgileri yer almakta ve kütüphaneler de bu bilgileri kendi iş süreçlerinde kullanmaktadırlar.

Günümüzde ise büyük veri ile birlikte kütüphanecilik mesleği başka bir yapıya dönüşmeye başlamıştır. Kütüphaneler yalnızca bilginin toplandığı bilgi merkezleri değil, aynı zamanda her türlü verinin toplandığı, erişiminin sağlandığı ve bu verilerin enformasyona ve daha sonra belki kendi süreçlerinde belki de yapay zekalara sahip makineler ile birlikte bilgiye dönüşeceği merkezler haline gelecektir. Tabi bu bağlamda mesleki yeterlikler de değişecek ve bilgi yöneticilerinin ileride de konularda da değinilecek pek çok yeni özellikler kazanması gerekecektir.

Üniversite kütüphaneleri yapıları gereği bu konusunda öncü olabilir. Üniversite kütüphanelerinin kullanıcıları genellikle üniversite de akademik personeller, öğrenciler ve araştırma geliştirme desteği duyan bireyler ve kurumlardır. Eğer bir kütüphanenin üniversitenin tüm verisini ve bilgisini yönetebileceğini varsayarsak; bütün kullanıcıları hakkında şu anda kinden çok daha fazla bilgiye sahip olabilecektir. Bu bilgiyi kendi otomasyon sistemi ile öğrenci bilgi sistemi, personel bilgi sistemi, maaş bilgi sistemi, elektronik belge yönetim sistemi vb. gibi kaynaklardan çok rahatlıkla derleyebilecektir. Bunun dışında bilgi işlem sistemleri vasıtası ile kullanıcılarının ne tür arama yaptıklarını, kendi kütüphaneleri dışında başka ne gibi yerlerde arama yaptıklarını keşfedebilecek ve bu bilgiler doğrultusunda kendilerini yeniden gözden geçirerek bu bilgileri hem kendi süreçlerini iyileştirmek için kullanabilecek hem de bu bilgileri araştırmacılarla paylaşarak, onların araştırmalarına destek olabilecektir. Kendi iç çalışma ve sistemleri tarafından üretilen verileri kolaylıkla toplayabilmenin yanı sıra araştırmacılarının çalışmalarında kullanmak istedikleri ya da isteyebilecekleri veri setlerini bulma, toplama ve hazırlama gibi konularda da yardımcı olabilecektir. Bu yeni dermedeki verileri, dış kaynaklardan (sosyal medya, açık kaynaklı içerik sunan çeşitli web siteleri, çeşitli istatistikler ve çeşitli başka dış veriler vb.) sağlanacak verilerle entegre ederek kendi dermesine ekleyecek; bu verileri her türlü iş sürecinde kullanarak ve/veya kullanıcısına bu verileri ulaştırarak önemli faydalar elde edeceği açıktır. Ancak yaşanacak dönüşüm sırasında toplanacak verinin tamamı yapılandırılmış bir şekilde otomasyon sistemlerine entegre edilemeyebilecek ya da yapılandırılamamış verinin süreçlere dahil edilmesi ile birlikte hacim, hız ve çeşitlilik konusunda önemli bir büyüklük ile karşı karşıya kalınacaktır. Bu durum ise bilgi yöneticilerinin önemli bir büyük veri problemi ile karşılaşmalarını

ortaya çıkaracaktır. Bu problem bilgi yöneticilerinin ancak mesleki yeterlilik ve bilgi birikimlerini bu yeni dönüşüme adapte etmeleri ile çözülebilecektir.

Günümüzde Büyük Veri ile bambaşka bir dönüşüm yaşanmaktadır ve bu dönüşümün pek çok bilimde olduğu gibi kütüphanecilikte de çok önemli etkileri olmaktadır. Bu dönüşüm mesleğin yönetimsel ve yöntemsel biçiminde bir takım değişiklikleri zorunlu kılacağı gibi kütüphanecinin ya da bir başka değişle bilgi yöneticisinin de tanımında ve içeriğinde de bir takım değişiklikler yaratmaktadır.

Bu bölümde kütüphane işlem ve hizmetlerinde büyük veriden nasıl faydalanılabileceği ile ilgili örnekler verilecek, sonraki alt başlıkta ise dünyada büyük veriyi farklı işlem ve hizmetlerinde kullanan kütüphanelerden örnekler verilecektir.

Kütüphaneciler için Büyük Veri üzerinde çalışılması gereken önemli bir konudur. Kütüphaneler açısından büyük veri iki başlık altında incelenebilir:

i. Kütüphane işlem ve hizmetlerinin geliştirilmesi için kullanılacak büyük veri
ii. Bilgi kaynağı olarak kütüphane dermesine katılacak büyük veri

Anlaşılabileceği üzere birinci grupta yer alan büyük veri kütüphane işlem ve hizmetlerini geliştirmek için toplanan büyük veriyi ifade etmektedir. Bu veriler toplandıktan sonra bilgi merkezi kendi amaç ve vizyonu doğrultusunda seçtiği, daha önceki bölümde belirtilen veri madenciliği, metin madenciliği gibi teknikleri kullanarak kendi uygulamalarına yönelik karar önerileri üretebilecektir.

İkinci grup ise büyük veri ile ilgili araştırma ve geliştirme yaparak kullanıcılar için büyük verinin toplanması ve kategorize edilerek değişik veri setlerinin hazırlanarak hizmete sunulması anlamındadır.

III.1 Kütüphane İşlem ve Hizmetlerinde Büyük Verinin Kullanımı

Bilgi merkezlerinin pek çoğunda veri genellikle ödünç verme, derme geliştirme ve danışma gibi çeşitli işlevlerin yer aldığı alanlarda üretilmektedir (Arslantekin 376).

Ödünç Vermede Büyük Veri:

Ödünç verme ve danışma, bilgi merkezi kullanıcısı ile bilgi merkezi arasında oluşan hareketlere dayalı hizmetlerdir. Ödünç verme işlemlerinde, ödünç verme, süre uzatma, kütüphaneler arası ödünç verme istekleri ya da bilgi kaynağı dağıtımındaki hareketler vb. işlemler ölçülebilir (Arslantekin 376). Bu hareketlerin ölçülmesi ve hareketler sonucu kendiliğinden oluşan veri, kütüphanelere kullanıcısı hakkında oldukça önemli veriler sağlayabilecektir.

Danışma Hizmetlerinde Büyük Veri:

Danışma hizmetlerinde, örneğin verilen hizmetlerin kullanıcı gereksinimleri için yararlı olup olmadığının karşılaştırılması ile veri toplanabilir. Bu şekilde verinin gereksinimlere doğru yanıt verip vermediği; kullanılan yöntem ve kaynakların yeterli olup olmadığı şeklinde veriler toplanabilmektedir (Arslantekin 376).

Derme Geliştirmede Büyük Veri:

Derme geliştirmede ise veri kullanıcı gereksinimlerine yönelik kaynakların dermede bulunması, bunların geliştirilmesi ve devamının sağlanması sırasında bir araya getirilir. Derme geliştirme verileri genellikle ödünç verme ve danışma işlemlerinin sonucunda oluşturulan verilerin kümelenmesi şeklinde oluşmaktadır (Arslantekin 376).

Veri Dışatımının Kütüphanelerde Faydaya Dönüştürülmesi:

Kütüphaneler yalnızca basılı kaynaklarla değil elektronik kaynaklar ve web sayfaları aracılığı da hizmet vermektedirler. Bu durumda kullanıcıların bu kaynakları kullanımının hemen her noktasında kütüphane için veri üretmeleri söz konusudur.

Kütüphanelerin Sosyal Medya ile İlişkilerinde Büyük Veri:

Kütüphanelerin kullanıcıları ve kendisi de sosyal medyada sürekli veri üretmektedirler.

Kütüphane kapsamında üretilen burada saydıklarımız ve sayamadığımız pek çok veri, kütüphane işlem ve hizmetlerini iyileştirmek için kütüphane yönetiminde kullanılabilir. Büyük veri ile bugüne kadar yapılmayan kapsamlı analizler yapılabilir ve bu analizler kütüphanenin karar alma süreçlerinde daha çok katma değer yaratarak kullanılabilir.

Bilgi merkezlerinin sahip oldukları büyük verinin gerçek değeri ve opsiyon değeri bakımından değerlendirilmesi ise kütüphanelerin büyük veri sonrası karşılaşacakları önemli bir durumdur. Kütüphaneler dermelerinden ve kullanıcılarında

elde ettikleri verileri, gelecek yıllarda etik sınırlar içerisinde faydaya da dönüştürebileceklerdir.

Stanton (Stanton), eskiden beri Kütüphanecilerin enformasyon yönetimi ve organizasyonunda diğer bilimlere göre çok önde olduklarını, bu konun Veri Biliminin özü olduğunu ve bu yüzden büyük veri konusunda kütüphanecilerin aşağıdaki disiplinler ile ilgili eğitim alarak onlarla ortak çalışma yapılabilecek duruma gelmeleri gerektiğini, "Veri Bilimi: Yeni kütüphaneci için içinde ne var?" (Data Science: What's in it for the New Librarian?) başlıklı yazısında belirtmiştir.

Bu disiplinler ve kazanmaları gereken yeterlilikler şunlardır (Stanton):

i. Bilgisayar Programcılığı: Bilgi çıkarımı konusu ile ilgilenen her kütüphaneci çeşitli yazılım araçlarının verileri nasıl dönüştürdüğüne aşina olmalıdır.
ii. Veritabanı Yönetimi: Her kütüphanesi bilgi erişim araçlarının temellerini, veri tabanları bağlamında anlamalıdır.
iii. İstatistik: Her kütüphaneci, sayısal verilerin açıklayıcı özetlerinin ve temel testlerinin nasıl kullanılıp kullanılamayacağı konusunu çok iyi anlamalıdır.
iv. Grafik Tasarımı: Her kütüphaneci etkili veri ekranlarının özelliklerinin önemini anlamalıdır, vb.

Kütüphaneler Büyük Veri konusunda gelişme göstermek istiyorlarsa öncelikle aşağıdaki vb. soruları cevaplamalarında fayda vardır:

- Kütüphanelerde hangi veriler, ne gibi şekillerde, nerelerde üretilmektedir?

- Bu veriler, başka hangi hizmetlerde kullanılabilir?
- Dışarıdan hangi veriler alınırsa, hangi hizmetlerinde iyileştirmelere neden olabilir?
- Hangi veriler dış kaynaklardan dermeye eklenmelidir (Kütüphanenin teknolojik kapasitesi ve kaynakları bağlamında değerlendirilmelidir)?
- Dış kaynaklardan edinilen veriler, erişime nasıl sunulacaktır?
- Bu veriler hizmet verilen araştırmacılara nasıl bir katkıda bulunacaktır? vb.

III.2 Kütüphanelerde Büyük Verinin Derme İçinde Yer Alması

Kütüphane kullanıcı gruplarının araştırmaları için kullanacağı verilerin toplanması, bunların neler olabileceği, nerelerden sağlanabileceği ve nelere dikkat edilmesi gerektiği bu bölümde incelenecektir.

III.2.1 RDA, Linked Data, VIAF ile Büyük Veri İlişkisi

Teets ve Goldner (Teets ve Goldner 430), kütüphane verilerinin gerçekten büyük veri olup olmadığı sorusu üzerine yaptıkları yorumda; kütüphane dermesindeki tüm basılı ve elektronik yayınların üst verilerinin bugünkü anlamıyla büyük veri olarak kabul edilemeyeceğine vurgu yapmaktadırlar. Hatta bu dermenin tam metinlerinin bile veri yönetimi mantığıyla büyük veri sayılamayacağını belirtmektedirler ((Hessman)'dan alıntı yapan (Teets ve Goldner 430). Bununla birlikte, eğer bu dermeye bütün insan bilgisinin bir eğitim kümesi olarak bakılacak olursa, büyük dermelere ait ilişki savları oluşturmak için belli veri yolları takip edilebilir. Tek bir

işten, yani tek bir dermeden; ortak yazarlar, alıntılar, konum belirteçleri, tarihler, isim verilmiş varlıklar, konu başlıkları, yazarlar, geçmiş ödünç verme bilgilerini kapsayan, pek çok ilişki çıkarılabilir. Bu ilişkilerle ise insanlar, işler, patentler, organizasyonlar birbirine bağlanabilir. Teets ve Goldner (Teets ve Goldner 431), bu ilişkilendirilmiş yapının ölçeklendirilerek yeni bir sav halinde yaratılması, işleme konulması, kullanılabilir hale getirilmesini ise kütüphanelerde büyük veri şeklinde yorumlamaktadır.

Gelişen teknoloji ile birlikte, kütüphane materyalleri de biçim değiştirmiş ve yeni formatlarda üretilmeye başlanmıştır. Teknolojik gelişmeler kataloglama kurallarının ihtiyaçlar doğrultusunda geliştirilmesini zorunlu kılmıştır. Özellikle internet üzerindeki kaynakların kontrolsüz bir şekilde artması ile de yeni çözümlere ihtiyaç duyulmuştur. Anglo Amerikan Kataloglama Kuralları 2 ile yeni gelişen ve çeşitlenen bilgi kaynaklarının kataloglanması zorlaşmış ve sonrasında ise RDA[3] (Resource Description and Access) oluşturulmuştur (Bayter, Bilgi Kaynaklarının Tanımlanması ve Erişimi: AACR2'den RDA'ya geçiş 129,130).

RDA, çevrimiçi ağlardaki her türlü enformasyon tanımlanabilecek ve erişilebilecek bir standartır. RDA ile birlikte enformasyona erişim ve elektronik medya ve web uygulamalarının arasındaki sınırların ortadan kalkması sağlanmıştır (Atılgan, Özel ve Çakmak 664).

Özellikle RDA ile birlikte bilgi kaynakları arasında mantıksal ve hiyerarşik ilişkilerin kurulmasında ve kütüphane verilerinin farklı kaynaklardaki verilerle bütünleştirilmesinde (Özel) büyük verinin kullanılması da kütüphanelere önemli

[3] RDA: Resource Description and Access (Kaynak Tanımlama ve Erişim)

faydalar sağlayabilecektir. Çünkü ilişkisel veri tabanları ile geleneksel çözümler ile bu ilişkilerin kurulması ve çok çeşitli kaynaklardan alınan verilerin aynı sistemde yönetilmesi çok zordur.

Uluslararası Sanal Otorite Dosyasının (Virtual International Authority File – VIAF) oluşturulması sonucunda otorite dosyaları ve bunlara bağlı veriler arasındaki ilişkiler de önemli bir hale gelmiştir (Gür 42).

Web ortamında sosyal medya, bloglar, vb. her bir yazışma veya metin bir yazarlık çalışması olarak ele alınabilir. Dolayısıyla bütün bunlar dünya kültürel mirası kapsamında da ele alınabilecektir. İnternet ve sosyal ağ kullanımının günümüzdeki hızlı artışı, gelecekte bu tarzda yazarlık çalışmalarının milyarlarca kişiyi kapsayacağı gerçeğini ortaya koymaktadır. Bütün bu yazarların ve birbirlerinden ayırt edici unsurların da yer aldığı bir yazar otorite dosyasının ise geleneksel veri tabanı yönetim sistemleri ile yönetilemeyeceği de ortadadır. Bu otorite dosyalarının oluşturulmasının ve bu otorite dosyalarının birbirlerine bağlanmasının ne kadar zor olacağı önemli bir sorun olarak karşımıza çıkabilecektir.

Bu konu, büyük veri teknolojileriyle düşünülmesi ve üzerine çalışılması gereken önemli bir konudur.

Geleneksel veri tabanı yönetim sistemleri ile tüm dünyadaki bibliyografik kayıtları düzenlemenin ve otorite dizinlerinin oluşturulması gerçekten çok zordur. Anlık hızla çoğalan bu büyüklükte ve çeşitlikteki verinin tam anlamı ile büyük verinin konusunu oluşturduğu söylenebilir ve büyük veri için geliştirilen yöntem ve uygulamalar ile çözüme ulaştırılması gerekmektedir. Geleneksel veri tabanı yönetim

sistemleri üzerinde çalışan kütüphane otomasyon sistemlerinin bu bağlamda çözüm üretemeyecekleri açıktır.

III.2.2 Açık Veri ve Kütüphanecilik

Daha önceki bölümde açık veri kavramı üzerinde tartışmıştık. Bu bölümde, daha önce bahsedilen açık verinin kütüphaneler açısından nasıl değerlendirildiğini, büyük veri ilişkisini de gösteren şekilde bazı örneklere yer vereceğiz.

III.2.2.1 LOC , Twitter ve Büyük Veri

2000 yılında Amerikan Kongre Kütüphanesi LOC, hükümet enformasyonunun ve politik olaylarının dokümanlarını bulunduran web sitelerini derleyerek, büyük hacimli dijital dermenin yönetimi ile ilgili önemli bir tecrübeye kavuşmuştur. 2013 yılı Ocak ayında bu dermenin büyüklüğü 300 terabayttan fazladır ve on binden fazla web sitesini içermektedir. Milli kütüphaneler bu dijital materyalleri kültürel miras kapsamında dermelerine eklemektedirler (Library of Congress 3).

Twitter ise kütüphane açısından yeni bir kaynak türüdür. Twitter dermesi sadece hacim olarak büyük değil, aynı zamanda her gün hızla artmaktadır. Tweetler, retweetler ve tweetlerin içine eklenen resim, link vb. materyallerle çeşitliliği de oldukça fazla sayıya ulaşmaktadır. (Library of Congress 3).

2010 yılında Twitter en çok 140 karakterlik tweetlerden oluşan bütün dijital arşivini LOC'ye bağışlayacağını duyurmuştur (Library of Congress). Bu haber Amerika Birleşik Devletleri'nin ünlü gazetelerinde de yer almış (Lohr, Library of

Congress Will Save Tweets) ve bu kapsamda 2010 yılında Amerikan Kongre Kütüphanesi LOC ile Twitter 2006 yılından 2010 yılına kadar olan bütün gönderilerin LOC dermesine katılması ile ilgili bir anlaşma yapmıştır (Library of Congress 1). 133,2 Terabayt hacminde 170 milyar tweet bu şekilde 2012 yılının Aralık ayında kütüphane dermesine eklenmiştir (Library of Congress 2). LOC, 2011 yılının Şubat ayında günlük 140 milyon tweet i dermesine ekliyorken, 2012 Kasım ayında bu rakam günlük 500 milyon tweete yükselmiştir (Library of Congress 3).

LOC'den James H. Billington konu ile ilgili Twitter dijital arşivinin yaşamımızın çağdaş yolu içinde araştırma yapan araştırmacılar için muazzam bir potansiyeli olduğunu belirtmiştir. Stanford Üniversitesi Enerji Kaynakları Mühendisliği Bölümü Profesörü Margot Gerritsen ise Twitter' ın günümüz modern kültürü için, ekonomik, sosyal, politik eğilimler ve müşteri ilişkileri ile davranışlarını içeren en açıklayıcı bilgilendirici kaynakları sunduğundan bahsetmiştir (Library of Congress).

Tweetlerin arşivlenmesi ve muhafaza edilmesi, gelecekteki araştırmacıların günümüzün kültürel normları, diyalogları, olayların ve eğilimlerin bilimsel enformasyonu, yasama süreci, yeni yazarlık çalışmaları, eğitim, vb. amaçlarda gerekli enformasyona erişilmesini sağlayacaktır (Library of Congress 1).

Tweetlerin kültürel miras kapsamında derlenmesi, gelecek nesillerin yaptıkları araştırmalar bağlamında enformasyon ihtiyaçlarını karşılaması bakımından son derece önemlidir.

III.2.2.2 CiteSeer^{x} Dijital Kütüphanesinde Büyük Verinin Entegrasyonu

CiteSeer^{x} akademik dokümanlar için bir dijital kütüphane ve arama motorudur. Bilgisayar, enformasyon ve bunlarla ilgili bilimlere odaklanmaktadır. Önemli özelliklerinden biri dermeye eklenen bütün dokümanlardan otomatik olarak enformasyon çıkarması ve otomatik alıntı indekslemesi – bağlaması (linklemesi) yapmasıdır (Williams, Wu ve Choudhury 69). Otomatik alıntı indekslemesi ve kendi kendine alıntı indekslemesi tarafından yapılan alıntı bağlaması konusunda ilk dijital kütüphanedir (CiteSeerX).

Günde ortalama 50.000–100.000 doküman, crawler adı verilen ve internet üzerinde sayfa sayfa dolaşarak bu sayfalardan bilgi toplayan bir uygulamayla toplanmaktadır. Bunların ortalama %40' ı akademik dokümandır ve dermesine günde ortalama 10.000 doküman ve ayda ortalama 300.000 doküman katabilmektedir (Williams, Wu ve Choudhury 73). Dermesinde toplam 3,5 milyon doküman vardır, 80 milyon alıntı vardır, 3-6 milyon arası yazarı, 800.000 bireysel kullanıcısı vardır ve günde ortalama 2 ile 4 milyon ziyaret almaktadır (Williams, Wu ve Choudhury 69). Bu bağlamda hızla büyümeye devam eden CiteSeer^{x} büyük veriyi en iyi kullanan ve yöneten bilgi merkezlerinden biridir.

III.2.2.3 OCLC ve Büyük Veri

OCLC (Online Computer Library Center)'nin çeşitlilik ve hacim konusunda diğer kütüphanelerle ve kütüphane ile ilgili kuruluşlarla kıyaslandığında büyüklük konusunda önemli bir yeri vardır.

Çeşitlilik açısından incelenecek olursa aşağıdaki başlıklar OCLC çeşitliliğini oluşturmaktadır (Green ve Panzer):

- MARC ortamındaki genel derme kayıtları
- Bir dermedeki MARC kayıtları
- Otorite dosyalarının MARC kayıtları
- Yayıncı kayıtları
- WorldCat bilgi tabanı
- Kurumsal abonelik verileri
- Kurum – özel sağlama, ödünç verme ve kütüphaneler arası ödünç verme verileri vb.

OCLC, kullanıcılarına; gazete ve dergiler, veri tabanları, e-kitaplar, özel dermeler, açık erişim içerikleri, yayıncı kayıtları vb. verileri ve çeşitli formatlarda 1,304 milyardan fazla öğenin arama kabiliyetini sunmaktadır (Green ve Panzer).

Ancak bilinmektedir ki MARC bir veri yapısıdır ve bu veri yapısına kaydedilmiş veriler yapılandırılmış veri olma özelliğini taşımaktadır. Bu nedenle bu verinin büyük veri olarak nitelendirilip nitelendirilemeyeceği bize göre tartışma konusudur.

OCLC'nin verileri hacim olarak incelendiğinde ise, 300 milyonu aşkın bibliyografik verisi, 2 milyarı aşkın derme verisi ve otorite verisi olarak da 26 milyon Kongre Kütüphanesi konu başlığı ve 24.2 milyon küme ve 21 milyon kayıtlar arası linkten oluşan VIAF'yla, hacim olarak çok büyük veriden oluşmaktadır (OCLC).

OCLC, bu derece büyük olması yüzünden yeni büyük veri uygulamalarına ve teknolojilerine ihtiyaç duymuştur. Tıpkı Facebook, Adobe, Salesforce.com, vb. pek

çok küresel enformasyon sağlayıcılarında olduğu gibi, 6 Haziran 2013 tarihinde, HBase ve Hadoop'a geçiş, dünyanın en büyük bibliyografik veri tabanına da performans artışı ve yeni teknik olanaklar sağlamıştır. Hadoop'a yapılan dönüşüm, kayıt eşleştirilmesi-birleştirilmesine, işbirliği yapılan kurumların verilerinin yeni gösterimlerine ve kullanımlarına olanak sağlanması konusunda yürütülen kalite yükseltme çabalarına yardım etmektedir. HBase ve Hadoop, OCLC'nin kütüphanelere ve kullanıcılarına daha tutarlı, daha güvenilir, daha hızlı hizmet sunmasını sağlamaktadır. Bu programların kullanımı ile başlıca veri işlemlerinin performanslarında büyük artışlar görülmüş; böylece çalışma süreleri günlerden saatlere düşürülmüştür. Bu dönüşüm, işbirliği yapılan bütün kütüphane verilerinin değerini de arttıracak ve detaylı analitik ve zenginleştirilmiş ilişkiler gibi yeni alanların keşfedilmesine olanak sağlayacaktır (OCLC).

III.2.3 Araştırma Verilerinin Kütüphane Dermesinde Yer Alması

Araştırma verilerini sağlayan unsurlar için aşağıdakiler örnek olarak verilebilir (The European Organisation for International Research Information 3):

- Araştırmacılar
- Organizasyonlar (Araştırma yürüten, fonlama yapan)
- Fon sağlama programları, çağrılar
- Projeler (önerilen, yürütülen, tamamlanmış)
- Yayınlar, patentler, veri, ürünler
- Tesisler, ekipmanlar, hizmetler
- Adresler, coğrafi bağlar, diller vb.

- Ve bunların birbirleri ile ilişkileri

Araştırma verilerine ihtiyaç duyanların da bir bölümü aşağıda verilmiştir (The European Organisation for International Research Information 4):

- Araştırmacılar
- Karar vericiler
- Fonlama Organizasyonları
- Araştırma Organizasyonları
- Proje Müdürleri
- Kütüphaneler
- Yayıncılar
- Eğitimciler
- Kamu
- Medya
- İşletmeler
- Brokerlar / Aracılar vb.

Bu konu ele alınırken öncelikle kavramsal olarak birkaç sorun ortaya çıkmaktadır:

i. Araştırma verileri gerçekten açık olmalı mıdır? : Bu kapsamda önemli olan soru, okuyucular bilimsel eseri kullanarak kendi bilimsel gelişimlerine katkıda mı bulunmaktadırlar, yoksa bu bilimsel yayınların verilerini de kullanmalı mıdırlar? Çok büyük sponsorlarla, büyük emeklerle veri toplanan araştırmaların verileri paylaşıldığında, diğer

bilim adamları hiç çaba, emek, sermaye vb. harcamadan bu verileri kendi bilimsel eserinde kullanabilecektir. Bu şekilde bir açık veri politikası ise yeni araştırmalarda yeni sponsorların bulunması, bilim adamlarının yeni veriler üretmek için kaynaklarını harcamaması ve bu gibi konuları beraberinde getirebilecektir. Bu problemlerde açık veri ve telif hakları sorununu ortaya çıkaracaktır.

ii. Telif hakları açısından konunun değerlendirilmesi: Araştırma verilerinin kütüphane dermesine eklenip eklenmeyeceği konusu, araştırmacının iznine tabi olmalı mıdır, yoksa diğer görüş olan kamu kaynakları ile toplanan verinin kamunun malı olması gibi bir yaklaşım telif hakları konusunda yapılacak yeni düzenlemeler ışığında incelenmeli midir?

Kavramsal olarak ortaya çıkan bu sorunlar çözüldüğünde, bu sefer de bu verilerin gerçek bir büyük veri problemi olarak ortaya çıkması konusu önemlidir. Çünkü bir arkeolog için yaptığı araştırmanın verileri ile bir tıp doktorun yaptığı araştırmanın verileri hangi formatlarda kütüphane dermesine eklenmelidir? Bu kadar çok çeşitte veriye gelecekte teknolojinin değişmesi ile birlikte nasıl erişim sağlanacaktır? Araştırma verileri kütüphane dermesinde eş zamanlı olarak bulunabilecek midir? Kütüphaneler veri ile ilgili projeler parçalara ayrıldığında, ana projenin veri merkezi olabilecekler midir? gibi sorunlar büyük veri kapsamında çözülmesi gereken sorunlardır.

Veri aynı amaç için birçok defa kullanılabilir ve birden fazla amaç için kontrol edilebilir. Nasıl ki Amazon firması müşterilerine tavsiyede bulunurken; geçmiş işlemlerden gelen veriyi kullanarak, bu verileri pek çok müşteri için de tekrar tekrar kullanabiliyorsa veya nasıl ki Walmart eski satış veri tabanlarını araştırarak,

kasırgalarla pop-tarts satışları arasındaki ilişkiyi bulabiliyorsa (Schönberger ve Cukier, Büyük Veri 108, 109), bilim adamları da yaptıkları araştırmalarda elde ettikleri verileri pek çok farklı araştırmada da kullanabilirler. Dolayısı ile araştırma verilerini paylaşmaları diğer araştırmacıların hiçbir çaba ya da maddi-manevi kaynak harcamadan bu verileri elde ederek, verileri toplayan bilim adamının olumsuz etkilenmesine neden olur mu? Yoksa araştırmasında bu verileri topladığı bilim adamına ya da araştırmaya atıf yapmak yeterli olacak mıdır? Şüphesiz burada öncesinde yapılan araştırmanın kişilere maddi kazanç sağlayıp sağlamadığı da önemli bir kriter olarak karşımıza çıkmaktadır.

IV . BÖLÜM

ANKARA ÜNİVERSİTESİ KÜTÜPHANELERİNDE ÇALIŞAN BİLGİ YÖNETİCİLERİNİN BÜYÜK VERİ FARKINDALIĞI

Çalışmamızda bu bölüme kadar günümüzde oldukça önem kazanan ve hatta bir fenomen haline dönüşen büyük veri ve kütüphane uygulamalarında kullanımına kısaca değinmeye çalıştık. Büyük verinin kütüphaneler için yeni bir kavram olduğu, literatürde kütüphane uygulamalarına ilişkin çok fazla araştırmayla karşılaşılmadığı ortadadır. Bu durum meslek uzmanlarının (bilgi yöneticiler/kütüphaneciler) kendilerini bu konuda yetiştirmelerinin ötesinde, konunun öneminin farkındalığına varılması açısından da olumsuz bir ortam yaratmaktadır.

Yukarıda bahsedilen farkındalığın ülkemizde var olup olmadığı, Ankara Üniversitesi Kütüphanelerinde çalışan meslek mezunlarına bir anket uygulamasıyla ölçülmeye çalışılmıştır. Üniversite kütüphaneleri ülkemizdeki kütüphanecilik işlem ve hizmetlerinin geliştirilmesi ve yeniliklerin uygulanması bağlamında diğer kütüphane türlerine hep önderlik etmiştir. Bilimsel çalışmalara destek vermesi açısından da büyük veri konusuna eğilmesi gereken kütüphane türlerinin başında gelmelidir. Ankara Üniversitesi, gerek kurumsal yapısı ve geçmişi, gerekse de dağınık yapıda çok sayıda ihtisas kütüphanesini bünyesinde barındırması nedeniyle çalışmamıza iyi bir örnek olacağı düşüncesiyle seçilmiştir.

Bu çalışmada önce büyük veri hakkında fikir sahibi olup olmadıklarını ortaya koyabilmek için anket yapılmış; hemen ardından büyük veri ve kütüphanelerde kullanımına yönelik yaklaşık 15 dakikalık bir bilgilendirici sunum gerçekleştirilmiştir. Sunum sonrasında tekrar bir anket uygulanarak büyük veri konusunda farkındalığın sağlanıp sağlanamadığı ortaya konulmaya çalışılmıştır. Yapılan anketlerin ve sunum neticesinde meslek elemanlarının büyük veriye "kendilerini hazır hissedebilmeleri için beklentilerinin ne olduğunun" belirlenmesi de amaçlanmıştır.

IV.1 Anket Çalışması, Kapsam ve Yöntem

Araştırma kapsamındaki Ankara Üniversitesi Kütüphaneleri, Ankara Üniversitesi Kütüphane ve Dokümantasyon Daire Başkanlığı çatısı altında, fiziki yerleri birbirlerinden farklı olan 13 fakülte, 16 enstitü, yüksekokul, birim kütüphanesinden ve bir öğrenme merkezinden oluşmaktadır (Ankara Üniversitesi Kütüphane ve Dokümantasyon Daire Başkanlığı). Bunlar;

1. Tandoğan Yerleşkesi Öğrenme Merkezi
2. Dil ve Tarih-Coğrafya Fakültesi Kütüphanesi
3. Diş Hekimliği Fakültesi Kütüphanesi
4. Eczacılık Fakültesi Kütüphanesi
5. Eğitim Bilimleri Fakültesi Kütüphanesi
6. Hukuk Fakültesi Kütüphanesi
7. İlahiyat Fakültesi Kütüphanesi
8. İletişim Fakültesi Kütüphanesi
9. Sağlık Bilimleri Fakültesi Kütüphanesi
10. Siyasal Bilgiler Fakültesi Kütüphanesi

11. Spor Bilimleri Fakültesi Kütüphanesi
12. Tıp Fakültesi Kütüphanesi
13. Veteriner Fakültesi Kütüphanesi
14. Ziraat Fakültesi Kütüphanesi
15. Türk İnkılap Tarihi Enstitüsü Kütüphanesi
16. Devlet Konservatuvarı Müdürlüğü Kütüphanesi
17. Beden Eğitimi ve Spor Yüksekokulu Kütüphanesi
18. Başkent Meslek Yüksekokulu Kütüphanesi
19. Beypazarı Meslek Yüksekokulu Kütüphanesi
20. Elmadağ Meslek Yüksekokulu Kütüphaesi
21. Haymana Meslek Yüksekokulu Kütüphanesi
22. Hukuk Fakültesi Adalet Meslek Yüksekokulu Kütüphanesi
23. Kalecik Meslek Yüksekokulu Kütüphanesi
24. Nallıhan Meslek Yüksekokulu Kütüphanesi
25. Sağlık Hizmetleri Meslek Yüksekokulu Kütüphanesi
26. Afrika Çalışmaları Uygulama ve Araştırma Merkezi Kütüphanesi
27. Fikri ve Sınai Hakları Araştırma Merkezi Kütüphanesi
28. Latin Amerika Çalışmaları Araştırma ve Uygulama Merkezi Kütüphanesi
29. Türkçe ve Yabancı Dil Araştırma ve Uygulama Merkezi Kütüphanesi
30. Avrupa Toplulukları Araştırma ve Uygulama Merkezi Kütüphanesi

Araştırma kapsamında Ankara Üniversitesi Kütüphanelerinde çalışan, Bilgi ve Belge Yönetimi\Kütüphanecilik Bölümü mezunu 35 kütüphaneciye, anketlerin ve sunumun tam sayım şeklinde uygulanması düşünülmüş, ilk ankete 28 kütüphaneci

katılmış, sunum ve ikinci ankete ise 22 kütüphaneci katılmıştır. Ankete verilen yanıt oranı ilk anket için %80, sunum ve ikinci anket için ise %63'dür.

Örneklem sayısı yeterli büyüklükte olmadığı için araştırma kapsamında yapılan hipotez testlerinde, parametrik olmayan istatistiksel testler tercih edilmiştir (Gamgam ve Altunkaynak), (Sümbüloğlu, Akyüz ve Sümbüloğlu), (Durmuş, Yurtkoru ve Çinko 183). Parametrik olmayan testler, ilgilenilen değişkenin dağılımıyla ilgili normallik varsayımı sağlanamadığı, gözlem sayısının az olduğu ve değişkenin sınıflı veya sıralı ölçekle toplanmış olduğu durumlarda kullanılan istatistiksel yöntemlerdir (Durmuş, Yurtkoru ve Çinko 183).

Birinci anketin birinci bölümünde, araştırmada kullanılmak üzere kütüphanecilerin eğitim durumları ile büyük veri hakkında genel bilgileri olup olmadığına yönelik bulgular elde edilmeye çalışılmıştır.

Birinci anketin ikinci bölümünde ise büyük veri hakkındaki bilgi düzeyleri ölçülmüş, büyük veri hakkında kavramsal bilgilerinden bağımsız olarak, büyük veri teknolojilerini kullanım durumları ile bu kavramlara yakın olup olmadıkları konusu saptanmaya çalışılmıştır.

Birinci anketin güvenilirliğini sınamak amacıyla hesaplanan Cronbach Alfa değeri 16. Soru çıkarıldıktan sonra 0,783 olarak bulunmuştur. Bu değer anketin oldukça güvenilir olduğunu göstermektedir.

16. Soru, ikinci anketteki 7. soru ile aynıdır ve yapılan sunumda bu konu ile ilgili bilgi verilmemiş, bu soru diğer sorularla birlikte değerlendirildiğinde anketi yanıtlandıranların "Araştırma verilerin yapılandırılmamış olması ve çok çeşitli biçimlerde olması nedeniyle erişimi çok zordur" şeklindeki önermeye iki ankette de

“Kararsızım/Yeterli Bilgim Yok” yanıtını vermesi beklenmiş ve bu beklenti karşılanmıştır. Bu yanıtlarla anketi dolduranların, yanıtlarını samimi ve doğru olarak verdikleri söylenebilir. Aynı ölçüme sahip iki sorunun da değerlendirilmeye alınması doğru bulunmadığı ve gerçekte ankete verilen yanıtlardaki “samimiyeti” ölçmek amacıyla konulmuş bu sorular sınamada ve değerlendirmede dikkate alınmamıştır.

İkinci ankette ise birinci anketle ortak olan altı soru ile “ön-test/son-test kontrol gruplu model” kullanılarak 22 kütüphanecinin,

- büyük veri konusundaki bilgilendirici sunumdan önce ve sunumdan sonra büyük veri hakkındaki bilgi düzeylerindeki farklılık
- bilgi düzeylerinde artış meydana geldi ise bu bilgiyi kütüphane işlem ve hizmetlerinde kullanılması hakkındaki tutumları
- büyük verinin kütüphanecilik bilimde kullanımı ile ilgili tutumları

ölçülmüştür. İkinci anketin son bölümünde ise katılımcılara eğitim sunumunun katkısı ve daha detaylı ve geniş çerçevede eğitim almak isteyip istemedikleri ile ilgili tutumları ölçülmüştür.

İkinci anketin güvenilirliğini sınamak amacıyla hesaplanan Cronbach Alfa değeri ikinci ankette 7. Soru çıkarıldıktan sonra (birinci ankette 16. Sorunun çıkarılması ile aynı nedenle) 0,897 olarak bulunmuştur. Bu değer anketin yüksek derecede güvenilir olduğunu göstermektedir.

Araştırmaya katılan kütüphanecilerinin görev yaptıkları Ankara Üniversitesi Kütüphaneleri aşağıdaki gibidir:

i. Tandoğan Yerleşkesi Öğrenme Merkezi
ii. Dil ve Tarih-Coğrafya Fakültesi Kütüphanesi
iii. Diş Hekimliği Fakültesi Kütüphanesi
iv. Eczacılık Fakültesi Kütüphanesi
v. Eğitim Bilimleri Fakültesi Kütüphanesi
vi. Hukuk Fakültesi Kütüphanesi
vii. İletişim Fakültesi Kütüphanesi
viii. Sağlık Bilimleri Fakültesi Kütüphanesi
ix. Spor Bilimleri Fakültesi Kütüphanesi
x. Siyasal Bilgiler Fakültesi Kütüphanesi
xi. Ziraat Fakültesi Kütüphanesi
xii. Avrupa Toplulukları Araştırma ve Uygulama Merkezi Kütüphanesi

Yukarıdaki bilgilerden de anlaşılacağı gibi daha önce sıralanan 30 kütüphane ve öğrenme merzinin hepsinde kütüphaneci\bilgi uzmanı bulunmamaktadır. Ankara Üniversitesi ölçeğinde evrenimiz, kütüphanecilerden oluşturmaktadır.

IV.1.1 Ankara Üniversitesi Kütüphanecilerinin Büyük Veri Konusunda Genel Bulguları ve Değerlendirmesi

Birinci anketi yanıtlayan 28 kütüphaneci için bu bölümde teknoloji, sosyal medya kullanım düzeylerine yönelik ve büyük veri bilgisi ile ilgili sorulan sorularla ilgili genel değerlendirmeler bu bölümde yer alacaktır.

Ankete katılan kütüphanecilerin “mezuniyet yılları” ile ilgili sorunun yanıtlarına ilişkin tablo aşağıda verilmiştir:

Tablo 2. Kütüphanecilerin Mezuniyetlerinden Günümüze Kadar Geçen Yıllara Göre Sıklığı

Kaç Yıllık Mezun Olduğu	**n**	**%**
1 ile 5 yıl arası	3	10,71
6 ile 10 yıl arası	7	25,00
11 ile 15 yıl arası	3	10,71
16 ile 20 yıl arası	2	7,15
21 ile 25 yıl arası	7	25,00
26 ile 30 yıl arası	6	21,43
Toplam	28	100,00

Araştırmaya katılan kütüphanecilerin mezuniyetlerinden, içinde bulunduğumuz 2015 yılına kadar, 1 ile 5 yıl arasında geçmiş olanlar %10,71’ ini, 6 ile 10 yıl arasında geçmiş olanlar %25,00’ ini, 11 ile 15 yıl arasında geçmiş olanlar %10,71’ ini, 16 ile 20 yıl arasında geçmiş olanlar %7,15’ini, 21 ile 25 yıl geçmiş olanlar %25,00’ini, 26 ile 30 yıl arasında geçmiş olanlar ise %21,43’ünü oluşturmaktadır.

Ankete katılan kütüphanecilerin “eğitim seviyeleri” ile ilgili sorunun yanıtlarına ilişkin tablo aşağıda verilmiştir:

Tablo 3. Kütüphanecilerin Eğitim Seviyelerine Göre Sıklığı

Eğitim Durumları	**n**	**%**
Lisans Mezunu	26	92,86
Yüksek Lisans Mezunu	1	3,57
Doktora Mezunu	1	3,57
Toplam	28	100,00

Araştırmaya katılan kütüphanecilerin %92,86'sı (26 kişi) lisans mezunu, %3,57'si (1 kişi) yüksek lisans mezunu, %3,57'si (1 kişi) ise doktora mezunudur.

Ankete katılan kütüphanecilerin "Büyük Veri kavramını daha önce duyup duymadıkları" ile "bu konuda herhangi bir yazı okuyup okumadıkları" konusundaki soruların yanıtlarına ilişkin tablolar aşağıda verilmiştir:

Tablo 4. Kütüphanecilerin Büyük Veri Kavramını Daha Önce Duyup/Duymama Durumlarına Göre Sıklığı

Büyük Veri Kavramını Duydum/Duymadım	**n**	**%**
Duydum	15	53,57
Duymadım	13	46,43
Toplam	28	100,00

Tablo 5. Kütüphanecilerin Büyük Veri Konusunda Herhangi Bir Yazı Okuyup/Okumamalarına Göre Sıklığı

Büyük Veri Konusunda Bir Yazı Okudum/Okumadım	**n**	**%**
Okudum	4	14,29
Okumadım	24	85,71
Toplam	28	100,00

Büyük veri kavramını daha önce duyduklarını söyleyen kütüphanecilerin oranı % 53,57 ve duymadığını söyleyen kütüphanecilerin oranı %46,43 olmasına rağmen kütüphanecilerin yalnızca 14,29'u büyük veri ile ilgili bir kaynak okuduklarını, %85,71'i ise okumadıklarını belirtmişlerdir. Bu iki anket sorusu ile kütüphanecilerin büyük çoğunluğunun büyük veri kavramını daha önce duymuş olmalarına rağmen, yine büyük çoğunluğunun bu konu ile ilgili herhangi bir kaynak okumadıkları, dolayısıyla derinlemesine bir bilgilerinin olmadığı söylenebilir.

Yukarıdaki bulgulardan da anlaşılacağı üzere çalışanların yarısından fazlası duymuş görünmekle birlikte, bu durum özellikle eğitim seviyeleri ile karşılaştırılmak istendiğinde yüksek lisans ve doktora yapmışlar sadece birer kişi olduğu için doğru bir genellemeye varmamızı engellemektedir.

IV.1.2 Kütüphanecilerin, Kendileri ile İlgili Veri Toplayan Araçları Kullanma Düzeylerine İlişkin Değerlendirme ve Bulgular

Bu bölümde Ankara Üniversitesi Kütüphanelerinde çalışan kütüphanecilerin kendileri ile ilgili veri toplayan araçlardan bir kısmını kullanma düzeylerine ilişkin bulgular ve değerlendirmeler bulunmaktadır.

Ankete katılan kütüphanecilerin "sosyal medyayı kullanmaları" ile ilgili sorunun yanıtlarına ilişkin tablo aşağıda verilmiştir:

Tablo 6. Kütüphanecilerin Sosyal Medya Uygulamalarını Kullanım Sıklığı

Sosyal Medya Uygulamalarını Kullanım Sıklıkları	**n**	**%**
Hiç Kullanmayanlar	1	3,57
Bazen Kullananlar	12	42,86
Sıklıkla Kullananlar	15	53,57
Toplam	28	100,00

Birinci ankette sosyal medyayı kullanma sıklıkları ölçülen kütüphanecilerin %3,57'si (yalnızca 1 kütüphaneci) sosyal medyayı hiç kullanmadıklarını, %42,86'sı bazen kullandıklarını, yüzde 53,57'si sıklıkla kullandıklarını belirtmişlerdir. Kütüphanecilerin 96,43' ünün, yani bir kişi hariç neredeyse tamamına yakınının sosyal medyayı kullandıkları söylenebilir.

Kütüphanecilere, Facebook, Twitter, Instagram, Google+, LinkedIn, Four Square, Tumblr, Youtube, Wikipedia, Flickr, Picasa, Blogger, Wordpress, ve diğer sosyal medya uygulamalarından hangilerini kullandıkları sorulmuştur ve onlarda diğer kısmına Whatsapp, Pinterest, Vine' ı eklemişlerdir.

Tablo 7. Kütüphanecilerin Kullandıkları Birbirlerinden Farklı Sosyal Medya Uygulamaları Sıklığı

Sosyal Medya Uygulamaları Sayısı	**n**	**%**
0	1	3,57
1 ile 4 arası	13	46,43
5 ile 9 arası	13	46,43
10 ile 13 arası	1	3,57
Toplam	28	100,00

Kütüphanecilerden bir kişi önceki soruda olduğu gibi hiçbir sosyal medya uygulamasını kullanmadığını söylemiş, 1 ile 4 sosyal medya uygulaması arasında kullananlar 13 kişi yani örneğin % 46,43'ü, 5 ile 9 sosyal medya uygulaması kullananlar 13 kişi yani %46,43'ü, 10 ile 13 sosyal medya uygulaması kullananlar 1 kişi yani %3,57'si olarak bulunmuştur. Bu sonuçlarla ankete katılan Ankara Üniversitesi Kütüphanecilerin sosyal medya uygulamalarını büyük oranda kullandıkları söylenebilir.

Ankete katılan kütüphanecilerin "e-posta kullanıp kullanmadıkları" ile ilgili sorunun yanıtlarına ilişkin tablo aşağıda verilmiştir:

Tablo 8. Kütüphanecilerin e-Posta Kullanım Durumlarına Göre Sıklığı

e-Posta Kullanıyorum/Kullanmıyorum	**n**	**%**
Kullananlar	28	100,00
Kullanmayanlar	0	0
Toplam	28	100,00

Yapılan ankete çalışması ile ankete katılan kütüphanecilerin tamamının e-posta kullandığı saptanmıştır.

Ankete katılan kütüphanecilerin "bulut bilişim hizmetlerini kullanıp kullanmadıkları" ile ilgili sorunun yanıtlarına ilişkin tablo aşağıda verilmiştir:

Tablo 9. Kütüphanecilerin Bulut Bilişim Hizmetlerini Kullanım Sıklığı

Bulut Bilişim Hizmetlerini Kullanıyorum/Kullanmıyorum	**n**	**%**
Kullananlar	7	25,00
Kullanmayanlar	21	75,00
Toplam	28	100,00

Kütüphanecilerin %25'inin bulut bilişim hizmetlerini kullanmakta olduğu, %75'i kullanmadıkları saptanmıştır. Kütüphanecilerin büyük çoğunluğunun bulut bilişim hizmetlerini kullanmadıklarını söylenebilir.

Ankete katılan kütüphanecilerin "kredi kartı vb.ni kullanıp kullanmadıkları" ile ilgili sorunun yanıtlarına ilişkin tablo aşağıda verilmiştir:

Tablo 10. Kütüphanecilerin Kredi Kartı veya Alışveriş Kartı Kullanım Durumlarına Göre Sıklığı

Kredi Kartı vb. Kullanıyorum/Kullanmıyorum	**n**	**%**
Kullananlar	27	96,43
Kullanmayanlar	1	3,57
Toplam	28	100,00

Kütüphanecilerin %96,43'ünün kredi kartı veya sadakat kartlarını kullandıkları, %3,57'sini oluşturan sadece bir kişinin kullanmadıkları saptanmıştır.

Kütüphanecilerin *"Güvenlik kameraları, sensörler, kullandığınız telefonlar, alışveriş kartları, kredi kartları, sosyal medya uygulamaları vb. cihazlar ve uygulamaların sizin hakkınızda bilgi topladığını biliyor musunuz? Daha önce hiç düşündünüz mü?"* sorusunu 28 kişi yanıtlamıştır.

Ankete katılan kütüphanecilerin "veri toplayan cihazlar ve uygulamalar hakkında bilgi toplaması" ile ilgili sorunun yanıtlarına ilişkin tablo aşağıda verilmiştir:

Tablo 11. Kütüphanecilerin Kendileri Hakkında Veri Toplayan Cihazların Farkındalık Durumlarına Göre Sıklığı

Kullandıkları Cihazların ve Uygulamaların, Kendileri Hakkında Veri Toplayabilmeleri Konusunda Farkındalıkları	**n**	**%**
Evet	24	85,71
Kısmen Evet	3	10,72
Hayır	1	3,57
Toplam	28	100,00

Anketi dolduranların %85,71'i yani 24 kişi "Evet", % 10,72'si yani 3 kişi "Kısmen Evet", %3,57'si yani 1 kişi "Hayır" yanıtını vermiştir. Bu sonuçlara göre katılımcıların %96,43'ü güvenlik kameraları, sensörler, kullandığınız telefonlar, alışveriş kartları, kredi kartları, sosyal medya uygulamaları vb. cihazlar ve uygulamalarının kendileri ile ilgili bilgi topladığını düşünmektedir sonucuna varılabilir.

IV.1.3 Kütüphanecilerin, Meslekleri ile İlgili Yeni Teknolojileri Takip Etme Düzeylerine İlişkin Değerlendirme ve Bulgular

Kütüphanecilerin, Büyük Veri ile ilgili meslekleri kapsamındaki teknolojik gelişmeleri takip etmeleri için, öncelikle meslekleri ile ilgili yeni teknolojileri takip etmeleri gerekmektedir.

Bu bölümde Ankara Üniversitesi Kütüphanelerinde çalışan kütüphanecilerin meslekleri ile ilgili yeni teknolojileri takip etme düzeylerine ve meslekleri ile ilgili yeni teknolojileri hangi ortamlarda takip ettiklerine ilişkin bulgular ve değerlendirmeler bulunmaktadır.

Ankete katılan kütüphanecilerin "yeni teknolojileri takip edip etmedikleri" ile ilgili sorunun yanıtlarına ilişkin tablo aşağıda verilmiştir:

Tablo 12. Kütüphanecilerin Meslekleri İle İlgili Yeni Teknolojileri Takip Edip Etmemelerine Göre Sıklığı

Meslekleri ile İlgili Yeni Teknolojileri Takip Etme Sıklıkları	**n**	**%**
Bazen	19	67,86
Sıklıkla	9	32,14
Toplam	28	100,00

Araştırma kapsamında soru sorulan 28 kütüphanecinin % 67,86 sı meslekleri ile ilgili teknolojik gelişmeleri “Bazen” takip ettiklerini, %32,14’ü ise “Sıklıkla” takip ettiklerini belirtmişlerdir. Hiç takip etmeyen kütüphaneci bulunmamaktadır.

Kütüphanecilerin meslekleri ile ilgili yeni teknolojileri takip ettikleri kaynaklar ve araçların durumu incelendiğinde kütüphanecilere birden fazla yanıtı işaretleme olanağı verilmiştir. Bu sorunun yanıtlarını özetleyen tablo aşağıda verilmiştir.

Tablo 13. Kütüphanecilerin Meslekleri ile İlgili Yeni Teknolojileri Takip Ettikleri Kaynaklar ve Araçların Durumu

Kaynaklar, Araçlar	**n**	**%**
Basılı kaynaklar	15	53,57
Web siteleri	24	85,71
Açık erişim dergileri	4	14,29
Veri tabanları	14	50,00
Forumlar	9	32,14
Tartışma listeleri	10	35,71
Sosyal Medya	19	67,86
Diğer	1	3,57

Meslekleri ile ilgili gelişmeleri takip etmek için ankete katılan 28 kütüphaneciden %53,57’sinin (15 kişi) basılı kaynaklardan, %85,71’inin (24 kişi) web sitelerinden, %14,29’unun (4 kişi) açık erişim dergilerinden, %50,00’sinin (14 kişi) veri tabanlarından, %32,14’ünün (9 kişi) forumlardan, %35,71’inin (10 kişi) tartışma listelerinden, % 68,86’sının (19 kişi) sosyal medyadan yararlandıkları görülmüştür. Diğer yanıtını işaretleyen bir kişi ise hangi kaynağı kullandığını belirtmemiştir.

Araştırma kapsamındaki Kütüphanecilerin büyük çoğunluğunun meslekleri ile ilgili gelişmeleri web sitelerinden takip ettikleri, bu konuda açık erişim dergilerini ise en az sayıda kullanıldıkları belirlenmiştir.

Ankete katılan kütüphanecilerin "yeni teknolojileri kaç farklı ortamda takip ettikleri" ile ilgili yanıtlar aşağıdaki tabloda verilmiştir:

Tablo 14. Kütüphanecilerin Meslekleri ile İlgili Yeni Teknolojileri Birbirlerinden Farklı Ortamlarda Takip Etme Sıklığı

Yeni Teknolojiler Kaç Farklı Ortamda Takip Edilmektedir	**n**	**%**
1	2	7,14
2	3	10,71
3	11	39,29
4	7	25,00
5	4	14,29
7	1	3,57
Toplam	28	100,00

28 kütüphaneciden meslekleri ile ilgili gelişmeleri bir kaynaktan takip eden 2 kişinin oranı %7,14, iki farklı kaynaktan takip eden 3 kişinin oranı %10,71, üç farklı kaynaktan takip eden 11 kişinin oranı %39,29, dört farklı kaynaktan takip eden 7 kişinin oranı %25,00, beş farklı kaynaktan takip eden 4 kişinin oranı %14,29, yedi farklı kaynaktan takip eden 1 kişinin oranı %3,57 olarak saptanmıştır.

Araştırmaya katılan kütüphanecilerden üç veya daha fazla ortamda meslekleri ile ilgili gelişmeleri takip eden 23 kişinin oranı %82,14'dür ve hiç takip etmeyen kütüphanecinin bulunmaması araştırma kapsamında olumlu olarak değerlendirilmektedir.

IV.1.4 Kütüphanecilerin Büyük Veri Konusunda Temel Seviyede Bilgilerinin Ölçülmesine İlişkin Değerlendirme ve Bulgular

Birinci anket uygulandıktan sonra büyük veri ile ilgili bilgilendirici bir sunum yapılmış, sunum sonunda ikinci anket uygulanmış ve büyük veri konusundaki bilgi düzeylerinde artış olup olmadığı saptanmak istenmiştir. Bu kapsamda ilk ankette ve ikinci ankette aynı olan sorular aşağıdaki tabloda görülmektedir.

Tablo 15. İşlem Öncesi ve Sonrası Büyük Veri Bilgi Düzeyinin Ölçülmesinde Kullanılacak Sorular.

1. Anket Soru No:	2. Anket Soru No:	Sorular	Kesinlikle katılmıyorum	Katılmıyorum	Kararsızım / Yeterli bilgim yok	Katılıyorum	Kesinlikle katılıyorum
S1	S1	Büyük Veri, verinin hacminin büyüklüğünü ifade eden bir kavramdır.	1	2	3	4	5
S2	S2	Büyük Veri, sadece verinin hacminin büyüklüğünü değil, aynı zamanda hızının ve çeşitliliğinin de büyüklüğünü ifade etmektedir.	1	2	3	4	5
S3	S3	Büyük Veri popüler bir kavram olduğu için gündeme gelmiştir. Popülerlik geçici olmasına rağmen, bu kavram geçerliliğini koruyacaktır.	1	2	3	4	5
S4	S4	Büyük Veri, yeni bir devrin başlangıcıdır.	1	2	3	4	5
S13	S5	Bir teknoloji şirketim olsa, ben de güvenlik kameraları, sensörler, telefonlar, alışveriş kartları, kredi kartları, sosyal medya uygulamaları vb. cihazlardan ve uygulamalardan edinilen bilgilerle, özel hayatın gizliliğine dikkat edilerek yeni teknolojiler geliştirmek isterdim.	1	2	3	4	5
S14	S6	Kullandığım Cep Telefonu uygulamaları ile takip edildiğimi ya da izlendiğimi düşünüyorum.	1	2	3	4	5

Birinci anketteki ilgili 6 sorunun güvenilirliğini sınamak amacı hesaplanan Cronbach Alfa değeri 0,660 olarak bulunmuştur. İkinci anketteki ilgili 6 sorunun güvenilirliğini sınamak amacı hesaplanan Cronbach Alfa değeri ise 0,762 olarak bulunmuştur. Bu değerler anketteki ilgili soruların oldukça güvenilir olduğunu göstermektedir.

Bu bölümde kütüphanecilerin eğitim sonrasında, eğitim öncesine göre bilgi düzeylerinde artış olup olmadığı ile ilgili aşağıdaki hipotez oluşturulmuş ve iki bağımlı örnek için uygulanan testlerden en uygun olanı Wilcoxon İşaretli Sıra Sayıları Testi %95 güven düzeyinde analiz edilmiştir.

H_0: İşlem öncesi ve sonrası bilgi düzeyi puanlarının farklar yığınının medyanı sıfırdır.

H_1: İşlem öncesi ve sonrası bilgi düzeyi puanlarının farklar yığınının medyanı sıfırdan büyüktür.

Tablo 16. İşlem Öncesi ve Sonrası Bilgi Düzeyi Puanlarının Farklarına İlişkin İki Bağımlı Örnek için Wilcoxon İşaretli Sıra Sayıları Testi Sonucu.

Yokluk Hipotezi	**Test**	**N**	**Test İstatistiği**	**Standart Hata**	**Standartlaştırılmış Test İstatistiği**	**P değeri (2 yanlı test)**	**Karar**
İşlem öncesi ve sonrası bilgi düzeyi puanlarının farklar yığınının medyanı sıfırdır.	İki bağımlı örnek için Wilcoxon İşaretli Sıra Sayıları Testi	22	231	28,703	4,024	,000	Yokluk Hipotezi red edilmiştir.
Anlamlılık düzeyi 0,05'dir.							

Ölçülen 0,000 P değeri (Anlamlılık), 0,05'den küçük olduğundan yokluk hipotezi red edilmiştir. % 95 Güven düzeyinde işlem öncesi ve sonrası bilgi düzeyi puanlarının farklar yığınının medyanının sıfırdan büyük olduğu söylenebilir.

Diğer bir anlatımla büyük veri konusunda ki bilgilendirici sunuma katılanların, sunumdan sonra Büyük Veri konusunda temel bilgi düzeylerinde istatistiksel olarak anlamlı bir artış olduğu söylenebilir.

Bu sonuca göre büyük veri konusunda yeterli temel bilgiye sahip olmayan kütüphanecilere, temel bilgileri içeren bir sunum yapıldıktan sonra, bilgi düzeylerindeki artış olumlu olarak yorumlanabilir. Yine bu sonuçlara göre kütüphanecilerin mesleklerinin doğası gereği, büyük veri konusuna ne kadar çabuk adapte oldukları görülebilir ve büyük verinin bu yüzden kütüphanecilik mesleği içinde de değerlendirilmesi gerektiği söylenebilir.

IV.1.5 Kütüphanecilerin Veri Toplayan Nesneler Hakkındaki Farkındalıklarının Ölçülmesine İlişkin Değerlendirme ve Bulgular

Kütüphanecilerin veri toplayan nesneler hakkındaki farkındalıklarının ölçülmesine ilişkin değerlendirmede bulunmak için birinci anketteki 5., 6., 7., 8., 9., 10., 11., 12., 15., 17., 18. sorulardan yararlanılmıştır.

Bu kapsamda kütüphanecilere birinci ankette sorulan sorulara verilen yanıtların sıklıkları aşağıdaki tabloda görülmektedir.

Tablo 17. Kütüphanecilerin Veri Toplayan Nesneler Hakkındaki Farkındalıklarının Ölçülmesinde Yanıtların Sıklığı

1. Anket Soru No:	Sorular	Kesinlikle katılmıyorum		Katılmıyorum		Kararsızım / Yeterli bilgim yok		Katılıyorum		Kesinlikle katılıyorum		Toplam n
		N	%	N	%	N	%	N	%	N	%	
S 5	Gün içinde güvenlik kameraları, mobese kameraları vb. cihazlar benimle ilgili veriler toplamaktadır.	1	4	3	11	2	7	12	43	10	36	28
S 6	Telefonum ve gün içinde karşılaştığım çeşitli sensörler benimle ilgili veriler toplamaktadır.	1	4	1	4	7	25	10	36	9	32	28
S 7	Alışveriş kartlarım, kredi kartlarım benimle ilgili veriler toplamaktadır.	0	0	3	11	1	4	14	50	10	36	28
S 8	Kullandığım sosyal medya uygulamaları benimle ilgili veriler toplamaktadır.	0	0	4	14	0	0	13	46	11	39	28
S 9	Kullandığım e-posta servisleri benimle ilgili veriler toplamaktadır.	0	0	4	14	1	4	15	54	8	29	28
S10	5 – 9 Sorulardaki araçlar benimle ilgili veriler topluyorsa bunları toplumun yararı (hastalıkların tedavisinde, ulaşımın geliştirilmesinde, suç oranının azaltılmasında vb.) için kullanmalarını destekliyorum.	0	0	6	21	3	11	13	46	6	21	28
S11	5 – 9 Sorulardaki araçlar benimle ilgili veriler topluyorsa bunları özel hayatımın gizliliğine dikkat edilerek diğer firmalarla paylaşmalarını olumlu buluyorum.	7	25	7	25	8	29	5	18	1	4	28
S12	Hakkımda toplanan verilerin, özel hayatımın gizliliğine dikkat edilerek başka sektörlerde bilim ve teknolojinin gelişmesi için kullanılmasını olumlu buluyorum.	4	14	0	0	12	43	9	32	3	11	28
S15	Kişisel verilerimi kullanan cep telefonu ya da bilgisayar uygulamaları beni rahatsız etmektedir.	0	0	4	14	7	25	11	39	6	21	28
S17	Günümüzde her cihazın birbirleriyle etkileşimi ve anlık veri yaratması, veri patlamasının ana nedenidir.	0	0	1	4	7	25	16	57	4	14	28
S18	Sadakat Kartları (mağaza promosyon kartları), yalnızca müşterilere promosyon yapmak veya çeşitli avantajlar sağlamak amacıyla kullanılmaktadır.	0	0	9	32	12	43	5	18	2	7	28

Birinci anketteki yukarıda verilen 11 sorunun güvenilirliğini sınamak amacı hesaplanan Cronbach Alfa değeri 0,671 olarak bulunmuştur. Bu değerler anketteki ilgili soruların oldukça güvenilir olduğunu göstermektedir.

Genel olarak kütüphanecilerin büyük çoğunluğunun ankette 5., 6., 7., 8. ve 9. sorularda sorulan veri toplayan nesnelerin farkında oldukları söylenebilir.

Bu tabloya göre varılan diğer sonuçlar aşağıdaki gibidir:

- 10. Soruda veri toplayan nesnelerin toplumun yararı için kullanılmasını, anketi dolduran kütüphanecilerin %68'i desteklediğini, %21'i desteklemediğini, %11'i ise kararsız olduğunu ya da fikrinin olmadığını belirtmiştir.
- 11. Soruda veri toplayan nesneler vasıtasıyla toplanan verilerin özel hayatın gizliliğine dikkat edilerek diğer firmalarla paylaşılmasını, anketi dolduran kütüphanecilerin %22'si olumlu bulduğunu, %50'si olumlu bulmadığını, %28'u ise kararsız olduğunu ya da fikrinin olmadığını belirtmiştir.
- 12. Soruda veri toplayan nesneler vasıtasıyla toplanan verilerin özel hayatın gizliliğine dikkat edilerek bilim ve teknolojinin gelişmesi için kullanılmasını, anketi dolduran kütüphanecilerin %43'ı olumlu bulduğunu, %14'si olumlu bulmadığını, %43'u ise kararsız olduğunu ya da fikrinin olmadığını belirtmiştir.
- 15. Soruda kişisel verilerin, cep telefonu ya da bilgisayar uygulamalarında kullanılmasından, anketi dolduran kütüphanecilerin %61'i rahatsız olduğunu, %14'ü rahatsız olmadığını, %25'i ise kararsız olduğunu ya da fikrinin olmadığını belirtmiştir.
- 17. Soruda günümüzde her cihazın birbirleriyle etkileşimi ve anlık veri yaratması, veri patlamasının ana nedeni olmadığına, anketi dolduran kütüphanecilerin %71'i katıldığını, %4'ü katılmadığını, %25'i ise kararsız olduğunu ya da fikrinin olmadığını belirtmiştir.

- 18. Soruda Sadakat Kartları (mağaza promosyon kartları), yalnızca müşterilere promosyon yapmak veya çeşitli avantajlar sağlamak amacıyla kullanılmakta olduğuna, anketi dolduran kütüphanecilerin %25'i katıldığını, %32'si katılmadığını, %43'i ise kararsız olduğunu ya da fikrinin olmadığını belirtmiştir.

Bu sonuçlara göre ankete katılan kütüphanecilerin büyük çoğunluğunun veri toplayan nesnelerin, veri patlamasının nedeninin birbirleriyle etkileşimli anlık veri yaratan bu cihazlar olduğunun farkında olduklarını, fakat genel olarak kendileri ile ilgili veri toplayan cihazlardan memnun olmadıkları söylenebilir. Bu durumda toplanan verilerin hangi durumlarda, kimler ile ve ne kadar ölçekte paylaşılabileceği konusunda ayrıntılı bilgi sahibi olmaları beklenebilir. Gerek etik açıdan, gerekse de yasal olarak hangi haklara sahip oldukları konusunda bilgilendirilmeleri gerekmektedir. Bu bilgilere sahip kütüphanecilerin kullanıcılarını da bilgilendireceği öngörülebilir.

Ankete katılan kütüphanecilerin çoğunluğunun, kişisel veri toplayan cihazları sadece toplumun yararı (hastalıkların tedavisinde, ulaşımın geliştirilmesinde, suç oranının azaltılmasında vb.) için destekleyebilecekleri, bilim ve teknolojinin gelişmesinde bile uygun görmedikleri ve diğer tüm koşullarda da kişisel verilerinin toplanmasını olumsuz buldukları söylenebilir. Sadakat kartlarının kişisel bilgileri toplaması ile ilgili ise yeterli bilgileri olmadıkları söylenebilir.

IV.1.6 Anket Bazında Genel Değerlendirme

IV.1.6.1 Birinci anket

Tablo 18. Kütüphanecilerin Birinci Anketin İkinci Bölümüne Verdikleri Yanıtların Sıklıkları

No	Sorular	Kesinlikle katılmıyorum		Katılmıyorum		Kararsızım / Yeterli bilgim yok		Katılıyorum		Kesinlikle Katılıyorum		Toplam n
		n	%	n	%	n	%	n	%	n	%	
1	Büyük Veri, verinin hacminin büyüklüğünü ifade eden bir kavramdır.	0	0,00	5	17,86	14	50,00	7	25,00	2	7,14	28
2	Büyük Veri, sadece verinin hacminin büyüklüğünü değil, aynı zamanda hızının ve çeşitliliğinin de büyüklüğünü ifade etmektedir.	0	0,00	1	3,57	16	57,14	9	32,14	2	7,14	28
3	Büyük Veri popüler bir kavram olduğu için gündeme gelmiştir. Popülerlik geçici olmasına rağmen, bu kavram geçerliliğini koruyacaktır.	0	0,00	4	14,29	14	50,00	5	17,86	5	17,86	28
4	Büyük Veri, yeni bir devrin başlangıcıdır.	0	0,00	1	3,57	16	57,14	8	28,57	3	10,71	28
5	Gün içinde güvenlik kameraları, mobese kameraları vb. cihazlar benimle ilgili veriler toplamaktadır.	1	3,57	3	10,71	2	7,14	12	42,86	10	35,71	28
6	Telefonum ve gün içinde karşılaştığım çeşitli sensörler benimle ilgili veriler toplamaktadır.	1	3,57	1	3,57	7	25,00	10	35,71	9	32,14	28
7	Alışveriş kartlarım, kredi kartlarım benimle ilgili veriler toplamaktadır.	0	0,00	3	10,71	1	3,57	14	50,00	10	35,71	28
8	Kullandığım sosyal medya uygulamaları benimle ilgili veriler toplamaktadır.	0	0,00	4	14,29	0	0,00	13	46,43	11	39,29	28
9	Kullandığım e-posta servisleri benimle ilgili veriler toplamaktadır.	0	0,00	4	14,29	1	3,57	15	53,57	8	28,57	28

10	5 – 9 Sorulardaki araçlar benimle ilgili veriler topluyorsa bunları toplumun yararı (hastalıkların tedavisinde, ulaşımın geliştirilmesinde, suç oranının azaltılmasında vb.) için kullanmalarını destekliyorum.	0	0,00	6	21,43	3	10,71	13	46,43	6	21,43	28
11	5 – 9 Sorulardaki araçlar benimle ilgili veriler topluyorsa bunları özel hayatımın gizliliğine dikkat edilerek diğer firmalarla paylaşmalarını olumlu buluyorum.	7	25,00	7	25,00	8	28,57	5	17,86	1	3,57	28
12	Hakkımda toplanan verilerin, özel hayatımın gizliliğine dikkat edilerek başka sektörlerde bilim ve teknolojinin gelişmesi için kullanılmasını olumlu buluyorum.	4	14,29	0	0,00	12	42,86	9	32,14	3	10,71	28
13	Bir teknoloji şirketim olsa, ben de güvenlik kameraları, sensörler, telefonlar, alışveriş kartları, kredi kartları, sosyal medya uygulamaları vb. cihazlardan ve uygulamalardan edinilen bilgilerle, kişilerin özel hayatın gizliliğine dikkat ederek yeni teknolojiler geliştirmek isterdim.	5	17,86	2	7,14	5	17,86	14	50,00	2	7,14	28
14	Kullandığım Cep Telefonu uygulamaları ile takip edildiğimi ya da izlendiğimi düşünüyorum.	0	0,00	5	17,86	9	32,14	10	35,71	4	14,29	28
15	Kişisel verilerimi kullanan cep telefonu ya da bilgisayar uygulamaları beni rahatsız etmektedir.	0	0,00	4	14,29	7	25,00	11	39,29	6	21,43	28
16	Araştırma verilerinin yapılandırılmamış olması ve çok çeşitli biçimlerde olması nedeniyle erişimi çok zordur.	0	0,00	5	17,86	12	42,86	9	32,14	2	7,14	28
17	Günümüzde her cihazın birbirleriyle etkileşimi ve anlık veri yaratması, veri patlamasının ana nedenidir.	0	0,00	1	3,57	7	25,00	16	57,14	4	14,29	28
18	Sadakat Kartları (mağaza promosyon kartları), yalnızca müşterilere promosyon yapmak veya çeşitli avantajlar sağlamak amacıyla kullanılmaktadır.	0	0,00	9	32,14	12	42,86	5	17,86	2	7,14	28

IV.1.6.2 İkinci anket

Tablo 19. Kütüphanecilerin İkinci Ankete Verdikleri Yanıtların Sıklıları

No	Sorular	Kesinlikle katılmıyorum		Katılmıyorum		Kararsızım / Yeterli bilgim yok		Katılıyorum		Kesinlikle Katılıyorum		Toplam n
		n	%	n	%	n	%	n	%	n	%	
1	Büyük Veri, verinin hacminin büyüklüğünü ifade eden bir kavramdır.	1	4,55	3	13,64	0	-	11	50,00	7	31,82	22
2	Büyük Veri, sadece verinin hacminin büyüklüğünü değil, aynı zamanda hızının ve çeşitliliğinin de büyüklüğünü ifade etmektedir.	0	-	0	-	0	-	11	50,00	11	50,00	22
3	Büyük Veri popüler bir kavram olduğu için gündeme gelmiştir. Popülerlik geçici olmasına rağmen, bu kavram geçerliliğini koruyacaktır.	1	4,55	0	-	0	-	9	40,91	12	54,55	22
4	Büyük Veri, yeni bir devrin başlangıcıdır.	0	-	0	-	0	-	11	50,00	11	50,00	22
5	Bir teknoloji şirketim olsa, ben de güvenlik kameraları, sensörler, telefonlar, alışveriş kartları, kredi kartları, sosyal medya uygulamaları vb. cihazlardan ve uygulamalardan edinilen bilgilerle, özel hayatın gizliliğine dikkat edilerek yeni teknolojiler geliştirmek isterdim.	0	-	1	4,55	1	4,55	9	40,91	11	50,00	22
6	Kullandığım Cep Telefonu uygulamaları ile takip edildiğimi ya da izlendiğimi düşünüyorum.	0	-	1	4,55	0	-	8	36,36	13	59,09	22
7	Araştırma verilerinin yapılandırılmamış olması ve çok çeşitli olması nedeniyle erişimi çok zordur.	2	9,09	4	18,18	5	22,73	8	36,36	3	13,64	22
8	Kütüphane işlem ve hizmetlerinde Büyük Veri yeni ve önemli bir devrin başlangıcıdır.	0	-	0	-	2	9,09	13	59,09	7	31,82	22
9	RDA ile kütüphanelerde büyük veri uygulamaları önemli bir hale gelmiştir.	0	-	0	-	4	18,18	13	59,09	5	22,73	22
10	Bağlı veri (Linked Data) le kütüphanelerde büyük veri uygulamaları önemli bir hale gelmiştir.	1	4,55	0	-	6	27,27	10	45,45	5	22,73	22
11	Amerikan Kongre Kütüphanesinin (LOC) bütün Twitter verilerini dermesine eklemesi, kültürel mirasın gelecek nesillere aktarılması nedeniyle kütüphanecilik mesleği bakımından çok önemlidir.	0	-	0	-	3	13,64	9	40,91	10	45,45	22
12	Citeseerx gibi dijital kütüphaneler, kütüphencilik mesleği konusunda yeni bir devrin başlangıcı olarak kabul edilebilir.	1	4,55	0	-	2	9,09	10	45,45	9	40,91	22
13	Araştırma verileri, büyük veri uygulamaları ile kütüphane dermesinde yer alabilecektir.	0	-	0	-	1	4,55	12	54,55	9	40,91	22
14	Kütüphaneler de zamanla veri simsarı (komisyoncusu) olabilecek, özel hayatın gizliliği ve telif haklarını koruyarak verilerini satabileceklerdir.	1	4,55	3	13,64	5	22,73	9	40,91	4	18,18	22

15	Kütüphaneler zamanla bağlı olduğu kurumun her türlü enformasyon ihtiyacını karşılayabileceklerdir.	0	-	0	-	0	-	14	63,64	8	36,36	22
16	Web 2.0 ve Web 3.0 konusunda büyük veri bir dönüm noktasıdır.	0	-	1	4,55	5	22,73	11	50,00	5	22,73	22
17	Yapay zekalar gelecekte kütüphanelerin önemli bir parçası olacaktır.	0	-	0	-	0	-	16	72,73	6	27,27	22
18	Yapay zekalar gelecekte kütüphanelerde pek çok hizmette kullanılabilecektir.	0	-	0	-	1	4,55	15	68,18	6	27,27	22
19	Süper bilgisayarlar ile birlikte kütüphanecilik mesleği başka bir boyut kazanacaktır.	1	4,55	0	-	0	-	13	59,09	8	36,36	22
20	Günümüzde her cihazın birbirleriyle etkileşimi ve anlık veri yaratması, kütüphanecilik mesleği açısındanda yeni fırsatlar yaratmaktadır.	1	4,55	0	-	1	4,55	14	63,64	6	27,27	22
21	Açık veri konusunda ABD, Avrupa Birliği vb. ülkeler ve toplulukların geliştirdikleri uygulamalar ve bunların erişime sunulması, kütüphanecilik biliminin konusunu içinde değerlendirilmelidir.	0	-	0	-	2	9,09	13	59,09	7	31,82	22
22	Üniversite Kütüphaneleri, büyük veri ile ilgili hizmetlerinin ve dermesinin geliştirilmesinde, diğer türdeki kütüphanelere önderlik etmelidir	0	-	0	-	2	9,09	12	54,55	8	36,36	22
23	Kütüphaneler, kullanıcılarının verilerini izleyerek, onlara daha iyi hizmet verebilirler.	0	-	0	-	3	13,64	9	40,91	10	45,45	22
24	Kütüphaneler iç yapılarını (yazılım , donanım ve eğitim açısından) büyük veri toplayacak, başka kurum/kuruluşlar tarafından toplanan büyük veriye erişimi sağlayacak şekilde değiştirmelidir.	0	-	0	-	3	13,64	11	50,00	8	36,36	22
25	Büyük Veri toplama ve değerlendirme yöntemleri konusunda bir eğitim almak isterim.	0	-	0	-	1	4,55	13	59,09	8	36,36	22
26	Büyük verinin kütüphanlerde kullanılması konusunda sadece kütüphaneciler değil, bilgisayar donanım ve yazılım uzmanları, istatistikçiler, gerektiğinde konu uzmanları vb. birlikte çalışmalıdır.	0	-	0	-	0	-	10	45,45	12	54,55	22
27	Yapılmış olan sunum bana Büyük Veri kavramını algılamamda, olumlu katkıda bulunmuştur.	0	-	0	-	1	4,55	11	50,00	10	45,45	22

II. BÖLÜM

S1 : Büyük Verinin kütüphanenizde hangi hizmet ve işlemlerde nasıl kullanılabileceği ile ilgili düşünce ve önerileriniz varsa yazınız?

- Tüm Hizmetler. Okuyucuya daha iyi hizmet sunabilmek açısından geliştirmekte kullanılabilir.
- Büyük veri konusunda eğitim almak isterim.
- Kullanıcıların hesapları takip edilerek ilgi alanlarına yönelik çalışmalar yapılabilir.
- Kullanıcıların ilgi duydukları kaynakların tespit edilmesi ve sağlanmasında, kullanıcı hizmetlerinin geliştirilmesinde kullanılabilir
- Akademik araştırma, deney, gözlem verilerinin kütüphaneye bağlı yazılımlarla açık veri olarak sonlanması gerek diye düşünüyorum.

V BÖLÜM

SONUÇ VE ÖNERİLER

Araştırmanın bu bölümünde kavramsal bölümler ile ilgili bulgulara ve uygulanan anket çalışması sonuçlarına yer verilecektir.

Büyük verinin günümüzde bu kadar ünlü bir konu haline gelmesinde çeşitli devletlerin, toplulukların, kurum ve kuruluşların bu konuya çok büyük yatırımlar yapmalarının ve dolayısıyla da medyada çokça yer almasının sonucu olduğu söylenebilir.

Büyük veri ile de yeni pek çok kavram ortaya çıkmıştır ve bazı kavramlar da günümüzde, eskiden hiç olmadığı kadar popüler hale gelmiştir. Çalışmamızda bu kavramların pek çoğuna yer verilmiştir.

Büyük veri ile birlikte yalnızca teknoloji alanında değil, düşünme, algılama biçimlerimizde, araştırma yöntemlerimizde, yapay zekaların da devreye girmesiyle daha pek çok farklı alanda büyük değişiklikler yaşanmaktadır. Kütüphanelerin de bu değişimlerin dışında kalamayacağı sonucuna varılmıştır.

Kütüphanelerin büyük veri konusuna ne kadar hazır olduğunu ölçmek için Ankara Üniversitesi Kütüphaneleri genelinde yapılan araştırmamızda varılan sonuçlar aşağıdaki şekilde özetlenebilir:

- Kütüphanecilerin yarıdan fazlası büyük veri kavramını daha önceden duymuşlardır. Fakat büyük çoğunluğunun büyük veri ile ilgili daha önce bir kaynak okumamalarına rağmen, büyük veriye kaynak oluşturabilecek cihazların veri ürettiklerinin farkındadırlar ve bu cihazları genel olarak kullanmaktadırlar. Kütüphanecilerin genelinin veri toplayan cihazların kendileri ile ilgili veri toplamalarını uygun görmedikleri, yalnızca toplumun yararı için belki bu verilerin toplanabileceği yönünde görüşlerinin olduğu sonucuna varılmıştır.
- Yapılan bilgilendirici sunuma katılanların, sunumdan sonra Büyük Veri konusunda temel bilgi düzeylerinde istatistiksel olarak anlamlı bir artış olduğu görülmüştür. Kütüphanecilerin mesleklerinin doğası gereği, büyük veri konusuna ne kadar çabuk adapte oldukları görülebilir ve büyük verinin bu yüzden kütüphanecilik mesleği içinde de değerlendirilmesi gerektiği söylenebilir.
- Kütüphanecilerin büyük çoğunluğu kütüphane işlem ve hizmetlerinde büyük verinin kullanımına katıldıkları, ayrıca yine büyük çoğunluğunun, büyük veriyi kütüphanecilik bilimi açısından önemli gördükleri sonucuna varılmıştır.
- İkinci ankette 5 kütüphanecinin, büyük verinin kütüphanelerinde hangi hizmet ve işlemlerde nasıl kullanılabileceği ile ilgili düşünce ve önerilerine, büyük veriyi kullanıcılara daha iyi hizmet vermek için bir araç olarak kullanabilecekleri, kütüphanede pek çok hizmetin geliştirilmesinde kullanılabileceğini, araştırma verilerinin kütüphane dermesinde yer almasında kullanılabileceği gibi yanıtlar alınmıştır. Bu

yanıtlara göre kısa bir sunumdan sonra bile kütüphanecilerin meslekleri gereği konuya ne kadar çabuk adapte olduklarını göstermiştir.

- Kütüphanecilerin büyük çoğunluğu büyük veri toplama ve değerlendirme konusunda kapsamlı bir eğitim almak istedikleri sonucuna varılmıştır.

Araştırma kapsamında elde edilen bu sonuçlara göre aşağıdaki öneriler sıralanabilir:

- Özellikle üniversite kütüphaneleri büyük verinin kütüphanelerde kullanımı konusunda öncü olmalı, üniversitelerin ve bilimle ilgili diğer kurum ve kuruluşların büyük veri ile ilgili teknoloji ve uygulama geliştirilmesine destek vermeleri gerekmektedir. Bunu gerçekleştirebilmeleri için de üniversite kütüphanecileri teşvik edilmelidir.
- Kütüphaneciler büyük veriyi kendi meslekleri içinde değerlendirmek istiyorlarsa, onlara bu konuda gerekli eğitimi almaları için gerekli ortam sağlanmalı, kütüphaneci yetiştiren bölümler ise müfredatlarını bu doğrultuda düzenlemelidir.
- Mezun kütüphaneciler için büyük veri konusunu ve araçlarını anlatan uygulamalı seminerler ve hizmet içi eğitimleri düzenlenmelidir.
- Verilecek eğitimlerde büyük veriyi işleme sonucunda elde edilecekler somut olarak ortaya konmalı, böylelikle kütüphanecilerin konunun önemini daha gerçekçi kavramalarına olanak sağlanmalıdır.
- Kütüphanecilerin özellikle büyük veri konusunda diğer ilgili disiplinlerle işbirliği içinde çalışmaları sağlanmalıdır.

- Büyük veri konusunda yurt dışı çalışmalar (CiteseerX, LOC'nin sosyal medya verilerini saklaması vb.) örnek alınıp incelenmeli benzeri modeller yurtiçinde üniversite kütüphanelerinin öncülüğünde gerçekleştirilmelidir.
- Büyük veri ve etik konusu üzerine gerekli tartışma platformları oluşturulmalı, bu konuda politikalar geliştirmek için üniversiteler, kamu kurum ve kuruluşları ve özel sektöründe katılımıyla gerekli çalışmalar yapılmalıdır.

KAYNAKÇA

Ankara Üniversitesi Kütüphane ve Dokümantasyon Daire Başkanlığı. *Ankara Üniversitesi Kütüphane ve Dokümantasyon Daire Başkanlığı Web Sitesi*. tarih yok. Web. 15 06 2015. <http://kutuphane.ankara.edu.tr/>.

Arslantekin, Sacit. «Veri Madenciliği ve Bilgi Merkezleri.» *Türk Kütüphaneciliği* 2003: 369-380. 6 3 2015. <http://www.tk.org.tr/index.php/TK/article/view/303/295>.

Atılgan, Doğan, Nevzat Özel ve Tolga Çakmak. «Awareness, Perceptions, and Expectations of Academic Librarians in Turkey about Resource Description and Access (RDA).» *Cataloging & Classification Quarterly* 22 08 2014: 660-676. web. 01 06 2015. <http://dx.doi.org/10.1080/01639374.2014.945023>.

Bayter, Mustafa. «Bilgi Kaynaklarının Tanımlanması ve Erişimi: AACR2'den RDA'ya geçiş.» *Bilgi Kksenli Kuram ve Uygulamalar, Sorgulayıcı ve Çözümleyici Yaklaşımlar Sempozyumu*. Dü. Fatoş Subaşıoğlu, Oya Gürdal Tamdoğan ve Tülay Fenerci. Nevşehir: Ankara Üniversitesi Basımevi, 2012. 119-131.

—. *Türkçe Web Belgelerinin Kataloglanması: Bir İşbirliği Modeli Önerisi*. Ankara, 2008. Web. 01 06 2015. <https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=7d53ed97e31a8bd343165919e7dff23bdcdb4045a56396ed4ef2bfaaa49722ab63cf256a52fccf68>.

Brazilian Federal Government. *All results*. tarih yok. 05 04 2015. <http://www.copa2014.gov.br/en/copa/tabela>.

Cackett, Doug. *Information Management and Big Data, A Reference Architecture*. white paper. Redwood Shores: Oracle Corporation, 2013. web. 20 11 2014. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.398.7632&rep=rep1&type=pdf>.

Can, Ahmet. «Veriyi Petrol gibi İşleyin.» *Hürriyet* (2015). Web. 12 06 2015. <http://www.hurriyet.com.tr/ekonomi/29039191.asp>.

CiteSeerX. *CiteSeerX Hisstory*. tarih yok. The Pennsylvania State University, The College of Information Sciences and Technology. web. 20 11 2014. <http://csxstatic.ist.psu.edu/about/history>.

Davis, Kord ve Doug Patterson. *Ethics of big data: Balancing risk and innovation*. Sebastopol: O'Reilly, 2012.

Dean, Jeffrey ve Sanjay Ghemawat. «MapReduce: Simplified Data Processing on Large Clusters.» 12 2004. *Google, Inc. Web Sitesi.* 06 04 2015. <http://static.googleusercontent.com/media/research.google.com/tr//archive/mapreduce-osdi04.pdf>.

Dede, Melih Bayram. «Büyük Veri'yle İlgili 5 Yanlış.» *Yeni Şafak* (2015). Web. 12 06 2015. <http://www.yenisafak.com.tr/teknoloji/buyuk-veriyle-ilgili-5-yanlis-2152669>.

Demirkan, Haluk ve Bülent Dal. «Büyük Veri, Büyük Fırsatlar, Büyük Kararlar.» *Harvard Business Review, Türkiye* (2014). Web. 12 06 2015.

<http://www.hbrturkiye.com/x/editorun-secimi-buyuk-veri-buyuk-firsatlar-buyuk-kararlar>.

Diebold, Francis X. «A personal perspective on the origin(s) and development of "big data": the phenomenon, the term, and the discipline.» 26 11 2012. *University of Pennsylvania web Sitesi.* 01 05 2014. <http://www.ssc.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf>.

Durmuş, Beril, E. Serra Yurtkoru ve Murat Çinko. *Sosyal Bilimlerde SPSS'le Veri Analizi*. 5. İstanbul: Beta, 2013.

Elmas, Çetin. *Yapay Zeka Uygulamaları - Yapay Sinir Ağları, Bulanık Mantık, Genetik Algoritma*. 2. Ankara: Seçkin, 2011.

Fenerci, Tülay. «Veri Tabanı Tasarımının Önemi ve Normalizasyon Süreci.» *Türk Kütüphaneciliği* 15 2 2001: 123-135.

Gamgam, Hamza ve Bülent Altunkaynak. *SPSS Uygulamalı Parametrik Olmayan Yöntemler*. 5. Ankara: Seçkin, 2013.

Green, Rebecca ve Michael Panzer. «The Interplay of Big Data, WorldCat, and Dewey.» *Big Data, Linked Data: Classification Research at the Junction 24th ASIS&T SIG/CR Classification Research Workshop*. OCLC, 2013. Web. 19 06 2015. <https://www.oclc.org/events/2013/the-interplay-of-big-data-worldcat-and-dewey.en.html>.

Gülle, M. Tayfun. «Büyük Veri ya da İçgörü.» *Türk Kütüphaneciliği* 27 04 2013: 581-582.

Gür, Emine H. «Kütüphanelerde Linked Data'ya Geçiş, RDA, VIAF, BibFramework.» 26 03 2013. *RDA Türkiye Web Sitesi.* Web. 28 06 2015. <http://www.rdaturkiye.org/wp-content/uploads/2013/03/Emine.pdf>.

Gürdal, Oya. *Tekstil Endüstrisinde Enformasyon Olgusu*. Ankara: Türk Kütüphaneciler Derneği, 2000.

Gürsakal, Necmi. *Büyük veri*. Bursa: Dora, 2013.

Henkoğlu, Türkay ve Bülent Yılmaz. «Avrupa Birliği (AB) Bilgi Güvenliği Politikaları.» *Türk Kütüphaneciliği* 27 3 2013: 451-471.

Hessman, Travis. «Putting big data to work: Dell masters the art of information to help pull itself back to the top, one byte at a time.» *Industry Week* 04 2013: 14-18. Web. 19 11 2014. <http://content.ebscohost.com/ContentServer.asp?T=P&P=AN&K=87306829&S=R&D=a9h&EbscoContent=dGJyMNXb4kSep7c4zOX0OLCmr0yeprdSs6u4TLaWxWXS&ContentCustomer=dGJyMPGstkyyr7FQuePfgeyx43zx>.

Huang, Lung. «Be mine: Data-driven Valentine' Day wishes.» 06 02 2014. *AdExchanger*. 10 04 2015. <http://adexchanger.com/data-driven-thinking/be-mine-data-driven-valentines-day-wishes/>.

Hürriyet. «'Büyük veri'yi anında okuyup fırsat yaratmak.» *Hürriyet* (2013). Web. 12 06 2015. <http://www.hurriyet.com.tr/ekonomi/23466283.asp>.

—. «Artık güç tüketicide.» *Hürriyet* (2015). Web. 12 06 2015. <http://www.hurriyet.com.tr/ekonomi/29124805.asp>.

—. «Büyük veri çağı.» *Hürriyet* (2015). Web. 12 06 2015. <http://www.hurriyet.com.tr/ekonomi/29115095.asp>.

—. «Büyük Veriyle İlgili Doğru Bilinen Yanlışlar.» *Hürriyet* (2014). Web. 12 06 2015. <http://www.hurriyet.com.tr/teknoloji/26101086.asp>.

IBM. *What is Watson?* tarih yok. Web. 23 06 2015. <http://www.ibm.com/smarterplanet/us/en/ibmwatson/what-is-watson.html>.

Jeffery, Keith . «Data is the New Oil.» *Best Practices for Data Management & Sharing*. Dü. The Joint Research Centre (JRC). Ispra, Italy, 2014. Sunum. 09 04 2015. <http://indico.cern.ch/event/313634/material/slides/5.pdf>.

Johnson, Jeanne E. «Big Data + Big Analytics = Big Opportunity.» *FinancialExecutive* 07/08 2012: 50-53. <http://www.financialexecutives.org/KenticoCMS/Financial-Executive-Magazine/2012_07/Big-Data---Big-Analytics---Big-Opportunity.aspx#axzz3937tfvaD>.

Laney, Doug. «3D data management: Controlling Data Volume, velocity, and variety.» 06 02 2001. *Gartner Web Sitesi.* Dü. Application Delivery Strategies. 01 05 2014. <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.

Library of Congress. «Twitter Donates Entire Tweet Archive to Library of Congress.» 15 04 2010. *News from the Library of Congress.* 25 05 2015. <http://loc.gov/today/pr/2010/10-081.html>.

—. «Update on the Twitter Archive.» 01 2013. *Library of Congress Web Sitesi.* 25 05 2015. <http://www.loc.gov/today/pr/2013/files/twitter_report_2013jan.pdf>.

Lohr, Steve. «Library of Congress Will Save Tweets.» 14 04 2010. *The New York Times Web Sitesi.* 25 05 2015. <http://www.nytimes.com/2010/04/15/technology/15twitter.html?_r=0>.

—. «The Age of Big Data.» *The New York Times* (2012). Web. 12 06 2015. <http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?_r=1>.

Manyika, James, ve diğerleri. «Big data: The next frontier for innovation, competition and productivity.» Mayıs 2011.

Mashey, John R. «Big data and the next wave of infrastress.» 25 04 1998. *Usenix Web Sitesi.* 01 05 2014. <https://www.usenix.org/legacy/event/usenix99/invited_talks/mashey.pdf>.

McAfee, Andrew ve Erik Brynjolfsson. «Big Data: The Management Revolution.» *Harvard Business Review* (2012). Web. 12 06 2015. <https://hbr.org/2012/10/big-data-the-management-revolution/>.

Microsoft. «Microsoft SQL Server 2012 Parallel Data Warehouse - A Breakthrough Platform for Next-Generation Data Warehousing and Big Data Solutions.» 2013. *Microsoft Web Sitesi.* Web. 23 06 2015. <http://download.microsoft.com/download/5/0/1/5015A62E-06BF-4DEC-B90A-

37D52E279DE5/SQL_Server_2012_Parallel_Data_Warehouse_Breakthrough_Platform_White_Paper.pdf>.

Nabiyev, Vasif V. *Yapay Zeka - Problemler, Yöntemler, Algoritmalar*. 2. Ankara: Seçkin, 2005.

OCLC. *OCLC Completes Major Technical Upgrade of Core WorldCat Infrastructure*. 10 06 2013. Web. 23 06 2015. <https://www.oclc.org/news/releases/2013/201329dublin.en.html>.

Oğuz, Şeref. «Büyük veri mi, bilgi çöplüğü mü?» *Sabah* (2014). Web. 12 06 2015. <http://www.sabah.com.tr/yazarlar/oguz/2014/12/13/buyuk-veri-mi-bilgi-coplugu-mu>.

Oğuzlar, Ayşe. *Temel Metin Madenciliği*. Bursa: Dora, 2011.

Özel, Nevzat. «Bilginin Düzenlenmesine Yönelik Yeni Bir Standart: Resource Description and Access (RDA).» *Bilgi Dünyası* 05 2015. Web. 11 06 2015. <http://bd.org.tr/index.php/bd/article/view/473>.

Palmer, Michael. «Data is the New Oil.» 03 11 2006. *ANA Marketing Maestros Blog.* 10 04 2015. <http://ana.blogs.com/maestros/2006/11/data_is_the_new.html>.

Pektaş, Ali Osman. *SPSS ile Veri Madenciliği*. İstanbul: Dikeyeksen, 2013.

PriceStats. *History: PriceSstats*. tarih yok. 05 03 2015. <http://www.pricestats.com/about-us/history>.

Prytherch, Ray. *Harrod's librarians' glossary and reference book : a dictionary of over 10,200 terms*. 10. Hampshire: Ashgate Publishing Limited, 2005.

Rotella, Perry. «Is Data The New Oil?» 02 04 2012. *Forbes.* 2015 04 2015. <http://www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/>.

Sankur, Bülent. *İngilizce - Türkçe Ansiklopedik Bilişim Sözlüğü*. İstanbul: Pusula, 2004.

SAP SE. *SAP News Center: SAP and the German Football Association (DFB) Turn Big Data Into Smart Decisions to Improve Player Performance at the World Cup in Brazil*. 11 06 2014. 05 04 2015. <http://global1.sap.com/corporate-en/news.epx?category=ALL&articleID=23079>.

Schönberger, Viktor Mayer ve Kenneth Cukier. *Big Data: A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt, 2013.

—. *Büyük Veri - Yaşama, Çalışma ve Düşünme Şeklimizi Dönüştürecek Bir Devrim*. Çev. Banu Erol. 1. İstanbul: Paloma Yayınevi, 2013.

Stanton, Jeffrey M. «Data Science: What's in it for the New Librarian?» *Information Space, School of Information Studies, Syracuse University* (2012). Web. 12 06 2015. <http://infospace.ischool.syr.edu/2012/07/16/data-science-whats-in-it-for-the-new-librarian/>.

Sümbüloğlu, Vildan, Kürşat Akyüz ve Kadir Sümbüloğlu. «Probleme Uygun Biyoistatistik Yöntemlerinin Seçilmesi: II. Özel Yöntemler.» *İç Hastalıkları Dergisi* 1997: 420-423. Web. 19 06 2015. <http://www.ichastaliklaridergisi.org/fulltext.aspx?issue_id=68&ref_ind_id=545>.

Şeker, Şadi Evren. *İş Zekası ve Veri Madenciliği*. İstanbul: Cinius Yayınları, 2013.

Tanış, Tolga. «Big data çağı.» *Hürriyet* (2013). Web. 12 06 2015. <http://www.hurriyet.com.tr/yazarlar/23207665.asp>.

—. «http://www.hurriyet.com.tr/yazarlar/23207665.asp.» *Hürriyet* (2013). Web. 12 06 2015. <http://www.hurriyet.com.tr/yazarlar/23566118.asp>.

Teets, Michael ve Matthew Goldner. «Libraries' Role In Curating And Exposing Big Data.» *Future Internet* 5.3 (2013): 429-438. Web. 19 11 2014. <http://www.mdpi.com/1999-5903/5/3/429>.

Tezbaşaran, A. Ata. «Likert Tipi Ölçek Hazırlama Kılavuzu.» 2008. *Academia.edu Web sitesi.* Web. 01 06 2015. <https://www.academia.edu/1288035/Likert_Tipi_%C3%96l%C3%A7ek_Haz%C4%B1rlama_K%C4%B1lavuzu>.

The European Organisation for International Research Information. «CERIF Tutorials.» 13 11 2013. *CERIF Web Sitesi.* Sunum. 09 04 2015. <http://www.eurocris.org/Uploads/Web%20pages/CERIFtutorial/CERIF%20tutorial_Porto_13.11.2013%20-%20Jan%20Dvorak.pptx>.

The White House, Executive Office of the President. «Big Data: Seizing Opportunities, Preserving Values.» 01 05 2014. *The White House Web Site.* 01 06 2015. <https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf>.

The White House, Office of the Press Secretary. «We Can't Wait: Obama Administration Unveils Blueprint for a "Privacy Bill of Rights" to Protect Consumers Online.» 23 02 2012. *The White House Web Sitesi.* Web. 18 06 2015. <https://www.whitehouse.gov/the-press-office/2012/02/23/we-can-t-wait-obama-administration-unveils-blueprint-privacy-bill-rights>.

Türk Dil Kurumu. «Büyük Türkçe Sözlük.» tarih yok. *T.C. Başbakanlık, Atatük Kültür, Dil ve Tarih Yüksek Kurumu, Türk Dil Kurumu.* Web. 24 06 2015.

Uysal, Mithat. *Access 2003 ile Veri Tabanı Yönetimi*. 1. İstanbul: Beta, 2006.

Watson and the Jeopardy! Challenge. IBM. 2013. Web. 23 06 2015. <https://www.youtube.com/watch?v=P18EdAKuC1U>.

Williams, K., ve diğerleri. «Scholarly big data information extraction and integration in the CiteSeerχ digital library.» *Data Engineering Workshops (ICDEW) , 2014 IEEE 30th International Conference on Data Engineering (ICDE).* Chicago: IEEE, March 31 2014-April 4 2014. 68-73. Web. 20 11 2014. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6818305>.

World Economic Forum. «Big data, big impact: new possibilities for international development.» 2012. *World Economic Forum web sayfası.* 01 05 2014. <http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf>.

Yıldızoğlu, Ergin. «'Büyük Veri', Büyük Tehlike.» *Cumhuriyet* (2013). Web. 12 06 2015.

<http://www.cumhuriyet.com.tr/koseyazisi/424142/_Buyuk_Veri___Buyuk_Tehlike.html>.

Yılmaz, Malik. «Enformasyon ve Bilgi Kavramları Bağlamında Enformasyon Yönetimi ve Bilgi Yönetimi.» *Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi* 2009: 95-118.

Zikopoulos, Paul C., ve diğerleri. *Harness the power of big data, the IBM big data platform*. McGraw-Hill, 2013.

EKLER

Ek-1: Anket Öncesi Bilgilendirme Metni

Değerli Katılımcı,

Bu anket Ankara Üniversitesi, Dil ve Tarih-Coğrafya Fakültesi, Bilgi ve Belge Yönetimi Bölümü'nde yürütmekte olduğum Yüksek Lisans Tezi kapsamında yapmış olduğum araştırmaya veri oluşturması beklentisiyle, tez konusuyla ilgili yapacağım sunumdan önce doldurulacak şekilde hazırlanmıştır.

Anketimize vermiş olduğunuz yanıtlar gizlilik prensibi çerçevesinde genel araştırma verileri olarak değerlendirilecek ve bilimsel çalışma amaçlı kullanılacaktır. Hiçbir şekilde üçüncü şahıslarla paylaşılmayacaktır.

Anket sunumdan önce ve sunumdan sonra olmak üzere iki farklı şekilde uygulanacaktır. Sonuçların doğru bir şekilde değerlendirilebilmesi amacıyla ilk anketin ve ikinci anketin ilk bölümüne aynı unutmayacağınız kodların girilmesi iki ankette de tekrarlayan yedi sorunun değerlendirilmesi açısından son derece önemlidir.

Araştırmamızın temeli olan verileri elde etmek için bu anketin sorularına vereceğiniz yanıtlar son derece önem taşımaktadır. Ankete vereceğini doğru, samimi, içten yanıtlar anketin ve araştırmanın doğru sonuçlar vermesini sağlayacaktır.

İlginiz ve ayırdığınız zaman için teşekkür ederim.

Uzm. Korcan Doğan
Ankara Üniversitesi
Dil ve Tarih-Coğrafya Fakültesi
Bilgi ve Belge Yönetimi Bölümü
dogank@ankara.edu.tr
+90 533 3545094

Ek-2: Birinci Anket

"BÜYÜK VERİNİN ÜNİVERSİTE KÜTÜPHANELERİ AÇISINDAN ÖNEMİ VE KÜTÜPHANE HİZMETLERİNDE KULLANIMI: ANKARA ÜNİVERSİTESİ KÜTÜPHANELERİ ÖRNEĞİ" KONULU YÜKSEK LİSANS TEZİ İÇİN ARAŞTIRMA ANKETİ 1

Sayın Katılımcı

Anketimize vermiş olduğunuz cevaplar gizlilik prensibi çerçevesinde genel veriler olarak değerlendirilecek ve bilimsel çalışma amaçlı kullanılacaktır. İşbirliğiniz ve desteğiniz için öncelikle teşekkür ederiz. **Size uygun seçeneklerin numaralarını lütfen (O) daire içine alarak işaretleyiniz.**

İkinci ankette de kullanacağınız, unutmayacağınız bir kod yazınız
(Sıralı ve Tekrarlayan rakamlardan oluşmamalıdır)

I. BÖLÜM

S1	Lisans mezuniyet yılınız	S2	Kütüphanecilik veya Bilgi ve Belge Yönetimi Bölümünden en son aldığınız derece hangisidir?
	.....	1	Lisans
		2	Yüksek Lisans
		3	Doktora
		4	Hiçbiri

S3	Büyük Veri kavramını daha önce duydunuz mu?	S4	Büyük Veri ile ilgili daha önce bir kaynak okudunuz mu?	S5	Sosyal medya uygulamalarını hangi sıklıkla kullanıyorsunuz?
1	Duydum	1	Okudum	1	Hiç kullanmıyorum
2	Duymadım	2	Okumadım	2	Bazen
				3	Sıklıkla

S6	Sosyal medya uygulamalarından hangilerini kullanıyorsunuz? (Birden fazla seçenek işaretlenebilir)		
1	Facebook	9	Wikipedia
2	Twitter	10	Flickr
3	Instagram	11	Picasa
4	Google+	12	Blogger
5	LinkedIn	13	Wordpress
6	Four Square	14	Hiçbiri
7	Tumblr	15	Diğer .....
8	Youtube		

S7	E-posta kullanıyor musunuz?	S8	Bulut Bilişim hizmetlerini kullanıyor musunuz?	S9	Kredi kartı veya alışveriş kartları kullanıyor musunuz?
1	Evet	1	Evet	1	Evet
2	Hayır	2	Hayır	2	Hayır

S10	Güvenlik kameraları, sensörler, kullandığınız telefonlar, alışveriş kartları, kredi kartları, sosyal medya uygulamaları vb. cihazlar ve uygulamaların sizin hakkınızda bilgi topladığını biliyor musunuz? Daha önce hiç düşündünüz mü? (Birden fazla seçenek işaretlenebilir)
1	Evet
2	Kısmen Evet
3	Hayır

S11	Mesleğinizle ilgili yeni teknolojileri takip ediyor musunuz?	S12	Mesleğinizle ilgili yeni teknolojileri hangi ortamlarda takip ediyorsunuz? (Birden fazla seçenek işaretlenebilir)		
1	Hiçbir zaman	1	Basılı kaynaklar	5	Forumlar
2	Bazen	2	Web siteleri	6	Tartışma listeleri
3	Sıklıkla	3	Açık erişim dergileri	7	Sosyal Medya
		4	Veri tabanları	8	Diğer .....

	II. BÖLÜM	Kesinlikle katılmıyorum	Katılmıyorum	Kararsızım / Yeterli bilgim yok	Katılıyorum	Kesinlikle katılıyorum
S1	Büyük Veri, verinin hacminin büyüklüğünü ifade eden bir kavramdır.	1	2	3	4	5
S2	Büyük Veri, sadece verinin hacminin büyüklüğünü değil, aynı zamanda hızının ve çeşitliliğinin de büyüklüğünü ifade etmektedir.	1	2	3	4	5
S3	Büyük Veri popüler bir kavram olduğu için gündeme gelmiştir. Popülerlik geçici olmasına rağmen, bu kavram geçerliliğini koruyacaktır.	1	2	3	4	5
S4	Büyük Veri, yeni bir devrin başlangıcıdır.	1	2	3	4	5
S5	**Gün içinde güvenlik kameraları, mobese kameraları vb.** cihazlar benimle ilgili veriler toplamaktadır.	1	2	3	4	5
S6	**Telefonum ve gün içinde karşılaştığım çeşitli sensörler** benimle ilgili veriler toplamaktadır.	1	2	3	4	5
S7	**Alışveriş kartlarım, kredi kartlarım** benimle ilgili veriler toplamaktadır.	1	2	3	4	5
S8	Kullandığım **sosyal medya uygulamaları** benimle ilgili veriler toplamaktadır.	1	2	3	4	5
S9	Kullandığım **e-posta servisleri** benimle ilgili veriler toplamaktadır.	1	2	3	4	5
S10	5 – 9 Sorulardaki araçlar benimle ilgili veriler topluyorsa bunları **toplumun yararı** (hastalıkların tedavisinde, ulaşımın geliştirilmesinde, suç oranının azaltılmasında vb.) için kullanmalarını destekliyorum.	1	2	3	4	5
S11	5 – 9 Sorulardaki araçlar benimle ilgili veriler topluyorsa bunları özel hayatımın gizliliğine dikkat edilerek **diğer firmalarla** paylaşmalarını olumlu buluyorum.	1	2	3	4	5
S12	Hakkımda toplanan verilerin, özel hayatımın gizliliğine dikkat edilerek başka sektörlerde **bilim ve teknolojinin gelişmesi** için kullanılmasını olumlu buluyorum.	1	2	3	4	5
S13	Bir teknoloji şirketim olsa, ben de güvenlik kameraları, sensörler, telefonlar, alışveriş kartları, kredi kartları, sosyal medya uygulamaları vb. cihazlardan ve uygulamalardan edinilen bilgilerle, kişilerin özel hayatın gizliliğine dikkat ederek yeni teknolojiler geliştirmek isterdim.	1	2	3	4	5
S14	Kullandığım Cep Telefonu uygulamaları ile takip edildiğimi ya da izlendiğimi düşünüyorum.	1	2	3	4	5
S15	Kişisel verilerimi kullanan cep telefonu ya da bilgisayar uygulamaları beni rahatsız etmektedir.	1	2	3	4	5
S16	Araştırma verilerinin yapılandırılmamış olması ve çok çeşitli biçimlerde olması nedeniyle erişimi çok zordur.	1	2	3	4	5
S17	Günümüzde her cihazın birbirleriyle etkileşimi ve anlık veri yaratması, veri patlamasının ana nedenidir.	1	2	3	4	5
S18	Sadakat Kartları (mağaza promosyon kartları), yalnızca müşterilere promosyon yapmak veya çeşitli avantajlar sağlamak amacıyla kullanılmaktadır.	1	2	3	4	5

Ek-3: İkinci Anket

"BÜYÜK VERİNİN ÜNİVERSİTE KÜTÜPHANELERİ AÇISINDAN ÖNEMİ VE KÜTÜPHANE HİZMETLERİNDE KULLANIMI: ANKARA ÜNİVERSİTESİ KÜTÜPHANELERİ ÖRNEĞİ" KONULU YÜKSEK LİSANS TEZİ İÇİN ARAŞTIRMA ANKETİ 2

Sayın Katılımcı

Anketimize vermiş olduğunuz cevaplar gizlilik prensibi çerçevesinde genel veriler olarak değerlendirilecek ve bilimsel çalışma amaçlı kullanılacaktır. İşbirliğiniz ve desteğiniz için öncelikle teşekkür ederiz. **Size uygun seçeneklerin numaralarını lütfen (O) daire içine alarak işaretleyiniz.**

Birinci ankette kullandığınız kodu yazınız

	I. BÖLÜM	Kesinlikle katılmıyorum	Katılmıyorum	Kararsızım / Yeterli bilgim yok	Katılıyorum	Kesinlikle katılıyorum
S1	Büyük Veri, verinin hacminin büyüklüğünü ifade eden bir kavramdır.	1	2	3	4	5
S2	Büyük Veri, sadece verinin hacminin büyüklüğünü değil, aynı zamanda hızının ve çeşitliliğinin de büyüklüğünü ifade etmektedir.	1	2	3	4	5
S3	Büyük Veri popüler bir kavram olduğu için gündeme gelmiştir. Popülerlik geçici olmasına rağmen, bu kavram geçerliliğini koruyacaktır.	1	2	3	4	5
S4	Büyük Veri, yeni bir devrin başlangıcıdır.	1	2	3	4	5
S5	Bir teknoloji şirketim olsa, ben de güvenlik kameraları, sensörler, telefonlar, alışveriş kartları, kredi kartları, sosyal medya uygulamaları vb. cihazlardan ve uygulamalardan edinilen bilgilerle, özel hayatın gizliliğine dikkat edilerek yeni teknolojiler geliştirmek isterdim.	1	2	3	4	5
S6	Kullandığım Cep Telefonu uygulamaları ile takip edildiğimi ya da izlendiğimi düşünüyorum.	1	2	3	4	5
S7	Araştırma verilerinin yapılandırılmamış olması ve çok çeşitli olması nedeniyle erişimi çok zordur.	1	2	3	4	5
S8	Kütüphane işlem ve hizmetlerinde Büyük Veri yeni ve önemli bir devrin başlangıcıdır.	1	2	3	4	5
S9	RDA ile kütüphanelerde büyük veri uygulamaları önemli bir hale gelmiştir.	1	2	3	4	5
S10	Bağlı veri (Linked Data) le kütüphanelerde büyük veri uygulamaları önemli bir hale gelmiştir.	1	2	3	4	5
S11	Amerikan Kongre Kütüphanesinin (LOC) bütün Twitter verilerini dermesine eklemesi, kültürel mirasın gelecek nesillere aktarılması nedeniyle kütüphanecilik mesleği bakımından çok önemlidir.	1	2	3	4	5
S12	Citeseerx gibi dijital kütüphaneler, kütüphencilik mesleği konusunda yeni bir devrin başlangıcı olarak kabul edilebilir.	1	2	3	4	5
S13	Araştırma verileri, büyük veri uygulamaları ile kütüphane dermesinde yer alabilecektir.	1	2	3	4	5
S14	Kütüphaneler de zamanla veri simsarı (komisyoncusu) olabilecek, özel hayatın gizliliği ve telif haklarını koruyarak verilerini satabileceklerdir.	1	2	3	4	5
S15	Kütüphaneler zamanla bağlı olduğu kurumun her türlü enformasyon ihtiyacını karşılayabileceklerdir.	1	2	3	4	5
S16	Web 2.0 ve Web 3.0 konusunda büyük veri bir dönüm noktasıdır.	1	2	3	4	5
S17	Yapay zekalar gelecekte kütüphanelerin önemli bir parçası olacaktır.	1	2	3	4	5
S18	Yapay zekalar gelecekte kütüphanelerde pek çok hizmette kullanılabilecektir.	1	2	3	4	5
S19	Süper bilgisayarlar ile birlikte kütüphanecilik mesleği başka bir boyut kazanacaktır.	1	2	3	4	5
S20	Günümüzde her cihazın birbirleriyle etkileşimi ve anlık veri yaratması, kütüphanecilik mesleği açısındanda yeni fırsatlar yaratmaktadır.	1	2	3	4	5
S21	Açık veri konusunda ABD, Avrupa Birliği vb. ülkeler ve toplulukların geliştirdikleri uygulamalar ve bunların erişime sunulması, kütüphanecilik biliminin konusunu içinde değerlendirilmelidir.	1	2	3	4	5
S22	Üniversite Kütüphaneleri, büyük veri ile ilgili hizmetlerinin ve dermesinin geliştirilmesinde, diğer türdeki kütüphanelere önderlik etmelidir	1	2	3	4	5
S23	Kütüphaneler, kullanıcılarının verilerini izleyerek, onlara daha iyi hizmet verebilirler.	1	2	3	4	5
S24	Kütüphaneler iç yapılarını (yazılım , donanım ve eğitim açısından) büyük veri toplayacak, başka kurum/kuruluşlar tarafından toplanan büyük veriye erişimi sağlayacak şekilde değiştirmelidir.	1	2	3	4	5
S25	Büyük Veri toplama ve değerlendirme yöntemleri konusunda bir eğitim almak isterim.	1	2	3	4	5
S26	Büyük verinin kütüphanlerde kullanılması konusunda sadece kütüphaneciler değil, bilgisayar donanım ve yazılım uzmanları, istatistikçiler, gerektiğinde konu uzmanları vb. birlikte çalışmalıdır.	1	2	3	4	5
S 27	Yapılmış olan sunum bana Büyük Veri kavramını algılamamda, olumlu katkıda bulunmuştur.	1	2	3	4	5

II. BÖLÜM

S1 Büyük Verinin kütüphanenizde hangi hizmet ve işlemlerde nasıl kullanılabileceği ile ilgili düşünce ve önerileriniz varsa yazınız?

.....

.....

Ek-4: Sunum İçeriği

"Büyük Verinin Üniversite Kütüphaneleri Açısından Önemi ve Kütüphane Hizmetlerinde Kullanımı" başlıklı sunumun içeriği aşağıdaki başlıklardan oluşmaktadır.

1. Medya ve şirketlerde büyük veri
2. Nesnelerin interneti
3. Büyük verinin tanımı ve içeriği
4. Büyük veriyi kullanan kütüphane örnekleri
5. Kongre Kütüphanesi
6. CiteseerX
7. Kütüphaneler ve büyük veri
8. Mesleki Yeterlilikler
9. Soru, görüş ve öneriler

(Sunum süresi toplam 15-20 dakika)

Ek-5: Sunum

BÜYÜK VERİNİN ÜNİVERSİTE KÜTÜPHANELERİ AÇISINDAN ÖNEMİ VE KÜTÜPHANE HİZMETLERİNDE KULLANIMI

(SUNUM SÜRESİ TOPLAM 15-20 DAKİKA)

UZM. KORCAN DOĞAN
ANKARA ÜNİVERSİTESİ
DİL VE TARİH-COĞRAFYA FAKÜLTESİ
BİLGİ VE BELGE YÖNETİMİ BÖLÜMÜ

SUNUMUN İÇERİĞİ

1. Medya ve şirketlerde büyük veri
2. Nesnelerin interneti
3. Büyük verinin tanımı ve içeriği
4. Büyük veriyi kullanan kütüphane örnekleri
   i. Kongre Kütüphanesi
   ii. CiteseerX
5. Kütüphaneler ve büyük veri
6. Mesleki Yeterlilikler
7. Soru, görüş ve öneriler

(Sunum süresi toplam 15-20 dakika)



https://www.ucl.ac.uk/big-data/bdi

1. MEDYADA BÜYÜK VERİ



http://www.hurriyet.com.tr/ekonomi/29115095.asp



http://www.hurriyet.com.tr/teknoloji/26101086.asp

1. MEDYADA BÜYÜK VERİ



http://www.cumhuriyet.com.tr/koseyazisi/424142/_Buyuk_Veri___Buyuk_Tehlike.html



http://www.hbrturkiye.com/x/editorun-secimi-buyuk-veri-buyuk-firsatlar-buyuk-kararlar

1. MEDYADA BÜYÜK VERİ


MENU
Harvard Business Review

DECISION MAKING

Big Data: The Management Revolution

by Andrew McAfee and Erik Brynjolfsson

SUMMARY SAVE SHARE COMMENT TEXT SIZE PRINT $8.95 BUY COPIES

"You can't manage what you don't measure."

There's much wisdom in that saying, which has been attributed to both W. Edwards Deming and Peter Drucker, and it explains why the recent explosion of digital data is so important. Simply put,

https://hbr.org/2012/10/big-data-the-management-revolution


HOME SEARCH The New York Times

The Agency | 2. Last Task After Layoff at Disney: Train Foreign Replacements | 3. Instagram to Open Its Photo Feed to Ads | 4.

Gmail for Work
Daha profesyonel E-posta hizmeti

SundayReview | NEWS ANALYSIS

The Age of Big Data

By STEVE LOHR FEB. 11, 2012

Email Share Tweet Save

GOOD with numbers? Fascinated by data? The sound you hear is opportunity knocking.

Mo Zhou was snapped up by I.B.M. last summer, as a freshly minted Yale M.B.A., to join the technology company's fast-growing ranks of data consultants. They help businesses make sense of an explosion of data — Web traffic and social network comments, as well as software and sensors that monitor shipments, suppliers and customers — to guide decisions, trim costs and lift sales. "I've always had a love of numbers," says Ms. Zhou,

http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?_r=0

1. MEDYADA BÜYÜK VERİ


SEÇİM 2015
ZİRAAT TÜRKİYE KUPASI

SABAH
TÜRKİYE'NİN EN İYİ HABER SİTESİ

ANA SAYFA | +PLUS | SPOR | GÜNAYDIN | GÜNDEM | EKONOMİ | YAŞAM | DÜNYA | SABAH TV

YAZARLAR

ÖNERİ-YORUM
ŞEREF OĞUZ
Büyük veri mi bilgi çöplüğü mü?

Paylaş tweetle paylaş

Bilgi güçtür ve daima güç olagelmiştir. **Yeni ekonomide** ise bilgi, üretim faktörleri **emek**, **sermaye** ve **hammadde** gibi belirleyici öneme ulaşmıştır. **Adam Smith**, milletleri zengin kılan **üretim faktörlerini biriktirmenin önemine** vurgu yapar. **Paylaşılınca azalan** bu üretim faktörleri sadece geçen yüzyılda **100 milyon insana** mal olan savaşların da ana sebebi oldu.

Bilgi ise **farklı davranan** üretim faktörü... **Paylaşılınca çoğalıyor**, paylaşmayınca **bilgi kirlenmesi** dediğimiz **veri çöplüğüne** dönüşüyor.

Bilişim teknolojileri, yapay zekâ ve internet, küresel boyutta bilgi birikimine yol açtı.

http://www.sabah.com.tr/yazarlar/oguz/2014/12/13/buyuk-veri-mi-bilgi-coplugu-mu


Yeni Şafak

4 HAZİRAN

GÜNDEM | POLİTİKA | EKONOMİ | SPOR | DÜNYA | SAĞLIK | YAZARLAR | AKTÜEL | EĞİTİM
TEKNOLOJİ | SİNEMA | YAZI İŞLERİ | KÜLTÜR-SANAT | RÖPORTAJ | YORUM | PAZAR | KİTAP | OTOMOTİV

Paralel Yapı'nın ABD 'kulağı' | BUGÜN NE OLDU?

Büyük Veri'yle ilgili 5 yanlış

Büyük Veri (Big data) adı verilen yöntemle devasa verileri işleyerek anlamlı konuda doğru yöntemler kullanmak ve zamanı verimli kullanmak şart.

Yeni Şafak | Mesih Bayram Dede | 30 Mayıs 2015, 4:00

Haberler » Teknoloji » Yazdır

http://www.yenisafak.com.tr/teknoloji/buyuk-veriyle-ilgili-5-yanlis-2152669

2. KURUMLAR VE BÜYÜK VERİ

Aşağıdaki kurumların büyük veri için açtıkları özel sayfalar, bu konuya ne kadar önem verdiklerinin göstergesidir.

- ABD, Beyaz Saray, Büyük Veri Web Sayfası
  https://www.whitehouse.gov/issues/technology/big-data-review
- IBM Büyük Veri Web Sayfası
  http://www.ibm.com/big-data/us/en/
- Microsoft Büyük Veri Web Sayfası
  https://msdn.microsoft.com/en-us/big-data-development-msdn.aspx
- Oracle Büyük Veri Web Sayfası
  https://www.oracle.com/big-data/index.html
- EMC Büyük Veri Web Sayfası
  http://www.emc.com/big-data/marketing.htm
- SAP Büyük Veri Web Sayfası
  http://go.sap.com/solution/big-data.html

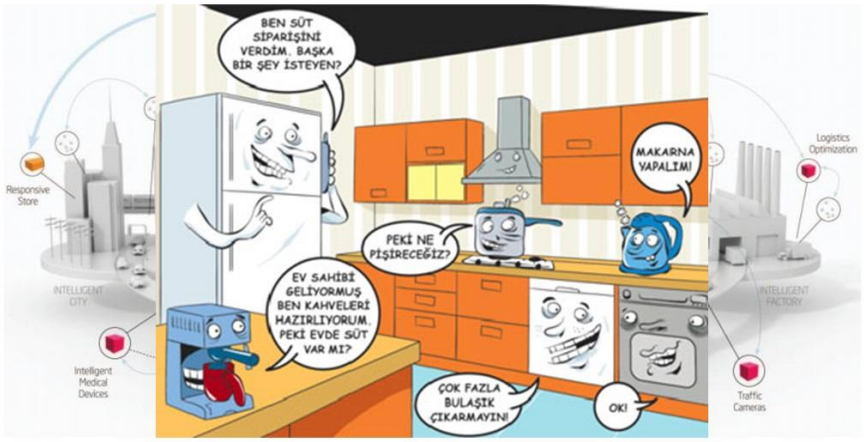
2. NESNELERİN İNTERNETİ



http://www.nesnelerininterneti.co/internet-of-things-iot-nedir/

- Nesnelerin interneti, her gün kullanılan **nesnelerin içine çipler, sensörler ve iletişim modülleri yerleştirilerek**, kısmen çevrimiçi ağ oluşturmakla ama daha çok insanları çevreleyen her şeyi verileştirmekle ilgili bir kavramdır (Schönberger ve Cukier 103,104).

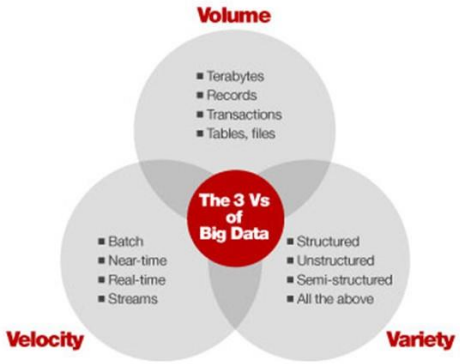
2. NESNELERİN İNTERNETİ



http://www.milliyet.com.tr/teknoloji-gelecek-cihazlar-diye-pazarlanacak/haberdetay/03.08.2014/1920141/default.htm

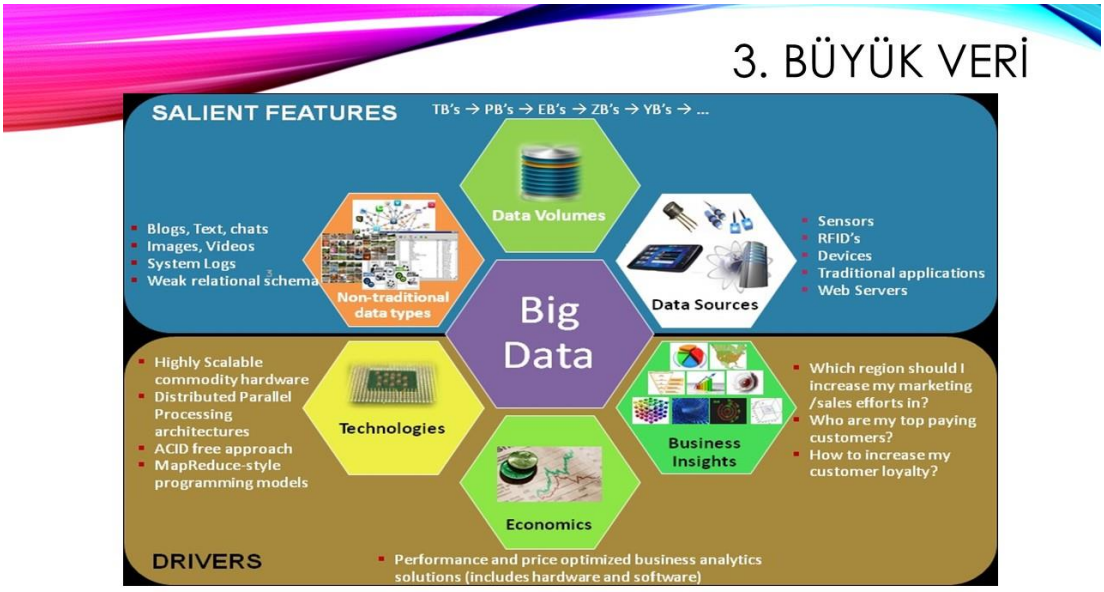
http://www.vedalabs.com/nesnelerin-interneti/

3. BÜYÜK VERİ

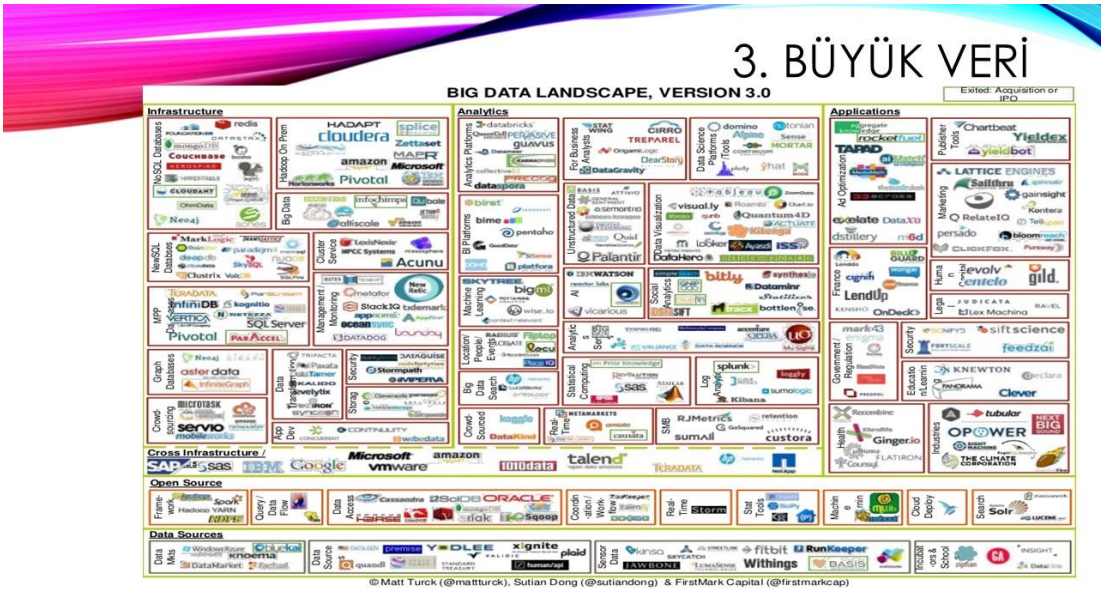


http://foresightinvestor.com/articles/411823-the-abc-s-of-big-data

- McKinsey Global Institute'nün 2011 yılında tanımladığı (Manyika, Chui ve Brown 1) büyük veri, **tipik ve geleneksel veri tabanı yazılımlarının kabiliyetlerinin ötesinde,** yapamayacağı şekilde, veri kümelerini alan, saklayan, yöneten, erişime sunan ve analiz eden araçları tanımlamak için kullanılmaktadır.



https://shhrota.files.wordpress.com/2012/01/big-data2-shreedandekar1.jpg



http://www.ongridventures.com/wp-content/uploads/2012/10/Big-Data-Landscape1.jpg

4. KONGRE KÜTÜPHANESİ ÖRNEĞİ

- 2010 yılında Twitter en çok 140 karakterlik tweetlerinden oluşan 2006-2010 yıllarına ait bütün dijital arşivini LOC'ye bağışlayacağını duyurmuştur (Library of Congress).
  - 133,2 Terabayt hacminde 170 milyar tweet bu şekilde 2012 yılının Aralık ayında kütüphane dermesine eklenmiştir (Library of Congress 2).
  - LOC, 2011 yılının Şubat ayında günlük 140 milyon tweeti dermesine ekliyorken, 2012 Kasım ayında bu rakam günlük 500 milyon tweete yükselmiştir (Library of Congress 3).

4. KONGRE KÜTÜPHANESİ ÖRNEĞİ

- **Kütüphaneler Açısından Önemi:** Tweetlerin arşivlenmesi ve muhafaza edilmesi, gelecekteki araştırmacıların günümüzün kültürel normları, diyalogları, olayların ve eğilimlerin bilimsel yönleri, yasama süreci, yeni yazarlık çalışmaları, eğitim, vb. amaçlarda gerekli enformasyona erişilmesini sağlayacacağını düşündüğü için dermesine ekleme kararı almıştır. (Library of Congress 1).
- Bunların da ne kadar önem taşıdığı sizlerce daha iyi değerlendirilebilir.

4. CİTESEERX ÖRNEĞİ

- CiteSeerx akademik dokümanlar için bir dijital kütüphane ve arama motorudur.
- Bilgisayar, enformasyon ve bunlarla ilgili bilimlere odaklanmaktadır.
- Önemli özelliklerinden biri dermeye eklenen bütün dokümanlardan otomatik olarak enformasyon çıkarması ve otomatik alıntı indekslemesi - linklemesi yapmasıdır (Williams, Wu ve Choudhury 69).

4. CİTESEERX ÖRNEĞİ

- Günde ortalama 50.000–100.000 doküman crawl edebilmektedir.
- Bunların ortalama %40' ı akademik dokümandır.
- Dermesine güne ortalama 10.000 doküman ayda ortalama 300.000 doküman katabilmektedir (Williams, Wu ve Choudhury 73).
- Dermesinde toplam 3,5 milyon doküman vardır,
- 80 milyon bibliyografik atıf vardır,
- 3-6 milyon arası yazarı, 800.000 bireysel kullanıcısı vardır ve günde ortalama 2 ile 4 milyon ziyaret almaktadır (Williams, Wu ve Choudhury 69).

Bu bağlamda hızla büyümeye devam eden CiteSeerx büyük veriyi en iyi kullanan ve yöneten bilgi merkezlerinden biridir.

5. KÜTÜPHANELER VE BÜYÜK VERİ

- Kütüphanelerde Büyük Veri iki ana başlık altında ele alınabilir.

  i. Büyük Verinin Kütüphane hizmetlerini geliştirmek için kullanılması

  ii. Kütüphanelerin Büyük Veri yi dermesinde bulundurması

5. KÜTÜPHANELER VE BÜYÜK VERİ

Kütüphaneler Büyük Veri konsunda gelişme göstermek istiyolarsa öncelikle aşağıdaki soruları vb. cevaplamalarında fayda vardır:

- Kütüphanelerde hangi veriler, ne gibi şekillerde, nerelerde üretilmektedir?
- Bu veriler, başka hangi hizmetlerde kullanılabilir?
- Dışarıdan hangi veriler alınırsa, hangi hizmetlerimizi iyileştirmelere neden olabilir?
- Hangi veriler dış kaynaklardan dermemize eklenmelidir (Teknolojik kapasitemiz ve kaynaklarımız bağlamında değerlendirilmelidir)?
- Dış kaynaklardan edinilen veriler, erişime nasıl sunulacaktır ?
- Bu veriler hizmet verdiğimiz araştırmacılara nasıl bir katkıda bulunacaktır ?
- …

6. BÜYÜK VERİ İÇİN GEREKEN MESLEKİ YETERLİLİKLER

- Kütüphaneciler için büyük Veri konusunda önemli olan aşağıdaki disiplinler ile ilgili eğitim alarak onlarla ortak çalışma yapılabilecek duruma gelmeleridir.
- Bu disiplinler (Stanton):

a) Bilgisayar Programcılığı
b) Veritabanı Yöneticiliği
c) İstatistik
d) Grafik Tasarımcı
e) vb.

7. SORU, GÖRÜŞ VE ÖNERİLER

8. KAYNAKÇA

- Cackett, Doug. *Information Management and Big Data, A Reference Architecture*. white paper. Redwood Shores: Oracle Corporation, 2013. web. 20 11 2014. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.398.7632&rep=rep1&type=pdf>.
- CiteSeerX. *CiteSeerX Hisstory*. tarih yok. The Pennsylvania State University, The College of Information Sciences and Technology. web. 20 11 2014. <http://csxstatic.ist.psu.edu/about/history>.
- Demir, Murat Can. "Nesnelerin İnterneti (IOT) Nedir ?" 19 05 2012. Murat Can Demir, Articles on Mobile, Telecommunication, Tech. Entrepreneurship. 03 06 2015. <http://www.muratcandemir.com/nesnelerin-interneti-nedir-iot.html>.
- Library of Congress. «Twitter Donates Entire Tweet Archive to Library of Congress.» 15 04 2010. *News from the Library of Congress*. 25 05 2015. <http://loc.gov/today/pr/2010/10-081.html>.

  —. «Update on the Twitter Archive.» 01 2013. *Library of Congress Web Sitesi*. 25 05 2015. <http://www.loc.gov/today/pr/2013/files/twitter_report_2013jan.pdf>.
- Manyika, James, ve diğerleri. «Big data: The next frontier for innovation, competition and productivity.» Mayıs 2011.
- Schönberger, Viktor Mayer ve Kenneth Cukier. *Büyük Veri - Yaşama, Çalışma ve Düşünme Şeklimizi Dönüştürecek Bir Devrim*. Çev. Banu Erol. 1. İstanbul: Paloma Yayınevi, 2013.
- Stanton, Jeffrey M. "Data Science: What's in it for the New Librarian?" 16 07 2012. *Information Space, School of Information Studies, Syracuse University*. 03 06 2015. <http://infospace.ischool.syr.edu/2012/07/16/data-science-whats-in-it-for-the-new-librarian/>.
- Williams, K., ve diğerleri. «Scholarly big data information extraction and integration in the CiteSeerx digital library.» *Data Engineering Workshops (ICDEW) , 2014 IEEE 30th International Conference on Data Engineering (ICDE)*. Chicago: IEEE, March 31 2014-April 4 2014. 68-73. Web. 20 11 2014. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6818305>.
- Zikopoulos, Paul C., ve diğerleri. *Harness the power of big data, the IBM big data platform*. McGraw-Hill, 2013.

İLETİŞİM

Uzm. Korcan Doğan
Ankara Üniversitesi
Dil ve tarih-coğrafya fakültesi
Bilgi ve Belge Yönetimi Bölümü

dogank@ankara.edu.tr

ÖZET

Büyük veri günümüzde yeni bir devrin başlangıcı olarak yorumlanmaktadır. Büyük veri ile birlikte dünyada büyük bir dönüşüm gerçekleşirken, kütüphanelerinde bu dönüşümün dışında kalmaları mümkün görünmemektedir. Kütüphaneler açısından büyük veri iki temel başlık altında incelenmiştir. Bunlar;

i. Kütüphanelerin dermesinde büyük verinin yer alabilmesi,
ii. Kütüphane işlem ve hizmetlerinde büyük verinin, bu işlem ve hizmetleri geliştirmek için kullanılabilirliğidir.

Bu dönüşüm sürecinde ise üniversite kütüphanelerine de yapılarından dolayı büyük görevler düşmektedir ve bu dönüşümde öncü olmaları gerekmektedir.

Fakat büyük verinin kütüphaneler için yeni bir kavram olduğu, literatürde kütüphane uygulamalarına ilişkin çok fazla araştırmayla karşılaşılmadığı ortadadır. Bu durum meslek uzmanlarının (bilgi yöneticiler/kütüphaneciler) kendilerini bu konuda yetiştirmelerinin ötesinde konunun kendilerince öneminin farkındalığına varılması açısından da olumsuz bir ortam yaratmaktadır.

Yukarıda bahsedilen farkındalığın var olup olmadığı, Ankara Üniversitesi Kütüphanelerinde çalışan meslek mezunlarına bir anket uygulamasıyla ölçülmeye çalışılmıştır. Bu uygulamada önce 28 kütüphaneciye, büyük veri hakkında fikir sahibi olup olmadıklarını ortaya koyabilmek için küçük bir anket yapılmış hemen ardından büyük veri ve kütüphanelerde kullanımına yönelik bilgilendirici sunum gerçekleştirilmiştir ve sunuma 22 kütüphaneci katılmıştır. Sunum sonrasında 22 kütüphaneciye tekrar bir anket uygulanarak büyük veri konusunda farkındalığın sağlanıp sağlanamadığı ortaya konulmaya çalışılmıştır.

Çalışmamızın sonucuna göre eğitim sunumuna katılan kütüphanecilerin, sunumdan sonra büyük veri konusunda temel bilgi düzeylerinde istatistiksel olarak anlamlı bir artış olduğu söylenebilir. Yine bu sonuçlara göre kütüphanecilerin, mesleklerinin doğası gereği, büyük veri konusuna ne kadar çabuk adapte oldukları görülebilir. Büyük verinin bu yüzden kütüphanecilik mesleği içinde de değerlendirilmesi gerektiği ve büyük veri ile ilgili hazırlanacak eğitim programlarının kütüphanelere olumlu katkılar sağlayacağı söylenebilir.

ABSTRACT

Big data can be considered as the beginning of a new era. With the effects of the big data while world is in a big transition, it is not seen as possible to keep libraries out of this big transition. Big data is evaluated for libraries under two basic chapters within the scope of this study:

i. Place of big data in libraries` collection
ii. Usage of big data in the library operations and services, and improvements of these operations and services

In this transition, university libraries have a lot of responsibilities due to their structure and they should lead this transition.

However it is very clear that big data is a new concept for libraries and in literature there are not many researches on the execution of big data in the libraries. This situation results in a negative environment in which occupation experts (data managers and librarians) neither have the awareness on the importance of the subject nor incentive and desire to develop themselves.

To understand the level of the mentioned awareness, a research was conducted with the graduates of the occupation who works in Ankara University Library. In this conduct, a small survey was made with 28 librarians to measure the level of their information on big data and then, just after the survey, an informative presentation about big data was made in which 22 librarians attended. After the presentation another survey was conducted to these 22 librarians to understand whether awareness on big data were raised on not.

Based on our study, it can be said that librarians who attended the presentation statistically significantly increased their knowledge on the basic levels of big data. Based on these results, it can be also seen how quickly librarians get adapt to the big data which may be also thought as a natural consequences of their occupation. Thus big data should be considered in the occupation of the librarianship and it can be said that education curriculums about big data would have positive contribution to libraries.