

İlişkisel Veri Tabanlarında Anahtar Kelime Arama

Serap Demircioğlu, Suat Özdemir

Bilgisayar Mühendisliği, Mühendislik Fakültesi, Gazi Üniversitesi, Ankara, Türkiye
karadag.serap@hotmail.com, suatozdemir@gazi.edu.tr
 (Geliş/Received:20.05.2012; Kabul/Accepted: 17.12.2012)

Özet— İlişkisel veri tabanlarında anahtar kelime aramak için birçok çalışma yapılmıştır. Son kullanıcının veri tabanı yapısından habersiz ve SQL sorgu dilini kullanmadan veri tabanından sorgu yapabilme ihtiyacı ilişkisel veri tabanlarında anahtar kelime arama uygulamaları ihtiyacını doğurmuştur. Yapılan çalışmaların ortak özelliği metin alanlarda metin arama işlemini gerçekleştirmeleridir. Bu çalışmaların amacı kullanıcının dışarıdan girdiği anahtar kelimeleri kullanarak veri tabanı sorgusu oluşturmak ve elde edilen sonuçları kullanıcıya sunmaktır. Burada önemli olan tablolar arası ilişkilerin doğru tanımlanabilmesi ve sonuç olarak verilecek veri setinin doğru sıralanabilmesidir. Yapılan çalışmalarda performans ikinci planda tutularak doğru sırada doğru sonuç üretebilme hedeflenmiştir. Çalışmalar veri tabanına özel geliştirildiklerinden genele hitap etmemektedirler. Bu durum yapılan çalışmaların dezavantajı gibi görünse de aslında olması gereken bir durumdur.

Yapılan çalışmaların tamamında tablolar arası ilişkinin tanımlanmasında dış anahtar, birincil anahtar ilişkisi kullanılmıştır. Bu çalışmada tablolar arası ilişkinin tanımlanmasında sadece dış anahtar, birincil anahtar ilişkisinin kullanılmasının yeterli olmayacağını göstererek kayıtlar için tanımlayıcı olabilecek diğer alanlar üzerinden de ilişkiler tanımlanmıştır. Tanımlanan bu ilişkiler elde edilen bilginin detaylanmasını sağlamıştır. Uygulanan yöntemin adımları ve elde edilen sonuçlar detaylı bir şekilde sunulmaktadır.

Anahtar Kelimeler— Anahtar kelime arama, ilişkisel veri tabanları

Keyword Search in Relational Databases

Abstract— As most of end users are not aware of database structure and want to make query without using SQL, keyword search in relational databases have been studied in the literature extensively. The common idea of the existing studies is to search keywords in text areas of the databases. These studies focused on two main points. First, the relation between tables should be well defined and the second, the results should be sorted in logical manner. However, run time performance of these systems is overlooked by the existing studies. In addition, current studies are developed for specific database schemas and they could not be extended for general purpose databases. Although this appears to be disadvantage, it is a necessity. In this study, we show that primary key, foreign key relation is not enough to construct relation between tables. In addition to this relation we also define new relations by using other fields which hold unique data like e-mail address or identity number. Performance analysis shows that, by using these newly introduced relations, query results are enriched.

Keywords— Keyword search, Relational database

1. GİRİŞ

İnternette anahtar kelime araması, arama motorlarına verilen kelimenin tüm internet dokümanlarında aranması ve yakınlığı olan dokümanların belli bir mantık sırasında kullanıcıya sunulması şeklinde gerçekleştirilir.

Veritabanında anahtar kelime araması internet üzerindeki aramalardan farklıdır. Veri tabanlarında istenilen bir bilgiye ulaşılması internette metin üzerinde yapılan aramaya göre daha karmaşıktır. Çünkü istenilen bir bilgi tek bir tabloda değil bir kaç tabloya dağıtılmış şekilde

tutulmaktadır. Veri tabanlarında gerçekleştirilen bu normalizasyon işlemi, veri tabanlarında anahtar kelime aramasını güçleştirmektedir. İstenilen bilgiye ulaşmak için aranan kelimenin bulunduğu ve bu kayıt ile ilişki içinde bulunan tablolardan arama yapılması gerekir. Bundan dolayı arama işlemi bize aranan kelimenin bulunmasının yanında kelimenin bulunduğu satırın ilişkilerinin de bilinmesi ve arama işleminin o satırlara da genişletilmesi iş yükünü getirir. Peki, veritabanında anahtar kelime ile arama yapılması gerçekten gerekli midir [1]?

Günümüzde birçok firma bilgilerini ilişkisel veritabanları üzerinde tutmaktadır. Bu durumda kullanıcıların veritabanındaki veriye ulaşmaları önem kazanmaktadır. Çünkü firmanın her personeli ve personelin her ihtiyaç duyduğu bilgiye yönelik bir rapor hazırlanması işlemi daha büyük bir iş gücü kaybına neden olmaktadır. Bu ihtiyacı gidermek için ilişkisel veritabanlarına uygun, firmaya özel arama motorları tasarlanmalıdır. İlişkisel veri tabanları bir firma için çok gerekli ve yararlıdır. Fakat bunun yanında ilişkisel veri tabanlarında arama yapmak internete arama yapmak gibi kolay olmadığından veri tabanları geliştirilme esnasında kısıtlı olarak kelime arama desteği vermiştir. Fakat veri tabanlarının anahtar kelime arama desteği çok kısıtlıdır. Aynı zamanda internet üzerinde arama yapan arama motoru teknikleri ilişkisel veri tabanlarında doğrudan kullanılamaz. İlişkisel veritabanlarında arama yapmayı sağlayabilmek için veritabanının yapısının ve ilişkilerinin bilinmesi gerekmektedir. Ayrıca veritabanının normalizasyonu da bu işlemi güçleştirmektedir.

İlişkisel veri tabanlarında anahtar kelime araması önemli bir ihtiyaçtır. Çünkü ilişkisel veritabanında anahtar kelime araması sıradan bir kullanıcı için çok zordur. Bunun nedeni, ilişkisel veritabanlarında istenilen bilgiye ulaşılabilmesi için SQL sorgu yapısının bilinmesi ve sorgulanacak veri tabanında hangi bilginin hangi tabloda bulunduğu ve bu tabloların hangi tablolar ile ilişki içerisinde olduğunun bilinmesi gerekmektedir. Oysaki sıradan bir kullanıcı SQL sorgu dilini, verinin şema ve ilişkisel yapısını bilmez. Tek istediği anahtar kelimeleri verip ilgili sonuçları almaktır.

2. İLGİLİ ÇALIŞMALAR

İlişkisel veri tabanlarında anahtar kelime arama konusunda çeşitli çalışmalar gerçekleştirilmiştir [2,3,4,5,6,7,8,14,16]. Literatür çalışmalarından da görüleceği gibi yapılan çalışmalar bazı ortak yöntemlere sahiptirler. Örneğin çalışmaların çoğu veri tabanında arama gerçekleştirirken bir çizge yapısından yararlanmışlardır. Bir kısmı kendilerinin tasarlamış olduğu bir indeksleme sistemini kullanırken bir kısmı da veri tabanının indeks yapısını kullanmışlardır. Sonuçların sıralanmasında da her uygulama geliştirilen uygulamaya uygun olarak bir sıralama kriteri belirlemiş ve onu kullanmıştır. Yapılan çalışmalarda en büyük farkı indeksleme yöntemleri ve sonuçların sıralanmasında kullanılan kriter oluşturmuştur. Aynı zamanda her çalışma veri tabanı aramasında karşılaşılabilecek bir soruna çözüm aramıştır. İlişkisel veri tabanlarında anahtar kelime arama uygulamalarını diğer bir ortak özelliği de geliştirilen uygulamaların veri tabanına özel uygulamalar olmasıdır. Çünkü veri tabanının indeks yapısı, tablo ilişkileri, tablolarda tutulan veri tipleri, veri tabanının büyüklüğü, veri tabanındaki tutulan verinin düzgünlü vs. geliştirilen uygulamayı etkileyecektir.

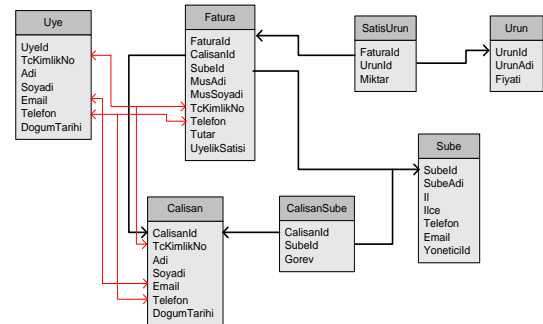
Literatürde yapılan çalışmalardan bazıları BANKS, Bidirectional Expansion, Blinks, Discovery, Querying Communities, Rsearch, DBXplorer, Proximity şeklindedir [2,3,4,5,6,7,8,9,10]. Çalışmaların tümünde dış anahtar →

birincil anahtar ilişkisi kullanılmıştır. Aynı zamanda çalışmaların tümü oluşturulan bir çizge üzerinde metin alanlarda arama yapmaktadır. Çalışmaların bazıları veri tabanlarında anahtar kelime arama problemini bir indeksleme problemi olarak görmüş indeksleme algoritmalarından yararlanarak kendi indeksleme tablolarını oluşturarak çözmeye çalışmışlardır [4,8]. Bazı çalışmalar ise veri tabanı indeks tablolarından yararlanarak arama işlemini gerçekleştirmişlerdir [2,5,6,9,10]. Literatürde bir diğer çalışmada da aynı tablo içerisindeki çift kayıtlar tespit edilerek arama işlemini gerçekleştirilmiştir [7].

Çalışmalar, ilişkisel veri tabanlarında arama yaparken SQL veri tabanı sorgu dilinin gösterdiği başarıyı gösteremediğini ortaya koymuştur. Fakat veri tabanı bilgisine ihtiyaç duyan her kişinin veri tabanı yapısını ve SQL sorgu dilinin bilmesinin imkânsızlığı ilişkisel veri tabanlarında anahtar kelime arama çalışmalarını devam ettirmektedir.

3. ÖNERİLEN YÖNTEM

Literatürdeki çalışmalarda tablolar arası ilişkiler veri tabanı tasarımı sırasında tanımlanmış dış anahtar → birincil anahtar ilişkisi üzerinden kurulmaktadır. Ancak birçok veri tabanında tasarım sırasında tanımlanmasa bile kullanım sırasında ortaya çıkan mantıksal ilişkiler vardır. Örnek vermek gerekirse veri tabanı tasarımı sırasında telefon numarası içeren alanlar genellikle kayıtları ilişkilendirmek için kullanılmazlar. Bu alanlar kayıtları ilgili ek bilgi olarak düşünülürler, ancak mantıksal olarak bir telefon numarası aslında bir kişiyi ya da aile gibi ilişkili kişileri ifade eder. Bu alanlar üzerinde de tanımlanacak ilave ilişkiler ile ilişkisel veri tabanlarında anahtar kelime arama sonuçları genişletilebilir. Bu amaçla bu çalışmada önerilen yöntem, ilişkilerin çıkarılmasında ayırt edici özelliğe sahip bu alanlar üzerinden de ilişki tanımlayarak aramanın genişletilmesidir. Uygulamada kullanılan veri tabanı şeması üzerine ayırt edici özelliklerin bulunduğu alanlar arasındaki ilişkiler de eklenirse tasarım sırasında ortaya çıkmayan ilişkiler de elde edilmiş olur. Ayırt edici alanlar üzerinden kurulan ilişkiler Şekil 1'de kırmızı oklarla ifade edilmiştir.



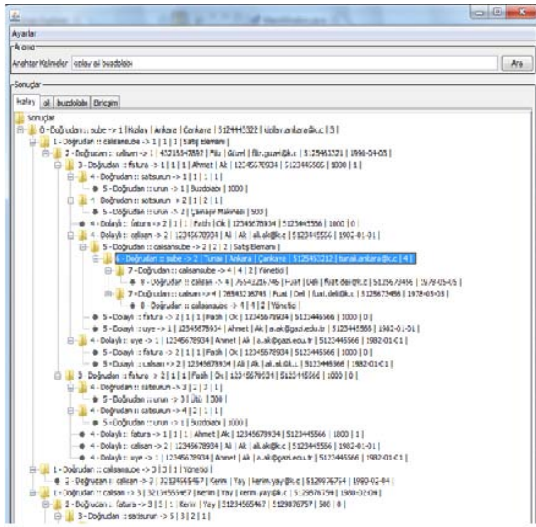
Şekil 1. Anahtar olmayan alanlar üzerinden ilişkilerin tanımlanması

Önerilen bu yöntem dış anahtar → birincil anahtar ilişkisinin kullanılarak anahtar kelimelerin bulunduğu

Şekil 4'de görüldüğü gibi iki veya daha fazla anahtar kelime için ortak kayıtlar bulunmaktadır. Aynı şekilde Şekil 4'de k_0 , k_1 ve k_2 anahtar kelimeleri için T_{01} , T_{11} ve T_{21} ağaç yapılarının ortak dalı olan r_{1211} dalı üzerinden ilişkilidir.

4. ANALİZ

Önerilen metodun faydalarının anlaşılması için örnek veri tabanı üzerinde “kızılay”, “ali” ve “buzdolabı” anahtar kelimeleri kullanarak bir analiz senaryosu oluşturulmuştur. Bu senaryoda, kullanıcı "kızılay" şubesinde "ali" adındaki bir kişiye satılan ya da "ali" adlı bir kişi tarafından satılan bir "buzdolabı" olup olmadığını öğrenmek istemektedir. Bu amaçla uygulama öncelikle girilen anahtar kelimeleri veri tabanı içerisinde bağımsız olarak aramakta daha sonra bulunduğu sonuçları birleştirmektedir.



Şekil 5. "Kızılay" anahtar kelimesi için uygulama sonuç ekranı

Şekil 5'te “kızılay” anahtar kelimesi için bulunan sonuçlar görülmektedir. Sonuçlar incelendiğinde uygulama “kızılay” anahtar kelimesinin geçtiği tek tablo ve kayıt olan Sube tablosundaki kayda ulaşmıştır. Bu kayıt R kümesinin elemanı olan R_0 'ı ifade etmektedir. Bu durumda R_0 kümesi de $R_0 = \{r_{01}\}$ şeklinde tek elemanlıdır. Veri tabanında tanımlı dış anahtar → birincil anahtar ilişkilerinin ve veri tabanı yöneticisi tarafından tanımlanmış olan ilişkiler kullanılarak Şekil 5'te görülen T_{01} ağaç yapısı oluşmaktadır. T_{01} ağaç yapısında görüldüğü gibi Sube tablosunda bulunan “kızılay” anahtar kelimesi dış anahtar → birincil anahtar ilişkisinden birinci seviyede 3 farklı kayda ulaşmıştır.

İlk bulunan Sube tablosundaki kayıttan Calisansube tablosundaki “Satış Elemanı” kaydına bu kayıttan da Calisan tablosundaki “Filiz Güzel” kaydına ulaşmıştır. Bir sonraki ilişkide ise yani 3. seviyede 2 farklı kayda ulaşmıştır. Bu kayıtlar Fatura tablosunun “Ahmet Ak” ve “Fatih Ok” kayıtlarıdır. “Ahmet Ak” kaydına bakıldığında bu kayda ait 4. seviyeden 2'si doğrudan 3'ü dolaylı ilişki olarak tanımlanan 5 kayıt mevcuttur.

2 doğrudan ilişki SatışUrun tablosu ile kurulmuştur ve bu iki kayıt da Urun tablosundaki “buzdolabı” ve “çamaşır makinesi” kayıtları ile doğrudan ilişkilidir. Bu aşamaya kadar olan doğrudan ilişkiler değerlendirildiğinde Filiz Güzel'in Kızılay şubesinde çalışan bir satış elemanı olduğu; buzdolabı ve çamaşır makinesi alan Ahmet Ak adlı kişinin faturasını düzenlediği bilgisine ulaşılır. Doğrudan ilişkilerin kullanılması ile CalisanSube tablosundan sadece bu bilgilere ulaşılabilir.

Fakat 4. seviyede oluşmuş 3 dolaylı ilişkiden daha fazla bilgiye ulaşılabildiği görülür. 4. seviyede oluşmuş ilk dolaylı ilişki Fatura tablosundaki “Fatih Ok” kayıttır. Bu kayıt ile ilişkili kayıtlar aynı ağaç içerisinde daha önceden bulunduğu için ağaç devam etmemiştir. Bu dolaylı ilişki tekimlikno ve telefon alanları üzerinden kurulmuştur. Dolaylı ilişki “Fatih Ok” ve “Ahmet Ak” kayıtları arasındaki ilişkiyi ortaya çıkarmıştır.

İkinci dolaylı ilişki Calisan tablosundaki “Ali Ak” kaydı ile gerçekleşmiştir. Kurulan bu dolaylı ilişki tekimlikno ve telefon alanları üzerinden kurulmuştur. Dolaylı ilişki ile ulaşılan bu kayıt doğrudan ilişki olarak tanımlanan veri tabanı dış anahtar → birincil anahtar ilişkileri kullanılarak CalisanSube ve Sube tablolarına erişilmiştir. Dolaylı ilişkinin kullanılması “Ahmet Ak” kaydındaki tekimlikno ve telefon alanlarındaki bilginin Ali Ak'a ait olduğu ve Ali Ak'ın da Tunali şubesinde satış elemanı olduğu bilgisini ortaya çıkarmıştır. Bu kayıt üzerinden ilerlendiğinde tekrar CalisanSube ve Calisan tablolarına erişilmiş ve Tunali şubesinin yöneticisi bilgisine erişimi sağlamıştır.

3. dolaylı ilişki ise aynı telefon ve tekimlikno alan değerlerini içeren bir fatura kaydının daha olduğunu “Fatih Ok” kaydını ortaya çıkarmıştır.

Dolaylı ilişkilerin kullanılması ile arama sonucu genişletilmiş, doğrudan ilişkiden ulaşılmayacak sonuçlar elde edilmiştir. Eğer dolaylı ilişki kullanılsaydı 4. seviyede sadece 2 ilişki elde edilecekti ve dolaylı ilişkilerin çıkarmış olduğu sonuçlar elde edilemeyecekti. Dolaylı ilişkinin de kullanılması ile 4. seviye ilişki sayısı 5 olmuştur.

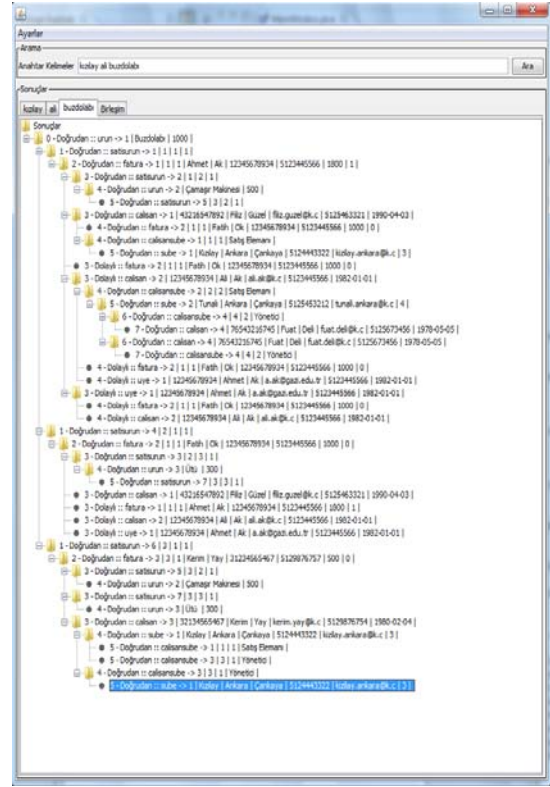
Aynı şekilde bütün seviyeler veri tabanında tanımlı dış anahtar → birincil anahtar ilişkisi ve veri tabanı yöneticisi tarafından tanımlanmış olan ilişkiler kullanılarak en uç yapıya kadar gider.



Şekil 6. "Ali" anahtar kelimesi için uygulama sonuç ekranı

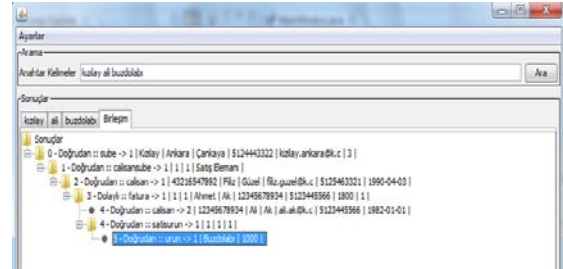
"Ali" anahtar kelimesi için veri tabanında gerçekleştirilen arama sonucu bulunan kayıtlar Şekil 6'da gösterildiği gibidir. Görüldüğü gibi daha ilk seviyelerde dolaylı ilişki aramayı genişletmiştir. "Ali" anahtar kelimesi Calisan ve Sube tablolarının birer kaydında bulunmuş ve Calisan tablosundaki kayıt ile ilişkili diğer kayıtlara ulaşılmak istenildiğinde 1. seviyeden 1 doğrudan 3 dolaylı ilişkiye ulaşılmıştır yine aynı şekilde Sube tablosundaki kayıt ile ilişkili diğer kayıtlara ulaşılmak istenildiğinde de 1. seviyeden 2 doğrudan 1 dolaylı ilişkiye erişilmiştir.

Eğer "Ali" anahtar kelimesini aramada dolaylı ilişki kullanılmamış olsa idi müşteri "Fatih" in ve üye "Ahmet" in çalışan "Ali" ile olan ilişkisi ortaya çıkmayacak ve Calisan tablosundan ulaşılan 1. seviye ilişki sayısı 4 değil 1 olacak, Sube tablosundan ulaşılan 1. seviye ilişki sayısı da 3 değil 2 olacaktır.



Şekil 7. "Buzdolabı" anahtar kelimesi için uygulama sonuç ekranı

Şekil 7'de "buzdolabı" anahtar kelimesi için bulunan sonuçları göstermektedir. Şekil 7'de görüldüğü gibi arama sonucunda kurulan ilişkilerde dolaylı ve doğrudan ilişkiler mevcuttur. Şekil 7'de görüldüğü gibi dolaylı ilişkinin kullanılması bulunan ilişkili kayıt sayısını artırmış bu sayede arama genişletilmiştir.



Şekil 8. "Kızılay", "ali", "buzdolabı" anahtar kelimeleri için uygulama sonuç ekranı

Şekil 8'de kullanıcı tarafından girilmiş olan "kızılay", "ali" ve "buzdolabı" anahtar kelimeleri için ortak arama sonucunu göstermektedir. Şekil 8'de de görüldüğü gibi sonuç ekranını doğrudan ve dolaylı ilişkiler oluşturmaktadır. Bu ekranda gösterilen doğrudan ve dolaylı ilişkiler her bir anahtar kelimenin aranması ile elde edilen ilişkileri ifade etmektedir. Sonuçlardan da anlaşılacağı gibi gösterilen sonuçlar 3 anahtar kelime için de ortak kayıtlardır.

Sonuç kayıtları değerlendirildiğinde Sube tablosunun "kızılay" kaydının kök yaprağı oluşturduğu görülmektedir. Bu kayıt üzerinden sırası ile CalisanSube, Calisan ve Fatura tablolarına erişilen kayıtlar mevcuttur.

Bu ilişkilerden CalisanSube ve Calisan tablolarına doğrudan ilişki ile ulaşılmışken Fatura tablosuna dolaylı ilişki ile ulaşılmıştır. Bu da gösteriyor ki eğer dolaylı ilişki kullanılmazdı sadece Kızılay şubesinin satış elemanı Filiz Güzel bilgisine ulaşılabilirdi. Dolaylı ilişkinin devreye girmesi ile Filiz Güzel'in Ahmet Ak adına düzenlediği fatura ile buzdolabı aldığı bilgisine; Ahmet Ak'ın TC kimlik numarası ve telefon numarasını kullandığı kişinin Ali Ak adında bir çalışan olduğu bilgisine erişilemeyecekti.

Dolaylı ilişki kullanılmamış iken arama sonucu 2. seviyeye kadar ilerleyebilirken dolaylı ilişkinin kullanılması ile arama seviyesi 5'e çıkmıştır. Bu durum da aramayı zenginleştirmiştir.

5. SONUÇ

Literatürdeki çalışmalarda kullanılan dış anahtar → birincil anahtar ilişkisi veri tabanındaki ilişkilerin çoğunu ortaya koyduğundan başka bir ilişki tanımlama ihtiyacı duyulmamıştır. Fakat bazı veri tabanlarında sadece dış anahtar → birincil anahtar ilişkisi ilişkilerin tamamını ortaya çıkaramamakta ve aynı zamanda kayıtlar arası ilişkiyi eksik bırakmaktadır. Bu tip veri tabanlarında diğer alanlar üzerinden ek ilişkilerin tanımlanması ilişkisel veri tabanlarında anahtar kelime arama kabiliyetini arttırmaktadır. Önerilen yöntemde veri tabanı yöneticisi tarafından tanımlanmış olan benzersiz alanlar üzerinden kurulan ilişkilerin örnek veri tabanı üzerindeki arama işlemini genişlettiği ve dış anahtar → birincil anahtar ilişkisi ile erişilemeyecek sonuçlara erişildiği görülmektedir.

İlişkisel veri tabanlarında anahtar kelime arama çalışmaları veri tabanına bağımlılık gerektirdiğinden farklı yapıya sahip veri tabanları için farklı uygulamaların gerçekleştirilmesi gerekecektir. Veri tabanlarında anahtar kelime arama çalışmalarında önerilen yöntemler ile henüz SQL sorgu dilinin başarısına erişilememiştir. Bu nedenle veri tabanlarında anahtar kelime arama çalışmaları devam etmektedir.

KAYNAKLAR

- [1] Wang, S., Zhang, K.L., "Searching databases with keywords", Journal of Computer Science and Technology, 20(1):55 - 62, 2005.
- [2] Bhalotia, G., Hulgeri, A., Nakhe, C., Chakrabarti, S., Sudarshan, S., "Keyword searching and browsing in databases using BANKS", 18th International Conference on Data Engineering, San Jose, 431 - 440, 2002.
- [3] Kacholia, V., Pandit, S., Chakrabarti S., Sudarshan, S., Desai, R., Karambelkar, H., "Bidirectional Expansion For Keyword Search on Graph Databases", Very Large Data Bases, 505-516, 2005.
- [4] He, H., Wang, H., Yang, J., Yu, P.S., "BLINKS: Ranked keyword searches on graphs", International Conference on Management of Data, New York, 305 - 316, 2007.
- [5] Hristidis, V., Papakonstantinou, Y., "DISCOVERY: Keyword search in relational Databases", Very Large Databases, 670 - 681, 2002.
- [6] Qin, L., Yu, J.X., Chang, L., Tao, Y., "Querying communities in relational databases", International Conference on Data Engineering, Shanghai, 724 - 735, 2009.

- [7] Yang, X., Wang, B., Wang, G., Yu, G., "RSearch: Enhancing keyword Search in relational databases using nearly duplicate records". Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2010.
- [8] Agrawal, S., Chaudhuri, S., Das, G., "DBXplorer: A system for keyword - based search over relational databases", International Conference on Data Engineering, San Jose, 5, 2002.
- [9] Balmin, A., Hristidis, V., Papakonstantinou, Y., "ObjectRank: Authority-based keyword search in databases", Very Large Data Bases, 564 - 575, 2004.
- [10] Goldman, R., Shivakumar, N., "Proximity Search in Databases", Proceeding of the 24th Very Large Data Bases Conference, New York, 26 - 37, 1998.