

Osmanlı Arşivleri İçerik-Bazlı Sorgulama (İBS) Sistemi

Content-Based Retrieval (CBR) System for Ottoman Archives

İsmail Sengör Altingövde¹, Ediz Şaykol¹, Özgür Ulusoy¹, Uğur Güdükbay¹, A. Enis Çetin²,
Muammer Göçmen³

¹ Bilgisayar Mühendisliği Bölümü, Bilkent Üniversitesi, Ankara

² Elektrik ve Elektronik Mühendisliği Bölümü, Bilkent Üniversitesi, Ankara

³ İslam Tarihi ve Sanatları Bölümü, Süleyman Demirel Üniversitesi, Isparta

{ismaila, ediz, oulusoy, gudukbay}@cs.bilkent.edu.tr,
cetin@ee.bilkent.edu.tr, mumen@sdu.edu.tr

Özetçe

Bu çalışmada sayısallaştırılmış Osmanlı arşiv belgeleri üzerinde içerik-bazlı sorgulama yapılabilmesini sağlayan bir sistem önerilmiştir. Sistemde belgelerden çıkarılan tanımlanmamış işaretlerin yönlendirilmiş bir yaklaşımla daha önceden simge kütüphanesine yerleştirilen simgelerden en çok benzeyene eşleştirilmesi sağlanmıştır. Kullanıcılar istedikleri bir belge üzerinden istedikleri simge grubunu seçerek belge kümesi içerisinde aratabileceklerdir. Sistemin prototipi şu an İnternet üzerinde kullanıcıların hizmetine açılmıştır.

Abstract

We propose a content-based retrieval (CBR) system for digital Ottoman archive documents. In this system, the symbols extracted from the documents are matched with the most similar one in the symbol library, which is created in a supervised manner. The users specify queries by marking a region on an example document and the system retrieves all documents that include the symbols found in the query region. A prototype of the system is currently available on the Web.

1. Giriş

Osmanlı arşivleri, yüzyıllarca hüküm sürmüş bir imparatorluğun milletimize bırakmış olduğu kültür mirası içerisinde hiç şüphesiz ki oldukça önemli bir yer tutmaktadır. Her ne kadar bu arşivler, geçtiğimiz on yıllar boyunca sayısız araştırmacının en temel kaynağı olmuşsa da, arşivlerin içerdiğiengin bilgi ve belgelerin bütün potansiyeliyle kullanılabilmesi ve araştırmaların daha hızlı, kapsamlı ve verimli yürütülebilmesi için günümüz teknolojisinin yardımına ihtiyaç vardır. Bilgisayar bilimindeki ilerlemeler sayesinde bugün başta Latin alfabesi olmak üzere çeşitli dillerde yazılmış eserlerin sayısallaştırılarak dijital ortama aktarılması ve üzerlerinde gelişmiş sorgular da dahil olmak üzere çeşitli işlemler yapılması mümkün olmaktadır (örneğin, bkz. [1]). Devlet Arşivleri Genel Müdürlüğü'nün çalışmalarıyla Osmanlı arşivleri de son birkaç yıl içerisinde sayısal ortama aktarılmaya başlanmıştır [2]. Şu an için (2005 yılı itibarıyla) bu kurumun İnternet sitesinden Türkçe anahtar sözcükler girilerek belgelerin çözümlemeleri (transkripsiyon) üzerinde arama yapılabilmekte ve sorguya cevap olan belgelerin kimlik bilgilerine ulaşılabilir.

Ancak tahmin edilebileceği gibi, bu durumda sadece çözümlemeleri yapılmış belgeleri yine sadece bu çözümlemelerde geçen Türkçe kelimelerle sorgulamak ve sonuçta da belgelerin kimlik bilgilerini almak mümkün olabilmektedir. Bu makalede geliştirilen Osmanlı Arşivleri İBS (İçerik-Bazlı Sorgulama) sistemi ile bu söylenenlerden daha farklı ve gelişmiş fonksiyonların sağlanması hedeflenmektedir: Herhangi gibi bir ön tasnife veya çözümlemeye ihtiyaç duyulmaksızın, varolan Osmanlıca belgeler direkt olarak dijital ortama aktarılıp kullanıcı erişimine açılabilir; kullanıcı Osmanlıca simgeler üzerinden örnek-tabanlı sorgular yapabilecek ve bulunan belgelerin sadece kimlik bilgilerine değil metin-resmi şeklindeki görüntüsüne de ulaşabilecektir. Böylece hem çözümlemesi yapılmamış belgeler de elektronik ortamda araştırılabilir, hem de Osmanlıca bilen araştırmacılar isterlerse çözümlemedeki kelimeler ile değil direkt Osmanlıca sembollerle arama yapabilmek imkanına kavuşacaktır. Bu durum, bir taraftan tasnif vb. aşamalarda belgelerin otomatik olarak bir ön gruplandırılmaya tutulmasını sağlayıp bu işlemlerle uğraşan uzmanlara ışık tutacak, diğer taraftan da bu belgeler üzerinde araştırma yapan kişilerin istedikleri belgeleri orijinal diliyle arayıp kolayca bulabilmesine imkan verecektir. Böylece Osmanlı arşivleri üzerinde bilimsel çalışmalar yapan tüm araştırmacıların işlerinin önemli ölçüde kolaylaşması ve erişimin hem kalitesinin hem de hızının artması sağlanacaktır.

Yukarıda bahsedildiği üzere Latin alfabesi ile yazılmış yeni ve düzgün baskılı materyalin sayısallaştırılıp, optik karakter tanıma sistemleri ile metine çevrilebilmesi kolay ve yüksek başarı oranlarında (%99'a varan) mümkün olsa da, farklı alfabeler ve basım kalitelerindeki belgeler için aynı başarıdan bahsetmek mümkün değildir [3]. Özellikle Osmanlıca belgeler düşünüldüğünde aşılması gereken önemli zorluklar şöyle sıralanabilir: (i) Osmanlıca dil özellikleri (bitişik yazım, harflerin kelimedeki yerlerine göre farklı şekillerde yazılması, kelime içerisinde boşluklar bulunabilmesi, vb.), (ii) Arşiv belgelerinin içerik çeşitliliği (arşivde matbu ve el yazısı belgelerin bulunması, ve özellikle bu ikinci gruptakilerin çok farklı yazı stilleri ile yazılmış olması), ve (iii) Arşivleri oluşturan belge kümesinin sayısal büyüklüğü. Bu zorluklar göz önüne alındığında tüm arşiv belgeleri için hızlı ve kesin olarak içerik-bazlı sorgulama yapabilecek bir sisteminin bir anda oluşturulabilmesi mümkün görünmemektedir. Ancak, bu zorluklar farklı araştırmacıların arşiv belgelerinin değişik alt kümelerine yoğunlaşarak yeni çözümler bulmasıyla aşılabılır.

Bu çalışma çerçevesinde, el yazısı ve matbu Osmanlıca belgeler için içerik tabanlı bir sorgulama sisteminin ilk adımları atılmış ve prototip bir sistem ortaya çıkarılmıştır. Aşağıdaki bölümde ilk önce literatürdeki benzer çalışmalar kısaca tartışılmaktadır. Üçüncü bölümünde, simge kütüphanesinin oluşturulmasında kullanılan yönlendirilmiş yalkaşım ve histogram bölütleme tekniği anlatılmaktadır, ve Web-tabanlı prototip sistem sunulmaktadır. Dördüncü bölümdeyse varılan sonuçlar ve olası gelecek çalışmalar belirtilmektedir.

2. Önceki Çalışmalar

Latince karakterler içeren diller için basılı eserlerin sayısallaştırılması ve optik karakter tanıma sistemleriyle elektronik ortamda metin belgelerine çevrilip sorgulanması için yapılan çalışmalar şu an olgunlaşmış denebilecek düzeydedir [4]. Diğer taraftan, Arap alfabesiyle bitişik yazılan belgeler için aynı ölçüde bir başarıdan söz etmek mümkün değildir. Amin'in onlarca makaleyi taradığı çalışmasında, Arapça belgeler için optik karakter tanıma sistemlerinde kullanılan çeşitli yaklaşımlar (karakter bölütleme, yapay sinir ağları, çizge teorisi tabanlı algoritmalar, Markov model tabanlı algoritmalar, vs.) sıralanmakta, ancak bu yaklaşımların hiç birinin karakterlerin doğru tanınması işleminde istenen başarı seviyelerine çıkamadığına dikkat çekilmektedir [3]. Her ne kadar Osmanlıca Arapça'dan farklı bir dil olsa da, genel yapısı itibarıyla Arap dillerine olan benzerliklerinden dolayı, bu konuda yapılacak çalışmalarda ilgili literatürün taranması faydalı görülmüştür.

Literatürde doğrudan Osmanlıca ya da Osmanlıca belgeler için yapılan çalışmaların sayısı ise çok daha azdır. Atıcı ve Yarman-Vural'ın çalışmasında [5] önce zincir kodu (chain code) dönüşümü karakter darbelerine uygulanmış, tanıma işlemi içinse gizli Markov modelinden yararlanılmıştır. Sistem başarılı olmakla beraber, bu başarı algoritmanın zayıflatma (thinning) aşamasının kesinliğine bağımlı durumdadır. Bir diğer makaledeyse Osmanlıca karakterlerin yapay sinir ağları ile tanınmasına çalışılmıştır [6]. Bu çalışmada tanıma aşamasında sistemin karakterleri bitişik bir sembol grubundan nasıl çıkardığı açıkça belirtilmemiş olmasına karşın, anlaşıldığı kadarıyla gerek sinir ağının öğrenmesi, gerekse tanıma testleri sırasında sistem sadece izole durumdaki harfleri kullanmıştır. Bu durumda harflerin doğru tanımlanma başarısının %94'lere kadar çıkabildiği görülmüştür. Ancak bu makalede öngörülen sistem için tanıma işleminin belgenin tümü üzerinden gerçekleşmesi gerektiği açıktır. Yapay sinir ağlarının bizim sistemimizde de kullanılabilmesi için çalışmalarımız sürmektedir.

Şaykol ve Ark.'nın [7] çalışmasındaysa amaç, öncelikle eldeki belge kümesinden bir "simge" kütüphanesi (alfabe) oluşturulması ve sonrasında da her belgenin içerdiği simgeleri listeleyen bir metin dosyasına (kod kitabı) dönüştürülmesidir. Simge çıkartıcı alt modülü temel olarak şu işlevleri gerçekleştirmektedir:

- Belge-resmi üzerinde bulunan tüm izole *işaretlerin* (birleşik siyah nokta gruplarının) bulunup çıkartılması. Bu iş için literatürde bilinen algoritmalar kullanılmıştır.

- Çıkarılan işaretlerin simge kütüphanesinde daha önce kaydedilmiş simgelerle karşılaştırılarak (eğer varsa) benzer simgenin tespit edilmesi. Bu amaçla bahsedilen çalışmada bir yönlendirilmeyen (unsupervised) yaklaşım kullanılmıştır. Bu yaklaşımda simge çıkartıcı modül Osmanlıca karakterler hakkında herhangi bir ön bilgiye sahip değildir, yani başlangıçta simge kütüphanesi boştur. Modül resimden her bir izole işareti çıkardığında kütüphaneye o ana dek kaydedilmiş simgelere bakar, eğer kayıtlı simgelerden biri eldeki işarete (veya işaretin *içerisinde* bir parçaya) belli bir eşik değerin üzerinde benziyorsa o işaretin (veya parçasının) bilinen bir simge olduğuna hükmeder, koordinatlarını ve hangi simge olduğuna kayıt ederek bir sonraki işarete geçer. Eğer eldeki işaret öncekilere benzer bulunmamışsa bunu yeni bir simge olarak kütüphaneye ilave eder. Böylece simge kütüphanesi sabit bir stildeki alfabe harflerini değil, belgelerde bulunma durumuna bağlı olarak harflerin çeşitli biçimlerdeki hallerini uyarlamalı bir şekilde kapsayabilecektir.

Dikkate değer bir diğer konu da belgeden çıkartılmış bir işaret ile kütüphanedeki bir simgenin benzerliğinin nasıl karşılaştırılacağıdır. Bir belgede büyük yazılmış bir harf (simge) bir başkasında çok küçük yazılmış olabilir. Sembollerin benzerlik karşılaştırması için bahsedilen çalışmada uzaklık ve açısız dağılım vektörleri kullanılmıştır. Bu vektörlerin simge büyüklüğünden bağımsız olmasının yukarıda belirtilen ölçeklenebilirlik sorununu çözdüğü belirtilmiştir.

Bu makalede genel olarak yukarıdaki bu son çalışmada sunulan çerçeve esas alınmış ve gerek simge kütüphanesinin belirlenmesi, gerekse simgelerinin karşılaştırılması sırasında değişik yaklaşımlar uyarlanarak performansın artırılmasına çalışılmıştır.

3. Osmanlı Arşivlerini İçerik-Bazlı Sorgulama

3.1. Simge kütüphanesinin yönlendirilen yaklaşımla oluşturulması

Simge kütüphanesinin oluşturulması bu makalede hedeflenen içerik tabanlı sorgu sisteminin gerek performansını gerekse sorgu cevaplayabilme başarısını belirleyen en önemli aşaması durumundadır. Burada, önceki çalışmadaki [7] yöntemine alternatif olarak bir yönlendirilen yaklaşım önerilmiştir. Bunun gerekçesi ise şöyle açıklanabilir:

Osmanlıca karakterlerin belgelerde oldukça farklı biçimlerde (ve bitişik şekilde) yazılabiliyor olması ve bu çeşit belgelerin harfler dışında işaretler de içerebilmesi (motif, çizik, leke, vs.) sebebiyle yönlendirilmeyen yaklaşım farklı belgelerden çıkarılan pek çok işareti yeni simgeler olarak kaydetmektedir. Böylece simge kütüphanesi istenildiği şekilde harfler ve bunların çeşitli yazım biçimlerini kapsadığı gibi, istenmeyecek şekilde bahsedilen harf-dışı pek çok sembolü de içermektedir. Bunun bir sonucu olarak da simge kütüphanesi çok büyümektedir. Simge çıkartımı sırasında o ana kadar kütüphaneye ilave edilmiş simgeler, eldeki yeni çıkarılan simgenin içerisinde aranmakta, bu arama sırasında eldeki simgenin içinden alınan her parça için uzaklık ve açısız dağılım vektörleri geri kalan tüm simgelerin vektörleriyle karşılaştırılmaktadır. Bu durum, simge kütüphanesinin

büyümesiyle sistem performansının düşmesine yol açabilmektedir.

Alternatif olarak önerilen yönlendirilen yaklaşımda Osmanlıca belgelerde bulunabileceği düşünülen karakterler (çeşitli yazı tiplerinde) daha önceden simge kütüphanesine elle ilave edilmiştir. Bu simgeler Osmanlıca harflerin çeşitli biçimleridir; çünkü harfler kelimenin başında, ortasında veya sonunda farklı biçimlerde yazılabilirler. Simge kütüphanesi bu şekilde oluşturulduktan sonra daha önce olduğu gibi eldeki belgeden çıkartılan her işaret kütüphanede hazır bulunan simgelerle karşılaştırılacaktır. Elbette ki bu yaklaşımın da dezavantajları bulunmaktadır: Sistemin adaptasyon kabiliyeti (farklı yazım biçimlerini tanıyabilmek açısından) simge kütüphanesine konulan harflerle sınırlı olacaktır. Eğer kütüphane yeterince kapsayıcı değilse belgelerde bulunan işaretler kütüphanede sınırlı sayıda simgeden ilgisi olmayan birine benzer olarak bulunabilir (ya da hiç bir simgeye benzer bulunmaz). Bu durum da sorgularda doğru belgeleri bulabilme oranını düşürecektir. Burada bir çözüm yolu sistemin başka simgelere benzetmekte zorlandığı simgeleri saklayıp, belirli aralıklarla kütüphaneye ilave etmek üzere değerlendirmesidir. Örneğin, her bir kaç bin belge işlendikten sonra, bu tanınmayan simgeler kendi aralarında karşılaştırılabilir. Böylece, çok sık rastlanılan simgeler otomatik olarak tespit edilebilir. Bu simgeler daha sonra bir uzmana sunulup kütüphaneye eklenmeden önce kullanıcı onayı alınabilir ve yarı-otomatik bir çözümle kütüphane içeriğinin doğruluğu korunabilir. Her şartta bir takım simgelerin kaçırılması mümkündür. Ancak, bu yaklaşım nispeten daha düşük tanıma başarısıyla da olsa daha büyük belge kümelerinin işlenmesine olanak sağladığı için cazip görünmektedir. Deneyler sırasında da bu yaklaşımın simge tanıma hızını ve buna bağlı olarak birim zamanda işlenebilecek belge sayısını önemli miktarda arttırdığı görülmüştür.

3.2. Histogram tabanlı bölütleme

Daha önce bahsedildiği üzere bir belgeden çıkartılan izole her işaretin (harf veya harf grubu olabilir) içerisinde simge kütüphanesinde bulunan tüm simgeler aranmaktadır. Bir diğer söyleyişle, eldeki işaretin her bir parçası için (belli bir eşik büyüklüğü geçtiği sürece) vektörler hesaplanıp bu parçanın kütüphanedeki simgelere benzerliği hesaplanmakta, en çok benzediği simge bulunmakta ve benzeşim oranı yine bir eşik değerin (genellikle %90 ya da %95) üzerindeyse bu parçanın o simge olduğuna karar verilip bu parça eldeki işaretten silinmektedir. Bu işlem eldeki işaretin her parçası bir simgeye eşleştirilene ya da yeni bir simge olarak kütüphaneye ilave edilene kadar sürmektedir.

Bu makalede yukarıda anlatılan “tanıma” işleminin performansını artırmak için literatürde daha önceden önerilen “histogram tabanlı bölütleme” algoritması kullanılmıştır. Böylece, eldeki belgeden çıkartılan bir işaret sadece algoritmanın bulduğu noktalardan (ve genellikle az sayıda) parçalara ayrılmaktadır. Burada elbette sistem performansının artması karşılığında tanıma başarısından fedakarlık yapılabilmektedir.

Bu bölütleme tekniği şöyle özetlenebilir [3]. Arapça (özellikle matbu) kelimelerde karakterlerin birleşme noktalarındaki genişliğin karakterin genişliğinden çok daha az olduğu görülmüştür. Buradan hareketle, kelimeyi oluşturan piksellerin düşey bir projeksiyonu elde edilirse, karakter

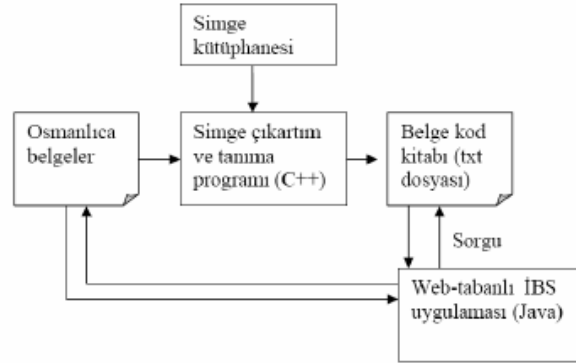
bağlantı noktalarındaki toplam değer aşağıdaki formülle gösterilen ortalamının altında olacaktır. Bu formülde (Denklem 1), N_c sütun sayısı, X_j ise j sütunundaki siyah piksel sayısıdır.

$$\text{Ortalama} = (1/N_c) \sum_{j=1}^{N_c} X_j \quad (1)$$

Böylece bu bağlantı noktalarından kelime parçalanarak her parçanın bir harfe (ya da bizim sistemimizde simgeye) benzerliğine (yine daha önce bahsedilen vektörler kullanılarak) bakılabilir. Ancak bu yaklaşımın uygulanabilmesi için histogramın içinde peş peşe gelen tepeler arasındaki uzaklığın karakter genişliğinin 1/3'ünü geçmemesi gibi bazı koşulların sağlanması gerekmektedir. Bu bakımdan bu yaklaşım kütüphanedeki simgeler için önceden tanımlanacak bir eşik karakter genişliği değerine bağlıdır ve esnek değildir. Diğer taraftan ise yapılan deneylerde bu yaklaşımın test veri kümesinde tanıma kesinliğinde fazla bir kayba yol açmadan simge tanıma işleminin hızını önemli oranda arttırdığı gözlenmiştir.

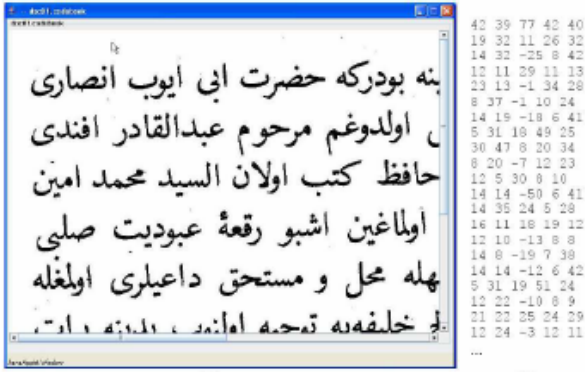
3.3. Prototip Sistem

Yukarıda anlatılan yaklaşımlar hali hazırda İnternet üzerinde çalışmakta olan bir prototip sistemde ([7]'de anlatılan mimari ve kodlamanın üzerine adapte edilerek) gerçekleştirilmiştir. Sistem mimarisi Şekil 2'de verilmiştir.



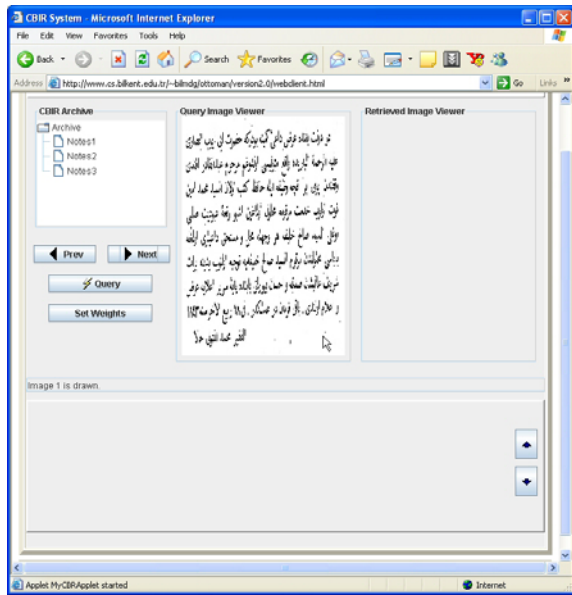
Şekil 1: Sistem mimarisi.

Belgelerden yönlendirilen yaklaşımla simgelerin çıkartılması ve her belgenin içerdiği simgeleri ve koordinatlarını listeleyen kod kitabı oluşturulması C++'da yazılan bir programla sağlanmaktadır. Şekil 2'de örnek bir belge ve bunun için oluşturulan kod kitabı gösterilmiştir. Kod kitabında her satırdaki ilk sayı belgedeki simgenin (kütüphaneye göre) numarasını, diğer dört sayı ise koordinatlarını (bir önceki sembole göre göreceli olarak x ve y ekseninde farklar, simge genişliği ve yüksekliği) vermektedir.

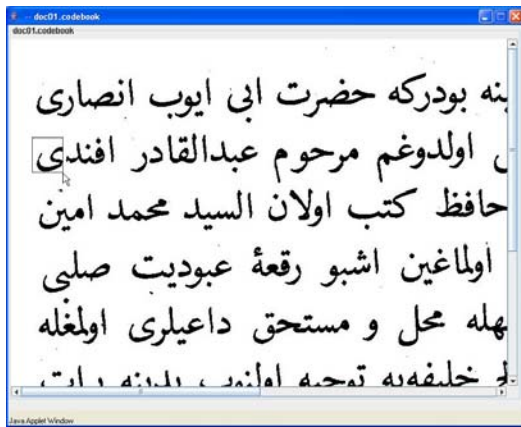


Şekil 2: Örnek Osmanlıca belge (solda) ve karşılık gelen kod kitabı (sağda).

Son kullanıcı İnternet arayüzü (Şekil 3) aracılığıyla sistemdeki Osmanlıca belgeleri tarayabilir ve seçtiği bir belge üzerinde dikdörtgen bir alanı sorgulamada kullanmak amacıyla belirleyebilir (Şekil 4).



Şekil 3: Osmanlı arşivleri İBS sistemi Web arayüzü.



Şekil 4: Seçilen belgede örnek tabanlı sorgu girişi.

Web-tabanlı İBS uygulaması seçilen sorgu alanının koordinatlarını belirleyerek o belgede seçili alanda bulunan simgeleri o belgenin kod kitabından bulur, ve diğer tüm kod kitaplarında aynı simgelerin peşleştiği yerleri belirler. Son olarak sorgu sonucunda tespit edilen kod kitaplarına karşılık gelen Osmanlıca (sayısallaştırılmış) belgeler kullanıcıya sunulur. Bu belgeler üzerinde sorguya cevap olan alanlar renkli olarak gösterilir.

4. Sonuçlar

Bu makalede Osmanlı arşivleri için sayısallaştırılmış belgeler üzerinde içerik-bazlı sorgulama sisteminin ilk adımları atılmıştır. Böyle bir sistemin gerçekleştirilmesinde gerek Osmanlıca'nın dil özellikleri gerekse arşivlerin büyüklüğü ve çeşitliliğinden kaynaklanan önemli zorluklar olduğu bir gerçektir. Bu çalışma çerçevesinde belgelerden Osmanlıca simgelerin çıkartılmaması ve tanınması için çeşitli yaklaşımlar geliştirilip uygulanmış ve İnternet üzerinde çalışan bir prototip sistem kullanıcılara sunulmuştur.

Elde edilen sonuçlar gelecek çalışmalar için cesaret vericidir. Bu makale için gelecek çalışmalar arasında şunlar sayılabilir: (i) özellikle yapay sinir ağlarının ve daha gelişmiş simge tanıma algoritmalarının sisteme adapte edilmesi, (ii) test veri kümesinin yüzlerce belge seviyesinden çok daha büyük rakamlara çıkarılması, (iii) sistemin makul bir tanıma yüzdesiyle gerçek belge kümeleri üzerinde çalışacak halinin bu alandaki uzmanların hizmetine sunulması, ve (iv) sistemin uzmanlardan gelecek geri-bildirimlere göre hatalarını düzeltecek şekilde esnek hale getirilmesi.

5. Teşekkür

Bu çalışma Devlet Planlama Teşkilatı 2004K120720 sayılı proje tarafından desteklenmiştir. Daha önce geliştirilen sistemin kodlarını sağlayan A. Kemal Sinop'a ve bu çalışmada bahsedilen yaklaşımların kodlanmasında yardımcı olan İsmet Zeki Yalnız ve Hakan Yılmaz'a teşekkür ederiz.

6. Kaynakça

- [1] Google Print Projesi Web Sitesi, <http://print.google.com/googleprint/about.html>
- [2] Devlet Arşivleri Genel Müdürlüğü Web Sitesi, www.devletarsivleri.gov.tr
- [3] Amin, A., "Off-line Arabic Character Recognition: The State of the Art", *Pattern Recognition, Vol. 31, No. 5, 1998*, pp. 517-530.
- [4] Witten, I. H., Moffat, A. and Bell, T. C., *Managing Gigabytes*, International Thomson Pub., 1994.
- [5] Atıcı, A. A. and Yarman-Vural, F. T., "A Heuristic Algorithm for Optical Character Recognition of Arabic Script", *Signal Processing, Vol. 62, 1, 1997*, pp. 87-99.
- [6] Öztürk, A., Güneş, S. and Özbay, Y., "Multifont Ottoman Character Recognition", In *Proceedings of the 7th IEEE Int. Conf. on Electronics Circuits and Systems (ICECS)*, December 17-20, 2000, pp. 945-949, Jounieh, Lebanon.
- [7] Şaykol, E., Sinop, A. K., Güdükbay, U., Ulusoy, Ö., Çetin, A. E., "Content-Based Retrieval of Historical Ottoman Documents Stored as Textual Images", *IEEE Transactions on Image Processing (TIP), vol. 13, no.3, 2004*, pp. 314-325.