BÜTÜN YÖNLERİYLE OSMANLICA VE MİRASI ULUSLARARASI SEMPOZYUMU BİLDİRİ KİTABI

26-27 Nisan KIRIKKALE/TÜRKİYE

Editörler:
Prof. Dr. Eyüp BAŞ
Yrd. Doç. Dr. Ayten EROL
Yrd. Doç. Dr. Adem YILDIRIM
Yrd. Doç. Dr. Fatıma Zeynep BELEN

Dizgi-Düzenleme:
Ali ÇELİK

Baskı:
ÖZEL OFSET Basın Yayın Mat. Rek. İnş. Tur. San. Ltd. Şti.
Matbaacılar Sanayi Sitesi 1514. Sok. No:6
Yenimahalle–ANKARA
Tel: 0.312.395 06 08
Sertifika No: 29514

ISBN 978-975-8626-13-7

Birinci Basım Aralık 2016
KIRIKKALE ÜNİVERSİTESİ İSLAMÎ İLİMLER FAKÜLTESİ
Kırıkkale Üniversitesi Kampüsü 71450 Yahşihan/KIRIKKALE

OSMANLICA BASKI METİNLER İÇİN ARAMA ALTYAPISI

C. OZAN CEYHAN
MİLETOS AR-GE AŞ.
MELİH TAŞDİZEN
MİLETOS AR-GE AŞ.
BERKİN MALKOÇ
İSTANBUL TEKNİK ÜNİVERSİTESİ FİZİK MÜHENDİSLİĞİ BÖLÜMÜ
ATABEY KAYGUN
İSTANBUL TEKNİK ÜNİVERSİTESİ MATEMATİK MÜHENDİSLİĞİ BÖLÜMÜ
KÜRŞAT AKER
ORTADOĞU TEKNİK ÜNİVERSİTESİ

ÖZET

Bu çalışmada, Arapça ve benzeri alfabelerin optik karakter tanımasındaki yeniliklerden yararlanarak, taranmış Osmanlıca baskı metinler için bir arama altyapısını tanıtıyoruz.

ABSTRACT

In this work, we outline a prototype for an information retrieval system for printed Ottoman Turkish materials by employing recent developments in OCR technology for cursive alphabets, such as Arabic.

GİRİŞ

Bu çalışmanın amacı, sayısallaştırılmış Osmanlı ve erken Cumhuriyet dönemi (1928 Harf inkılabı öncesi) matbu eserlerine araştırmacıların erişimini daha verimli kılmak icin geliştiril-

miş optik karakter tanıma tabanlı bir arama alt yapısını tanıtmaktır. Halihazırda pek çok kurum, büyük bir özen ve titizlikle ellerindeki Osmanlıca eserleri sayısallaştırmış ve görüntü olarak saklamaya ve sunmaya başlamıştır:

İstanbul Büyükşehir Belediyesi Atatürk Kitaplığı'nda,

- 383.600 gazete görüntüsü
- 3.000.000 kitap görüntüsü

Milli Kütüphane'de,

- 80.000 cilt matbu belge (yaklaşık 2.500.000 görüntü)

Marmara Universitesi'nde

- 1.000.000 matbu belge görüntüsü

Hakkı Tarık Us Koleksiyonu'nda,

- 1366 süreli yayın

bulunmaktadır. Başbakanlık Devlet Arşivleri Genel Müdürlüğü, İstanbul Üniversitesi Nadir Eserler Kütüphanesi, Erzurum Atatürk Üniversitesi, İslâm Arastırmaları Merkezi (ISAM) ve İslâm Tarih, Sanat ve Kültür Arastırma Merkezi (IRCICA) de milyonlarla Osmanlıca belgeye ev sahipliği yapan kuramlardır.

Bu çalışma, Osmanlı İmparatorluğu'nun son dönem baskı belgelerinin -resmi belgeler, süreli yayınlar vb. de dahil olmak üzere- sayısal beşeri bilimler (digital humanities) yaklaşımları ile incelenmesi için gerekli araçları geliştirecek bir araştırma-geliştirme sürecinin ilk adımı olarak kurgulanmıştır. Bu nedenle, arama/sorgulama altyapısının optik karakter tanıma üzerine kurulmasına karar verilmiştir. Osmanlıca ile ilgili bilgisayar bilimleri araştırmaları Türkiye'de 1990'lı yılların ortalarında başlamıştır. Metinlerin taranarak bilgisayar ortamına aktarımından sonra iki esas problem karşımıza çıkar: Tanıma, arama ve sorgulama. Tanıma -optik karakter tanıması- bilgisayar görüntülerine dönüşmüş kağıt üzerindeki yazılardaki harfleri ve sözcükleri, bilgisayar metinlerine çevirmeyi amaçlar. Arama ve sorgulama, görüntü içerisinde aranan bilgileri bulmayı amaçlar. Arama ve sorgulama, bir tanıma adımının ardından metin tabanlı olarak uygulanabileceği gibi, tanıma adımı olmaksızın, doğrudan görüntü üzerinden de (içerik-bazlı sorgulama) uygulanabilir.

Literaturde, el yazısı metinlerde aramaya da daha uygun olması nedeniyle, "İçerik-Bazlı Sorgulama"ya (Content-Based Retrieval, CBR) odaklanılmıstır [5, 6, 7, 9, 18, 19, 21]. Bu amada, öncelikle görüntü kümesi içerisinde, ilgilendiğimiz görüntü parçaları (bu çalışmalar durumunda birbirine bağlı harf grupları) çeşitli özellikleri üzerinden indekslenir; bir görüntü aramak istendiğinde, aranacak görüntünün de aynı özellikleri hesaplanır, daha sonra indekste bu özlliklere göre arama yapılır.

Bu çalışma gibi, arama/sorgulama adımından önce görüntüyü tanıyarak başlayan sistemlerin farkı, aramayı tanıma adımı sonucunda oluşturulan metinler içerisinde yapmalarıdır. Metin içerisinde arama yaparken, bu sefer görüntü parçacıkları yerine, metin merisindeki süzcükler indekslenir. Bir süzcük arandığı zaman, sözcük indeksi üzerinde arama yapılır. Yalnız

ve arkadaşları [13, 15], Osmanlı arşivlerinde sorgulama yapmak amaçlı bir yazılım prototipini sunarlar. Söz konusu sistem yukarıda özetlendiği şekilde, optik karakter tanımasından sonra, indeksleme adımıyla, belgeler üzerinde sorgulama yapmayı saglar.

Bu calısma, Yalnız ve arkadaslarının çalısmaları [13, 15]'in devamı ve yeni teknolojilerle güncellenmesi niteliğindedir. Arapça ve benzeri, harflerin birbirine bağlanarak yazıldığı diğer alfabeler, optik karakter tanıma açısından Latin alfabesi gibi harflerin ayrı yazıldığı alfabelere göre daha büyük zorluklar göstermektedir. 2012 yılına kadar Arapça ve benzeri alfabelerin optik karakter tanıması için, saklı Markov modelleri (Hidden Markov Models, HMM) en gözde yöntemlerken (orn. [2] ve [12]), 2012 yılından bu yana özyinelemeli yapay sinir ağı modelleri (Recurrent Neural Network) modeli one çıkmıstır. Yapay Sinir Ağları (Artificial Neural Networks ya da kısaca, Neural Networks, NN), insan beyninden esinlenir, makine oğrenmesinde (Machine Learning) karşılaştığımız yaklaşımlardan biridir. Makine öğrenmesi, yüz tanıma ya da yazı tanıma gibi, standart bilgisayar algoritmalarıyla çözümü zor problemleri çözmek üzere geliştirilmiş bir yaklaşımdır: Bilgisayar, belirli bir problem konusunda (örneğin yüz tanıma) veri ile beslendikçe, sistem içerisindeki bir takım parametreleri gozden geçirir, değiştirir ve böylece veriyi öğrenir.

Bü çalışmadaki tanıma aşaması, özyinelemeli nitelikte bir yapay sinir ağı kullanarak, kendisine Osmanlıca kitaplardan oluşturulan okuma verisini öğrenir. Okuma verisi, taranmış kitaptan alınmış satırlar (girdi) ve bu satırların bilgisayar metni olan karşılıklarından (çıktı) oluşur. Makine öğrenmesi literatüründe, girdi-çıktı ikilileri ile yapılan bu tarz eğitime yönlendirilmiş eğitim denir.

Bu çalışmanın Yalnız ve arkadaşlarının çalışmalarından [13, 15] ana farkı, aramaya yönelik bir tanıma sistemi geliştirmek yerine, Osmanlıca gibi bağlı yazılan (çursive) alfabeler için büyük başarı göstermiş yapay sinir ağlarını kullanmasıdır. Tanıma açısından bir diğer fark, bu çalışmada geliştirilen tanıma sisteminin, yalnızca alfabedeki harfleri değil, el yazısında olduğu gibi baskı metinlerde de görülen, ligatürleri de tanıyor olmasıdır. Uygulama açısından bir diğer farkı da, görsel verinin metin verisine çevrilmesinden sonra, sorgulama amacıyla [13]'te kullanılan Zettair arama motoru yerine, güncel bir arama motoru olan Apache Lucene/Solr kullanmasıdır.

LİTERATÜR ÖZETİ

Osmanlıca metinlerin sayısal ortama aktarılmış görüntülerinde saklı olan metin içeriğinin kullanıma sokulabilmesi için son 20 yılda bazı akademik çalışmalar yapılmıştır. Bu çalışmaları birkaç başlık altında toplayabiliriz. Bir grup çalışma, genel olarak Osmanlıca metin tanımaya degil, spesifik bir alandaki ihtiyaçı karsılamaya dünük sistemlerin gelistirilmesine yönelik yapılmış; bir diğer grup ise, OCR teknolojisinden kaçınarak metinlerde sözcük aramanın önünü açacak alternatif yaklaşımların sınanmasına yönelik yapılmıştır. Bunların dışında kalan çalışmaları ise, esas olarak, Osmanlıca metinlerde OCR yapabilmek için gerekli ön adımların gerçekleştirilmesi olarak nitelendirmek mümkündür. Aşağıda bu çalışmalar ozetlenecek; ardından, bu çalışmada kullanmayı düşündüğümüz derin öğrenme (deep learning) yaklaşımının OCR alanındaki uygulamalarını ele alan güncel birkaç çalışmaya değinilecektir.

OCR sistemlerinde, karakter tanıma aşamasından oünçe, görüntü bir takım ön islemlerden geçirilir. Bunlar, tam sayfa metindeki satır ve sözcükleri ayrıştırmak gibi problemi daha basit bir noktaya indirgemek için gerekli adımlardır. Osmanlıca metinlerde satır ayrıştırma problemine odaklanan ilk çalışmalar 90'lı yıllarda yapılmış [1, 2, 3, 4]; bu çaba yakın zamanlara kadar başka gruplarca da devam ettirilmştir [18, 19, 21]. Satır ayrıştırma probleminin yanı sıra, [8] ve [10, 14] sözcük içindeki karakterlerin verimli bir temsilini olusturmayı amaçlar. Arapça karakterler için bir veritabanı ise [23] numaralı referansta sunulmuştur. Bunların dışında, genel olarak Arapça ve benzeri diller için faydalı yaklaşımların derlendiği çerçeve oluşturma niteliğindeki çalışmalara örnek olarak [16, 17, 20, 22] verilebilir. [17]'de, yazarlar bir ölçme/değerlendirme yapısı sunmaktadırlar.

Osmanlı Alfabesi ile yazılmış metinlerin optik karakter tanıması değişik yöntemlerle denenmiştir. [10, 14]'te yazarlar, doğrusal ayırtaç ayrıştırma (Linear Disçriminant Analysis, LDA) yöntemi, [29]'te ise tek katmanlı bir yapay sinir ağı ile tek tek harfleri tanımayı amaçlarlar.

[30]'deki tanıma sistemi, Osmanlıca yazılmış bir baskı metinde sözcükleri bulduktan sonra, sözcükleri harflere ayırıp, destek vektör makinası (Support Vector Machine, SVM) ile tek tek harfleri tanımak suretiyle sözcükleri tanırlar.

[11]'deki tanıma sistemi bu sefer, sözcükleri harf harf böldükten sonra, her bir harfi tek katmanlı bir yapay sinir ağı ile [30], tanımak suretiyle, sözcükleri tanır.

Harflerin birbirine bağlanarak yazıldığı Osmanlıca gibi alfabelerde harfleri birbirinden ayırmak zordur. Tek tek harfleri tanımak konusunda yuksek basarı bile elde edilse, ayırma adımında oluşacak yanlışlar, toplam sistem başarısını olumsuz yönde kuvvetle etkiler. Sözcükleri harflere bölmeksizin tanıma yaklaşımı

[2]'de ve [12]'de saklı Markov modelleri kullanılarak işlenmiştir.

Bunlara ek olarak, [15] arama/sorgulama amacıyla optik karakter tanıması yapan bir sistem geliştirmiştir. Sozcükler, harflere bölündükten sonra, alternatif yöntemler izlenerek, karakterlerin tanınması amaçlanır. Bu yöntemler, [5] ile ortaklıklar gösteren görsel karşlaştırma, [29] ve [30] ile benzerlik gösteren yapay sinir ağı kullanarak karşılaştırma ve çizge temelli karşılaştırmadır.

Osmanlıca metinler içerisinde arama/sorgulamaya yönelik yaklaşımlar, icerik- bazlı sorgulama çerçevesinde geliştirilmiştir; içerik-bazlı sorgulama, görüntü içerisinde görüntü aranmasına dayanır.

Optik karakter tanıma kullanmayan yaklaşımları ise temelde ikiye ayırmak mümkündür: Bu yaklaşımların ilki, görüntü merisindeki bağlı şekiller kümesi üzerinden metinlerdeki sözcükleri ve alt sözcükleri bulan yaklaşım (codebook yaklasımı); diğeri ise, doğrudan resimsel arama yapmayı sağlayan sözcük esleme (word spotting, word matching) yaklaşımıdır. İlk yaklaşıma Şeykal ve arkadaşlarının çalışmaları ([5, 6]) ile Yalnız ve arkadaşlarının çalışması [13] örnek verilebilir. Sözcük eşleme yaklaşımına ise ([7] ve [9]) ve en yakın tarihli olarak [24] ve [25] örnek verilebilir.

Yukarıda özetlenen çalışmalar, makine öğrenmesinin görece geleneksel yakla- şımları kapsamında yürütülmüs çalışmalardır. Bununla beraber, son birkaç yıldır tüm dünyada çesitli problemlere çok başarılı bir şekilde uygulanmış olan derin öğrenme yaklaşımı OCR'a da başarıyla uygulanmıştır ve özel olarak Arapça'da oldukça iyi sonuçlar alınmıştır [26, 27, 28].

OSMANLI ALFABESİ İLE SORGULAMA

OPTİK KARAKTER TANIMASI

Tüm optik karakter tanıma sistemleri, bir takım standart aşamalardan oluşur. Bu aşamalar, belgelerin tanıma için hazırlandığı önişleme, tanıma, tanıma sonrasında çıkan metnin düzeltilmesi aşamalarıdır.

Aşamalardan Osmanlıca'ya özgü kısımlar içermeyenleri, OCR sistemlerinde genel olarak kullanılan yöntemlerden seçilmistir [31], [ ]. Diğer aşamalar içinse gerekli açıklamalar ilgili yerlerde verilmiştir.

1.1.1 ÖNİŞLEME (PRE-PROCESSING)

Bu aşamada metin icçeren görüntüler aşağı geçirgenli filtreden geçirilerek (low pass filter), hem kağıt üzerindeki lekeler, hem de tarama sürecinde oluşan gürültü temizlenir. Metnin taranması sürecinde oluşan diğer bir sorun, metin içindeki satırların görüntünün alt ve üst kısımlarına paralel olmamasıdır. Bir sonraki aşamada, metin görüntüsünün söz konusu eğikliği düzeltilir (de-skew). Daha sonra, metnin sayısallastırması sırasında ortaya çıkan aydınlatma sorunlarım çözmek ve görsel veriyi basitleştirmek için, renkli ya da gri görüntü siyah- beyaz görüntüye dönüştürülür (binarization). Bu hazırlık işlemlerinin ardından, görüntünün kavramsal yapısının oluşturulması aşaması gelir. Görüntü, iki sayfadan oluşuyorsa, iki ayrı sayfaya ayrılır. Ardından sayfa düzeni çözümlenir. Her bir sayfa içerisinde, yazı olan kısımlar belirlenir. Yazı olan kısımlardaki okuma sırasının çözümlenmesinden sonra, yazı içeren görüntü parçaları satırlara ayrılır.

1.1.2 TANIMA (RECOGNITION)

Tanıma asamasında, sözcüklerin harflere bolünmesini gerektirmeyen özyinelemeli yapay sinir ağı mimarisi kullanılmıştır. Osmanlı alfabesi benzeri diğer bitişik yazılı alfabeler için bu yaklaşımın daha iyi sonuç verdiği görülmüştür (bkz. [28]).

İleri beslemeli yapay sinir ağları ile tanıma veya sınıflama probleminde girdiler ağa tek tek beslenir ve ağ bu herbir girdi için bir çıktı verir. Bu yaklaşımda girdiler birbirinden bağımsızdır. Özyinelemeli sinir ağlarının temel özelliği, ileri beslemeli ağların aksine, sinir ağını oluşturan katmanların kendi kendilerine de bağlantılar taşımalarıdır. Bu sayede özyinelemeli ağa art arda veri parççaları verilebilir ve ağ, sırayla gelen bu veri parçaları dizisinin bütününü kendi içinde ilintirendirerek doğru sonuca ulaşır. Dolayısıyla özyinelemeli sinir ağları dizi türündeki verilerle ilgili problemler için uygundurlar.

Osmanlıva karakter tanıma aşamasında, Osmanlıca yazıyı içeren görseller, önceki paragrafta özetlenen yaklaşıma uygun olarak, bir dizi olarak ele alınmış ve dizi şeklinde ağa beslenmişlerdir.

1.1.3 DÜZELTME (POST-PROCESSING)

Bu aşamada tanıma aşamasının çıktısı birbirine alternatif iki şekilde düzeltilir. Bunlardan ilki, sozcük dağarcığı (lexiçon) kullanarak doğrulamadır; tanıma sonucunda elde edilen herbir sözcük, sözcük dağarçığındaki sözcüklere karşı bulanık sözcük tutturma (fuzzy string matçhing) ile sınanır. Tanıma çıktısı ek olarak, n-gram dil modeli ile kullanılarak doğrulanır. Oluşturulan belge dağarcığı (çorpus) içerisinde harf ikilisi (2-gram) ve harf üçlülerinin (3-gram) istatistiki analizi yapılmak suretiyle, n=2, 3 için n-gram dil modeli elde edilmiştir. Tanıma aşamasından çıkan sözcükler, belge dağarcığından elde edilen bu istatistiklere göre düzeltilir.

SORGULAMA

Üzerinde sorgulama yapmak istenen metinlerin görüntüleri, tanıma asamasından geçirilerek bilgisayar metinlerine çevrilir. Belgelerin görüntü halleri, belgeler ile ilgili metadata, belgelerin tam metin hallerini içeren bir veritabanı yapısında ya da bir dijital arşiv sisteminde tutulur. Bizim uygulamamızda belgeler Islandora/Fedora dijital arşiv sisteminde tutulmaktadır.

Tam metin aramada yüksek performans ve işlevsellik için bu amaca özelleşmiş araçlar olan arama motorları kullanılır. Bu çalışmada da güncel ve yaygın kullanımlı Apaçhe Solr arama motoru sisteme entegre edilmiştir. Solr, Osmanlıça belgelerin görüntülerinin tanınması ile oluşmuş bilgisayar metinlerini indeksler ve böylece sorgulamalara hızla yanıt döner. Ayrıca, esnek sorgulama yapmaya ve arama ifadesini içeren sonuçların, sonucun kalitesine göre sıralanmasına olanak sağlar.

SONUÇLAR

Tanıma kısmının eğitilmesi aşamasında, 1327 yılında Manastır'da Beynelmilel Ticaret Matbaası'nda basılan Subhi Edhem'in Darvenizm kitabı kullanılmıştır [33]. Kitaptan elde edilen veri kümesinin %85'lik kısmı eğitim verisi, %15'lik kısmı ise test verisi olarak ayrılmıştır. Darvenizm, hem Osmanlıca, hem de Fransızca sözcükler içeren bir metindir. Eldeki kopya, görsel olarak temizdir. Ligatörlerle birlikte, tanıma aşamasında sistem 200'den fazla sembolü tanımayı oğrenmiştir.

Tanıma başarısı, eğitim metinleri üzerinde %85.1 test metinleri üzerinde %82.3 olarak açıklanmıştır.

Sistemin toplam başarısını ölçen sorgulama başarısı, %89.7 olarak gözlemlenmistir.

TEŞEKKÜR

Bildiriye konu olan sistemin geliştirilmesi asamasında dil ve tarih bilgisi konularında yardımcı olan Tuncay Zorlu, Rahmi Deniz Ozbay, Deniz Taner Kılıncoglu ve Akile Zorlu Durukan'a teşekkür ederiz.

KAYNAKLAR

[1] F. T. Yarman Vural and A. Atici, A Heuristic Algorithm for Optical Character Recognition of Arabic Script, SPIE 2787, 725 (1996).

[2] A. Alper Atici, Fatos T. Yarman Vural, A heuristic algorithm for optical character recognition of Arabic Script, Signal Processing 62, 8799 (1997).

[3] E. Oztop, A. Y. Mülayim, V. Atalay and F.T. Yarman Vural, Repulsive Attractive Network for Baseline Extraction on Document Images, IEEE, 3184 (1998).

[4] E. Oztop, A. Y. Mülayim, V. Atalay and F.T. Yarman Vural, Repulsive Attractive Network for Baseline Extraction on Document Images, Signal Processing 75, 1 10 (1999).

[5] E. Saykol, A. K. Sinop, U. Gudukbay and O. Ulusoy, Content Based Retrieval of Historical Ottoman Documents Stored as Textual Images, IEEE Transactions on Image Processing 13, 314 (2004).

[6] I. S. Altıngüvde, E. Şaykol, O. Ulusoy, U. Güdükbay, A. E. Cetin and M. Gocmen, Content Based Retrieval (CBR) System for Ottoman Archives, IEEE, (2006).

[7] E. Ataer and P. Duygulu, Retrieval of Ottoman Documents, MIR'06 (2006).

[8] W. G. Al Khatib, S. A. Shabab and S. A. Mahmoud, Digital Library Framework for Arabic Manuscripts, IEEE, 458 465 (2007).

[9] E. Ataer and P. Duygulu, Matching Ottoman Words: An image retrieval approach to historical document indexing, CIVR'07, 341 347 (2007).

[10] Z. Kurt, H. I. Türkmen and M. E. Karslıgil, Ottoman Alphabet Character Recognition by LDA, (2007).

[11] N. Kilic, P. Gorgel, O. N. Ucan and A. Kala, Multifont Ottoman Character Recognition Using Support Vector Machine, IEEE, 328 (2008).

[12] A. Onat, F. Yildiz and M. Gündüz, Ottoman Script Recognition Using Hidden Markov Model, World Academy of Science, Engineering and Technology 2, 630 632 (2008).

[13] I. S. Yalniz, I. S. Altıngovde, U. Gudukbay and O. Ulusoy, Integrated segmentation and recognition of connected Ottoman script, Optical Engineering 48(11), 117205 (2009).

[14] Z. Kurt, H. I. Turkmen and M. E. Karsligil, Linear Discriminant Analysis in Ottoman Alphabet Character Recognition, (2009).

[15] I. S. Yalniz, I. S. Altıngovde, U. Güdükbay and O. Ulusoy, Ottoman Archives Explorer: A Retrieval System for Digital Ottoman Archives, ACM J. Comput. Cult. Herit. 2(3) , 8 (2009).

[16] K. Pramod Sankar, C. V. Jawahar and R. Manmatha, Nearest Neighbor based Collection OCR, Das'10, 207 (2010).

[17] I. Z. Yalniz and R. Manmatha, A Fast Alignment Scheme for Automatic OCR Evaluation of Books, IEEE, 754 (2011).

[18] E. F. Can and P. Duygulu, A line based representation for matching words in historical manuscripts, Pattern Recognition Letters 32, 11261138 (2011).

[19] H. Adiguzel, E. Sahin and P. Duygulu, A Hybrid Approach for Line Segmentation in Handwritten Documents, IEEE, 503 (2012).

[20] I. Z. Yalniz and R. Manmatha, An Efficient Framework for Searching Text in Noisy Document Images, IAPR, 48 (2012).

[21] H. Adıgüzel, P. Duygulu Şahin and M. Kalpaklı, Line Segmentation of Ottoman Documents, IEEE (2012).

[22] L. Chen, S. Kapoor and R. Bahatia, Intelligent Systems for Science and Information from the Science and Information Conference 2013, Springer (2014).

[23] R. M. Saabni and J. A. El Sana, Comprehensive synthetic Arabic database for on/off line script recognition research, IJDAR 16, 285 294 (2013).

[24] M. Khayyat, L. Lam and C. Y. Suen, Learning based word spotting system for Arabic handwritten do-

cuments , Pattern Recognition 47, 1021 1030 (2014).

[25] P. Duygulu, D. Arifoglu and M. Kalpakli, Cross document word matching for segmentation and retrieval of Ottoman divans, Pattern Anal Applic. (2014).

[26] T. Bluche, J. Louradour, M. Knibbe, B. Moysset, F. Benzeghiba, Christopher Kermorvant, The A2iA Arabic Handwritten Text Recognition System at the OpenHaRT2013 Evaluation. International Workshop on Document Analysis Systems (DAS) bildiri kitabı içinde (2014).

[27] V. Pham, T. Bluche, C. Kermorvant, J. Louradour, Dropout improves recurrent neural networks for handwriting recognition. International Conference on Frontiers in Handwriting Recognition (ICFHR) bildiri kitabı içinde (2014).

[28] B. Moysset, R. Messina, C. Kermorvant, A Comparison of Recognition Strategies for Printed / Handwritten Composite Documents. International Conference on Frontiers in Handwriting Recognition (ICFHR) bildiri kitabı icinde (2014).

[29] A. Oztürk, S. Günes, Y. Ozbay, Multifont Ottoman Character Recognition, 7th IEEE Int. Conf. on Electronics Circuits and System (ICECS) bildiri kitapçıgı içinde (2000).

[30] Pelin Gorgel, Niyazi Kilic, Birsen Ucan, Ahmet Kala, Osman N. Ucan, A Backpropagation Neural Network Approach For Ottoman Character Recognition, Intelligent Automation & Soft Computing 15, 3 (2009).

[31] H. Cao, P. Natarajan, Machine Printed Character Recognition, Handbook of Document Image Processing and Recognition kitabı içinde Bülum 10, Springer (2014);

[32] F. J. Omee, S.S. Himel, Md. A. N. Bikas, A Complete Workflow for Development of Bangla OCR, International Journal of Computer Applications, Vol 21. (2011).

[33] Subhi Edhem, Darvenizm, Manastır, Beynelmilel Ticaret Matbaası, 1327 (1911).