YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İNTERNET SİTELERİNDEN DERLENEN ADRES BİLGİLERİNİN COĞRAFİ KODLANARAK HARİTA ÜZERİNDE GÖSTERİMİ

Bilgisayar Mühendisi Abdullah Kutlu ERSOY

F.B.E. Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Programında Hazırlanan

YÜKSEK LİSANS TEZİ

Tez Danışmanı: Yrd.Doç.Dr. Banu DİRİ

İSTANBUL, 2010

İÇİNDEKİLER

SİMGE LİSTESİ

β	Geri Çağırma / Hassasiyet Oranı
E	Van Rijsbergen'in etkinlik ölçüsü
Fβ	Geri çağırmaya hassasiyetin β (beta) katı kadar önem atfeden kullanıcı açısından erişimin etkinliği
N	Erişilmiş sayı
P(r)	Belli bir kesme (cut-off) derecesinin hassasiyeti
P	Hassasiyet (precision)
R	Geri çağırma (recall)
r	Rank (derece)
rel()	Belli bir derecenin ilgisinin ikili (binary) bir fonksiyonu

KISALTMA LİSTESİ

CARIS	Bilgisayar Destekli Kaynak Bilgi Sistemi (Computer Aided Resource Information System)
CBS	Coğrafi Bilgi Sistemi (Geographic Information System)
CGIS	Kanada Coğrafi Bilgi Sistemi (Canadian Geographic Information System)
CRF	Şartlı Rastgele Alanlar (Conditional Random Fields)
ESRI	Çevre Sistemleri Araştırma Enstitüsü (Environmental Systems Research Institute)
HTTP	Hiper Metin Transferi Protokolü (Hyper Text Transfer Protocol)
IIMtCAR	İnternet İçeriğinde Metin tabanlı Coğrafi Arama Robotu
IE	Bilgi Çıkarımı (Information Extraction)
IR	Bilgiye Erişim (Information Retrieval)
İBB	İstanbul Büyükşehir Belediyesi
LCGSA	Bilgisayar Grafikleri ve Mekansal Analiz Laboratuarı (Laboratory for Computer Graphics and Spatial Analysis)
MOSS	Harita Yerleşimi ve İstatistiksel Sistem Projesi (The Map Overlay and Statistical System Project)
MUC	Mesaj Anlama Konferansları (Message Understanding Conferences)
NIST	National Institute of Standards and Technology
NLP	Doğal Dil İşleme (Natural Language Processing)
OGC	Açık Coğrafi Konsorsiyum (Open Geospatial Consortium)
TREC	Metin Erişimi Konferansı (Text Retrieval Conference)
USA-CERL	Amerikan Ordusu Mühendisler Birliği Araştırma Laboratuvarı (U.S. Army Corps of Engineering Research Laboratory)
W3C	Dünya Çapında Ağ Birliği (World Wide Web Consorsium)
WELUT	Batı Enerji ve Arazi Kullanımı Ekibi (Western Energy and Land Use Team)
XML	Genişletilebilir İşaretleme Dili (Extended Markup Language)

ŞEKİL LİSTESİ

ÇİZELGE LİSTESİ

ÖNSÖZ

Yüksek lisans tez danışmanlığımı üstlenerek, çalışmalarım esnasında yol göstermede ve bilgi vermede her zaman desteğini gördüğüm değerli hocam Yrd. Doç. Dr. Banu DİRİ' ye teşekkür ederim.

Ayrıca eğitim hayatım boyunca desteğini hiçbir zaman eksik etmeyen aileme sonsuz teşekkürler sunarım.

A. Kutlu ERSOY

İstanbul, 2010

ÖZET

Günümüzde İnternetin, içerdiği bilginin çeşitliliği ve boyutu göz önüne alındığında yapılan her türlü araştırmada en önemli kaynağı oluşturduğu söylenebilir. Veri hacminin bu denli büyük olması kullanıcıların alışkanlıklarını da etkileyerek, İnternet siteleri üzerinde arama işlemlerinde yeni yaklaşımların oluşturulmasını gerekli kılmaktadır. Adres bilgisine ulaşmaya odaklı yapılan İnternet sorgularında kullanıcının, işlenmiş bilgiye daha hızlı bir şekilde ulaşma ihtiyacının arttığı görülmektedir.

Çalışmada tanımlanan problem "internet içeriğinde coğrafi koordinat özniteliği bulunmayan fakat metin olarak adres bilgisi mevcut olan internetteki arama sonuçlarının kullanıcı tarafından görsel olarak algılanabilmesi" olarak özetlenebilir. Bu ihtiyaç, insanın adres içeren veriye ulaşarak onu görselleştirmesinin, ve bu işlemi birkaç kez tekrarlamak sureti ile aradığı neticeye ulaşmasının zaman ve odak kaybına yol açabilmesinden kaynaklanmaktadır. İzlenen bu adımlarla dahi bilgiye bir bütün olarak bakmak/değerlendirmek çoğu zaman mümkün olamamaktadır.

Tezde ortaya konan çalışma ile, İnternet kullanıcısının mevcut metin-tabanlı arama motorlarını kullanarak adres bilgisine ulaşma sürecini kolaylaştırması sağlanmak istenmiştir. Bilgisayar yazılımları sayesinde, arama işlemlerinde atılacak adımların otomatik gerçekleştirilerek sonuçların coğrafi koordinatları ile harita üzerinde sunumu hedeflenmektedir. Sistemin başarısı, çözümleyebildiği adres cümlesi sayısı ve arama yapılan konu, mekan, vb. adres bilgisinin coğrafi koordinatının harita üzerinde doğru noktada gösterilmesi ile ölçülecektir.

Anahtar Kelimeler : Bilgi Erişimi, Bilgi Çıkarımı, Adrese Dayalı Arama, Coğrafi Kodlama, Coğrafi Bilgi Sistemi (CBS)

ABSTRACT

Today, the Internet can be considered as the most important source of any research when taken into account the variety and the size of information it provides. Containing large volumes of data requires the formation of new approaches to the searches on the internet sites, affecting the users' habits. It is observed that there is an increasing need to access to processed information more quickly when address focused searches are done.

The problem defined in the study can be summarized as "visual perception of the results of available address searches which has no geographic coordinates, but has address information as text instead". This need arises when accessing and then visualising the result -as well as repeating this process several times- cause loss of time and focus. Even with these steps, most of the time, it is not possible to look into or assess the information as a whole.

The study put forward in the thesis aims to make it easier for the internet user to facilitate the process of reaching the address information by using the internet's existing text-based search engines. By the help of computer software, it is targeted to have the presentation of the search results on the map with their geographic coordinates, by performing the steps automatically which are necessary in a search process. The success of the system is measured by the number of the address sentence that can be resolved, and by displaying the geographic coordinates of searched addresses in the right spot on the map.

Keywords : Information Retrieval, Information Extraction, Address Based Search, Geocoding, Geographic Information System (GIS)

1. GİRİŞ

Günümüzde İnternet kullanıcılarının en sık başvurduğu işlem, arama motorlarında ilgilendikleri konu hakkında sorgu oluşturmaları ve araştırma yapmalarıdır. Bunun sonucunda İnternet üzerinde arama yapan siteler, İnternetin kullanılabilirliği ve İnternet üzerindeki bilgiye erişim hususunda büyük önem kazanmıştır.

Coğrafi Bilgi Sisteminin gelişimi ve haritaların dijital ortama taşınması ile beraber, arama motorları bu yeni alana yönelmişlerdir. Arama motorları, İnternet kullanıcılarının belirledikleri anahtar kelimeler ile yaptıkları aramalarda, bulunan sonuçların metin olarak gösterilmesinin yanı sıra harita üzerinde de adreslerini gösterebilecek şekilde özelleştirilmiştir.

Harita uygulamalarının günlük hayata girmesi ve yaygınlaşması ile ticari kullanım alanları oluşmuştur. İşletmeler kendi coğrafi konumlarını ve iş alanı ile ilgili anahtar kelimeleri arama motorlarının sağladığı altyapı içine kayıt ettirmişlerdir. Teknolojiye adapte olmakta yavaş kalan küçük ve orta ölçekli işletmeler, harita-tabanlı arama sitelerine konumlarını kayıt ettirmediklerinden dolayı kendi internet sitesinde veya çalıştığı iş alanındaki portal sitelerde adres bilgilerini, sadece metin olarak yayınlamaktadırlar.

Google Maps gibi Harita-Tabanlı Arama Motorları, coğrafi koordinatların ve anahtar kelimelerin kayıt altında olduğu bir veritabanından sorgulama sonuçlarını getirebilmektedir. Bu veritabanının içindeki kayıtların bir kullanıcı veya site yöneticisi tarafından önceden girilmiş olması, sorgu sonucunda gözükmesi için gereklidir. Arama motorları kullanıcının yaptığı sorgunun sonuçlarını, anahtar kelimelerin konum bilgisi ile beraber kayıt edilmesi sayesinde, harita üzerinde gösterebilmektedir. Bu tez çalışmasında ise, metin-tabanlı arama yapan motorların çıkardığı sonuç kümesindeki web sitelerinin içeriğinden adres bilgisi bulunarak, bilginin kullanıcıya harita üzerinde gösterilmesi hedeflenmektedir.

Bu çalışmada, web içeriğinde arama yapıldığında elde edilen yazılı adres metinlerinin, daha kolay anlaşılır olan görsel bilgiye çevrilerek kullanıcıya sunulması gerçekleştirilmektedir. Bu konuda, mevcut arama olanaklarını tamamlayıcı bir katkı sunulması ve adres bilgisi coğrafi koordinata çevrilmemiş internet içeriğinin kullanıcıya görsel olarak sunumu sağlanmıştır.

1.1. Problemin Tanımı

İnternet üzerinde yapılan bir arama ile sonuç kümesinden adreslere ulaşılmak istendiği takdirde, kullanıcının arama sonucunda gelen bağlantılardaki her bir siteye giriş yapması ve sitenin içeriğindeki adres cümlelerinin bulunarak harita üzerinde konumunun bulunması gerekmektedir. Bu işlem kullanıcı tarafından yapıldığında aynı adımların tekrarlanması ve her bir adres cümlesinin harita üzerinde yalnız başına görünmesine sebep olmaktadır. O halde problem, kullanıcının adrese ulaşmak için yaptığı aramalarda sonuçları bir bütün olarak görememesidir.

Çalışmanın Amacı

Adrese ulaşılmak istenilen aramalarda sonuçların birden fazla siteden derlenmesi ve ulaşılan tüm adres konumlarının tek bir harita üzerinde incelenmesi bu çalışmanın çıkış noktasını oluşturmaktadır.

İnternet içeriğinde bulunan bir verinin sorgulanması sonucu, oluşan sonuç kümesinin toplu olarak değerlendirilebilmesi ve görsel olarak sunulması, bilginin kullanışlılığını arttırmaktadır. Bu tez çalışması ile, acil bir durumda gerekli tıbbi yardımın alınabileceği

adreslerin harita üzerinden hızlı bir şekilde erişilebilmesi, daha önce gidilmemiş bir yere geç olmadan ulaşılması gibi acil durumlarda, İnternet kullanıcılarının arama sonuçlarını okumak ve adresleri harita üzerinde tek tek sorgulamak yerine, görerek karar vermesi sağlanabilmektedir.

Ayrıca küçük ve orta ölçekli işletmelerin de adres bilgileri, kendi sitelerinde veya başka herhangi bir İnternet sitesinde yer alabilmekte, fakat coğrafi koordinatları olarak harita sunucularına kaydettirilmemiş olabilmektedirler. Bu çalışma sayesinde, İnternet kullanıcısı tarafından bahsi geçen nitelikte bir firmayı içerecek metin-tabanlı arama yapıldığında, firmanın harita üzerinde konumu ile gösterilmesi sağlanabilmektedir.

Geliştirilecek yazılım ile kullanıcının birden fazla adımda ulaşabileceği veriyi, ve bu adımları birkaç defa tekrarlayarak ancak derleyebileceği sonuç kümesini, kullanıcı yerine İnternet robotlarının hazırlaması ve uygun şekilde sunması hedeflenmiştir.

1.2. Uygulama Yöntemi

Bu çalışmayı gerçekleştirmek için İnternet robotlarının web sitelerinden adres bilgilerini toplarken kullanacakları karar verme algoritmaları belirlenmiştir. Harita üzerinde sunum yapılabilmesi için yazı formatındaki adres bilgisinin coğrafi koordinatları tespit edilmektedir.

İnternet içeriğinin taranması ve arama sonuçlarına ulaşılması için yeni bir arama motoru yazılmamıştır. Mevcut arama motorlarından Google kullanılacak şekilde kodlar hazırlanmıştır.

Uygulama 4 aşamada gerçekleşmektedir:

Arama (Search&Information Retrieval): Tez kapsamında geliştirilecek olan web arayüzünde bir arama kutucuğu bulunmaktadır. Arama gerçekleştirildiğinde karşımıza gelecek olan site bağlantıları, Google/Yahoo gibi arama motorları kullanılarak sağlanacaktır. Sorgu sonucunda bağlantıları sonuç kümesinde yer alan sitelerin içeriklerinden adres bilgisi yer alan sayfalara ve kısımlara ulaşılması gerekmektedir.

Bilgi Çıkarımı (Information Extraction): İnternet sitelerindeki içeriğinde adres bilgisi içerebilecek kısımlara ulaşıldıktan sonra elde edilen metinler ayıklanacaktır. Adres çıkarımı için Türkiye'nin il, ilçe ve mahallelerinin yer aldığı veritabanı kullanılacaktır. Ayrıca cadde, sokak gibi kelimeler şablon olarak kullanılarak adres ayıklaması gerçekleştirilecektir.

Coğrafi Koordinatların İşaretlenmesi (Geocoding): Ayıklanan adres cümleleri Coğrafi Kodlama fonksiyonları kullanılarak koordinat bilgisine çevrilecektir.

Sonuçların Kullanıcıya Gösterilmesi: Coğrafi koordinatları tespit edilen adreslerin harita üzerinde noktalar olarak gösterilmesi için bir internet sayfası hazırlanacaktır. Harita üzerindeki noktaların üzerlerine tıklandığında açılacak bir kutucuk içerisinde o nokta ile ilişkilendirilen sitenin bağlantı adresi olacaktır. Bu adrese tıklandığında ise o noktaya ait internet sitesi kullanıcının ekranında açılacaktır.

2. COĞRAFİ BİLGİ SİSTEMİ

Coğrafi objelere ait grafik ve grafik olmayan verilerin elde edilmesi, depolanması, işlenmesi, yönetimi, analiz edilmesi, sorgulaması ve sunulması fonksiyonlarını yerine getirmek için oluşturulan sisteme Coğrafi Bilgi Sistemi denilmektedir. CBS sistemi kullanıcılar, donanım ve harita modülü ve veritabanı modülü içeren yazılım ve yöntemlerin bir araya gelmesi ile oluşmaktadır. CBS, dünya üzerindeki sosyal, ekonomik ve çevresel sorunlar ve bunların çözümüne yönelik coğrafi verilerin yönetimi ve analizi ile ilgilenen bir sistemdir.

CBS'ni kullanacak olan ekip, veritabanı yöneticileri ve CBS kullanıcılarından oluşmaktadır. Veritabanı yöneticileri, coğrafi verinin gereksiz tekrarlamalardan arındırılmış, farklı uygulamalarda tekrar kullanılabilen ve ilişkilendirilmiş şekilde organize edilmesini organize eden taraftır. Üretilen veritabanına CBS uygulamalarını kullanarak veri girişini yapan, güncelliğini sağlayan ve bu verileri işleyen taraf ise CBS kullanıcılarıdır.

Harita hizmeti veren Google, Yahoo ve Microsoft gibi firmalarının CBS altlığında kullanılacak veritabanını oluşturdukları, daha sonra da kendi personelleri ve anlaşmalı firmalar ile cadde sokak gibi mahalli sınırları veritabanlarına bilgi olarak kaydettikleri görülmektedir. Ayrıca uydu görüntülerini de bir katman olarak harita uygulamalarında sunabilmektedirler.

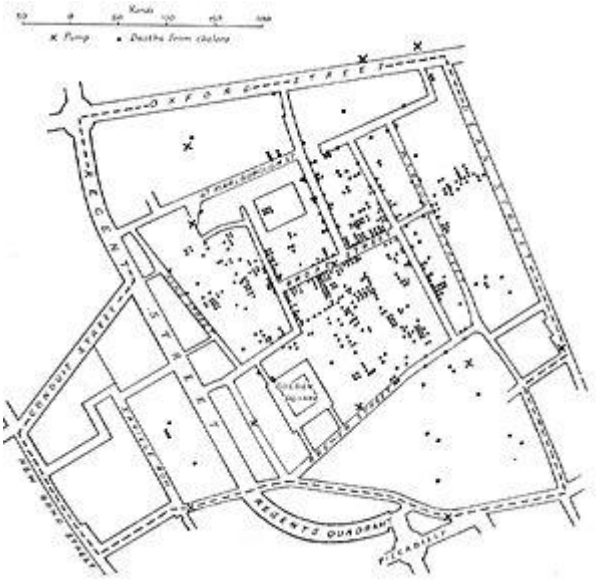
Aynı firmaların harita üzerinde arama işlemine olanak sağlayan web sayfalarında ise sorgulamalar mahalli birimler üzerinden yapılabildiği gibi, internet kullanıcılarının girdikleri bilgiler arasında da gerçekleştirilebilmektedir. İnternet kullanıcılarına bilgi girişi için hazırlanan sayfada coğrafi koordinatın harita üzerinde işaretlenmesi gerektiğinden dolayı, arama ekranlarında ilgili anahtar sözcüklerin geçtiği kayıtlar coğrafi olarak konumlandırılmış olmaktadır.

CBS, bilgisayar teknolojisinin ilerlemesi ile hız kazanmış ve dünyada çok çeşitli uygulama alanları bulmuştur. CBS'nin çok çeşitli uygulama alanlarının oluşmasında, CBS'nin kullanıcı dostu yazılımlar ile desteklenmesi ve coğrafi verinin görsel sunumunun yapılabilmesi etkili olmuştur.

2.1. Tarihsel Gelişim

John Snow 1854'te, coğrafi metodun belki de ilk kullanımı olarak, bazı bireysel vakaların yerlerini noktalarla gösterdiği Londra'da bir kolera salgınını ele almıştır (Stamp, 2006). Koleranın dağılımı üzerine yaptığı çalışma, onu hastalığın kaynağı olan bir kirli su pompasına

(kolun bağlantısını keserek salgını durdurduğu Broad Street Pump) götürmüştür. Şekil 2.1 1854'te Londra'daki Soho kolera patlamasında görülen vakaların kümelerini göstermektedir.



Şekil 2.1 1855'te John Snow'un çizdiği haritanın E.W. Gilbert versiyonu (1958)

Her ne kadar topografyanın temel unsurları ve teması daha önce kartografyada mevcut olsa da, John Snow'un coğrafi olarak bağımlı olaylar kümelerinin sadece tasviri için değil, ayrıca analizi için de kartografik metotların ilk kez kullanıldığı haritası benzersizdi.

20. yüzyılın ilk yılları, haritaların katmanlara bölündüğü fotolitografinin gelişimine sahne olmuştur. Nükleer silah araştırmalarının teşvik ettiği bilgisayar donanımı gelişmeleri, 1960'ların başlarında genel-amaçlı "harita yapımı" uygulamalarına yol açmıştır (Fitzgerald, 2007).

Dünyada gerçekten çalışan ilk CBS, 1962 yılında Kanada Ontario / Ottawa'da Federal Ormancılık ve Kırsal Kalkınma Bakanlığı (Federal Department of Forestry and Rural Development) tarafından geliştirilmiştir. Dr. Roger Tomlinson tarafından geliştirilen sisteme Kanada Coğrafi Bilgi Sistemi denmiştir ve topraklar, ziraat, rekreasyon, yaban hayatı, su akışı, ormancılık, ve arazi kullanımı hakkındaki bilgileri 1:50.000 ölçeğinde haritalandırarak kırsal Kanada'nın arazi kabiliyetini tespit etme çabasıyla Kanada Arazi Envanteri (Canada Land Inventory) için toplanan verilerin saklanması, incelenmesi, ve işlenmesi için kullanılmıştır. Analize imkan vermesi için derecelendirerek sınıflandırma faktörü de ilave edilmiştir.

CGIS, dünyadaki bu türden ilk sistemdir ve yerleşim, ölçüm, ve sayısallaştırma/tarama yetenekleri sağladığı için harita uygulamalarında bir ilerleme olarak kabul edilmektedir.

CGIS, tüm kıtaya yayılmış ve hatları gerçek bir gömülü topolojiye sahip "yaylar" olarak kodlayan bir ulusal koordinat sistemini desteklemiştir, ve öznitelikler ve konum bilgilerini ayrı dosyalarda tutmuştur. Bunun bir sonucu olarak Tomlinson, özellikle de yakınsak coğrafi verilerin mekansal analizini teşvikte uygulama krokileri kullandığı için "GIS'in babası" olarak tanınır (Tomlinson, 2007). CGIS, 1990'lara kadar sürmüştür ve Kanada'daki en geniş dijital arazi kaynak veritabanını oluşturmuştur. CGIS, federal ve il kaynak planlaması ve yönetimine destek veren sunucu-tabanlı bir sistem olarak geliştirilmiştir. Gücünü, kompleks veri setlerinin kıta çapında analizinden alır. CGIS'in ticari formu hiç olmamıştır.

Howard T. Fisher, Harvard Tasarım Enstitüsü'nde (Harvard Graduate School of Design) mekansal veri işleme alanındaki bir çok önemli teorik kavramların geliştirildiği ve 1970'lere gelindiğinde yeni ufuklar açan yazılım kodu ve dağıtık sistem dağıtmış olan Bilgisayar Grafikleri ve Mekansal Analiz Laboratuarı'nı (LCGSA 1965-1991) 1964 yılında kurmuştur (Lovison, 2007).

1980'lerin başlarında, adı sonradan Intergraph olan M&S Computing, Çevre Sistemleri Araştırma Enstitüsü (ESRI), Bilgisayar Destekli Kaynak Bilgi Sistemi (CARIS) ve ERDAS, CBS yazılımının ticari satıcısı olarak ortaya çıkmışlardır. Bunların ortaya çıkışı pek çok CGIS özelliğini başarıyla birleştirerek, mekansal ve öznitelik bilgilerini ayırma konusuna birinci nesil yaklaşımla öznitelik verilerinin veritabanı olarak düzenlenmesine ikinci nesil bir yaklaşımı birleştirerek gerçekleşmiştir. Buna paralel olarak, iki kamu sisteminin geliştirilmesi 1970'lerin sonlarında ve 1980'lerin başlarında başlamıştır.

Harita Yerleşimi ve İstatistiksel Sistem projesi (MOSS) 1977'de Batı Enerji ve Arazi Kullanımı Ekibi'nin (WELUT) ve ABD Balık ve Yaban Hayatı Servisi'nin himayesinde Fort Collins / Colorado'da başlamıştır. GRASS GIS, Amerikan ordusunun arazi yönetimi ve çevresel planlama yazılımı ihtiyacı için, Amerikan Ordusu Mühendisleri Birliği'nin bir kolu olan USA-CERL tarafından 1982'de Champaign / Illinois'de başlatılmıştır. Sonrasında 1980'lerdeki ve 1990'lardaki sanayi büyümesi, CBS ve Unix iş istasyonlarının ve kişisel bilgisayarların kullanımının artmasıyla gelmiştir. 20. Yüzyılın sonlarında, çeşitli sistemlerdeki hızlı büyüme az sayıda platformda sağlamlaştırılmış ve standardize edilmiştir, ve kullanıcılar internet üzerinden CBS verilerini görme, veri formatlamanın ve transfer standartlarının gerekmesi konseptini paylaşmışlardı. Yakın zamanda, bir dizi işletim sistemi üzerinde çalışabilen ücretsiz açık kaynak kodlu CBS paketlerinin sayısı artacaktır ve belirli işleri yapmak için özelleştirilebilir hale gelmektedirler.

2.2. Veri Modelleri

CBS, coğrafik ilişkisel veri modeline dayanır. İçerdikleri veriler sahip oldukları etkin veri tabanlarına ve yapılarına göre iki ana gurupta değerlendirilir.

2.2.1. Mekansal (Spatial) Veriler

Mekansal veriler yapılarına göre ikiye ayrılırlar:

Vektör Tabanlı Veriler

CBS'lerde vektör olarak tanımlanan veri, herhangi bir koordinat sisteminde nokta, çizgi veya alanlardan oluşan veridir. İdeal olan, tüm grafik verilerin tek bir koordinat sisteminde tanımlanmış olmasıdır. Farklı koordinat sistemlerindeki verilerin ise, tek bir koordinat sistemine dönüştürüleceği dönüşüm katsayılarının yeterli duyarlıkta belirlenmiş olması gerekir. Vektör olarak ele alınan gösterimde ve depolamada başlangıç ve bitiş noktaları ile tanımlı olan ve devamlılık gösteren çizgiler kullanılır. Vektörler, çalışma ortamını nokta, çizgi ve alan gibi topolojik özellikler takımına böler. Çizgilerin başlangıç ve bitiş noktalarının konumu, verinin topolojisini gösteren vektörleri tanımlar. Vektör CBS'ler verilerin yapısının gösterimini gerçeğe en yakın yapar, veri yapısı komplekstir, grafik yapısı hassas ve doğrudur, grafik ve niteliklerin güncellenmesi ve bilgiye erişimi oldukça kolaydır. Diğer taraftan veri yapısının karmaşık olması, vektör poligonları ile raster poligonların çakıştırılmasında güçlüklerin çıkması, renkli tarama ve çizim işlemleriyle özel yazılım ve donanım gerektirdiğinden teknolojisinin pahalı olması dezavantaj oluşturmaktadır.

Raster Tabanlı Veriler

CBS'lerde raster olarak tanımlanan veri, belirli sayısal, harf veya renk olarak değerleri olan hücrelerin (piksel) bir araya gelmesiyle oluşan görsel bilgiyi kapsamaktadır. Hücre, noktalardan oluşur. Hücre içinde kalan her noktanın kod değeri aynıdır. Rasterde çalışma alanı sıralı olarak tanımlanmış düzenli hücreler takımına bölünür. Her türlü topoloji bu hücrelerle tanımlanır. Vektöre nazaran veri yapısı basittir, haritalanmış veri ile uzaktan algılama ile elde edilen verinin çakıştırılması kolaydır. Boyutsal analiz imkanı daha fazladır ve teknolojisi ucuzdur. Diğer taraftan, veri yapısı çok hacimlidir. Veri hacmini küçültmek için büyük hücre kullanımı (çözünürlüğün düşürülmesi) bilgi kayıplarına neden olur. Harita olarak gösterimi hoş değildir. Projeksiyon dönüşümü güçtür.

2.2.2. Sözel (Non-spatial) Veriler

Sözel verilerde coğrafi nesnelerin nitelik bilgileri depolanır. Vektör nesnelerle aralarında ilişkisel bir bağ vardır. Genel olarak veri boyutunun büyük ve çeşitli olduğu projelerde tercih edilir.

2.3. Coğrafi Bilgi Sisteminde Temel İşlevler

Coğrafi bilgi sistemlerinin sağlıklı bir şekilde çalışması aşağıdaki 4 temel işlevlerin yerine getirilmesine bağlıdır. Bunlar;

Veri Toplama :

Coğrafik veriler toplanarak, CBS'de kullanılmadan önce mutlaka sayısal yani dijital formata dönüştürülmelidir. Verilerin kağıt ya da harita ortamından bilgisayar ortamına dönüştürülmesi işlemi sayısallaştırma (digitizing) olarak bilinir. Modern CBS teknolojisinde bu tür işlemler büyük boyutlu projelerde tarama tekniği kullanılarak otomatik araçlarla gerçekleşir. Küçük boyutlu projelerde daha çok masa tipi sayısallaştırıcılar kullanılarak elle sayısallaştırma yapılabilir. Bugün birçok coğrafik veri CBS'ne uyumlu formatta hazır halde piyasada mevcuttur. Bunlar üretici firmalardan sağlanarak doğrudan kurulacak sisteme aktarılabilir.

Veri Yönetimi :

Küçük boyutlu CBS projelerinde coğrafi bilgilerin sınırlı boyuttaki basit dosyalarda saklanması mümkündür. Ancak, veri hacimlerinin geniş ve kapsamlı olması, bunun yanında birden çok veri gruplarının kullanılması durumunda Veritabanı Yönetim Sistemleri (Data Base Management Systems) verilerin depolanması, organize edilmesi ve yönetilmesine yardımcı olur. Veritabanı yönetim sistemleri bir bilgisayar yazılımı olup veri tabanlarını yönetir veya birleştirir. Bir çok yapıda tasarlanmış veritabanı yönetim sistemi vardır, ancak CBS için en popüler olanı ilişkisel (relational) veritabanı sistemidir. Bu sistem tasarımında veriler tablo bilgilerinin elde edilişindeki düşünce yapısına uygun olarak bilgisayar belleğinde saklanır. Farklı bilgiler içeren tabloların birbiriyle ilişkilendirilmesinde bu tablolardaki ortak sütunlar kullanılır. Bu yaklaşım hem basit hem de esnek bir tasarım olup, geniş çapta CBS uygulamalarında kullanılmaktadır.

Veri İşlem :

Bazı durumlarda özel CBS projeleri için veri çeşitlerinin birbirine dönüşümü veya irdelenmesi istenebilir. Verilerin sisteme uyumlu olması bunu gerektirebilir. Örneğin, konumsal bilgiler

farklı ölçeklerde mevcut olabilir (yol verileri 1/100.000, nüfus dağılım verileri 1/10.000, bina verileri 1/1.000 gibi). Tüm bu bilgilerin birleştirilmeden önce aynı ölçeğe dönüştürülmesi gerekebilir. Bu dönüşüm görüntü amacıyla geçici olabileceği gibi bir analiz işlemi için sürekli ve kalıcı da olabilir. CBS, gerek bilgisayar ortamında obje üzerine imlecin (mouse) tıklanması ile basit sorgulama kapasitesine, gerekse çok yönlü konumsal analiz araçlarıyla (tools) yönetici ve araştırıcılara istenen süreçte bilgi sunar. CBS teknolojisi coğrafi verileri istatistiksel grafikler ve "eğer olur ise.." (if conditions) şeklindeki mantık sorgulamaları ve senaryolar şeklinde irdeleme aşamasına gelmiştir. CBS teknolojisi konumsal verilerin sorgulanması ve analizinde, yazılımlar sayesinde, birçok veri her türlü geometrik ve mantıksal işleme tabi tutulabilir. Eğer fonksiyonel coğrafi veriye sahip CBS mevcut ise, başlangıçta şu basit sorgulamalar yapılabilir.

Veri Sunumu:

Görsel işlemler yine CBS için önemli bir işlevdir. Birçok coğrafi işlemin sonunda yapılanlar harita veya grafik gösterimlerle görsel hale getirilir. Haritalar coğrafi bilgiler ile kullanıcı arasındaki en iyi iletişimi sağlayan araçlardır. Haritalar, yazılı raporlarla, üç boyutlu gösterimlerle, fotoğraf görüntüleri ve çok-ortamlı (multimedia) ve diğer çıktı çeşitleriyle birleştirebilmektedir.

Açık Coğrafi Konsorsiyum'un (OGC) ortaya koyduğu standartları sağlayan ArcGIS ve MapInfo gibi programların haricinde açık kaynak kodlu yazılımlar olan Chameleon, GeoNetwork opensource, GeoTools, OpenMap, MapGuide Open Source, OpenLayers, PostGIS, TerraView, vb. uygulamalar sunum için kullanılabilmektedir. Bu programların ortak özelliği standartlara uygun olarak kaydedilmiş coğrafi verinin bilgisayar destekli çizilmesi ve sorgulanmasıdır.

3. BİLGİ ERİŞİMİ

Bilgi Erişimi (Information Retrieval - IR) ile ilişkisel veritabanları ve İnternet üzerinde bilgi aramanın yanı sıra belge arama, belgelerin içinde bilgi arama ve belgeler hakkında metadata arama işlemleri gerçekleştirilir. Veri erişimi, belge erişimi, bilgi erişimi ve metin erişimi terimlerinin kullanımında bir çakışma, kesişim söz konusu olsa da her biri aynı zamanda kendi literatür, teori, süreç (praxis) ve teknolojisine sahiptir. Bilgi Erişimi bilgisayar bilimi, matematik, kütüphane bilimi, bilgi bilimi, bilgi mimarisi, bilişsel psikoloji, dilbilim ve istatistik'e dayalı disiplinler arası bir yaklaşım olarak ortaya çıkmaktadır.

Otomatik bilgi erişim sistemleri, "bilgi yüklemesi" (information overload) denen şeyi azaltmak için kullanılır. Birçok üniversite ve halk kütüphaneleri kitaplara, dergilere ve diğer belgelere erişim için Bilgi Erişim sistemlerini kullanmaktadır. Web arama motorları bunlar arasında en yaygın Bilgi Erişimi uygulamalarıdır.

3.1. Tarihsel Gelişim

İnsanlık, birbiriyle ilgili bilgileri aramak için bilgisayarların kullanılması düşüncesini ilk kez Vannevar Bush'un 1945 tarihli "As We May Think" makalesinden duymuştur. Buna rağmen ilk otomatik bilgi erişim sistemleri 1950'lerde ve 1960'larda ortaya çıkmıştır. 1970'e gelindiğinde, pek çok farklı tekniğin binlerce dokümandan oluşan Cranfield koleksiyonu gibi küçük metin gövdeleri (text corpora) üzerinde iyi sonuç verdiği gösterilmiştir (Singhal, 2001). 1992 yılında ABD Savunma Bakanlığı, TIPSTER adı verilen metin programının bir parçası olarak National Institute of Standards and Technology (NIST) ile birlikte Metin Erişimi Konferansı'na (TREC) sponsorluk yapmıştır. Bu konferansın amacı, çok geniş bir metin koleksiyonu üzerinde metin erişimi metodolojilerinin değerlendirilmesi için gereken altyapıyı temin ederek, bilgi erişimi topluluğunu araştırmaktır. Bu araştırmalar devasa boyutlara ulaşan veriler için yeni metotların geliştirilmesini teşvik etmiştir. Bugün ise, Web arama motorlarının ortaya çıkışı çok geniş ölçekli erişim sistemlerine olan ihtiyacı daha da arttırmış bulunmaktadır.

Bilginin saklanması ve erişilmesi için sayısal yöntemlerin kullanılması, çalışmaları bir dijital kaynağın fiziksel medyasının, o medyayı okumak için gereken okuyucunun, donanımın, onun üzerinde çalışan yazılımın artık mevcut olmadığı sayısal eskime (digital obsolescence) fenomenine götürmektedir. Bilgiye erişim, verinin kağıt üzerindeki halinden daha kolay olmasına karşın, bilginin sayısal ortamdan kaybolması da eskiye oranla daha etkili bir şekilde gerçekleşmektedir.

3.2. Kronoloji

Bilgi Erişimi konusundaki çalışmaların kronolojik bir listesi aşağıda verilmektedir:

1. 1900'lerden önce

1880:Herman Hollerith, makine tarafından okunabilen ortama veri kaydetmeyi icat etmiştir.

1890:Hollerith kartları, delgi makineleri ve tablolaştırıcılar, 1890-ABD Nüfus Sayımı verilerini işlemek için kullanılmıştır.

2. 1940'lar-1950'ler

1940'ların sonları: 2. Dünya Savaşı esnasında bilimsel araştırma belgelerinin ulaştırılması üzerine çalışılmıştır.

1945:Vannevar Bush'un "As We May Think" makalesi "Atlantic Monthly"de yayınlanmıştır.

1947:1941 yılından beri IBM'de araştırma mühendisi olan Hans Peter Luhn, kimyasal bileşikleri araştırmak için mekanize bir delikli kart-tabanlı sistem üzerinde çalışmaya başlamıştır.

1950'ler: Sovyet Sosyalist Cumhuriyet Birliği ile aralarındaki "bilimsel boşluk" konusunda ABD'de artan endişe, bilimsel projeler finansmanını teşvik etmiş ve mekanize literatür arama sistemleri (Allen Kent) ve atıf indeksinin icadı Eugene Garfield için uygun bir zemin sağlamıştır.

1950: "Bilgi erişimi" terimi Calvin Mooers tarafından ilk defa kullanılmıştır.

1951: Philip Bagley, MIT'de yüksek lisans tezi olarak ilk bilgisayarlı belge erişimi deneyini gerçekleştirmiştir (Doyle ve Becker, 1975).

1955: Allen Kent, Case Western Reserve Üniversitesi'ne katılarak ileride "Center for Documentation and Communications Research" diye anılacak merkezin müdür yardımcısı olmuştur. Aynı yıl Kent ve arkadaşları "American Documentation" da bir tebliğ yayınlamışlardır. Bilgi Erişimi sistemini değerlendirmek için gereken ve erişilmemiş ilgili dokümanların sayısını belirlemek için kullanılacak istatistiksel örnekleme metotlarını içeren bir "çerçeve" önerisi detaylandırılmış ve bu çalışmada hassasiyet (precision) ve geri-çağırma (recall) ölçütleri ilk defa tanımlanmıştır.

1958: Uluslararası Bilimsel Bilgi Konferansı Washington DC'de toplanmıştır. Bilgi Erişimi'nde tespit edilen sorunlar ele alınmıştır.

1959: Hans Peter Luhn "Auto-encoding of documents for information retrieval" adlı çalışması yayınlanmıştır.

3. 1960'lar

1960'ların başında: Gerard Salton, Harvard'da IR çalışmalarına başlamış ve daha sonra bu çalışmalar Cornell'e taşınmıştır.

1960: Melvin Earl Maron ve John Lary Kuhns (Maron, 2008) ortak çalışması olan "On relevance, probabilistic indexing, and information retrieval"'ı, Journal of the ACM'de yayınlamışlardır.

1962: Cyril W. Cleverdon, Bilgi Erişimi sisteminin değerlendirilmesi için bir model geliştirerek Cranfield çalışmalarının ilk bulgularını "Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems" adlı raporda yayınlamıştır.

1963: Bilimsel bilgi krizi fikrini tam olarak ifade eden Weinberg raporu "Science, Government and Information" yayınlanmıştır. Joseph Becker ve Robert M. Hayes'in, bilgi erişimi konusunda "Information storage and retrieval: tools, elements, theories" çalışması yayınlandı.

1964: Karen Spärck Jones, Cambridge'deki "Synonymy and Semantic Classification" tezini tamamladı ve hesaplamalı dilbilimin IR uygulamaları üzerine olan çalışmalar hız kazanmıştır. National Bureau of Standards, "Statistical Association Methods for Mechanized Documentation" başlıklı bir sempozyuma sponsor olmuştur.

1960'ların ortalarında: Amerikan Ulusal Tıp Kütüphanesi, ilk büyük makine tarafından okunabilen veritabanı ve toplu erişim sistemi olan "Medical Literature Analysis and Retrieval System (MEDLARS)"ı geliştirdi. MIT'deki Intrex Projesine başlanılmıştır.

1965: JCR Licklider'in "Libraries of the Future"adlı bir makalesi yayınlanmıştır.

1966: Don Swanson, University of Chicago'da "Requirements for Future Catalogs" isimli çalışmalarına başlamıştır.

1960'ların sonlarında: Wilfrid Lancaster, MEDLARS sisteminin değerlendirme çalışmalarını tamamlamış ve bilgi erişimi hakkındaki çalışmasının ilk baskısı yayınlanmıştır.

1968: Gerard Salton "Automatic Information Organization and Retrieval"ı yayınlamıştır. John W. Sammon'ın RADC Tech raporunda "Some Mathematics of Information Storage and Retrieval" adıyla yayınlanan çalışmasında vektör modelinin taslağını çizmiştir.

1969: Sammon'un "A nonlinear mapping for data structure analysis" (IEEE Transactions on Computers) adıyla yayınlanan çalışması, bir Bilgi Erişimi sistemine görselleştirme arayüzü için ilk öneri olarak ortaya atılmıştır.

4. 1970'lerde

1970'lerin başında: İlk online sistemler olarak NLM'e ait AIM-TWX, MEDLINE, Lockheed'e ait Dialog; SDC'ye ait ORBIT ortaya çıkmıştır. Theodor Nelson "hypertext" kavramını ortaya atmış ve "Computer Lib/Dream Machines" adıyla yayınlanmıştır.

1971: Nicholas Jardine ve Cornelis J. van Rijsbergen, “küme hipotezinden” (cluster hypothesis) ilk defa açıkça bahsetmişler, "The use of hierarchic clustering in information retrieval” isimli yayını çıkarmışlardır.

1975: Salton’un vektör işleme çerçevesinden ilk defa bahseden çok bilinen üç yayın:

1)A Theory of Indexing (Society for Industrial and Applied Mathematics),

2)A Theory of Term Importance in Automatic Text Analysis (JASIS v. 26),

3)A Vector Space Model for Automatic Indexing (CACM 18:11) yayınlanmıştır.

1978: İlk ACM SIGIR konferansı düzenlenmiştir.

1979: C. J. van Rijsbergen “Information Retrieval”ı adıyla bir yayın çıkarmıştır. Olasılıklı modellere ağırlıklı vurgu yapan bu eser döneminin en önemli eserlerindendir.

5. 1980'lerde

1980: Birinci uluslararası ACM SIGIR konferansı düzenlenmiştir.

1982: Nicholas J. Belkin, Robert N. Oddy ve Helen M. Brooks, bilgi erişimi için ASK (Anomalous State of Knowledge) bakış açısını önermiştir.

1983: Salton ve Michael J. McGill, vektör modellerine vurgu yapan “Introduction to Modern Information Retrieval”ı yayınlamıştır.

1980'lerin ortalarında: Ticari IR sistemlerinin son kullanıcı sürümlerini geliştirme çabaları artmaya başlamıştır.

1985-1993: Görselleştirme arabirimleri için anahtar tebliğler ve deneysel sistemler çalışmaları ortaya çıkarılmıştır.

1989: Tim Berners-Lee tarafından CERN’de ilk World Wide Web teklifi verilmiştir.

6. 1990'larda

1992: İlk TREC konferansı düzenlenmiştir.

1997: Korfhage’ye ait, görselleştirmeye ve çoklu-referans nokta sistemlerine vurgu yapan Information Storage and Retrieval (Korfhage, 1997) yayınlanmıştır.

1990'ların sonunda: Önceden sadece deneysel Bilgi Erişimi sistemlerinde bulunan birçok özelliğin web arama motorlarında uygulanması sağlanmıştır. Böylece, IR modellerinin araştırmanın ve uygulamaların yaygın kullanılması ve örneklerle en iyi desteklenmesi arama motorlarıyla gerçekleştirilmiştir.

3.3. Genel Bakış

Bir bilgi alma süreci, bir kullanıcının sisteme sorgu girmesiyle başlar. Sorgular, bilgi ihtiyaçlarının kurallara uygun ifadeleridir. Bilgi erişiminde sorgu, koleksiyondaki benzersiz tek bir nesneyi tespit etmez. Onun yerine, konuyla belki farklı seviyelerde ilgili olan birçok nesne bu sorguya uyabilir.

Nesne, veritabanındaki bilginin temsil ettiği bir varlıktır. Kullanıcı sorguları, veritabanı bilgisiyle eşleştirilir. Uygulamaya bağlı olarak veri nesneleri, mesela metin belgeleri, görüntüler (Goodrum, 2000), ses (Foote, 1999), zihin haritaları (Beel vd., 2009) veya video olabilir. Genellikle belgelerin kendileri doğrudan Bilgi Erişimi sisteminde tutulmaz veya saklanmaz, ancak onun yerine belge vekilleri veya öznitelikleri (metadata) tarafından sistemde temsil edilirler.

Çoğu Bilgi Erişimi sistemleri, veritabanındaki her nesnenin sorguya ne kadar uyduğuna dair bir sayısal puan hesaplar ve nesneleri buna göre sıralar. En üst sıralardaki nesneler daha sonra kullanıcıya gösterilir. Kullanıcı daha kesin sonuca varmak isterse bu süreci tekrar edebilir.

3.4. Performans Ölçüleri

Bilgi erişim sistemlerinin performansını değerlendirmek için birçok farklı ölçüler önerilmiştir. Bu ölçüler bir belge koleksiyonu ve bir sorgu gerektirir. Burada tarif edilen tüm ortak ölçüler şöyle bir ilgi ile hareket eder: her belgenin belirli bir sorguyla ilgili olup olmadığı biliniyordur ve pratikte sorgular sorunlu olabilir ve farklı ilgi tonları (Frakes, 1992) olabilir. Hassasiyet(Precision) ve Geri Çağırma(Recall) aşağıdaki şekilde ifade edilmektedir (Olson ve Delen, 2008)

3.4.1. Hassasiyet (*Precision*)

Hassasiyet, erişilen belgeler içinde kullanıcının bilgi ihtiyacına uyan kısmının, erişilen belgelerin tamamına oranıdır.

$$Hassasiyet\ (Precision) = \frac{|\{\text{ilgili dokümanlar}\} \cap \{\text{erişilen dokümanlar}\}|}{|\{\text{erişilen dokümanlar}\}|} \quad (3.1)$$

İkili sınıflandırmada hassasiyet, pozitif prediktif değer'e benzer. Hassasiyet erişilen tüm belgeleri hesaba katar. Ayrıca, yalnız sistemin döndürdüğü en tepedeki sonuçları dikkate alarak belirli bir kesim (cut-off) değerinde de değerlendirilebilir. Bu ölçüye n'deki hassasiyet (precision at n) veya P@n denir.

"Hassasiyet" (precision) teriminin Bilgi Erişimi alanındaki anlam ve kullanımının diğer bilim ve teknoloji dallarındaki doğruluk ve hassasiyet tanımlarından farklıdır.

3.4.2. Geri Çağırma *(Recall)*

Geri Çağırma, başarıyla erişilen ve sorgu ile alakalı olan belgelerin, örnek uzayda bulunan ve sorgu ile alakalı tüm belgelere olan oranıdır.

$$Geri\ Çağırma\ (Recall) = \frac{|\{\text{ilgili dokümanlar}\} \cap \{\text{erişilen dokümanlar}\}|}{|\{\text{ilgili dokümanlar}\}|} \quad (3.2)$$

İkili sınıflandırmada Geri Çağırma, Duyarlılık adını alır. Böylece Geri Çağırma, ilgili bir dokümana o sorguyla erişme ihtimali olarak da görülebilir.

Herhangi bir sorguya karşılık olarak tüm belgeleri geri döndürerek % 100 Geri çağırma elde edilmesi doğru değildir. Bu nedenle, Geri Çağırma tek başına yeterli bir ölçüt değildir, ancak araştırmacının ilgili olmayan belgelerin de sayısını, mesela hassasiyeti hesaplayarak ölçmesi gerekir.

3.4.3. Atık *(Fall-Out)*

Erişilmiş olan ilgisiz dokümanların, mevcut bütün ilgisiz dokümanlar içindeki oranı olarak tanımlanır eşitlik 3.3'deki gibi ölçülür:

$$Atık(Fall-Out) = \frac{|\{\text{ilgisiz dokümanlar}\} \cap \{\text{erişilen dokümanlar}\}|}{|\{\text{ilgisiz dokümanlar}\}|} \quad (3.3)$$

İkili sınıflandırmada Atık, özgüllük ile yakından ilgilidir ve ilgisiz bir dokümana o sorguyla erişme ihtimali olarak da görülebilir.

Herhangi bir sorguya karşılık olarak sıfır adet belgeyi geri döndürerek % 0 Atık elde edilmesi önemsizdir.

3.4.4. F-ölçüsü

Hassasiyetin ve Geri Çağırmanın ağırlıklı harmonik ortalaması, geleneksel F-ölçüsü veya dengeli F-puanı olarak bilinir ve eşitlik 3.4'deki gibi ölçülür:

$$F = \frac{2 \cdot P \cdot R}{(P+R)} \quad (3.4)$$

Bu ölçüm F1 ölçüsü olarak da bilinir, çünkü Geri Çağırma ve Hassasiyet eşit olarak ağırlıklandırılmıştır.

Negatif olmayan reel bir β için genel formül:

$$F_\beta = \frac{(1+\beta^2)\cdot P\cdot R}{\beta^2\cdot P+R} \tag{3.5}$$

Yaygın olarak kullanılan diğer iki F ölçüsü de geri çağırmayı hassasiyetin iki katı ağırlıklandıran F2 ve hassasiyeti geri çağırmanın iki katı ağırlıklandıran F0,5 ölçüsüdür.

F-ölçüsü Van Rijsbergen tarafından türetildiği (1979) için Fβ, "geri çağırmaya hassasiyetin β (beta) katı kadar önem atfeden bir kullanıcı açısından erişimin etkinliğini ölçer." Bu, Van Rijsbergen'in etkinlik ölçüsüne dayanır:

$$E = 1 - \left(1/(\alpha/P + (1-\alpha)/R)\right) \tag{3.6}$$

α = 1 / (β2 + 1) iken ilişkileri Fβ = 1 − E 'dir.

3.4.5. Ortalama Hassasiyet

Hassasiyet ve Geri Çağırma, sistem tarafından döndürülen belgelerin tam listesine dayanan tek değer ölçütleridir. Belgelerin sıralanmış olarak döndürüldüğü sistemler için, döndürülen belgelerin hangi sırada sunulduğunun göz önünde bulundurulması da arzu edilir. Ortalama Hassasiyet, ilgili dokümanların üst sıralarda verildiğini vurgular. Sıralanmış dizideki her bir ilgili dokümanın noktasında hesaplanmış hassasiyetlerin ortalamasıdır:

$$Ortalama\ Hassasiyet = \frac{\sum_{r=1}^{N}(P(r)\cdot rel(r))}{\text{ilgili dokümanların sayısı}} \tag{3.7}$$

Burada r derece (rank), N erişilmiş sayı, rel() belli bir derecenin ilgisinin ikili (binary) bir fonksiyonu, P(r) ise belli bir kesme (cut-off) derecesinin hassasiyetidir:

$$P(r) = \frac{\text{r veya daha az dereceli erişilen ve ilgili dokümanların sayısı}}{r} \tag{3.8}$$

Bu ölçüte bazen geometrik olarak, Hassasiyet-geri çağırma eğrisinin altında kalan alan da denmektedir.

Buradaki paydanın (ilgili dokümanların sayısının) tüm koleksiyondaki ilgili dokümanların sayısı olduğuna, böylece de bu ölçütün, bir erişim kesme değerinden bağımsız olarak tüm ilgili dokümanlar üzerinden performansı yansıttığına dikkat edilmelidir.

3.4.6. İndirimli Kümülatif Kazanç (Discounted cumulative gain - DCG)

İndirimli Kümülatif Kazanç, bir belgenin sonuç listesindeki konumuna göre kullanışlılığını veya kazancını değerlendirmek için, sonuç kümesindeki dokümanlara ait kademeli bir ilgi ölçek kullanır. DCG'nin önermesi, kademeli ilgi değeri arama sonucunun konumuna orantılı olarak logaritmik azaldığı için, bir arama sonucunda daha aşağıda görülen ama ilgi derecesi yüksek olan dokümanların cezalandırılmış olması gerektiğidir.

Belirli bir p derece konumundaki birikimli DCG, eşitlik 3.9'daki gibi hesaplanır:

$$DCG_P = \mathrm{rel}_1 + \sum_{i=2}^{p} \frac{\mathrm{rel}_i}{\log_2 i} \tag{3.9}$$

Farklı sorgular veya sistemleri arasında sonuç kümesi farklı ebatlarda olabildiğinden, DCG'nin normalleştirilmiş sürümü, performansları karşılaştırmak için – bir sonuç listesindeki dokümanları alakasına göre sıralamak suretiyle – ideal bir DCG kullanarak puanı normalleştirir:

$$nDCG_P = \mathrm{rel}_1 + \frac{DCG_P}{IDCG_P} \tag{3.10}$$

Bir sıralama (ranking) algoritmasının ortalama performansının bir ölçümünü elde etmek için, tüm sorguların nDCG değerlerinin ortalaması alınabilir. Mükemmel bir sıralama algoritmasında DCGp'nin, 1.0 değerinde bir nDCG üreten IDCGp ile aynı olacağına dikkat edilmelidir. O halde tüm nDCG hesaplamaları 0.0 – 1.0 aralığında nisbi değerlerdir ve çapraz sorgularla karşılaştırılabilirler.

3.5. Model tipleri

Bilgi erişiminin etkili olması için, belgeler genellikle uygun bir gösterime dönüştürülür. Birçok gösterim vardır.

3.5.1. Birinci boyut: Matematiksel Olarak

Küme teorisi modelleri, dokümanları sözcük veya sözcük kümeleri olarak temsil eder. Benzerlikler genellikle bu kümeler üzerinde yapılan küme teorisi işlemlerinden elde edilir. Ortak modeller şunlardır:

Standart İkili Modeli (Standard Boolean Model)

Genişletilmiş İkili Modeli (Extended Boolean Model)

Bulanık Erişim (Fuzzy Retrieval)

Cebirsel modeller, dokümanları ve sorguları genellikle vektör, matris veya üçlü (tuple)olarak gösterir. Sorgu vektörü ile belge vektörünün benzerliği, skaler bir değer olarak gösterilir.

Vektör Uzay Modeli (Vector Space Model)

Genelleştirilmiş Vektör Uzay Modeli (Generalized Vector Space Model)

(Geliştirilmiş) Konu Tabanlı Vektör Uzayı Modeli ((Enhanced) Topic-based Vector Space Model)

Genişletilmiş İkili Modeli (Extended Boolean Model)

Latent Anlamsal İndeksleme Modeli (Latent Semantic Indexing Model)

Olasılığa dayalı modeller, doküman erişim sürecini olasılıksal bir çıkarsama gibi ele alır. Benzerlikler, bir dokümanın belirli bir sorguyla ilgili olması ihtimali olarak hesaplanır. Bayes teoremi gibi olasılıksal teoremler, bu modellerde sık sık kullanılır.

İkili Bağımsızlık Modeli (Binary Independence Model)

Olasılıksal İlgi Modeli (Probabilistic Relevance Model)

Belirsiz Çıkarsama (Uncertain İnference)

Dil Modelleri (Language Models)

Rastgelelikten-ıraksama Modeli (Divergence-from-randomness Model)

Gizli Dirichlet Dağılımı (Latent Dirichlet Allocation)

Makine-öğrenimli/programlanmış derecelendirme modelleri, dokümanları (bazıları yukarıda bahsedilen diğer derecelendirme modellerini çoğu kez birleştiren) derecelendirme özelliklerinin vektörleri olarak görür ve bu özellikleri makine öğrenimi metotlarıyla tek bir alaka puanında toplamanın en iyi yolunu bulmaya çalışır.

3.5.2. İkinci boyut: Modelin Özellikleri

Birbirine bağımlı terimlerin olmadığı modeller, farklı terimleri / kelimeleri bağımsız olarak ele alır. Bu gerçek genelde vektör uzay modellerinde terim vektörlerinin ortogonalliği varsayımı ile veya olasılıksal modellerde terim değişkenleri için bir bağımsızlık varsayımı ile temsil edilir.

Terimlerin birbirine bağımlılığı için olan modellerde, terimler arasındaki karşılıklı bağımlılıklar gösterilebilir. Ancak iki terim arasındaki karşılıklı bağımlılığın derecesi,

modelin kendisi tarafından tanımlanır. Bu genelde, bütün doküman kümesinde o terimlerin birlikte bulunmasından (co-occurrence) doğrudan veya dolaylı olarak (mesela boyutsal indirgeme yoluyla) elde edilir.

Terimlerin birbirine bağımlılığı olan modellerde, terimler arasındaki karşılıklı bağımlılıklar gösterilebilir, ancak iki terim arasındaki karşılıklı bağımlılığın nasıl tanımlandığını söylemezler. İki kelime arasındaki karşılıklı bağımlılık derecesi için bir harici kaynakla bağlantı kurarlar. (Örneğin bir insan veya sofistike algoritmalar.)

4. BİLGİ ÇIKARIMI

Bilgi çıkarımı (*Information Extraction - IE*), hedefinde yapılandırılmamış ve makine tarafından okunabilir belgelerden otomatik olarak yapılandırılmış bilgiler (genellikle doğal dil işleme (NLP) yoluyla insan dilinde metinler) çıkarmak olan bir Bilgi Erişimi (*Information Retrieval*) türüdür.

Problemin zorluğu nedeniyle, IE için güncel yaklaşımlar dar biçimde sınırlandırılmış etki alanlarına odaklanır. Bunun bir örneği haberlerden veya şirket birleşmelerinden çıkarımdır. Örneğin yapısal ilişkide gösterilen *ArasındaBirleştirme*,(*firma*$_1$,*firma*$_2$,*tarih*), aşağıdaki haber cümlesindeki gibidir:

"Dün, New-York merkezli Foo Inc, Bar Corp'u devraldığını duyurdu."

IE'nin genel hedefi, öncesinde yapılandırılmamış olan veriler üzerinde hesaplama yapılmasını sağlamaktır. Daha özel hedefi de, girilen verilerin mantıksal içeriğine dayalı çıkarımların mantıksal akıl yürütme ile yapılmasına olanak sağlamaktır. Yapılandırılmış veriler, seçilmiş bir hedef etki alanından gelen anlamsal olarak (*semantically*) iyi tanımlanmış, kategori ve bağlam bakımından yorumlanmış verilerdir**.**

4.1. Tarihçe

Bilgi çıkarımı, Doğal Dil İşleme'nin ilk günlerinden 1970'lerin sonlarına kadar uzanır (Andersen, 1992). 1980'lerin ortalarındaki ilk ticari sistemlerden biri, finans tüccarlarına gerçek-zamanlı finans haberleri sağlamak için Carnegie Grubu tarafından Reuters için yapılmış olan JASPER'dır (Cowie ve Wilks, 1998).

Bilgi Erişimi, 1987 yılından başlayarak Mesaj Anlama Konferansları (Message Understanding Conferences - MUC) dizisiyle gündeme gelmiştir. Yarışma-tabanlı bir konferans olan MUC aşağıdaki alanlara odaklanır:

MUC-1 (1987), MUC-2 (1989): Donanma operasyonlarına ait mesajları.

MUC-3 (1991), MUC-4 (1992): Latin Amerika ülkelerindeki terörizm.

MUC-5 (1993): Ortak girişimler ve mikro-elektronik etki alanları.

MUC-6 (1995): Yönetim değişiklikleri hakkındaki haberler.

MUC-7 (1998): Uydu fırlatma raporları.

MUC, Muhtemel terör bağlantıları için gazetelerin taranması gibi, hükümet analistleri tarafından yürütülen sıradan görevleri otomatikleştirmek isteyen Amerikan savunma ajansı DARPA'dan ciddi bir destek alınmıştır.

4.2. Günümüzdeki Önemi

IE'nin günümüzdeki önemi, yapılandırılmamış formda artan miktarda bilgi olmasıyla ilgilidir. İnternetin mucidi Tim Berners-Lee, mevcut İnternet'in bir doküman ağı olduğunu söylemekte (Bizer vd., 2009) ve ileride bu içeriğin daha fazlasının veri ağına çevrileceğini savunmaktadır (Berners ve Lee, 2009). Bu gerçekleşinceye kadar web'in büyük ölçüde anlamsal öznitelikleri (metadata) eksik olan ve yapılandırılmamış dokümanlardan oluşacağı ifade edilmektedir.

Bu belgelerdeki bilgiler, ilişkisel (relational) forma dönüştürme yoluyla veya XML etiketleriyle işaretlenerek, işlemeyi yapacak olan makine için daha erişilebilir hale getirilebilir. Haber geçişlerini izleyen akıllı bir ajana, yapılandırılmamış bilgiyi akıl yürütülebilecek bir şekle dönüştürmek için IE gereklidir. Doğal bir dilde yazılmış bir doküman setini taramak ve bir veritabanını ayıklanmış bilgiyle doldurmak, tipik bir Bilgi Çıkarımı uygulamasıdır.

4.3. Metin Basitleştirilmesi ve Alt Görevler

Doğal dilde oluşturulmuş metinlerin, cümleleri ayıklayacak makinenin daha rahat okuyabilmesi için bir metin basitleştirme şekline / formuna gereksinimi olabilir.

Tipik Bilgi Çıkarımı alt görevleri şunlardır:

- **Adlandırılmış Varlık Tanıma (*Named Entity Recognition*):** (insanlar ve kurumlar için) varlık adlarının, yer adlarının, geçici ifadelerin ve belirli sayısal ifade türlerinin tanınması.
- **Ön-referans çözünürlüğü:** Metin varlıkları arasında ön-referans ve anaforik bağlantıların aranması. Bilgi Çıkarımı görevlerinde tipik olarak bu, önceden ayıklanmış ve adlandırılmış varlıklar arasındaki bağlantıları bulmada sınırlandırılmıştır. Örneğin, "International Business Machines" ve "IBM" gerçek dünyadaki aynı varlığı işaret etmektedir.
- **Terminoloji Çıkarımı:** Verilen bir metin gövdesi için alakalı terimlerin bulunması.
- **İlişki Çıkarımı:** Varlıklar arasındaki ilişkilerin tanımlanması. Örneğin:
    - KİŞİ KURUM'da çalışıyor (Kişi:Bill Kurum:IBM)
    - LOKASYON'daki KİŞİ (Kişi:Bill Lokasyon:France)

4.4. Bilgi çıkarımı ve Dünya Çapında Ağ

MUC konferanslarının odağını Bilgi Çıkarımı oluşturmaktadır. İnternetin yaygınlaşması, muazzam miktardaki mevcut çevirim içi veriyle başa çıkmada yardımcı olacak, Bilgi Çıkarımı sistemlerini geliştirme ihtiyacını yoğunlaştırmaktadır. Çevirimin, metinden bilgi çıkarımını gerçekleştiren sistemlerin düşük maliyet, geliştirmede esneklik ve yeni etki alanlarına kolay uyum gereksinimlerini karşılaması gerekmektedir. MUC sistemleri bu koşulları karşılamada başarılı olamamaktadır. Dahası, yapılandırılmamış metin için gerçekleştirilen dil analizi, çevirim içi metinde bulunan HTML / XML etiketlerini ve düzen formatından faydalanmamaktadır.

Sonuç olarak, belirli bir sayfanın içeriğini çıkaran, doğruluğu yüksek kurallar seti olan sarmalayıcılar kullanılarak, web üzerinde bilgi çıkarımı için dilbilimsel yönden daha az yoğun yaklaşımlar geliştirilmiştir. Sarmalayıcıların (wrappers) elle geliştirilmesinin yüksek düzeyde uzmanlık gerektiren zaman alıcı bir iş olduğu görülmektedir. Denetimli (supervised) veya denetimsiz (unsupervised) makine öğrenimi (machine learning) teknikleri, böyle kuralları otomatik olarak teşvik etmek için kullanılmaktadır.

Sarmalayıcılar genellikle, ürün katalogları ve telefon dizinleri gibi oldukça yapılandırılmış olan web sayfası koleksiyonlarını ele alır. Ancak, metin türü daha az yapılandırılmış olduğunda başarısız olurlar, ki bu da İnternette yaygındır. Bu konudaki yeni çalışmalar, iyi yapılandırılmış metinler ve sarmalayıcıların başarısız olduğu, serbest metinler üzerindeki Adaptif Bilgi Çıkarımı uygulamalarının geliştirilmesini motive eder. Bu sistemler sığ doğal dil bilgilerini sömürebilir ve böylelikle daha az yapılandırılmış olan metne uygulanabilir.

4.5. Şartlı Rastgele Alanlar

Şartlı Rastgele Alanlar (Conditional Random Fields - CRF) yaygın olarak, araştırma tebliğlerinden bilgi çıkarımından (Peng ve McCallum, 2006) navigasyon yönlendirmeleri çıkarımına (Shimizu ve Andrew, 2006) kadar uzanan çeşitli görevlerde Bilgi Çıkarımı ile birlikte kullanılırlar.

4.6. Ücretsiz veya Açık Kaynaklı Bilgi Çıkarımı Yazılım veya Hizmetleri

- **Metin Mühendisliği için Genel Mimari:** Ücretsiz bir Bilgi Çıkarımı sistemiyle çevrelenmiş "Metin Mühendisliği için Genel Mimari" dir.
- Thomson Reuters'dan otomatik bilgi çıkarımı web servisi (ücretsiz sınırlı sürüm).
- **Machine Learning for Language Toolkit (Mallet):** Bilgi çıkarımı dahil olmak üzere çeşitli doğal dil işleme görevleri için Java-tabanlı bir pakettir.

5. GELİŞTİRİLEN UYGULAMA (IIMtCAR)

İnternet İçeriğinde Metin tabanlı Coğrafi Arama Robotu (IIMtCAR), sorgunun oluşturulması, arama, bağlantıları bulma, site içeriklerine ulaşma, adres cümlelerini çıkarma/ayıklama ve sonuç kümesini görselleştirme aşamalarından oluşur ve aşağıdaki adımları içerir.

5.1. Sorgunun Oluşturulması ve Çalıştırılması

İlk olarak aranılan kelime(ler) ile eşleşen ve adres/iletişim bilgilerini içeren web sayfalarına erişim gerçekleştirilmektedir. Bu adımda arama motorlarından bir tanesi tercih edilerek sorgu gönderilir ve gelen sorgu neticesinde sayfa içerisindeki adres bilgilerine erişim için gerekli işlemler gerçekleştirilir.

Arama motorları arasından tercih yapılırken, indekslenen web içeriğinin fazla olması ve İnternet kullanıcılarının tarafından yaygın kullanımı dikkate alınmıştır. Aynı zamanda harita hizmetinin İnternet üzerinden verilmesinde de ilklerden olması, kullanım yaygınlığı ve standartları desteklemesi bakımından ön plana çıkan Google, tez çalışması içerisinde bütünlük teşkil etmesi için arama, coğrafi kodlama ve sunum adımlarında kullanılmıştır.

Sorguları çalıştırmak için, arama motorlarının sağladığı kod kütüphaneleri kullanılabildiği gibi, web üzerinden arama motoruna parametre gönderilebilmektedir. Google'ın sunduğu arama hizmetinin web üzerinden kullanılabilmesi için aşağıdaki kodlar kullanılır :

```
string google = "http://www.google.com.tr/search?search?hl=tr&num=" + searchResultCount + "&q=";
Uri ur = new Uri(google + txtSearch.Text.Trim().Replace(' ', '+') + "+ileti%C5%9Fim+adres");
```

Burada web sayfasına yerleştirilen txtSearch isimli metin alanında girilen arama kelimelerine "iletişim" ve "adres" kelimelerini de ekleyerek aranan sayfaların içerisinde iletişim bilgisinin bulunması sağlanmaktadır.

```
HtmlNodeCollection linkler = doc.DocumentNode.SelectNodes("//h3[@class='r']");
```

ile ulaşılan bağlantılar GetAddressDetails fonksiyonu ile tekrarlamalı olarak taranır. Bu fonksiyon ile ilgili sitelerdeki adres cümlelerine erişilecektir. GetAddressDetails fonksiyonu aşağıda belirtildiği şekilde kullanılmaktadır.

```
try
 {
   foreach (HtmlNode link in linkler)

       try

       {
          GetAddressDetails(link.InnerText, link.ChildNodes[0].Attributes["href"].Value);
        }
       catch { }
 }
```

5.2. Adres Cümlelerine Erişim İçin Kullanılan Kalıplar

Her bir bağlantı içindeki tüm parçaları gözden geçirirken adres içerebilecek parçaların filtreleme işlemine tabi tutulması gerekmektedir. Bunun için arama dili göz önünde bulundurularak anahtar kalıplar seçilir. Anahtar kalıplar seçilirken web sayfasında adres niteliği taşıyabilecek bilgileri bulmak hedeflenmektedir. Türkçe için bir adres bilgisi içerisinde yer alan anahtar kalıpları "X cadde", "X cd.", "X cad.", "X sokak", "X sok.", "X sk.", "X mahalle", "X mah.", "X mh.", "X bulvarı", " X alışveriş merkezi", " X apartmanı", " X apt.", " X sitesi", " X sit.", " X adres :", " X adres:", " X mevki", " X karayolu", " X otoyolu" olarak tespit edilmiştir. Eğer, web sayfası içerisindeki metin alanı (html tag) bu kalıplardan birini içeriyorsa, ilgili cümle daha sonra adres bilgisinin çıkarılması için kayıt altına alınır.

Örnek olarak, İnternet sayfasının içeriği anahtar kalıplar ile filtrelendiğinde, sonuç kümesindeki adres cümlelerinden bir tanesi *"- telefon numarası) konusunu görüntülüyorsunuz; varan turizm / /merkez adres : merkez mah. çınar cd.no:16 bahçelievler / istanbul / telefon : +90 ..."* şeklinde gözükmektedir. Bu cümle, uygulama kodu tarafından adres içerebilecek cümle olarak algılanmış ve bir sonraki adımda adres çıkarımının yapılabilmesi için kaydedilmiştir.

5.3. Adres Cümlelerinin Parçalara Bölünmesi

Anahtar kalıpları kullanarak filtrelemeye tabi tutulan bilgilerin cadde, sokak, mahalle, il ve ilçe olarak bölümlere ayrılması coğrafi kodlama için önemlidir. Adres içerisinde yer alan fazlalıklar temizlenmediği takdirde coğrafi koordinatı çıkaracak fonksiyonun gerçeğe yakın sonuçlar verme ihtimali düşmektedir.

Bu amaçla Türkiye'nin adres veritabanına1 başvurulması gerekmektedir. Öncelikli olarak il, ilçe ve semt adlarının sırası ile ilgili adres cümlesinde yer alıp almadığı kontrol edilir.

"- telefon numarası) konusunu görüntülüyorsunuz; varan turizm / /merkez adres : merkez mah. çınar cd.no:16 bahçelievler / istanbul / telefon : +90 …" örneğinde Türkiye adres veritabanı ile karşılaştırma yapıldığında "İSTANBUL" ve "BAHÇELİEVLER" tespit edilmiş ve bu kelimeler ayıklanan adres parçaları olarak kayıt edilmiştir.

Bu adımdan sonra metin içerisinde daha önce belirlenmiş anahtar kalıplardan (sok, cad, gibi) önce gelen kısımlar da ayıklanmış adres cümlesine ilave edilmiştir. Örneğin yukarıdaki cümlede "cd." ve "mah." anahtar kalıplarından önceki adres bilgisi ayıklanmış ve ayıklanan adres parçalarına ilave edilmiştir. Sonuç itibari ile "merkez mah. çınar cd. BAHÇELİEVLER İSTANBUL" adresi elde edilmiştir. Bu sayede temiz bir adres cümlesi çıkarılmış olur.

5.4. Coğrafi Kodlama Fonksiyonunun Kullanılması

Arama sonucunda dönen web sayfasından adres olamayacak bilgilerin temizlenerek, kabul edilebilir adres bilgisi çıkarıldıktan sonra, ilgili adresin coğrafi koordinata çevrilmesi gerekmektedir. Bu işlemi gerçekleştirmek için harita sunucusuna ve harita verisine ihtiyaç duyulmaktadır. Coğrafi koordinatın çıkarılması için Google'ın veya diğer harita altlığı sunan firmaların servisleri kullanılabilir. Aşağıda internet üzerinden Google'ın sunduğu coğrafi kodlama fonksiyonuna erişim ve kullanım şekli belirtilmektedir.

```
string urlAdd = "http://maps.google.com/maps/geo?q=" + address +
 "&output=csv&oe=utf8&sensor=false&key=……=tr";
```

Yukarıda verilmiş olan bağlantıda *address* parametresi olarak önceki adımlarda elde edilen ayıklanmış adres bilgisi kullanılmaktadır. Bağlantıyı çağırdığımızda geri gönen http sayfasında virgüllerle ayrılmış dört adet değer yer almaktadır. Bu değerler sırasıyla:

1 : HTTP durum kodu

2 : Doğruluk (*Accuracy*)

3 : Enlem (*Latitude*)

4 : Boylam (*Longitude*)

bilgilerini içerir.

[1] http://www.ptt.gov.tr/tr/interaktif/pkodu1.zip bağlantısından indirilebilmektedir. Microsoft Excel formatında elde edilen verinin norrmalize edilmesi ile oluşturulan ilişkisel veritabanı ek 1 olarak tezde yer almaktadır.

Çizelge 5.1 HTTP Durum Kodları

Durum İsmi	DurumKodu
Başarılı - Success	200
Başarısız İstek - BadRequest	400
Sunucu Hatası - ServerError	500
Eksik Sorgu - MissingQuery	600
Eksik Adres - MissingAddress	601
Bilinmeyen Adres - UnknownAddress	602
Mevcut Olmayan Adres - UnavailableAddress	603
Bilinmeyen Yön - UnkownDirections	604
Sorgu Cümlesinde Hatalı Anahtar - BadKey	610
Çok Fazla Sorgu - TooManyQueries	620

Çizelge 5.2 Adres Doğruluğu Kodları

Adres Bilgisi İsmi	Doğruluk Kodu
Bilinmeyen Konum - UnkownLocation	0
Ülke - Country	1
Bölge - Region	2
Alt Bölge - SubRegion	3
Şehir - Town	4
Posta Kodu - PostCode	5
Cadde/Sokak - Street	6
Yol Kavşağı - Intersection	7
Adres - Address	8
Bina - Premise	9

http://www.msxlabs.org/forum/mk-rehber/260306-varan-turizm-istanbul-adres-telefon-numarasi.html bağlantısından ulaşılan "- telefon numarası) konusunu görüntülüyorsunuz; varan turizm / /merkez adres : merkez mah. çınar cd.no:16 bahçelievler / istanbul / telefon : +90 ..." alandan ayıklanarak elde edilen "merkez mah. çınar cd. BAHÇELİEVLER İSTANBUL" adresinin koordinatı "41.0204, 28.81272" olarak tespit edilmiştir.

http://www.bamtur.com/Iletisim adresinden ulaşılan ": şaşkınbakkal kazım özalp sk. kulan apt. b blok no:22/2 suadiye kadıköy istanbul türkiye 34740// telefon : +90 216 444 0 157// fax : +90 216 355 02 99// e-posta : info@bamtur.com//" cümlesinden "şaşkınbakkal kazım özalp sk SUADİYE KADIKÖY İSTANBUL" adresi çıkarılmıştır. Bu adres bilgisi Google fonksiyonuna gönderildiğinde koordinat çözümlemesi gerçekleşmemiştir. Adres cümlesi http://maps.google.com sayfasında da aratıldığında koordinat noktası çıkarılamamıştır. Site üzerinde "kazım özalp sk SUADİYE KADIKÖY İSTANBUL" arattırıldığında ise, Google konumlandırma işlemini gerçekleştirebilmektedir.

Aynı bağlantıdan elde edilen ": acıbadem mahallesi acıbadem caddesi sayın apt. no:145 k:1d:3 acıbadem/istanbul// telefon : 0216 428 92 70// fax : 0216 428 95 35//" cümlesinden adres olarak "acıbadem mahallesi acıbadem cad ACIBADEM İSTANBUL" bilgisi çıkarılmış, fakat bu adresin de konumlandırılma işlemi Google fonksiyonu tarafından gerçekleştirilememiştir. Ancak "acıbadem cad ACIBADEM İSTANBUL" adresi koordinat olarak çözümlenebilmiştir.

5.5. Adreslerin Coğrafi Koordinatlarının Harita Üzerinde Gösterilmesi

Geliştirilen uygulamanın sunum kısmı internet sayfası olarak hazırlanacağı için internet-tabanlı harita istemci yazılımına ihtiyaç duyulmaktadır. Openlayers, ArcGIS, GMap Api, Microsoft Bing Maps, vb. uygulamaları arasından Google'ın sunduğu harita altyapısı seçilerek uygulamaya eklenmiştir.Haritanın gösterileceği internet sayfasının *head* kısmında aşağıdaki kodlar eklenmiştir. *Key* ile belirtilen alanda sitenin etki alanı belirtilerek Google'dan alınan şifre girilmektedir.

```
<script
src="http://maps.google.com/maps?file=api&amp;v=2.x&amp;key=ABQIAAAAWKkKqkMq
Q5iunGSRXeMAEhSokcyqauxLCyIcQKbf1aqZhru6WRQ1WJgctbQGn-
QENQYIuEMsq5fhUg" type="text/javascript"></script>
```

Bu *script* içerisinde harita fonksiyonları olan haritayı oluşturan resimlerin yüklenmesi, yakınlaştırma/uzaklaştırma, kaydırma, vb. işlemlerin yerine getirilmesi için gerekli olan kodlar da yer almaktadır.

Sayfanın *Body* alanında sayfa açılışında devreye girmesi istenilen *initialize* fonksiyonu aşağıdaki kod ile çağrılmaktadır.

```
<body onload="initialize()" onunload="GUnload()">
```

"*initialize*" fonksiyonunun çağırılması ile *map_canvas* ismi ile belirtilen sunum sayfası içerisindeki *div* parçası haritayı gösterebilmek için hazırlanmıştır. Fonksiyon içeriği aşağıdaki kodda belirtilmiştir.

```
function initialize() {
  if (GBrowserIsCompatible()) {
    map = new GMap2(document.getElementById("map_canvas"));
    map.setCenter(new GLatLng(41.017, 28.967), 9);
    geocoder = new GClientGeocoder();
    ShowAdres();
  }
}
```

Mevcut fonksiyonların üzerine, verilen koordinatları haritada göstermek için bir fonksiyon daha eklenmelidir. Aşağıda belirtilen metod ile parametre olarak verilen (x, y) koordinatına bir işaret koyulması ve bu işarete fare (mouse) ile tıklandığında, ilgili web sayfasının başlığının ve bağlantı adresinin gösterilmesi sağlanmaktadır.

```
function showAddress(x, y, address, title_, link) {
    if (geocoder) {
       var marker = new GMarker(new GLatLng(x, y));
       GEvent.addListener(marker, 'click', function() {
         marker.openInfoWindowHtml('<a href= "' + link + '" target="_blank">' + title_ +
'</a><br/>' + address);
         });
       map.addOverlay(marker);
    }
  }
```

Son olarak coğrafi koordinatların çıkarıldığı fonksiyondaki değerlerin showAddress fonksiyonuna gönderilmesi ile harita üzerinde adres gösterimi sağlanır.

6. SİSTEM BAŞARISININ ÖLÇÜMÜ

Sistem başarısının ölçümü yapılırken iki yol izlenmektedir. Bunlardan birincisi, arama yapılan sorguya göre kaç adet adres çözümlemesinin yapıldığıdır. İkinci yol ise yapılan sorgulamanın sonucunda ortaya çıkan adres eşleşmelerinin beklenen değer kümesine ulaşılıp ulaşılmadığının gözlemlenmesidir.

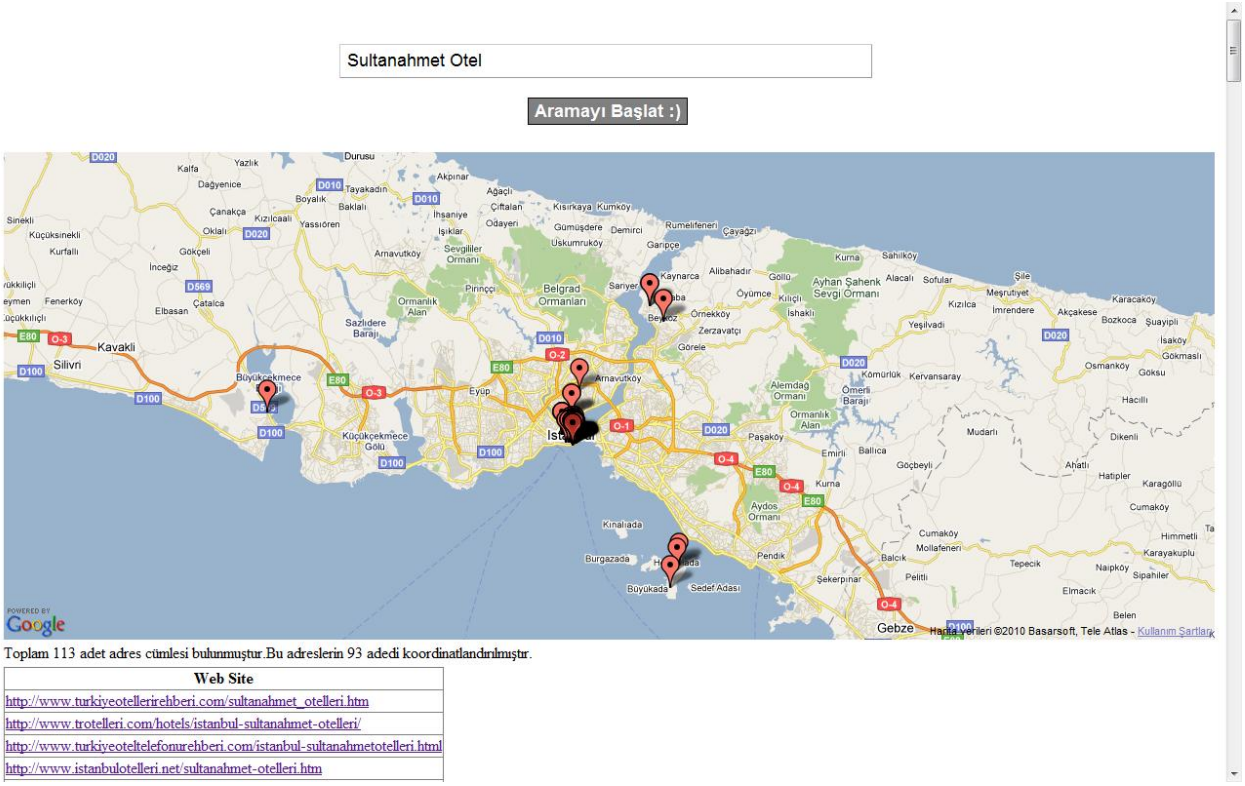
6.1. Sayısal Veriler ile Ölçüm

Birinci yöntemde sorgu cümlesi olarak "Beşiktaş Eczane", "Sultanahmet Otel" ve "Yurtiçi Kargo" örnek sorgularını vermiş olduğumuzu kabul edelim. Sorgu sonucunda gelen bağlantılar Google arama motorunun verdiği sonuç kümesini oluşturmaktadır. Arama yapılırken geri döndürülecek bağlantı sayısı bu çalışmada 10 ile sınırlandırılmıştır. Bu sayı arttırıldığı takdirde elde edilen adres cümleleri de artış gösterecektir. Bu bağlantıların içindeki adres bilgilerinin çıkarılarak, filtrelenen adres parçalarının (x, y) koordinatına dönüştürülmesindeki başarı ölçülmektedir. C# yazılım dili ile geliştirilen uygulama ek 2'de yer almakta olup, www.akutlu.com/default.aspx adresi yardımıyla da uygulama çalıştırılabilinir.

Adres bulmaya yönelik bu çalışmada hassasiyet (precision), "Adresi çözümlenen ve sorgu ile alakalı olan adres sayısı"nın "toplam çözümlenen adres sayısı"na oranını vermektedir. Geri-Çağırma (recall) ise "Adresi çözümlenen ve sorgu ile alakalı olan adres sayısı"nın "arama sonucunda elde edilen bağlantıların içindeki tüm adresler"e oranını vermektedir. Geri çağırmanın hassasiyete göre iki kat önem atfedildiği erişimin etkinliğini ölçmek için hesaplanan F-ölçüsü değeri (F2), 1'e yaklaştıkça daha iyi sonuçlar elde edildiğini ifade etmektedir.

Sorgu Kelimeleri: “Sultanahmet Otel”

İlgili sorgudan toplam 113 adet adres cümlesi elde edilmiştir. Bu adreslerin sadece 93 tanesi Şekil 6.1’de koordinatlandırılmıştır.



Şekil 6.1 Sultanahmet Otel sorgusu görseli

“Sultanahmet Otel” kelimeleri ile yapılan aramada adresi çözümlenen sorgu ile alakalı adres sayısı 87, çözümlenen adres sayısı 93, sorgu sonucu gelen bağlantılardaki anahtar kelime ile alakalı tüm adreslerin sayısı 167 olarak görülmüştür. Bu değerler ile,

$$P = 87 / 93$$

$$R = 87 / 167$$

$$F2 = \frac{\left(1+2^2\right)\cdot P \cdot R}{2^2 \cdot P + R} = 0.571$$

sonuçlarına ulaşılmıştır.

Sorgu Kelimeleri : “Beşiktaş Eczane”

İlgili sorgudan toplam 223 adet adres cümlesi elde edilmiştir. Bu adreslerin sadece 76 tanesi Şekil 6.2’de koordinatlandırılmıştır.

Şekil 6.2 Beşiktaş Eczane sorgusu görseli

“Beşiktaş Eczane” anahtar kelimesi verildiğinde, Beşiktaş ilçesinin sınırlarının dışındaki eczanelerin gelmesinin sebebi Google arama motorundan gelen sonuç kümesindeki sitelerde Beşiktaş ilçesinde olmayan eczane adreslerinin de yer almasıdır.

“Beşiktaş Eczane” kelimeleri ile yapılan aramada adresi çözümlenen sorgu ile alakalı adres sayısı 8, çözümlenen adres sayısı 76, sorgu sonucu gelen bağlantılardaki anahtar kelime ile alakalı tüm adreslerin sayısı 23 olarak görülmüştür. Bu değerler ile,

$$P = 8 / 76$$

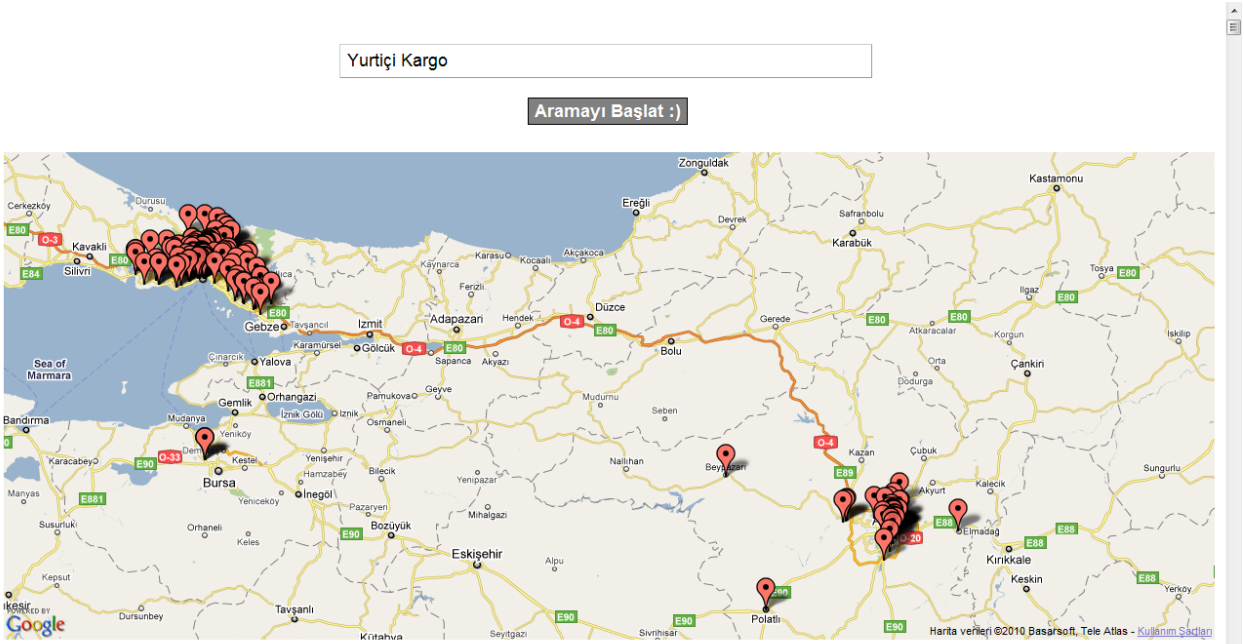
$$R = 8 / 23$$

$$F2 = \frac{(1+2^2)\cdot P \cdot R}{2^2 \cdot P + R} = 0.215$$

sonuçlarına ulaşılmıştır. Arama motorundan gelen bağlantılardaki adreslerin, Beşiktaş’ın dışında yer alması hassasiyet (P) değerinin düşmesine sebebiyet vermiştir.

Sorgu Kelimeleri: “Yurtiçi Kargo”

İlgili sorgudan toplam 411 adet adres cümlesi elde edilmiştir. Bu adreslerin sadece 169 tanesi Şekil 6.3’de koordinatlandırılmıştır.



Toplam 411 adet adres cümlesi bulunmuştur.Bu adreslerin 169 adedi koordinatlandırılmıştır.

Web Site
http://www.yurticikargo.com/ne_kadara_ana_test.asp
http://www.harbiforum.org/rehber/114816-yurtici-kargo-istanbul-telefon-numarasi-adres.html
http://www.kargosubesi.com/yurtici-kargo-ankara-kecioren-subesi.html
http://www.turkbusinesscenter.com/bilgimerkezi/kargo_sirketleri_listesi-222-0-0.html

Şekil 6.3 Yurtiçi Kargo sorgusu görseli

“Yurtiçi Kargo” kelimeleri ile yapılan aramada adresi çözümlenen sorgu ile alakalı adres sayısı 87, çözümlenen adres sayısı 93, sorgu sonucu gelen bağlantılardaki anahtar kelime ile alakalı tüm adreslerin sayısı 169 olarak görülmüştür. Bu değerler ile,

$$P = 87 / 93$$

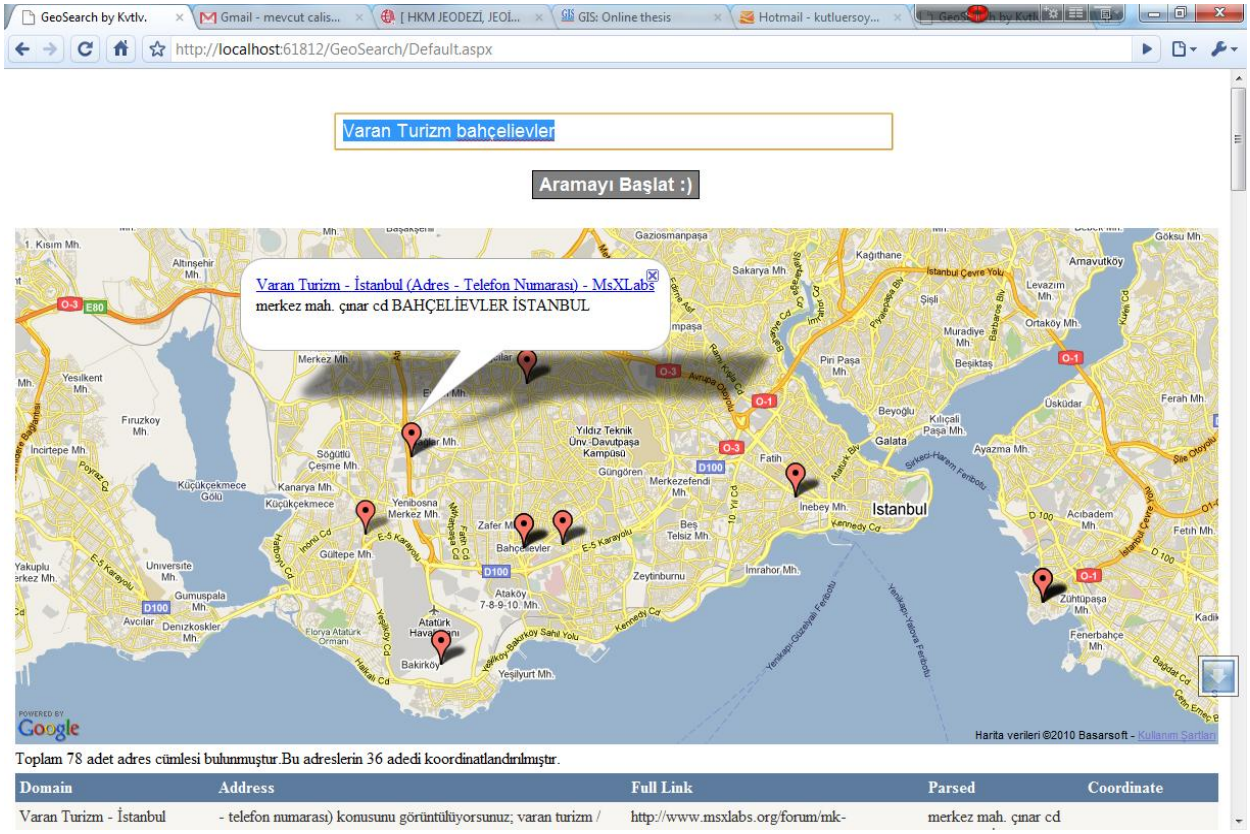
$$R = 87 / 169$$

$$F_2 = \frac{(1+2^2)\cdot P \cdot R}{2^2 \cdot P + R} = 0.566$$

sonuçlarına ulaşılmıştır.

6.2. Sonucu Bilinen Sorgunun Doğrulanması

İkinci yöntemde sonuç kümesi bilinen bir sorgunun doğru sonuçları üretmesi incelenmektedir. Örnek olarak Bahçelievler Yenibosna'da bulunan Varan Turizm'in otobüs terminaline ait adresin haritada tespit edilmesi için anahtar kelimeler olarak "Varan Turizm Bahçelievler" seçilmiştir.

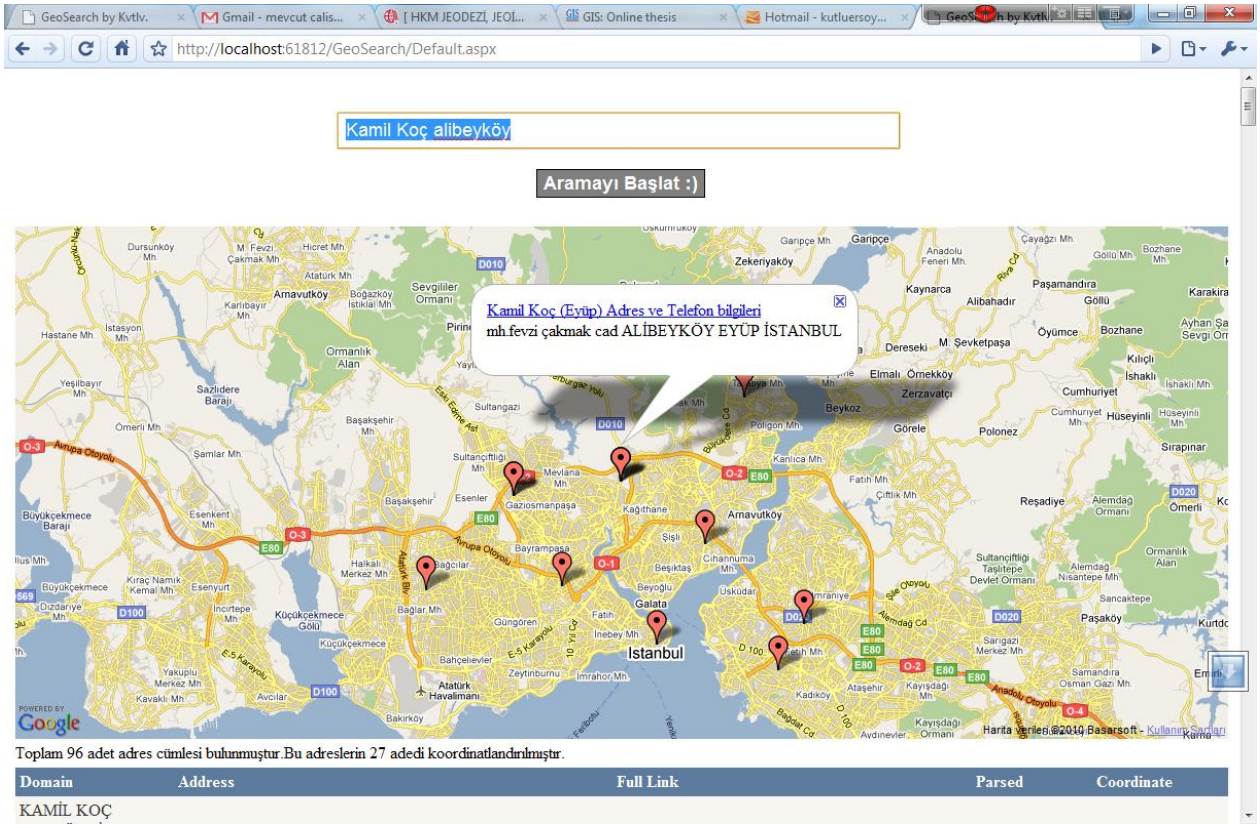


Şekil 6.4 Varan Turizm Bahçelievler sorgusu görseli

"Merkez mah. Çınar Caddesi" nde gösterilen adres Varan Turizm'in otobüs terminaline ait olan yeri göstermektedir. Şekil 6.4'de işaretlenen diğer adresler aynı firmanın ilgili sayfalarda bulunan diğer şubelerinin yerlerini belirtmektedir. Bu sonuç ile yeri bilinen bir arama sonucunun haritada sağlamasının yapılması gerçekleştirilmiştir.

Arama kelimelerinde terminale ait özel bir kelime seçilemediğinden, şube adresleri de sonuç kümesine dahil olmuştur.

Bir başka örnek olarak “Kamil Koç Turizm”e ait Alibeyköy’de bulunan otobüs terminali için arama kelimeleri olarak “Kamil Koç alibeyköy” anahtar kelimeleri tercih edildiğinde, sonuç kümesinden gelen adresler Şekil 6.5’deki gibi konumlandırılmıştır.



Şekil 6.5 Kamil Koç Alibeyköy sorgusu görseli

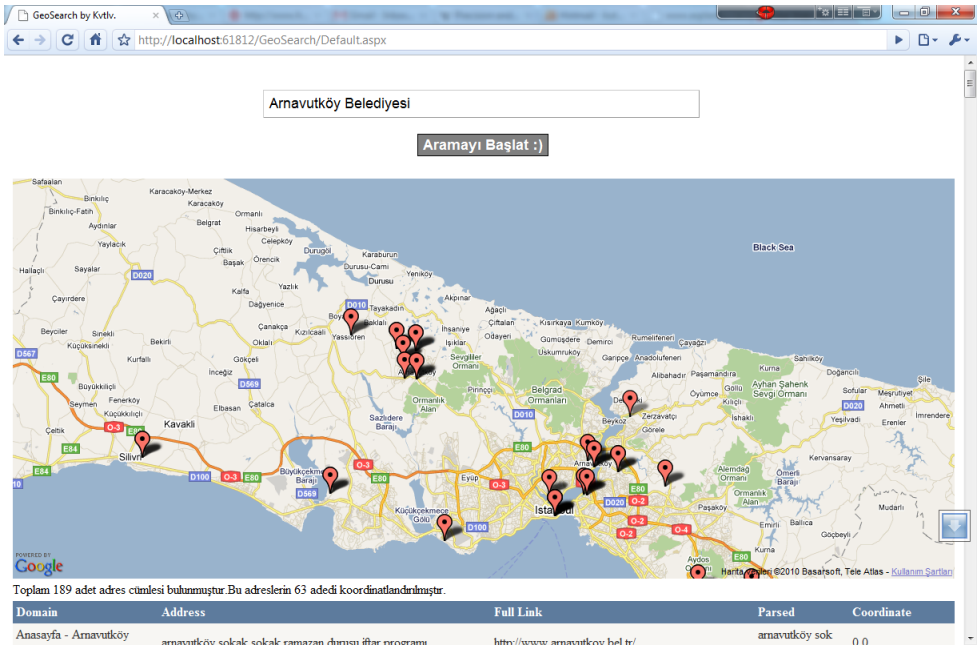
Bu sonuç kümesinde de aranılan otobüs terminali için “mh.fevzi çakmak cad ALİBEYKÖY EYÜP İSTANBUL” adresine ulaşılmıştır. Verinin doğruluğu için “http://tubiba.turkcebilgi.com/kamil-koc-guzeltepe/iletisim.html” bağlantısına girildiğinde adres bilgisi olarak “Kamil Koç (İstanbul) Mh.Fevzi Çakmak Cad. No: 230 Alibeyköy Eyüp, İstanbul,İstanbul, Türkiye 34726” yazısı ile karşılaşılmıştır. Veri kalitesinin düşük olduğu bu tip durumlarda da cadde bilgisinin yeterli olması sayesinde adresin koordinatının tespiti sağlanabilmektedir.

7. DEĞERLENDİRME

Aramanın anahtar kelimeleri olarak "Arnavutköy Belediyesi" seçildiğinde elde edilen sonuç kümesinin bir kısmı aşağıdaki Çizelge 7.1' ve görseli Şekil 7.1'de yer almaktadır.

Çizelge 7.1 Arnavutköy Belediyesi adres çıkarımı ve coğrafi koordinatları

İnternet Sitesinde Erişilen Adres Cümlesi	Bilgi Çıkarımı Yapıldıktan Sonra Elde Edilen Adres Cümlesi	Geocoding Fonksiyonu İle Elde Edilen Coğrafi Koordinatlar
arnavutköy sokak sokak ramazan durusu iftar programı	arnavutköy sok DURUSU	0
ramazan sokakta başka güzel	ramazan sok	40.7792, 29.96408
ilk sokak iftarı bollucada gerçekleşti	ilk sokak iftarı bollucad	0
anadolu mahallesi belnet	anadolu mah	36.97982, 35.35361
21 hadımköy toki / yassıören sokak iftar programı	yassıören sok	0
22taşoluk toki evleri / hacımaşlı köyü sokak iftar programı	hacımaşlı köyü sok TAŞOLUK	0
23arnavutköy merkez i.ö.o okulu bahçesi sokak iftar programı	23arnavutköy merkez i.ö.o okulu bahçesi sok	0
24anadolu mah.32-26-41-42 sok. kesiştiği nokta sokak iftar programı	24anadolu mah.32-26-41-42 sok	41.2725, 36.35965
25haraççı merkezi köyiçi sokak iftar programı	25haraççı merkezi köyiçi sok	0
26ömerli köyiçi // tayakadın sokak iftar programı	tayakadın sok ÖMERLİ	0
27arnavutköy salı pazarı sokak iftar programı	27arnavutköy salı pazarı sok	0
28boğazköy köy içi sokak iftar programı	28boğazköy köy içi sok	0
29imrahor köyiçi sokak iftar programı	29imrahor köyiçi sok	0
30arnavutköy suatlar pazar yolu sokak iftar programı	30arnavutköy suatlar pazar yolu sok	0
merkez mah. genç osman cad. no: 19 arnavutköy / istanbul	merkez mah. genç osman cad ARNAVUTKÖY İSTANBUL	41.18161, 28.73874
// // m. akif ersoy mah. atatürk cad. no: 79 arnavutköy / istanbul// //	m. akif ersoy mah. atatürk cad ARNAVUTKÖY İSTANBUL	41.21789, 28.73692
arnavutköy belediyesi resmi sitesi	ARNAVUTKÖY	41.23842, 28.62648
merkez mahallesi	merkez mah	41.04272, 29.16609
/ / yavuz selim mahallesi/ cemal çendek/ /	yavuz selim mah	39.94578, 32.87126
sarıyer ask.şb.bşk.	sarıyer ask	0
// // silivri ask.şb.bşk. // // // 212 // // // 727 31 60 // //	silivri ask	0



Şekil 7.1 Arnavutköy Belediyesi sorgusu görseli

Çizelge 7.1'de "Arnavutköy Belediyesi" sorgusunun sonucunda İnternet sitelerinden filtrelenen 189 adet cümlenin 21 tanesi örnek olarak çizelgeye taşınmıştır. Çizelgedeki cümlelerin adres çıkarımı işlemi sırasında, site içerisinde geçen yazıların adres olabilme ihtimaline karşın ayıklanma işlemine tabi tutulduğu görülmüştür. Adres bilgisi açık bir şekilde belirtilmeyen cümlelerde coğrafi kodlamanın başarısız olması ile bu cümleler IIMtCAR yazılımı tarafından göz ardı edilmektedir.

Sonuçların içerisinde adres bilgisi içerebilecek yazıların geçmesi, arama için kullanılan anahtar kelimelerin özelleşmiş bir aramadan çok, genel ifadeleri içeriyor olmasından kaynaklanmaktadır. Belirtilen anahtar kelimeler ile Arnavutköy Belediyesinin faaliyetlerini içerecek şekilde gelen sonuç kümesi harita üzerinde gösterilmektedir.

Şekil 7.1'de "Arnavutköy Belediyesi" kelimeleri ile yapılan aramada Arnavutköy'ün idari sınırları dışında coğrafi koordinatların çözümlendiği görülmektedir. Bu noktalardan iki tanesi Beşiktaş semtine bağlı olan Arnavutköy semtine aittir, geri kalan noktalar ise aynı aramada Google tarafından sonuç olarak dönen bağlantılardan gelmektedir. İstanbul Büyükşehir Belediyesi'nin sosyal tesislerinin bulunduğu site sonuç kümesine girmiş ve İBB'nin sosyal tesislerinin adresleri çözümlenerek haritada konumlandırılmıştır.

Arama ifadesinin kapsamı geniş tutulduğunda "Arnavutköy Belediyesi" örneğinde olduğu gibi sonuç kümesinden elde edilen konumlar da belirli bir alanda odaklanamamaktadır.

Arama sonuç kümesinde yer alan bağlantıların web standartlarına uyumsuzluğu adreslerin koordinat çözümlemesinde zorluklara yol açmaktadır. Web sayfalarında içerik taraması yaparken karakter seti aşağıdaki kod ile göz önünde bulundurulmaktadır. Uygulama geliştirilirken karakter seti bildirimi olmayan sayfalarda varsayılan olarak, Türkçe'ye uyumlu olan ve Türkiye'deki sitelerde de kullanılmakta olan, *"windows-1254"* değeri alınmıştır.

```
if (encoding != "UTF-8" && html.ToLower().Contains("charset=utf-8"))
        return GetWebContent(link, "UTF-8");
else
    return html;
```

Web sayfasının içerisinde "charset" parametresinin bulunmadığı durumlarda, Türkçe karakter problemi ile karşılaşılmakta ve temiz adres çıkarımı yapılmasını zorlaştırmaktadır. Aşağıda verilen örnekte karakter kümesi belirtilmeyen web sayfasının örneği Şekil 7.2'de görülmektedir.

```
KAMİL KOÇ OTOBÜS FİRMASI GENEL MERKEZ VE TERMİNAL TELEFON
NUMARA
halä°de edä°p adıvar cad. no:2/e
http://memurdostu.blogcu.com/kamil-koc-otobus-firmasi-genel-merkez-ve-terminal-
telefon-numara/1088327
p adıvar cad
0,0
```

Şekil 7.2 Karakter kümesi belirtilmeyen web sayfası içeriği

Adres cümlelerinin temizlenerek doğru adres bilgisinin çıkarım başarısı web sitelerinde kullanılan karakter setinin sayfa öznitelikleri içerisinde belirtilmesi ve adres cümlelerinin standartlara uygun olarak yazılması ile doğru orantılıdır.

Tez kapsamında hazırlanan uygulamada, mevcut internet içeriğinde metin üzerinden adres çıkarımını daha sağlıklı yapabilmek için arama için kullanılan anahtar kelimeler hedefe uygun olarak tercih edilmelidir. Arama kelimeleri olarak genel ifadeler kullanıldığında ve cadde, sokak gibi kelimelerin geçtiği fakat adres bilgisi olmayan içeriğe ulaştıran sorgulamalar yapıldığı takdirde sistem başarısını negatif yönde etkileyecektir.

8. SONUÇ VE ÖNERİLER

Kullanıcıların adrese ulaşmaya odaklı internet arama işlemlerinde, sorgularının sonuçlarını genişletmek ve derinleştirmek için metin olarak mevcut olan internet içeriğine de ulaşması gerekmektedir. Bu çalışma İnternette metin olarak mevcut olan adres bilgilerine de ulaşarak, kullanıcıya harita üzerinde bir sonuç döndürmeyi sağlamıştır. Harita üzerinde sunulan sonuç kümesi sayesinde kullanıcının görsel algısına hitap edilmiştir.

İnternet siteleri hazırlanırken, sayfa içeriğine eklenen adres cümleleri tek bir parça (tag) altında toplanmalı ve standartlara uygun olarak yazılmasına özen gösterilmelidir. Adres bilgisinin yer aldığı sayfa parçaları için bir W3 Standardı oluşturularak (W3C tarafından) coğrafi koordinat bilgisinin sayfanın içeriğinde tutulabilmesi internet içeriğinin CBS ile entegre olmasında büyük kolaylıklar sağlayacaktır.

İleriki çalışmalarda, sitelerden adres çözümlemesi ile elde edilen koordinatların kayıt altına alınması gerçekleştirilebilir. Sitelerde bulunan adresler sitenin isim alanı (domain) ile ilişkili olacaktır. Sitelerin kendi isim alanlarına ait elektronik posta adresleri ile kullanıcılar oluşturulabilir. Bu kullanıcılar kendi sitelerindeki adreslerin koordinatlarını düzenleyebilir. Bu sayede harita üzerindeki koordinatların doğruluk oranı arttırılabilir.

KAYNAKLAR

Andersen, P.M., (1992), "Automatic Extraction Of Facts From Press Releases To Generate News Stories", Proceedings Of The Third Conference On Applied Natural Language Processing, ANLC '92, 31 March-3 April, Trento, Italy.

Beel, J., Gipp, B. ve Stiller, J.O., (2009), "Information Retrieval On Mind Maps - What Could It Be Good For?", Proceedings Of The 5th International Conference On Collaborative Computing: Networking, Applications and Worksharing, CollaborateCom'09, 11-14 November, Washington.

Berners, L.T., (2009), "On The Next Web", TED Conference, 3-4 February, USA.

Bizer, C., Heath, T. ve Berners, L.T., (2009), "Linked Data – The Story So Far. International Journal On Semantic Web And Information Systems", International Journal On Semantic Web And Information Systems International Journal, IJSWIS, 7:4, 278-286.

Cowie, J., Wilks, Y., (1998),. "Information Extraction", Journal Of Documentation, 54:1, 70–105.

Doyle, Lauren; Becker, Joseph, (1975),. Information Retrieval and Processing, Melville, Los Angeles.

Goodrum, A.A., (2000), "Image Information Retrieval: An Overview of Current Research", Informing Science, 3:2, .63-7.

Fitzgerald, J.H, (2007), "Map Printing Methods", Archived from the original on 2007-06-04. http://web.archive.org/web/20070604194024

Foote, J., (1999), "An Overview Of Audio Information Retrieval", Journal Multimedia Systems - Special issue on audio and multimedia, 7:1, 307-328.

Frakes, W. B., (1992), "Introduction to Information Storage and Retrieval Systems", Information Retrieval: Data Structures and Algorithms, Englewood Cliffs, NJ: Prentice-Hall.

Korfhage, R.R., (1997), Information Storage and Retrieval, Wiley.

Lovison, Lucia., (2007), "Howard T. Fisher". Harvard University.

Olson, D.L., Delen, D., (2008), Advanced Data Mining Techniques, Springer.

Peng, F., McCallum, A., (2004), "Accurate. Information Extraction From Research Papers Using Conditional Random Fields", .In: Proceedings of HLT-NAACL, Boston, Massachusetts.

Shimizu, N., Hass, A., (2006), "Extracting Frame-based Knowledge Representation From Route Instructions", HLT-NAACL workshop on computationally hard problems and joint inference in speech and language processing, New York City.

Singhal, A., (2001), "Modern Information Retrieval: A Brief Overview", IEEE Data Engineering Bulletin, 24:4, 35-43.

Stamp, L.D., (1964), The geography of life and death, Ithaca, New York: Cornell University Press.

Tomlinson, R., (2007), Thinking about GIS: geographic information system planning for managers, ESRI Press, USA.

INTERNET KAYNAKLARI

[1] http://en.wikipedia.org/wiki/Information_extraction Erişim, Ağustos 2010

[2] http://wiki.osgeo.org/wiki/Open_Source_GIS_History Erişim, Ağustos 2010

[3] http://en.wikipedia.org/wiki/Information_retrieval Erişim, Ağustos 2010

EKLER

Ek 1 Türkiye’nin adres veritabanı

Ek 1 Türkiye'nin adres veritabanı

AdresNo	IlNo	Ilce No	SmtBck No	MhlKoy No	AdresAciklama	Posta kodu	EskiPosta kodu	UstAdres No
340000000	34	0	0	0	İSTANBUL	34000	34000	0
340100000	34	1	0	0	BAKIRKÖY	34140	34740	340000000
340101000	34	1	1	0	ZEYTİNLİK	34140	34740	340100000
340101014	34	1	1	14	ZEYTİNLİK MAH.	34140	34710	340101000
340102000	34	1	2	0	CEVİZLİK	34142	34720	340100000
340102006	34	1	2	6	CEVİZLİK MAH.	34142	34720	340102000
340102009	34	1	2	9	SAKIZAĞACI MAH.	34142	34720	340102000
340102011	34	1	2	11	YENİ MAH.	34142	34720	340102000
340103000	34	1	3	0	KARTALTEPE	34144	34740	340100000
340103007	34	1	3	7	KARTALTEPE MAH.	34144	34740	340103000
340103008	34	1	3	8	OSMANİYE MAH.	34144	34730	340103000
340104000	34	1	4	0	ZUHURATBABA	34147	34740	340100000
340104015	34	1	4	15	ZUHURATBABA MAH.	34147	34740	340104000
340105000	34	1	5	0	YEŞİLKÖY	34149	34800	340100000
340105012	34	1	5	12	YEŞİLKÖY (ŞEVKETİYE) MAH.	34149	34800	340105000
340105013	34	1	5	13	YEŞİLYURT MAH.	34149	34800	340105000
340106000	34	1	6	0	FLORYA	34153	34820	340100000
340106005	34	1	6	5	BASINKÖY ZÜMRÜTYUVA MAH.	34153	34820	340106000
340106010	34	1	6	10	ŞENLİK MAH.	34153	34810	340106000
340107000	34	1	7	0	ATAKÖY	34156	34740	340100000
340107001	34	1	7	1	ATAKÖY 1. KISIM MAH.	34158	34710	340107000
340107002	34	1	7	2	ATAKÖY 2-5-6 KISIM MAH.	34158	34710	340107000
340107003	34	1	7	3	ATAKÖY 3-4-11. KISIM MAH.	34158	34710	340107000
340107004	34	1	7	4	ATAKÖY 7-8-9-10 KISIM MAH.	34156	34740	340107000
340200000	34	2	0	0	BAYRAMPAŞA	34030	34160	340000000
340201000	34	2	1	0	NUMUNEBAĞ	34030	34140	340200000
340201007	34	2	1	7	ORTA MAH.	34030	34020	340201000
340201010	34	2	1	10	YENİDOĞAN MAH.	34030	34140	340201000
340202000	34	2	2	0	ALTINTEPSİ	34035	34160	340200000
340202001	34	2	2	1	ALTINTEPSİ MAH.	34035	34160	340202000
340202008	34	2	2	8	TERAZİDERE MAH.	34035	34160	340202000
340202009	34	2	2	9	VATAN MAH.	34035	34160	340202000
340203000	34	2	3	0	MURATPAŞA	34040	34150	340200000

AdresNo	IlNo	Ilce No	SmtBck No	MhlKoy No	AdresAciklama	Posta kodu	EskiPosta kodu	UstAdres No
340203003	34	2	3	3	İSMETPAŞA MAH.	34040	34150	340203000
340203004	34	2	3	4	KARTALTEPE MAH.	34040	34150	340203000
340203006	34	2	3	6	MURATPAŞA MAH.	34040	34150	340203000
340204000	34	2	4	0	YILDIRIM	34045	34170	340200000
340204002	34	2	4	2	CEVATPAŞA MAH.	34045	34170	340204000
340204005	34	2	4	5	KOCATEPE MAH.	34045	34170	340204000
340204011	34	2	4	11	YILDIRIM MAH.	34045	34170	340204000
340300000	34	3	0	0	BEŞİKTAŞ	34330	80690	340000000
340301000	34	3	1	0	LEVENT	34330	80600	340300000
340301010	34	3	1	10	KONAKLAR MAH.	34330	80600	340301000
340301014	34	3	1	14	LEVENT MAH.	34330	80600	340301000
340302000	34	3	2	0	AKATLAR	34335	80630	340300000
340302002	34	3	2	2	AKATLAR MAH.	34335	80630	340302000
340303000	34	3	3	0	ETİLER	34337	80600	340300000
340303008	34	3	3	8	ETİLER MAH.	34337	80600	340303000

ÖZGEÇMİŞ

Doğum tarihi	07.01.1984	
Doğum yeri	Ankara	
Lise	1995-2001 2001-2002	Beşiktaş Atatürk Anadolu Lisesi Taksim Atatürk Lisesi
Lisans	2002-2006	İstanbul Kültür Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü
Yüksek Lisans	2007-…	Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Müh. Anabilim Dalı, Bilg. Müh. Programı