

# Şirketler için Veri Temin Yöntemleri ve Uygulanabilir Bilgi Çıkarım Teknikleri

## Data Procurement Methods and Applicable Information Extraction Techniques for Companies

Metin Turan  
Yıldız Teknik Üniversitesi  
turanmetin@gmail.com

A Coşkun Sönmez  
İstanbul Teknik Üniversitesi  
sonmez@itu.edu.tr

### Öz

Günümüzde şirketler ticaret faaliyetlerini sürdürebilmek ve daha da önemlisi mevcut müşterilerini elde tutabilmek ve potansiyellerini arttırmak üzere bilgiye eskisinden daha da fazla ihtiyaç duymaktadırlar. Bu bilgi şirketin altyapısını oluşturan bilgi işlem verilerinden elde edilebileceği gibi, çeşitli kaynaklardan sağlanabilecek ticari faaliyetlere yönelik verilerden de elde edilebilir. Uğraşılan ticari iş kolundaki mevcut yönelimler, talepler, tedarik, lojistik ve dağıtım yöntemlerinin doğru tespiti firmanın diğer rakiplerine göre üstünlük sağlaması ve rekabette bir adım öne geçmesi demektir. Bilgisayar teknolojisini alt yapısına entegre etmiş ve kendi bünyesindeki veya dış ortamdaki verilerden güncel ve doğru bilgi çıkarımını sağlayabilen, karar alma ve uygulama profesyonelliğini bünyesinde oluşturabilmiş bir şirket global düzeyde başarı sağlayabilecektir. Organizasyon içindeki yönetim ve idari yapılanma tamamen farklı bir uzmanlık alanıdır. Bu yazıda sadece, veriden bilgi çıkarımı teknolojileri ve şirketlerin bunları hangi amaçlar için kullanabileceği ele alınmaktadır.

**Anahtar Sözcükler:** Bilgi Çıkarım

### Abstract

Nowadays companies feel more need to knowledge than ever to sustain their trade activities and, more importantly, in order to retain existing customers and increase potential. This information may be obtained directly from the company's infrastructure, such as transactional data, it may also be data supplied from different resources pertaining to the commercial activities. Correct determination of the methods for

business trends on the commercial line, demands, supply, logistics and distribution means company outclasses other competitors and forges ahead one more step. The company in which computer technology has been integrated into the infrastructure and that can provide current and accurate information extraction from data its on-site or from the external environment that embodies decision making and implementation professionalism within the company, was able to create success on a global level. Management and administrative structure of the organization is a completely different area of expertise. In this paper, only information extraction technologies from data and companies that can use these for which objectives are addressed.

**Keywords:** Information Extraction

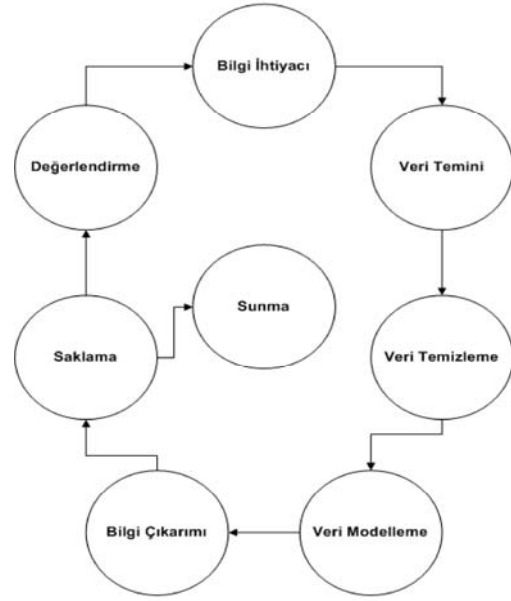
### 1 Giriş

Günümüz bilgi çağıdır. Bu doğrultuda organize olmamış ortalıkta dolaşan onca verinin içinden gürültünün ayrıştırılıp, “güncel ve doğru bilgi çıkarımı bilgisayar teknikleri kullanılarak nasıl yapılabilir” sorusuna şirketler yanıt aramaktadırlar. Daha da önemlisi şirketlerin ne istediklerini doğru ifade etmeleri, uygulanabilir bir modelin oluşturulup, hedef bilginin çıkarımı gereklidir. Bilgi basit istatistiksel bir yöntemle elde edilmekten öte, bilginin genelde veride fark edilemediği durumlara yönelik çözümler için gereklidir.

Basit düzeyde bir bilgi çıkarım yaşam döngüsü Şekil 1’de verilmiştir. Bu döngü öncelikle bir bilgi ihtiyacının ortaya çıkmasıyla başlar. Verinin temini, temizlenmesi ve modellenmesinden sonra bilgi çıkarım işlemi uygulanır.

Genderim ve kabul tarihi : 18.10.2015-30.12.2014

Veriye uygulanabilecek bilgi çıkarım teknikleri olarak bilgi getirme (Information Retrieval), metin özetleme (Document Summarization), veri madenciliği (Data Mining) ve Web madenciliği (Web Mining) önerilebilir. Bu tekniklerin ortak olan özelliği, çok miktarda dağınık veya organize olmuş verilerden istenilen bilgileri elde etme yöntemleri içermeleridir. Diğer bir deyişle, bu tekniklerin uzmanlığı gizli veya açık bilgiye erişim sağlamaktır.



Şekil-1: Bilgi Çıkarımı Yaşam Döngüsü

Bununla birlikte bu teknikler uygulama alanları açısından farklılıklar içermektedirler. Bilgi getirme sistemleri [1] genellikle veri kaynağı içinde sorgulanan verinin olup olmadığını, varsa nerede/nerelerde bulunduğu bilgisini bulmak üzere geliştirilen sistemleri içermektedir. Diğer bir deyişle mevcut verinin varlığı ve konumu hakkında bilgi elde eden sistemlerdir. Bu alanda uygulanan tekniğe bağlı olarak çeşitli yöntemler mevcuttur. Örneğin veri getirme (Data Retrieval) aranan veriyle ilişkili diğer bilgilere ulaşma amacıyla kullanılan sistemlerdir [2]. Veri kısmen veya tamamen organize olduğu için bir ilişkisel veritabanı mantığıyla aranan veriyle ilişkili diğer veri ve bilgilere ulaşmak mümkün olabilmektedir. Bulanık bilgi getirme (Fuzzy Information Retrieval) ise aranan bilginin kısmen

varlığı hakkında bilgi döndürebilen sistemlerdir [3]. Diğer bir deyişle aranan veriye tamamen eş veya belirli bir yakınlıkta verilerin bulunabilmesine olanak sağlamaktadır. Dizinleme, diğer bir deyişle verinin anlamlı parçalar halinde gruplanabilmesi, bilgi getirme çalışmalarının başarısını doğrudan etkileyen bir faktördür.

Diğer taraftan metin özetleme konusu ise bir veya daha fazla dökümandan bir özet oluşturabilme çalışmasını kapsamaktadır [4, 5]. Özetleme yöntemleri, mevcut dökümanların içinden bilgi değeri yüksek metin parçalarının [6] (cümle, paragraf veya metin segmenti) doğrudan özeti oluşturmak üzere kullanılması (çıkarımsal özetleme – Extractive Summarization) veya bulunan bu metin parçalarının dil bilimi kullanılarak yeniden yapılandırılarak kullanılması (soyut özetleme – Abstractive Summarization) olarak karşımıza çıkmaktadır. İkinci yöntem doğal dil işleme (Natural Language Processing) ve yapay zeka (Artificial Intelligence) uzmanlık alanlarına ihtiyaç duyduğu için oldukça zor bir çalışma konusudur. Bu alanda yapılan çalışmalar henüz uzman özetleyicinin oluşturduğu özete yakın sonuçlar elde edememektedir. Çalışmalar uzman özetleri referans alınarak deneysel olarak sürdürülmektedir.

Veri madenciliği [7, 8] veya Web madenciliği [9] ise bağımsız veriler (organize değil) arasındaki ilişkilerin tespit edilebilmesi amacıyla geliştirilmiş algoritma ve tekniklerden oluşmaktadır. Amaç veri içinden bilgi keşfine dayalıdır ve bu bilgi insan gözüyle çıplak olarak fark edilemez. Bu çalışmalardan elde edilen sonuçlar bugün özellikle ticari değer taşımaktadır. Bir firmanın müşterilerinin profilini tespit etmesi ve buna uygun satış teknikleri geliştirebilmesi artık veri madenciliği ile mümkün olabilmektedir. Bunun dışında gelecek tahminlerinde bulunabilmek, kestirim yapabilme olanağı da sağlayabilmektedir. Bir banka müşterisinin bankadan mevduatını çekmeyi planladığı gibi derin bilgiler bile, müşteri hareketlerinin anlamlı hale dönüştürülmesiyle elde edilebilmektedir. Bu açıdan veri madenciliği diğer veri çalışma konularına göre daha yaygın kullanım ve araştırma alanına sahiptir diyebiliriz. Bu çalışmaların Web ortamında yapılması ise (Web verileri) Web madenciliği olarak bir alt çalışma grubu oluşturmuştur. Web'teki bilgi beraberinde HTML (Hyper Text Markup Language) işleme gerekliliğini de beraberinde getirmiştir. Bu çalışmaların başarılı olabilmesi için de hızlı ve doğru arama motorları (Web Search Engines) önem

kazanmıştır. Basit olarak arama motorları ise verinin lokasyonuna ulaşmak üzere Web’te kullanılan bir bilgi getirme uygulamasıdır.

Elde edilen bilgi daha sonra değerlendirmek üzere saklanmalıdır ve aynı zamanda Yönetim Bilişim Sistemleri (YBD)’nin [10, 11] içinde bir rapor olarak sunulmalıdır.

Makalenin bundan sonraki kısımları sırasıyla, kuramsal altyapı (veri temin etme yöntemleri ve veri temizleme ve modelleme yöntemleri), uygulama (bilgi çıkarım teknikleri ve kullanımı) ve sonuçlar ve öneriler olarak sıralanmaktadır.

## 2 Kuramsal Altyapı

### Veri Temin Etme Yolları

Veri temin edebilmek için aşağıda listelenen yöntemler kullanılabilir:

1. Şirketin veri tabanı ve dosya sistemi
2. Web ortamında ajan
3. Sosyal ağlar
4. Web servisleri
5. Satın alma

Bunlardan ilki şirketin kendi bünyesinde kullanmış olduğu bilgi sistemleri aracılığı ile veri tabanı (organize) ve dosya sisteminde (organize olmayan) birikmiş olan verileri içermektedir. Bu verilerin detay ve içeriklerinin istenen bilgi çıkarımına yeterli olup olamayacağı sorgulanmalıdır. Gelecekte yeni bilgi sistemlerinin oluşturulması gerekliliğine kadar gidebilecek bir durum oluşabileceği gibi, verinin yeniden organize edilmesi veya modellenmesi ile çözülebilecek durumlarla da karşılaşılabilir. Bu yöntem, genel anlamda tedarik, satışlar, müşteri ve personel yönetimi gibi şirketin günlük faaliyetlerini planlayan, hayati önem taşıyan verileri içermektedir.

Web ortamında ajanlar (Web Agent) [12], genel olarak aynı iş kolunda çalışan diğer rakip firmaların faaliyetleri hakkında veri temin etmeye yönelik çalışmalardır. Ajanlar Web sayfaları, Web logları gibi açık veri kaynaklarını kullanarak istenilen verileri toplayabilen uygulamalardır. Bu sosyal içerikli bir haber olabileceği gibi (ortaklık, yatırım), hangi ürünlere yönlendiği ve bunların satış fiyatları da rekabet açısından önemli olabilecek verilerdir. Bunun dışında şirket faaliyetleri hakkında oluşan olumlu veya olumsuz iletişim verileri (örneğin

bayiler ve alt grupların faaliyetleri) temin edebilmek amacıyla da kullanılabilir. Bu veriler daha çok stratejik, uzun vadeli kararlar alınabilmesi amacıyla üst yönetime sunulacak verileri içermektedir.

Facebook, Twitter, Instagram gibi sosyal ağlardan [13] dünya genelinde paylaşılan veriler temin edilebilir. Bunun yanı sıra şirket uygulamalarına oluşturacağı bir Facebook link sayfası aracılığı ile, şirket potansiyel müşteri ve reklam çalışmaları verileri toplayabilir.

Web servisleri [14] ise çeşitli kurum ve kuruluşlardan tarafından sağlanan açık bilgi kaynaklarıdır. Bu kaynaklardan bazıları ücret karşılığı olabilir. Sağlayıcının sunduğu hizmetlere şirket uygulaması aracılığı ile ulaşarak, veriler kendi ortamınıza alınıp işlenebilir. Bunlar genellikle en basit düzeyde döviz kurları, hava tahmin verileri gibi günlük sunulan bilgilerdir.

Satın alma yönteminde, bazı firmaların hazırlamış olduğu verilerin ücret karşılığı temin edilmesi mümkün olabilmektedir. Bu genellikle yaygın olarak kullanılabilen şahısların telefon ve posta adres bilgileri olabileceği gibi, sosyal ağları işleten büyük yazılım firmaları aldıkları sözleşme onayı çerçevesinde kayıtlı kullanıcılarının her türlü verisini sunabilmektedir.

Bu veri kaynakları tamamen şirketin gelecek beklenti ve stratejilerine uygun olarak kullanılmalıdır. Gereksiz veri kaynaklarından arındırılmış, bilgi çıkarımı yapılması hedeflenen veri, ayrı bir ortama modellenmek üzere taşınmalıdır. Bazen büyük şirketler uygulama gereği birden fazla veri tabanı kullanabilmektedir. Bu durumda bu veri tabanlarında müşterek olan verilerden güncelliği yüksek olanların veri kaynağı olarak seçilmesi hedeflenen doğru bilgiye ulaşım başarımını arttıracaktır.

## 3 Veri Temizleme ve Modelleme Yöntemleri

Bilgi çıkarımı için kullanılacak verinin tespiti aşamasını, verinin gürültüden ayırıştırma aşaması takip etmelidir. Gürültü giderme çalışması hem verinin tutarlılığının sağlanmasını, hem de verinin içinde yer alan veri kayıplarının giderilmesini içerecektir.

Birincisi, özellikle ilişkisel veri yapısı varlığında söz konusudur ve var olmayan ilişki verilerinin tespiti ve yönetilmesini gerektirir. Bu durumda ya tutarsız veriler temizlenmeli, ya da bu verilerin karşılığı yeni

veri kayıtlarının ilişkili veri içinde tanımlanması gereklidir. Bu genellikle tutarlılık işleminin veri tabanında referans anahtarı (foreign key) veya tetikleyici (trigger) aracılığı ile veritabanı ortamında yapılmamasından, tutarlılığın uygulama katmanı sorumluluğuna verilmesinden kaynaklanan bir sorundur.

İkinci durum ise genellikle veri içinde bazı nitelik alanlarının boş bırakılması veya bir metnin içinde sözdizim (sözlükte bulunmayan bir kelime) hatalarından kaynaklandığı söylenebilir. Boş nitelik alanları için yaygın veriler kullanılabilir gibi, ilgili veri kayıtları silinebilir. Bu tamamen veriden elde edilecek bilginin hassasiyetine bağlıdır ve veri analisti bu doğrultuda karar verecektir. Metnin içindeki yanlış yazılan kelimeler için dile ait sözlük öneri sistemleri kullanılabilir.

Modelleme yöntemleri açısından ilişkisel ve metin verilerini farklı ele almak gereklidir. İlişkisel verinin modellenmesi amacıyla ilişkisel veri tabanı kullanılabilir gibi, daha profesyonel çok miktardaki verinin organize edilip amaca yönelik sorgulara hizmet edebilecek veri ambarları (data warehouse) oluşturulabilir. Veri ambarlarının oluşturulması [15] zahmetli ve oldukça zaman alıcı bir çalışmadır. Şirketlerin daha çok uzun vadeli planlama ve yönetsel kararlar için elde etmek istediği bilgilerin sağlanmasına yönelik, raporlamanın hızlı yapılabilmesi veri tabanı tasarımlarıdır. Kendi içinde çeşitli teknik ve yöntemler içermektedir. Genellikle kısmen ilişkisel ve oldukça fazla nitelikten oluşan tablolar içerebilir. Diğer tarafta metin verileri için yaklaşım benzer konulara göre gruplama veya tarih, yazar gibi niteliklere göre tasnifleme stratejileri içerebilir. Odaklanılan bilgi doğrultusunda yeniden düzenlenmeyi gerektirebilir. Örneğin web logları konu bazında ele alınabilecek metinler olabileceği gibi, yazarlarına göre de tasniflenmesi mümkün verilerdir.

İlişkisel veriler üzerinde yukarıda bahsedilen işlemleri gerçekleştirmek üzere çeşitli ETL (Extract-Transfer-Load) araçları vardır (ticari ürünler Ab Initio, Business Objects, IBM, Informatica ve SAS) ve bunların kullanılması bu süreci hızlandıracaktır. Metin kontrol ve düzeltme amacıyla Word türü metin düzenleme araçları içinde yer alan yazım ve gramer denetleme araçları kullanılabilir veya geliştirilecek bir yazılım içinde dil için hazırlanmış kütüphaneler

(örneğin Türkçe için Zemberek) kullanılarak da bu işlem gerçekleştirilebilir.

## 4 Uygulama

### Bilgi Çıkarım Teknikleri ve Kullanımı

Bilgi çıkarımı amacıyla daha önce söz edildiği gibi genel anlamda ya sorgu sistemleri kullanılarak veride bilgi lokasyonunun tespiti, ya veriden özet bilgi, ya da gözle fark edilemeyen bilginin elde edilmesi hedeflenebilir.

Sorgu amaçlı işlemlerde bilgi getirme sistemlerinin kullanılması mümkün olabileceği gibi, doğrudan veri tabanından yapılabilecek sorgular da yeterli olabilir. Bu tür işlemlerin yaygın olarak yapıldığı durumlar için veri tabanında genellikle veri tabanı yordamları geliştirmek ve uygulamak daha verimlidir. Bilgi çıkarım sistemlerinin başarısı genellikle tutarlılık (precision), anımsama (recall) ve anımsama ve tutarlılığın harmonik ortalaması olan F-ölçümü (F-measure) ile belirlenir. Bunun dışında alakasız bilgi getirme olasılığının hesaplandığı dışarıda (fall-out) ölçümünde yaygın olarak kullanılmaktadır. Sorgu amaçlı kullanılacak sistemler ilgili Web sayfasında [16] kategorize olarak listelenmiştir.

Sorgu yoluyla bilgi tespiti şirket içinde YBD' nin bir parçası olarak, üst düzey yöneticilerin bağımsız veri üzerinde sorgu yapabilmeleri ve raporlamalarına yönelik çalışmalar için kullanılmalıdır. Genellikle yöneticiler belirli hedef bilgiye yönelik sorgu işletiminde bulunacaklarından, modelin sınırlarının çizilmesi de mümkün olabilecektir (belirli raporların alınması). Veriyi etkin kullanabilen bir yönetici için ise esnek sorgulama alt yapısının sağlanması mutlaka bir artı olarak şirkete geri dönecektir.

Diğer bir bilgi çıkarım yöntemi olan metinlerden özet elde etmede, kullanıcının tercihine göre mevcut metinlerin toplam uzunluklarının belirli bir yüzdesi veya kesin olarak belirlenecek kelime sayısı kadar bir özet oluşturabilmek mümkündür. Bu alanda DARPA (Defense Advanced Research Projects Agency) tarafından uzun zamandır geliştirilen TIPSTER uygulaması tam olarak ticari olmamakla beraber en geniş kapsamlı uygulamadır. Bunun dışında soru-cevaba dayalı Amazon ve Google tarafından geliştirilen sırasıyla Amazon's Mechanical Turk [17] ve Google Search [18] uygulamaları kullanılabilir.

Metin özetleme daha çok bir haber veya konuya ilişkin kısa ve özlü bir yazı elde etmek üzere kullanılabilir. Habercilik alanı dışında bugün arama motorları başta olmak üzere birçok uygulama içinde özetleme artı olarak sunulmaya çalışılmaktadır. Metin özetleme, şirket içinde özellikle Web logları ve CRM (Customer Relationship Management) verileri üzerinde uygulanabilirliği olan bir uygulamadır. Bu bilgiler YBD raporları olarak orta düzey ve üst düzey yönetiminin bilgisine sunulmalıdır. İleriye dönük stratejik kararların (bayi konumlama, bayi açma veya kapatma, tedarikçi yönetimi, v.b.) alınması açısından önem taşıyacaktır.

Son bilgi çıkarım yöntemi, en gizemli olanıdır. Çünkü gözle fark edilemeyen gizli bilginin ve ilişkilerin bulunabilmesine olanak sağlar. Veri madenciliği veya Web madenciliği bu anlamda şirketler için bulunmaz bir fırsattır. Bu yöntem içinde çeşitli algoritmalar barındırmaktadır. Bu algoritmalar elde edilecek bilgiye yönelik özel kullanıma sahiptirler. Örneğin Apriori algoritması kullanılarak en sık kayıtlar (kelime, nesne) bulunabilir ve aralarındaki ilişkiler tespit edilebilir. İlişkiler örneğin alışveriş ortamlarında hangi ürünlerin yanyana konulmasının satışları arttıracığı gibi gözle görülemeyen önemli bir bilgi sunmaktadır. Bir başka yaygın olarak kullanılan K-Means algoritması ise, veri içinde yer alan sınıfların tespiti ve ayrıştırılmasını gerçekleştirir.

Veri madenciliği algoritmaları bugün veri tabanlarının bir uygulaması olarak sunulmaktadır. Örneğin Oracle Advanced Analytics SQL API ve MSSQL Server Sql Server Analysis Services Data Mining desteği sunmaktadır. Bunun dışında ticari olarak uygulamalara ilişkin olarak ilgili Web sayfasında [19] listelenmiştir. En önemlisi büyük çalışmalarda veri madenciliği uygulaması kullanılmadan önce verinin organize olarak veri ambarında istenilen modele uygun düzenlenmesi gereklidir.

Veri madenciliği şirketlerin beklentilerine yön verecek, üst yönetimin planlama ve stratejik kararlarına destek olacak uygulamaların geliştirilmesine olanak vermektedir. B2C'dan B2B'a geçen her şirket için artık CRM ne kadar önemini yitirirse, veri madenciliği ve Web madenciliği yapmak da o kadar önemli olmaya başlayacaktır. Bunu öngörüp önlem alan ve bilgi işlem yapısını düzenleyen şirketlerin gelecek kaygıları da o kadar azalacaktır.

## 5 Sonuçlar ve Öneriler

Günümüzde veri değerli bir araçtır ve bu yoğun veriden güncel ve doğru bilgiye ulaşmak ise ancak bilgi teknolojileri kullanarak oluşturulacak sistemlerle mümkün olabilir. Şirketler, bilgisayar teknolojilerini sadece günlük ticari faaliyetlerini sürdürmek için kullanma amacından öte, teknolojik gelişmelerin ışığında gelecek yatırım ve kararlarına yön tutacak bilgiye erişim amacıyla kullanmayı da başarabilmelidirler. Özellikle global düzeyde büyüyen veya büyümeyi hedefleyen şirketler mutlaka bütçelerinin yeterli bir kısmını bu yatırımlara ayırmalıdır. Bu yatırımlar, yeni yazılım maliyetlerinin yanısıra, bilgi çıkarımı ve bu teknolojilerin kullanımı alanında uzmanlaşmış personel istihdamını veya bu desteği alabileceği uzmanlaşmış kurumlarla işbirliğini içermelidir.

Uygulanan bilgi teknoloji projelerinin mutlaka değerlendirilmesi gereklidir. Değerlendirme süreci çalışanların geri bildirimleri şeklinde olabileceği gibi, daha rasyonel yaklaşımlarla işteki iyileşme (süre, maliyet, kazanç gibi parametrelerle) ölçülmelidir.

Veri kalitesi, bilgi çıkarım sürecini doğrudan etkileyen bir unsurdur. Veri kalitesinin de ölçülmesi, daha iyi verinin nasıl temin edilebileceği, verinin daha iyi nasıl modellenilebileceği sürekli araştırılmalıdır. Elde edilen bilginin ardışıl zaman sürelerinde varyansının çok büyük olması (sektörel bir eşik değeri tespitiyle), bilginin üretilme süresinin tahminden daha uzun sürmesi (raporlama süresi) yukarıda bahsedilen çalışmaların yapılmasını gösteren basit delillerdendir.

Çok miktarda veri üreten şirketlerin bu veriyi depolaması ayrı bir sorundur. Bugün gelinen noktada, bulut bilişim (Cloud Computing) yaklaşımı veri depolama ve bilgi paylaşımı sorunları için şirketlere oldukça faydalı bir yaklaşım sunmaktadır.

Sonuç olarak, bilgi çıkarma yatırımları şirketin maliyetlerini ve riskini azalttığı gibi, ticari faaliyetlerine de çeviklik getirecek ve başarıyı arttıracaktır.

## 6 Kaynakça

- [1] Manning, C. D., Raghavan, P., Schütze, H. *Introduction to Information Retrieval*, 2008, Cambridge University Press.
- [2] Capron, H. L., Johnson, J.A. *Computers: Tools for an Information Age*, 2001, Pearson/Prentice Hall.
- [3] Kang, B.-Y., Kim, D.-W., Kim, H.-J. *Fuzzy Information Retrieval Indexed by Concept Identification*, 2005, Springer.
- [4] Radev, D.R., Hovy, E., McKeown, K. Introduction to the Special Issue on Summarization, *J. Comput. Linguistics*, 2002, pp. 399-408.
- [5] Yatsko, V. A., Vishnyakov T. N. A method for evaluating modern systems of automatic text summarization, *In: Automatic Documentation and Mathematical Linguistic*, 2007, pp. 93-103.
- [6] Huang, A. Similarity Measures for Text Document Clustering, *New Zealand Computer Science Research Student Conference*, 2008, pp. 49-56.
- [7] Jiawei, H., Kamber, M. *Data mining: concepts and techniques*, 2011, Morgan Kaufmann.
- [8] Tan, P.-N., Steinbach, M., Kumar, V. *Introduction to Data Mining*, 2006, Pearson.
- [9] Markov, Z., Larose, D. T. *Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage*, 2007, Wiley.
- [10] O'Brien, J., Marakas, G. *Management Information Systems*, 2011, McGraw-Hill.
- [11] Chu, M. Y. *Blissful Data: Wisdom and Strategies for Providing Meaningful, Useful, and Accessible Data for All Employees*, 2004, American Management Association.
- [12] Kumar, S. *Agent-based Semantic Web Service Composition*, 2012, Springer.
- [13] Easley, D., Kleinberg, J. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*, 2010, Cambridge University Press.
- [14] Richardson, L., Ruby, S. *Restful Web Services*, 2007, O'Reily Media.
- [15] Kimball, R, Ross, M.. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 2013, Wiley.
- [16] <http://bit.csc.lsu.edu/~kraft/retrieval.html>, [24.12.2014].
- [17] <http://aws.amazon.com/mturk/>, [24.12.2014].
- [18] <https://www.google.com.tr/cse/>, [24.12.2014].
- [19] [http://www.dmoz.org/Computers/Software/Databases/Data\\_Mining/Tool\\_Vendors/](http://www.dmoz.org/Computers/Software/Databases/Data_Mining/Tool_Vendors/), [24.12.2014].