

Coğrafi Yer Bilgilerinin Elde Edilmesi ve Sorgu Genişlemesi Yöntemi ile Sorgulanması

Ömer Sevinç¹, Erdal Kılıç²

¹ Ondokuz Mayıs Üniversitesi, Bilgisayar Programcılığı Bölümü, Samsun

² Ondokuz Mayıs Üniversitesi, Bilgisayar Mühendisliği Bölümü, Samsun

osevinc@omu.edu.tr, ekilic@omu.edu.tr

Özet: Günümüzde coğrafi veriler üzerinden geliştirilen uygulamalar çok popüler hale gelmiştir. Örneğin GPS ile yön bulma, otel aramaları veya Google Maps gibi her türlü coğrafi konum arama uygulamaları insan hayatını kolaylaştırmaktadır. Bu çalışmada, sorgu genişlemesi metodunu kullanarak coğrafi konum arama için yeni bir yaklaşım sunulmuş, önerilen yöntem koordinatsal verileri arama sonuçlarını daha da geliştirmek için kullanılmıştır. Kelime tabanlı arama yöntemleri ile yapılan sorgulamalarda ne eş anlamlı kelimeler tespit edilebilmekte ne de aranan bir yere yakın konumda ki yerler bulunabilmektedir. Oysa sorgu genişlemesi yöntemi ile hem eş anlamlı kelimeler hem de yakındaki yerler sorgulanabilmektedir. Bu çalışmada önerilen yöntemde coğrafi konum bilgileri Vikipedi XML çıktı dosyalarından alınıp düzenlenmiş ve bunlardan istenilen veriler çekilmiştir. Geliştirilen bir program aracılığı ile veriler düzenlenmiş ve oluşturulan dosyanın SOLR ile indekslenip aratılabilmesi için istenilen XML formatında kaydedilmesi sağlanmıştır. Çalışmada son olarak, kelime tabanlı sorgulama ile sorgu genişleme tekniği kullanılarak zenginleştirilmiş kelime tabanlı sorgulama sonuçları karşılaştırılmıştır. Sonuçlar sorgu genişlemesi tekniğinin sorgu sonuçlarını önemli derecede iyileştirdiğini göstermiştir.

Anahtar Sözcükler: veri bulup getirme, sorgu genişlemesi, anlamsal ağ, anlamsal indeksleme.

Abstract: Nowadays the applications developing over geospatial data became very popular. For instance we can talk about GPS navigations, hotel searches, or any kind of geospatial search applications like Google Maps which make our life easier. In this paper, an approach is presented for geospatial searching that uses query expansion method which is used on geospatial data to make more improvement on search results. While a keyword-based search is neither able to detect synonyms nor able to detect the places closes to the searched place, query expansion allows both searching on synonyms and close places. Geospatial information is retrieved from Vikipedi XML output and eliminated to handle the expected data. We organized and saved data in a new XML format with an application that we developed, for the file being indexed and searched with SOLR. At the end we compared the performance of keyword based search systems with keyword based search system that is enriched with query expansion method. Results demonstrate that query expansion technique significantly improves search results.

Key Words: information retrieval, query expansion, semantic web, semantic indexing.

1. Giriş

Coğrafik verilerin internet ortamında aratılması toplam internet aramalarının önemli bir kısmını oluşturmaktadır. İnternet’de yapılan aramalar, verinin bulunup elde edilmesi ve görüntülenmesi anlamına gelmektedir. Verilerin elde edilmesi noktasında mevcut uygulama, metinsel veriler üzerinden kelime tabanlı arama yapılması prensibine dayanmaktadır ki bu da kelime torbası modeli olarak ifade edilmektedir. Bununla birlikte böyle bir model, metin içerisindeki anlamsal bilgiyi kaçırmaktadır. Bu durumun üstesinden gelebilmek için ontolojik yöntemler önerilmiştir[1]. Sorgu genişlemesi tekniği de arama sonuçlarını iyileştirilmesi sağlayabilen bir başka yöntemdir. Sorgu genişlemesi; verilerin bulunup getirilme işlemlerindeki bulup geri getirme performansının iyileştirilebilmesi için orijinal sorgunun yeniden düzenlenmesi işlemidir[2]. Arama motorları bağlamında, sorgu genişlemesi kullanıcının girdisinin değerlendirilmesini ve arama sorgusunun daha fazla dokümanla eşleşmesi için genişletilmesini gerektirmektedir[3].

Bu çalışmada, coğrafik verilerin sorgulanmasında sorgu genişlemesi tekniğini kullanılmıştır. Mevcut geleneksel kelime tabanlı arama tekniğini geliştirerek alternatif sonuçları ve eşanlamlı kelime gruplarını da listeleyen bir uygulama hedeflenmiştir. Bu metot ile önce Vikpedi’den elde edilen coğrafik veriler sorgulanmış olup mevcut kelime tabanlı arama yöntemlerinden daha iyi sonuçlar elde edilmiştir. Öncelikle coğrafik veriler açık bilgi kaynağı olan Vikpedi’den XML dosyası formatında alınmış, daha sonra da hedeflenen bilgileri yeni bir XML formatında düzenleyip kaydeden bir otomasyon geliştirilmiştir. Kaydedilen veriler, bir açık kaynak indeksleme ve arama programı olan SOLR ile indekslenerek hem eş anlamlı kelimeleri içeren sorgu sonuçların hem de aratılan bir coğrafik konuma belirli bir mesafedeki coğrafik konumları da içeren sorgu sonuçları görüntülenmiştir. Böylece normalde yalnızca kelime bazında arama yapan arama motorları ile tespit

edilemeyecek eş anlamlı ifadeler ve aratılan herhangi coğrafik yere yakın konumdaki diğer yerlerin de sorgu sonuçları içerisinde görüntülenebilme olanağı sağlanmıştır.

2.Önceki Çalışmalar

Sorgulama tekniklerinin geliştirilmesi üzerine bugüne kadar çok çeşitli çalışmalar yapılmıştır. Bu çalışmalar da ters indeksleme, vektör uzay modeli, sorgu genişlemesi, coğrafik indeksleme ve anlamsal indeksleme gibi teknikler kullanılmıştır[7, 8, 9]. Bazı çalışmalar ise arama sonuçlarını iyileştirmek için bu teknikleri birleştirmişlerdir[9,13]. Son çalışmalar ise sorgu genişlemesi ve anlamsal indeksleme üzerine yoğunlaşmıştır[11,13,14]. Bu çalışmada sorgu genişlemesi tekniğini coğrafik veriler üzerinde deneyerek sorgu sonuçlarını iyileştirmeye odaklanılmıştır. Sorgu genişlemesi kullanıcının sorgusuna ilgili yeni terimleri de ekleyerek daha çok sorgu sonucu elde edilebilmesini sağlamaktadır. Bu eklemeler otomatik şekilde veya kullanıcıya yapılan önerilerle etkileşimli bir şekilde yapılabilmektedir.

Ters indeksler klasik indeksleme tekniği olarak bilinmektedir. Ters indeksleme metin içerisindeki her kelimeyi işaretçiler ile kelimelerin dokümanlarda görüldükleri pozisyonlara eşlemektedir. Bu eşlemelerden oluşan listenin tamamı bulunanlar ya da ortaya çıkanlar olarak adlandırılmaktadır [4]. Ters indeksleme de coğrafik bilgilerle ilgili bir sınıflandırma bulunmamaktadır. Mesela bir ilçenin bir şehir içerisinde yer alması gibi bir durum göz ardı edilmektedir.

Bilinen geleneksel kelime tabanlı veri bulup getirme yaklaşımları Salton tarafından teklif edilen vektör uzay modeli üzerine kurulmuştur [5]. Bu yöntemim uygulanması kolay olmakla birlikte kesinlik düşüktür. Diğer taraftan konum bilgileri sorgulamak için koordinatsal indeksleme yapıları da kullanılmıştır. Bu yöntemlerden en çok bilineni ve öğretici olanı R-Tree’ dir [6].

Bazı çalışmalarda ters indeksleme ile konumsal indeksleme bir arada kullanılmıştır [7, 8, 9].

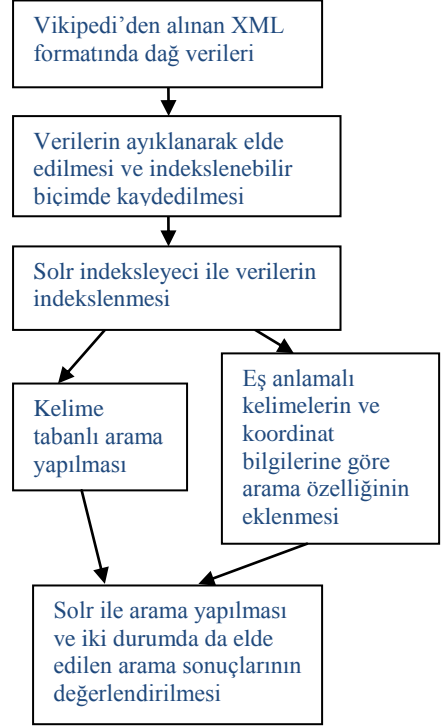
Ontolojiler, paylaşılan kavramların resmi, kesin özelliklerini belirlerler ve coğrafik alanların da karakteristik özelliklerini tanımlayabilirler [10]. Ontolojiler kullanılarak sınıflar, sınıflar arası ilişkiler, özellikler ve sınıf örnekleri tanımlanabilir. Bir sınıf bir şeyin genel özelliklerini tanımlar. İlişkiler ise ait olmak, içinde olmak gibi yüklemeldir.

Bir başka yeni yaklaşım ise sorgu genişlemesidir [11,12]. Sorgu genişlemesi kullanıcı sorgusunu elde edilmek istenen bilgi ile sorgu arasında ki boşluğu kapatacak şekilde zenginleştirerek elde edilebilecek sorgu sonuçlarını iyileştirir. Bu yöntem de arama sonuçlarını belirleyebilmek için bir ara yüz kullanıcıya ilgili olabilecek sonuçları göstermektedir. Bunlar yapılırken kullanıcıyı yönlendirmek yerine işlemler olasılık hesabı ile otomatik şekilde yazılım tarafından değerlendirilmekte ve ilgili sonuçlar sunulmaktadır. Bir başka yeni yaklaşım ise anlamsal indekslemedir. Bu tür çalışmaların sorgu genişlemesi yönteminden daha başarılı sonuçlar verebildiği ifade edilmiştir [13]. Anlamsal indeksleme metodu ile veriler kelime tabanlı arama ile aranabilmektedir. Bu yöntemde ontoloji örnekleri oluşturulur, veriler ontoloji veri dosyaları olan OWL dosyalarından elde edilir ve indekslenerek aratılır.

Bu çalışmada önerilen sorgu genişlemesi yöntemini coğrafik yer ve koordinat verileri üzerinde uygulanmıştır. Yöntemde eş anlamlı ve yakın anlamlı kelimelerin de aynı anda aratılabilmesi sağlanmış, bununla birlikte aratılan bir konumun koordinat bilgilerini kullanarak bu konuma 20 km uzaklık içerisinde bulunan diğer konum bilgilerini de sonuçlar içerisinde yer alması sağlanmıştır. Böylece bir konum bilgisi aratılırken kelime tabanlı arama ile elde edilen sonuçlar daha fazla ilgili sonuç görüntüleyerek iyileştirmiş olmuştur.

3. Önerilen Yöntem

Önerilen yöntem Şekil 1.'de grafiksel olarak gösterilmiştir.



Şekil 1. Önerilen Yöntem

Bu çalışmada sorgu genişlemesi üzerine odaklanılmıştır. Wikipedia'dan elde edilen veriler, ayrıştırılıp ayıklanmasında ve yeni bir formatta saklamasında Java kütüphanelerinden, verilerin indekslenerek aratılmasında ise SOLR indeksleme yazılımından faydalanılmıştır.

3.1 Sorgu Genişlemesi

Sorgu genişlemesi üzerine yapılmış çalışmalar bulunmaktadır. Sorgu genişlemesine, doğal dillerden kaynaklanan anlam bulanıklığından ve belli bir içeriğe ait bir bilginin tek bir terim olarak ifade edilmesindeki zorluklardan

dolayı ihtiyaç duyulmuştur. Sorgu genişlemesinde kullanıcı sorgunun tekrar düzenlenebilmesi için yönlendirilerek daha alakalı sonuçlar elde edilmeye çalışılır. Sorgu genişlemesinin temel hedefi başlangıçtaki sorguya yeni anlamlı terimler eklemektir. Terim ekleme işlemi elle, otomatik olarak veya kullanıcı yardımı alınarak yapılabilir. Elle yapılan sorgu genişlemesi tecrübeye dayanmaktadır. Otomatik yöntemde ağırlıklar bütün terimler için hesaplanır ve en yüksek ağırlık değerine sahip olan terimler başlangıç sorgusuna eklenirler. Farklı ağırlık hesaplama fonksiyonları farklı sonuçlar üretmektedir. Dolayısıyla verilerin elde edilme performansı ağırlıkların nasıl hesaplandığına bağlıdır. Kullanıcı destekli sorgu genişlemeyle, sistem olası sorgu genişlemesi terimlerini üretmekte ve kullanıcı bu terimler içerisinden hangisini aratmak istiyorsa onları seçmektedir [14].

3.2 Sax (XML için Basit Arayüz Uygulaması)

Sax; bir XML dokümanından verinin okunabilmesi için DOM (Doküman Nesne Modeli)' a alternatif bir yöntemdir. Sax XML dokümanını bir bütün olarak işlemek yerine her bir parçası üzerinde sırasıyla işlem yapar. Sax'ın resmi kuralları yok ise de Java uygulamaları kurallı olarak değerlendirilebilir. Daha az hafıza kullanarak nispeten daha detaylı XML dosyaları işleyebilmekte ve daha çok verinin aynı anda elde edilebilmektedir. [15]

Üzerinde çalışılan coğrafik konum verilerinden istenildiği gibi faydalanabilmek ve indeksleyebilmek için XML alanları içerisinden veriler çekilerek yeni bir formatta kaydedilebilmesi sağlandı. Bunun için de Java SAX kütüphanesinden faydalanıldı.

3.3 Solr

Solr Apache Lucene projesine

bağlı son derece hızlı, popüler bir açık kaynak arama platformudur. Genel olarak, güçlü tam metin araması, önemli kısımların vurgulaması, dinamik bölümlenme, veritabanı birleştirme, zengin doküman işleme ve coğrafik arama gibi özellikleri içermektedir. Solr yüksek derecede ölçeklenebilir olup, dağıtık arama ve indeks kopyalama gibi özellikler sağlar. Dünyaca büyük internet sitelerinin arama ve navigasyon özelliklerine güç katmaktadır.

Java ile kodlanmış olup içerdiği Tomcat sunucusu ile tek başına tam metin araması yapabilmektedir. Solr Lucene Java kütüphanesini tam metin arama ve indekslemede çekirdek olarak kullanır ve görsel olarak programlama dillerinin kullanımını kolaylaştıran HTTP/XML ve JSON uygulama ara yüzlerine de sahiptir. Solr'ün güçlü dışarıdan ayarlanabilme özelliği ile birçok türden uygulama Java kodlaması yapmadan gerçekleştirilmekte ve daha ileri düzeyde uygulamalara ihtiyaç duyulduğunda geniş eklentileri ile uygulama özelleştirilebilmektedir [16].

6. Değerlendirme ve Sonuçlar

Çalışmamızda Amerika Birleşik Devletlerinin New Mexico eyaletinde bulunan dağlara ve bu dağların buldukları il ve ilçelere ait isim, koordinat ve yön bilgileri örnek veri olarak seçilmiştir. Bu bilgiler Vikipedi'den XML formatında elde edildi. Bu veriler içerisinden ihtiyaç duyacağımız, isim, koordinat ve yön bilgileri ayıklanarak elde edildi. Ayrıca hangi dağın hangi il veya ilçe de bulunduğu bilgisi de veri kaynağından çekilerek dağ bilgisinin içeriğine eklendi.

Elde edilen verileri indekslemek ve aramak için kullandığımız açık kaynak kodlu Solr yazılımına verilerin uygun biçimde ve XML formatında kaydedilmesini sağlandı. Bunun için Java ile geliştirdiğimiz yazılım Vikipedi dosyası girdi olarak alınıp ilgili veriler Sax ile elde edilmiş ve bunlar uygun biçimde

XML dosyasına Solr yazılımın ekleyebileceği düzende kaydedilmiştir. Bu girdi dosyasına ait bir bölüm aşağıda görülmektedir.

Kod1 :

```
<doc>
<field name="id"> 16</field>
<field name="name"> Salinas Peak
</field>
<fieldname="is_in">Sierra County</field>
<fieldname="location2">33.2881,-
106.5319</field>
<field name="within">White Sands
Missile Range</field>
<field name="direction">south</field>
<fieldname="dir_place">New
Mexico</field>
<field
name="direction2">northwest</field>
<fieldname="dir_place2">Alamogordo
</field>
<field name="descript">lief. located
within the White Sands Missile Range,
and hence is closed to the public. </field>
</doc>
```

Solr yazılımına dışarıdan eklenecek veriler XML formatında dokümanlar ve alanlar olarak düzendi. Her bir dağ bilgisi bir doküman olarak <doc> etiketi içerisinde kaydedildi. Her bir dağa ait yer, yön koordinat bilgisi ise ilgili <doc> etiketi içerisinde alanlar içerisine yerleştirildi. Alanların her biri bir isme ve türe sahiptir. Alanlar <field> etiketi ile ifade edilir. Şema dosyasında alanlara ait özellikler önceden elle tanımlandı. Aşağıda bir örnek bu tanımlamayı göstermektedir.

Kod2 :

```
<field name="name" type="text_general"
indexed="true" stored="true"/>
```

Kod2’ de “name” olarak isimlendirilen isim alanının metinsel bir alan olduğu, indeksleneceği ve kaydedileceği, bu alana ait özelliklerin doğru veya yanlış olarak ayarlanması ile belirtilmiştir.

Bir dağın adı, enlem ve boylam

bilgileri, hangi bölgede yer aldığı, hangi bölgenin hangi yönünde (kuzey, güney) kaldığı, hangi iki bölge arasında kaldığı gibi bilgileri her bir dağ için ayrı birer doküman oluşturacak şekilde alanların içerisine ilgili bilgiler yerleştirildi. Bu, kodladığımız program ile otomatik olarak gerçekleştirildi. Elde ettiğimiz düzenli biçimdeki veri dosyası indekslenmek üzere Solr yazımlı içerisine eklendi. Ekleme işleminden sonra indekslenen terimlerin frekans değerleri Şekil-1 deki gibi oldu.

Top 10	Terms
term	frequency
peak	15
mountain	9
baldy	6
sierra	3
mount	3
volcano	2
black	2
crater	2
grande	2
blue	1

Şekil-1 Solr programının Şema Analiz ekranından bir bölüm

Yukarıdaki şekilde sadece isim alanlarında geçen terimlerin frekans değerleri görünmektedir. İsim alanı dışında yer bilgisi, yön bilgisi, tanım bilgisi gibi alanları da indekslendi. Arama sonuçlarında bu alanlar içerisinde geçen terimler de görüntülenmektedir. Herhangi bir dağ, tepe, sıradağ veya ilçe ismi arattığımızda doğrudan arattığımız kelimenin içinde yer aldığı dokümanlar sorgu sonucu olarak listelenmektedir.

Sorgu genişlemesinin uygulanması noktasında ise Solr yazılımının “Synonym.txt” dosyası içerisinde eş ve yakın anlamlı kelimelerin eşlenmesi sağlandı. Tepe ve dağ ifadeleri, dağ ve sıradağ ifadeleri, biri aratıldığında diğerine ait sonuçlar da listelenebilecek şekilde eşlendi. Dosya içerisinde aşağıdaki gibi bir eşleme oluşturuldu.

mountain => peak range
peak => mountain range

Böylece arama sonuçlarında listelenmeyen

eş ve yakın anlamlı ifadelerde de sorgu sonuçları içerisinde görüntüldü. Diğer bir sorgu genişlemesi ise dağ ve yer bilgilerine enlem ve boylamların sayısal değer türünde kaydedilebilmesi için Solr ayar dosyasında düzenlemeler yapıldı. Otomasyonumuzla, koordinat bilgileri Solr yazılımınca elde edip üzerinde matematiksel işlem gerçekleştirebilecek şekilde düzenlendi. Böylece indekslediğimiz dağ, yer vb. bilgilerinin yanında bu konumlara belli uzaklıklardaki yer ve dağ bilgilerini de içerecek şekilde sorgu sonuçları listelendi. Örneğin Baldy dağı 36.6269, -105.2133 koordinatlarında bulunmaktadır. Solr yazılımının koordinat verileri üzerinden de sorgulama yapabilme özelliğini kullanarak sorgumuza eklediğimiz bir filtre ile 25 km uzaklıktaki dağların ve ilçelerin listelenmesi sağlandı. Filtre eklentisi aşağıdaki gibidir.

{!geofilt%20sfield=location2}&pt=36.6269,-105.2133&d=25

Aramamızı bu şekilde genişlettiğimizde yakın bölgedeki iki dağ bilgisi daha sorgu sonuçları içerisinde listelenebilmektedir. Hem eş ve yakın anlamlı kelimelerin sorgu cümlesine eklenmesi ile hem de özellikle bu çalışmada üzerinde durduğumuz coğrafik konum aramalarında koordinat bilgilerinin kullanılarak belirli mesafelerdeki konum bilgilerinin de sorgu sonuçlarında listelenmesi ile sorgu genişlemesi sağlandı. Alakalı olabilecek sorgu sonuçları artırılmış ve mevcut kelime tabanlı arama yöntemi geliştirilerek daha iyi sonuç elde edildiği görüldü. 41 adet dağ bilgisi dokümanı üzerinden indeksleme yapıldı. Bunların 20 tanesinde yer bilgisi olarak doğrudan New Mexico ifadesi bulunmaktadır. Bu kayıtlar içinde 2 tanesinin ad bilgisi içinde "mountain", 9 tanesinin ad bilgisi "peak" ismi geçmektedir. Bu iki isim yakın anlamlı olarak kabul edilerek "synonym.txt" dosyasında eşleştirildiğinde arama her iki isimle de arama yapıldığında 11 adet sonuç döndürmektedir. Bu durum Tablo-1 de kesinlik(precision) ve veri çağırma (recall) formülleri ile hesaplanarak

gösterilmiştir. New Mexico bulunan dağlar "Mountain" kelimesi ile yapılan arama da 2 adet olarak elde edilirken, eşlemeden sonra 11 olarak elde edilmiştir. Peak kelimesi ile arama yapıldığında ise normalde 9 adet sonuç döndürürken eşlemeden sonra 11 adet sonuç döndürmüştür.

Yine Baldy ismindeki dağ koordinat verileri ile aratıldığında bu dağa ait bilgiler görüntülenirken koordinat verileri üzerinden filtreleme yapılarak 25 km mesafe için enlem ve boylam bilgileri genişletilerek aratılmış ve bu alan içerisinde bulunan 2 adet dağ bilgisi daha sorgu sonuç listesinde görüntülenmiştir. Toplamda 3 adet dağ bilgisi listelenmiştir. Bu durum ise Tablo-2 de gösterilmiştir.

S-1 : mountains in new mexico

S-2: peak in new mexico

S-3 : baldy mountain

S o r g u	Normal arama		Eş anlamlılar eklenerek arama	
	Veri Çağırma Recall	Kesinlik Precision	Veri Çağırma Recall	Kesinlik Precision
1	%100	%10	%100	%55
2	%100	%45	%100	%55

Tablo-1

Sorgu Numarası	Normal sorgu sonucu elde edilen kayıt sayısı	25 km mesafe araması sonucu elde edilen kayıt sayısı
3	1	3

Tablo-2

Elde edilen bu sonuçlarla arama sonuçlarının kesinlik değerlerinin iyileştiği anlaşılmıştır. Koordinat bilgisi kullanılarak yakın bölgedeki yerler aratıldığında ise elde edilen ilgili sorgu sonuç sayısının arttığı görülmüştür. Böylece sorgu genişleme yöntemleri kullanılarak, normal kelime tabanlı arama ile elde edilecek sonuçlardan daha iyi sonuç edilebildiği anlaşılmıştır.

7. İleri Çalışmalar ve Öneriler

Çalışmamızda sorgu genişlemesi yöntemi kullanılmış olup ilerleyen aşamaların Vikipedi'den elde ettiğimiz veriler de kullanılarak ontolojiler oluşturulacaktır. Ontolojiler sayesinde yer bilgisi ile ilgili yerlerde bulunan dağ bilgileri arasında anlamsal bir ilişki kurulabilecektir. Bir dağ bilgisi aratıldığında aynı bölge bulunan dağ bilgileri kolaylıkla listelenebilecektir. Bunun yanında çıkarsama yöntemleri kullanılarak bir birine yakın içe ve şehir bilgileri elde edilebilecek ve yakın çevredeki dağ bilgileri de bu bilgilerden faydalanılarak sonuçlarda listelenebilecektir.

Ayrıca anlamsal indekisleme tekniği ile frekans değerleri elde edilmiş verilerin matris tablosu SVD ile tekrar hesaplanacak ve eigen değerleri yüksek olan verilere ait doküman grupları sorgu sonuçlarında listelenebilecektir. Sonuçta bu iki yöntemle beraber sağladıkları katkılar de değerlendirilerek karma bir yöntem elde edilmeye çalışılacaktır.

Kaynaklar

[1] T.R. Gruber, Toward principles for the design o ontologies used for knowledge sharing, *International Journal of Human Computer Studies* 43, pp. 907-928, 1995.

[2] Vectomova, Olga; Wang, Ying (2006). "A study of the effect of term proximity on query expansion" (Abstract). *Journal of Information Science* 32 (4): 324–333.

[3]http://en.Wikipedi.org/wiki/Query_expansion, 12.09.2012.

[4] Baeza-Yates R, Ribeiro-Neto B, *Modern information retrieval*. Addison-Wesley, Harlow, 1999.

[5] G. Salton, A. Wong, C.S. Yang, A

vector space model for automatic indexing, *Communications of ACM* 18, pp. 613-620, 1975.

[6] Guttman A (1984) R-Trees: a dynamic index structure for spatial searching. In: Yormark B (ed) SIGMOD'84, proceedings of annual meeting, Boston, Massachusetts, pp. 47-57, June 18-21, 1984. ACM, New York.

[7] Amitay E, Har'El N, Sivan R, Soffer A (2004) Web-q-where: geotagging web content, pp. 273-280, In: SIGIR '04: proceedings of the 27th ACM, New York, <http://doi.acm.org/10.1145/1008992.1009040>

[8] Rauch E, Bukatin M, Baker K, A confidence-based framework for disambiguating geographic terms. In: *Proceedings of the HLT-NAACL 2003 workshop pn analysis of geographic refremces*. Association for Computational Linguistics, Morrystown, USA, pp 50-54,2003.

[9]Lieberman MD, Samet H, Sankaranarayanan J, Sperling J, STEWARD: architecture of a spatio-textual search engine. In: *Proceedings of the 15th ACM Int. Symp. on Advances in GIS (ACMGIS'07)*. ACM New York, pp. 186-193, 2007.

[12] Gruber TR, A translation approach to portable ontology specifications. *Knowl Acquis*5(2), pp. 199-200.

[11] A. Neda, P. Pallabi, M. Sheetal, K. Latifur, S. B. Steven, T. Bhavani, *Ontology-driven query expansion methods to facilitate federated queries*, IEEE, 2010.

[12] C. G. V. Guillermo, A. Lylia, C. Nadine, A query expansin method applied to water information system, 5th *International Conference on Signal Image Technology and Internet Based Systems*, 2009.

[13] K. Soner, A. Ozgur, S. Orkunt, A. Samet, C. K. Nihan, A. N. Ferda, "An ontology-based retrieval system using semantic indexing," Elsevier, Information Systems pp. 294-305, 37, 2012.

[14] J. Bhogal, A Macfarlane, P Smith, A review of ontology based query expansion, Information Processing Management, Science Direct, 2007.

[15] Simple API for XML, XML için Basit Uygulama Ara yüzü
http://en.wikipedia.org/wiki/Simple_API_for_XML , e.t : 19.09.2012.

[16]Apache Solr,
<http://lucene.apache.org/solr/>, e.t:
19.09.2012

[17] Latent Semantic Indexing,
http://en.wikipedia.org/wiki/Latent_semantic_indexing, 20.09.2012