



presented by Damla Arifoglu

Pinar Duygulu
Computer Science Department,
Hacettepe University, Ankara, Turkey

Damla Arifoglu
Computer Science Department
Sabanci University, Istanbul, Turkey

What is OTAP?

- OTAP is jointly sponsored by the University of Washington in Seattle and Bilkent University in Ankara, Turkey under the direction of Professor Walter G. Andrews (U.W.) and Professor Mehmet Kalpaklı (Bilkent).
- OTAP is a project employing computer technology to make transcribed Ottoman texts and resources for understanding Ottoman texts broadly accessible to international audiences.

Why Ottoman?

- The Ottoman Empire lasted for more than 6 centuries (1299-1922), spread over 3 continents
- More than 150 million historical documents were produced
- Huge collections in the archives, libraries, museums and private collections of almost forty nations
- Interest of scholars from many disciplines and from many different countries

Documents in Archives

- Number of documents in Başbakanlık Osmanlı archives is 150.000.000 (T.C. Başbakanlık Devlet Arşivleri, 2008; Özdemir, 2003; Bardakçı, 1998)
- Süleymaniye Library in Istanbul has 65.000 handwritten, and 45.000 old printed documents (Kalpaklı, 1991).

Challenges of Ottoman

- The Ottoman is a connected and cursive script
- The Ottoman alphabet is a subset of the Arabic alphabet
- Has additional vocals and characters from the Persian and Turkish scripts
- No explicit word boundaries
- Most of the characters are only distinguished by adding dots and zigzags (called as diacritics)

Ottoman Documents



Ottoman Documents



براه رو که در کجایم در شنبه چو یو که کیدم که یولی انگ پوسه سچ بنیسه ایلی قری مسکین برید و ادی
ورده در حرم نداشت مسکین کسدی دزه بیله قطع برتیه ایلیدی و یولی حربه طوقدی بیله ایلی قری حاقظ
بیر بر تو کوی سعادت که مدعی ای حافظ ابلت سعادت طوبیت کمدعل هجینش هنر نبود و خبر بر هم راست هیچ
انگ هنری یوغدی و خبر دخی طوقتا ایلی بیله سوزدن خبری دخی یوغدی خراد بونه علم فصاحت و علم
بلغت سعادت اید و کن بیانه و کند و ده موجود اولد و غنسه شکر در و کله ایضا. بی مهر رحمت روز را
نور خاندست سنگ رهک کونش سوزنم کونم نور خاندست در و عمرها جزب و بجز خاندست خوردن
بجا قره کونم کونم غیره قائم شد در هنگام وداع بوز زبوس کرم کرم سنگ و دایم و فتنه چوقه غلغله
دور اوز بوج چشمه نور خاندست سنگ رهکون اراجه بتم کوزنه نور خاندست میرفت خیال نور خاندست
و سیکست کیدرادی سنگ خیاک بتم کوزندن و دیرادی خیریه است از بتم کوزنه که معلوم خاندست مقول قول
یومر اعدر خراد و رک لسان حال کلا شنه حکایت کرم کوزدن کیده بو کوشه خراب اولور
و مسل تو ایل را رسم دور همی داشت سنگ و سلوک بیله بتم خاندست ابراه طو تراوی آرد و لست چوقه کونم
دور خاندست سنگ چکر که دولشندن شدی اراج قائم شد در نزدیک شد آن دکلر قیبه بو کونم تعیین
اولدی اولدو سنگ و سیکه دیه دور از درشته این خسته و بکچور خاندست سنگ فایکدن اراج اولی
خسته متروک قائم ریخینه اولمدر خسته نهان شتر روز وصل یار و سام حجی سینه کمانی ای ای جلا کسب
من بعد دیه سوزدن قری ریخه کند دوست من بعد نه فایده اگر قدم ریخه ایدر سه دوست کز جاقان رفیق
در من در کور خاندست درمن بقیه دوه در بر لریننه فایده بر بقیه خسته شنده قائم شد خرا بود که
چون اجل ایدر احبنا و اطینا فایده انتر در بچور کوز چشم مراب فایده سنگ چکر که بتم کوزنه صدقالبه
اغلیندن آب نام اولر کوه خوله چکر بر کوز خاندست در کوزنه چکر فایده که مغذور قائم شد
یعنی که خدر یوقدر خیر است خرا جاده چو این نو لیکن خبر در بجا سنگ فایده جاده سی اما چون خبر
نوان کرد که مقدور خاندست حافظ زخم از کرم پیر داشت بخنده حافظ اغلیندن فایده اولمدر کونم

Ottoman Documents

سید و اول مظہر انوار جلی
 اسدا لله ولی یعنی سیلے
 نہ کہ جریمہ تا بانگ
 یوزی بگزیدی مدوزایہ
 ذاتی یدینہ ایدی مولایہ
 ساغر غارض مہ سیمایہ
 بادہ عشق ایدی محبت
 عرض حسن ایسہ او محمد علی
 یوسفی گزیدی سراشک
 شویله بر نور ایدی اول وجه
 اگا باقلمز ایدی شوقندان
 عالی دولت ناکا کجی
 تابلک ایلردی ماہ کجی
 واضح ایتمشدی او یوزدن فلح
 نید و کنایت فیہا مصباح
 کعبہ وجہنہ ایندیکجه نظر
 سجدہ ایلر لرایدی شمس و مثر
 دیر عالمده اوروی پر شتاب
 اید مرتقلانی عقل و عقال
 معجزات رخنک بالاجمال
 اہناب فلک دین ایدی اول
 مردم چشم خدا بین ایدی اول

احداث تریک تریک سیتانہ سیتانہ سیتانہ سیتانہ
 برینین قمانت نام اولی ازہ و تریک تریک
 باریستہ ازہ تریک تریک تریک تریک تریک تریک
 ماہہ تریک تریک تریک تریک تریک تریک

برینین قمانتہ و باج تریک تریک تریک تریک تریک تریک
 رویت اولدی تریک تریک تریک تریک تریک تریک
 دوا ای تریک تریک تریک تریک تریک تریک
 فضل رویت اولدی ۲۲ صبح اول اول شنبہ ۱۲ صبح اول شنبہ

دوران حاکمیت
 رئیس القاتب

(مراتب امرت و کائنات علی افضل الترات)

اول صبح آسمان خزانہ	صدقہ اول نور کا سوزانہ
کر آدمی قیدی حوشن یاز	عدو کو سہ ایسٹن اول خاز
بریا پور واقع نونگارہ	ایتری رمنع سوزانہ
سومہ دکان اولی آدم	انسانی قیلوب کجی گرم
مرفوشری شمس پورانی	اولی کرین یوزانہ انسان
برداشتہ نونگارہ کلاکت	قیدی کادریس متعلق اولر کلاکت
سز کن ایتری حوشن اول	اولی تم رسا کلاکت خرمنا
اول سوزج اعقاب سب	اول سوزج راجیل ایسا سب
جم گرم کائنات قیدی	جم انرف سب جہانست قیدی

خوشید جانورہ سوزہ
 سوزہ قزاقیا محمد
 اوسون

Ottoman Documents

Handwritten Ottoman Turkish text, likely a legal or administrative document. The script is dense and written in a cursive style. The text is arranged in several columns, with some lines starting with large, decorative initial letters. The paper shows signs of age and wear.

۱۶۱۱ هجری قمری جو بد معنائی رنگین شہر اینک
اول کا رنگ اندر ارباب تعلیمی سب جو مناد **ابغا**
صفا بر نمندہ شول زین توح اکلمو بد کنسار باغ عشرتہ الحبش رو غبار
نخال نارون صحن جنبہ سالنویا میان عید محمد بزرگار سروالادر
مثال خون تبر اندر شقایق صحن گلشنده جھمان بر ششور غوغا شولیک بکر عید
دزلش لاله صحن چرم بنامچہ زین اش سر ای شامدھان کا کاس خون بنامادر
کل فرکس بینہ التون بنکلو جامہ لبش لبس لاله رنگ کنسندہ بر دیبا نیسار
اکردنیساودین امرندہ نابع بر عمل بک دعای دولت سلطان مقلدین دینادر
نہ رنگ بویہ شتون اول نہای سوریا بک بونایند بقا از ترنشا شولیک انا در

Handwritten Ottoman Turkish text, likely a legal or administrative document. The script is dense and written in a cursive style. The text is arranged in several columns, with some lines starting with large, decorative initial letters. The paper shows signs of age and wear.

Sample Ottoman Words

مقتضای غزل آیردی

حال پنهانمدن فلك

نمد غم نظر کیمسه مقتضا

طو پنها یاه فلكی عجب

ششمنده چپقا صل ممکن یتکی

Challenges of Ottoman

	initial	middle	final	isolated											
ba	ب	ب	ب	ب	ذ	د	خ	ح	چ	ج	ث	ت	پ	ب	ا
	ب	ب	ب	ب	z	d	h	h	ç	c	s	t	p	b	-
	ب	ب	ب	ب	dhal	dal	kha	ha	che	jim	tha	ta	pa	ba	alif
ta	ت	ت	ت	ت	غ	ع	ظ	ط	ض	ص	ش	س	ژ	ز	ر
	ت	ت	ت	ت	ğ	‘	z	ṭ	z	ṣ	sh	s	j	z	r
	ت	ت	ت	ت	ghayn	ayn	zah	tahn	dad	sadd	shiin	siin	ja	za	ra
tha	ث	ث	ث	ث	ی	ه	و	ن	م	ل	ک	گ	ک	ق	ف
	ث	ث	ث	ث	y	h	v	n	m	l	k	g	k	q	f
	ث	ث	ث	ث	ya	ha	waw	noon	miim	lam	kaf	kaf	kaf	qaf	fa
jim	ج	ج	ج	ج											
	ج	ج	ج	ج											
ha	ح	ح	ح	ح											
	ح	ح	ح	ح											
kha	خ	خ	خ	خ											
	خ	خ	خ	خ											
kaf	ک	ک	ک	ک											
	ک	ک	ک	ک											

OCR is difficult for Ottoman compared to Latin alphabet

Main difficulties with Ottoman :

Meaning is content dependent

Suffixes in Turkish cause mistakes

The language/vocabulary differences in time

Challenges of documents in archives

جناب حق ولی نعمت جهان خلیفہ
 Thuluth

جناب حق ولی نعمت جهان خلیفہ رسول رب
 Naskhi

جناب حق ولی نعمت جهان خلیفہ رسول رب
 Diwani

جناب حق ولی نعمت جهان خلیفہ
 Royal Diwani

جناب حق ولی نعمت جهان خلیفہ
 Ta'liq

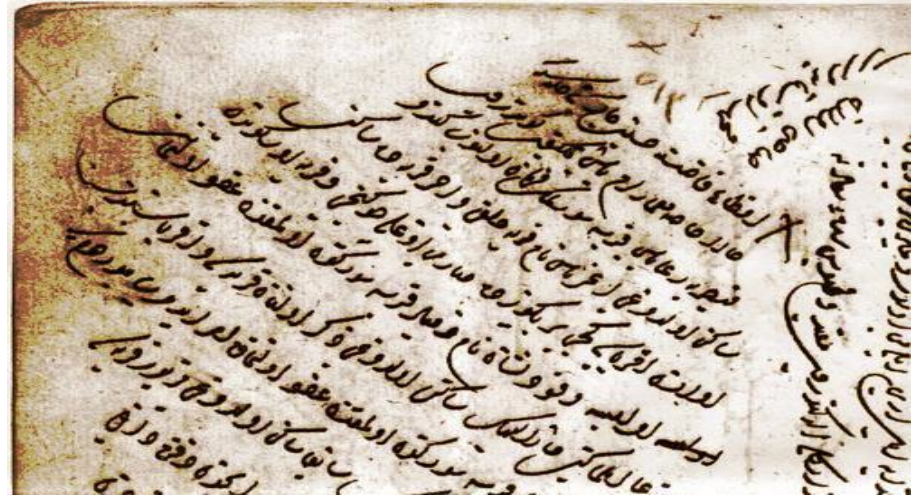
جناب حق ولی نعمت جهان خلیفہ
 Kufi

جناب حق ولی نعمت جهان خلیفہ رسول رب
 Rika

عشما تلو شیقہ لرینی او فوما یا کیر لیش
 Rayhani

حرف العریض جرایف کرسول ربنا
 Siyakat

کتبخانه لرده یا پیلاجق تدریجی اصلاحات
 Printed



Segmentation problems in handwritten documents

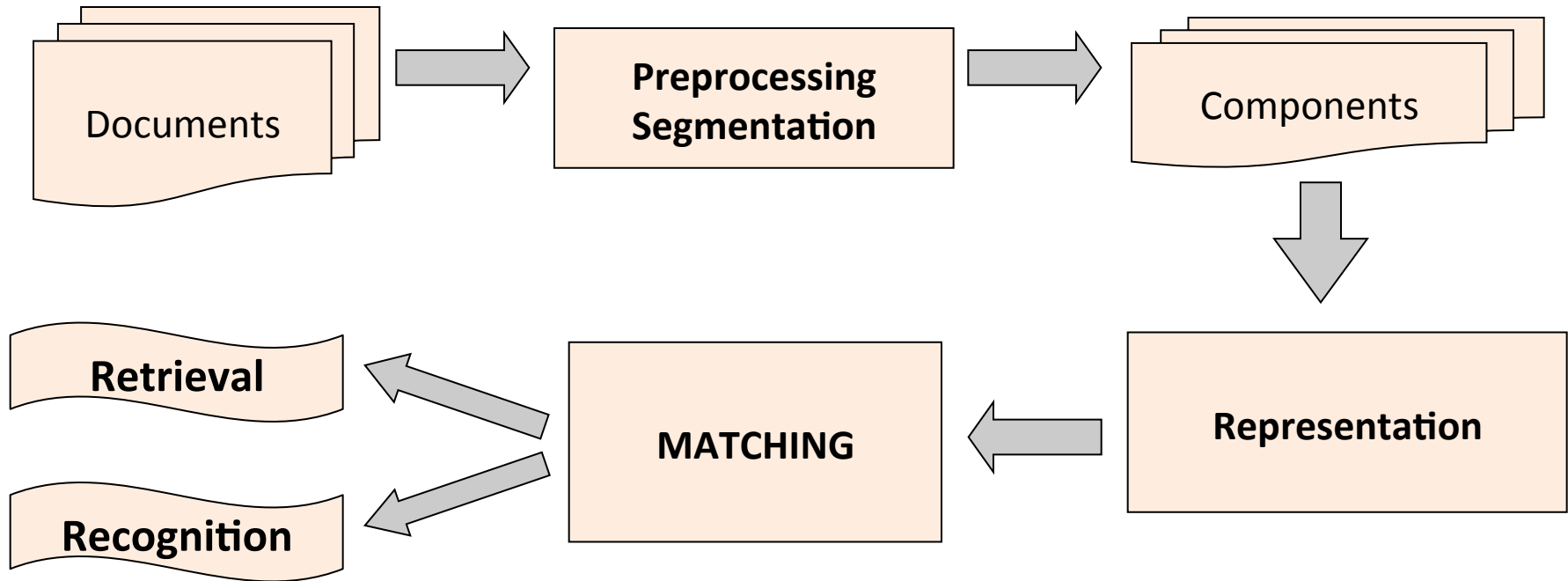
printed

نصل بویله بر ایمانی بو غار مدنت دید یکن تک دینی قامسه جانا وار

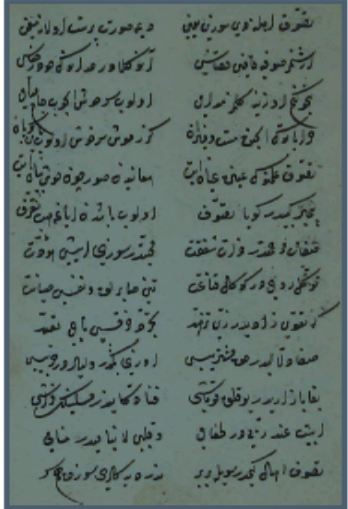
handwritten

۳۳۳ تاریخی و ۲۲۸ نومرولی امرنامه لر یله اشعار بیورلش ولدی التحقیق غلطه

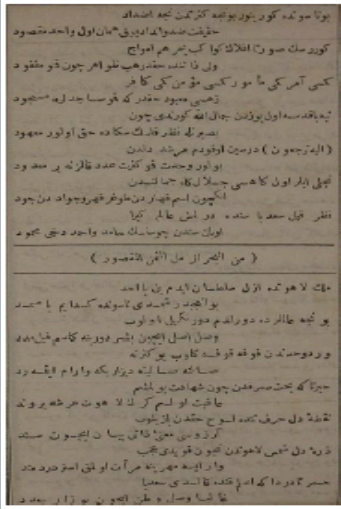
System Layout



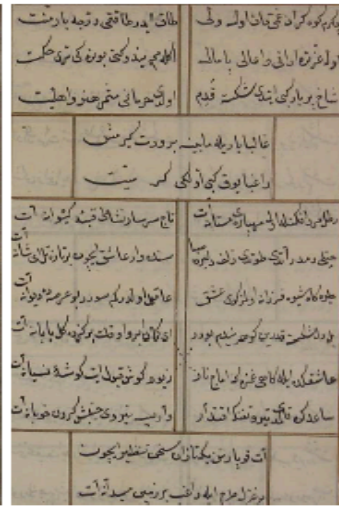
Classification of Calligraphy styles



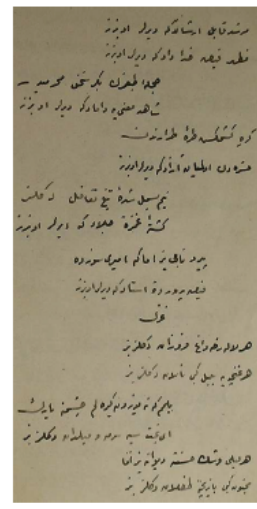
(a) Divani



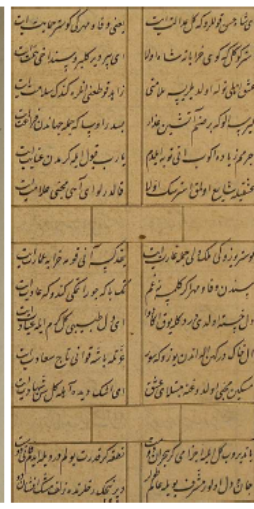
(b) Matbu



(c) Nesih



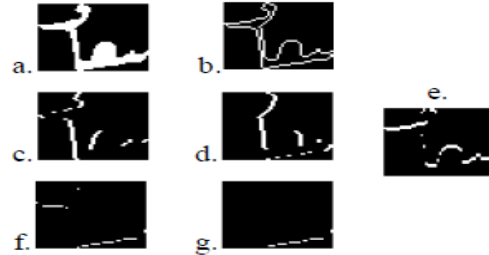
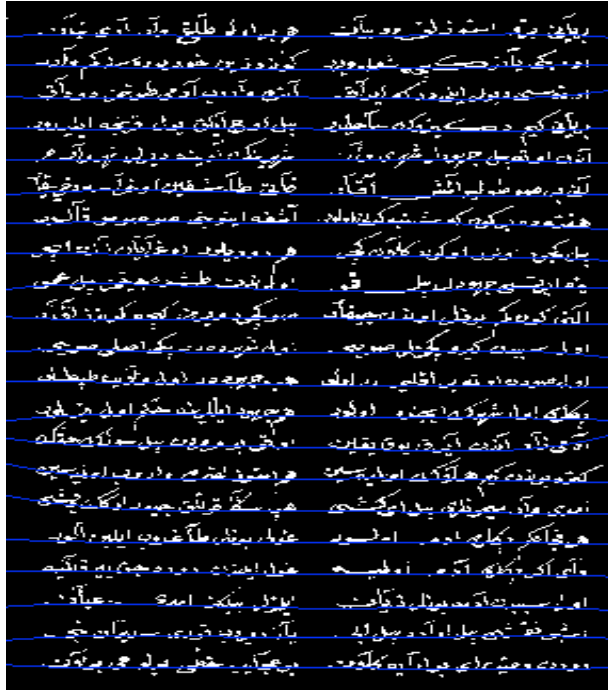
(d) Rika



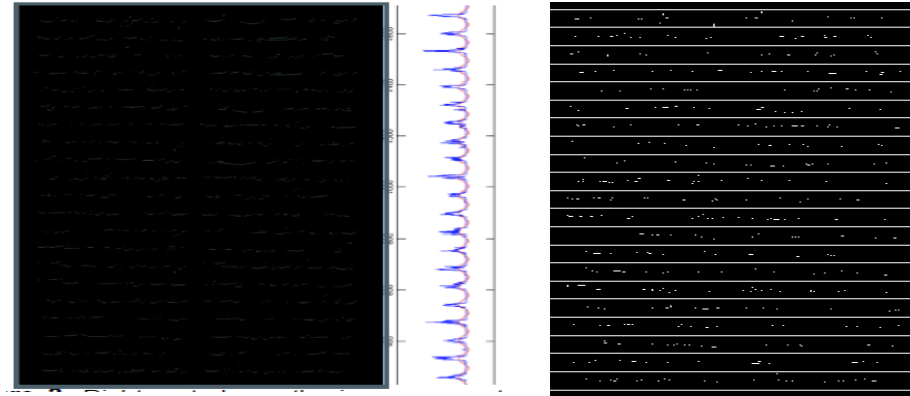
(e) Talik

True Style	Estimated Style				
	Divani	Matbu	Nesih	Rika	Talik
Divani	98.95	0.00	0.00	1.05	0.00
Matbu	0.00	100	0.00	0.00	0.00
Nesih	0.00	0.00	100	0.00	0.00
Rika	0.00	0.00	0.00	100	0.00
Talik	0.00	0.00	0.00	1.05	98.95

Line Extraction



0.90	Precision	Recall	F1 - Score
Printed	99.56	95.03	97.24
Handwritten	99.33	92.37	95.72



A Hybrid Approach for Line Segmentation in Handwritten Documents, Hande Adiguzel, Emre Sahin, Pinar Duygulu, 13th International Conference on Frontiers in Handwriting Recognition (ICFHR), Bari, Italy, September 18-20, 2012

Word matching

ج - آثار عتيقه موزه لری کتبخانه لری : (ولایت موزه لری کتبخانه لری) خلق کتبخانه لری ماهیتنی حاتر اولاجقدر .

د - اوقافدن معارفه دور اولنان کتبخانه لری : علمی کتبخانه ماهیتنی حاتردر . سرحدلرده بولونان و موزه لك قیمتی احتوا ایدن کتبخانه لری تامایله ویا قسماً آنقره بیه دولت کتبخانه سنه نقل اولوناجقدر .

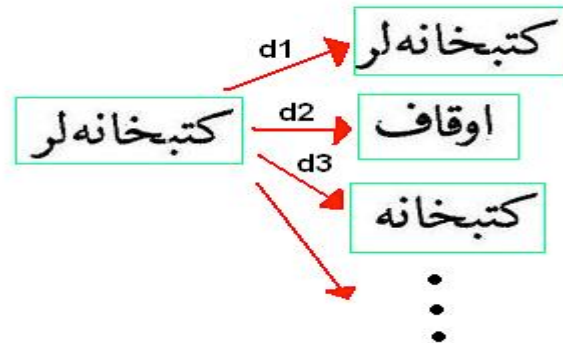
کتبخانه لری مفتش عمومیلری ویا کتبخانه لری انجمنی هر هانکی بر کتبخانه نك نقلنه لزوم کوستریسه بو کی کتبخانه لری آنقره ده کی دولت کتبخانه سنه یاخود دیگر بر کتبخانه بیه نقل اولوناجقدر . آنا کتبخانه بیه نقل اولونان کتبخانه لری اسکی عنواننی محافظه ایدرلر . قونیه ، بروسه ، قسطمونی کی مرکز لرده کتبخانه بنالری وجوده کتیرلده کده مختلف مدرسه ویا جامعلرده بولونان اوقاف کتبخانه لری علمی هیئتک نظارتی آلتنده توحید اولوناجقدر .

هیئت جلیله نك قراری موجبجه یازمه اثرلرک خارجه اخراجی ممنوعدر . بودجه نك مساعده سی درجه سننده اشخاص آلتنده بولونان قیمتدار یازمه اثرلرله ، اسکی مطبوعلر کتبخانه لری و موزه لری طرفندن جمع اولوناجقدر .

اوقافدن دور اولونان کتبخانه لری اثرلرک تفتیش و تصنیفی ایچون علمی بر هیئت استخدام اولوناجقدر .

خلق کتبخانه لری : علمی کتبخانه لری تامایله فرقل اولوب یوکسک مکتبلرده تحصیل کورمه مشن خلقک سویه سنی یوکسلته جک ، علم و عرفانه قارشنی ایستهک اویاندیراجق ، زمانی مطالعه ایله خوشبجه کچیرمکه سائق کتابلری احتوا ایدر .

segmentation

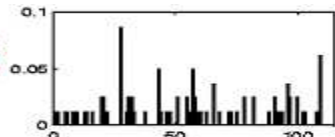


Word matching

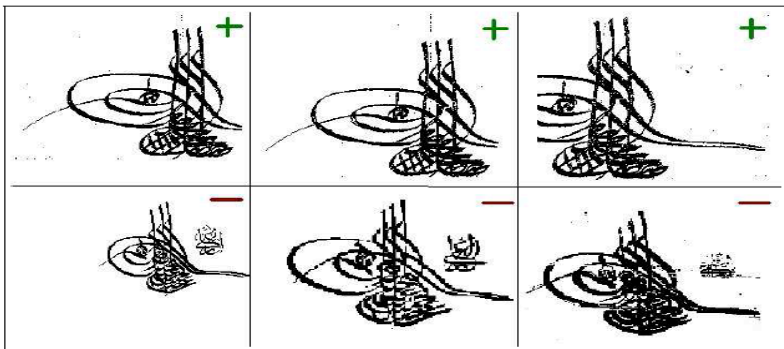
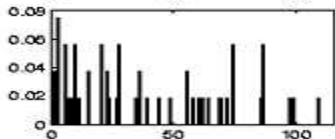
How to represent a word?

Keypoint-based representation

قسطنونی



اثر لرك



printed

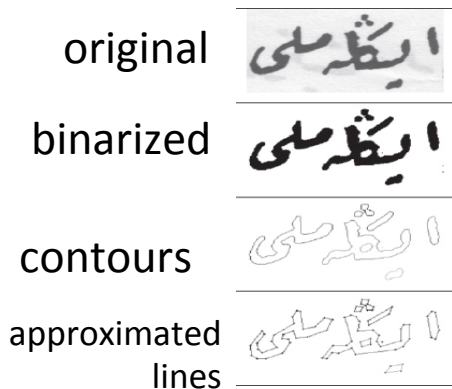
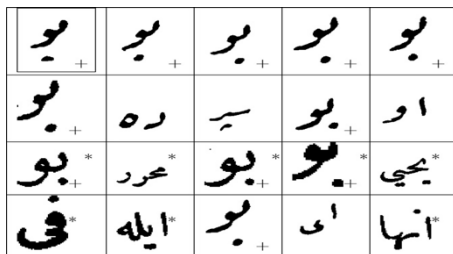
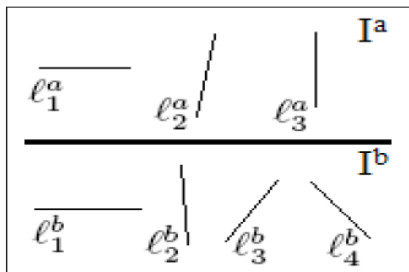
اولان	اولان	اولان	مستقبل استیالات اوزبک	مستقبل استیالات اوزبک
اولان	اولان	اولان	مستقبل استیالات اوزبک	مستقبل استیالات اوزبک
اولان	اولان	اوزبک	اولان	اولان
اردو	اردو	واردی	وافراددن	اردو
اردویه	اردویه	ذوات	درمیان	اردو
اردوبردی	اردو	برذات	واردی	فوق العاده
ملتك	ملتك	ملتك	ملتك	ملتك
مواصلتك	حركاتك	ملتك	مساعدین	آماسیهدن
ملینك	ملینك	ایته كندی كاپور	منسوبلرینك	مواصلتك

Rika

استقلال	استقلال	عالم	استقلال	صلوات
بنم	بنم	بنم	اوغراتما	یرلری
بنی	ای	بو	حیرجه سن	بر
وجدا	کیم	بنم	دینك	اولماز

Matching Ottoman Words: An Image Retrieval Approach to Historical Document Indexing, Esra Ataer, Pinar Duygulu
 In Proceedings of ACM International Conference on Image and Video Retrieval, (CIVR), July 9-11 2007, Amsterdam, The Netherlands

Line-based representation



Method	Data set	Precision	Recall
our approach	GW10	0.688	1.000
our approach	GW10	0.774	0.770
DTW (Rath and Manmatha [23])	GW10	0.653	0.711
DTW (Rath and Manmatha [22])	GW10	0.726	0.652
our approach	GW20	0.566	1.000
our approach	GW20	0.667	0.673
DTW (Rath and Manmatha [22])	GW20	0.518	0.550
our approach	OTM1	0.987	1.000
bag-of-words (Ataer and Duygulu [4])	OTM1	0.910	1.000
DTW (Ataer and Duygulu [4] ¹)	OTM1	0.940	1.000
our approach	OTM2	0.944	1.000
bag-of-words (Ataer and Duygulu [4])	OTM2	0.840	1.000
our approach	OTM1+2	0.957	1.000
bag-of-words (Ataer and Duygulu [4])	OTM1+2	0.810	1.000

A line based representation to match the words in historical manuscripts, Ethem Fatih Can, Pinar Duygulu, in Pattern Recognition Letters, Volume 32, Issue 8, Pages 1081-1222, June 2011

Redif extraction

1.1	Seni seyr itmek için reh-güzer-i gülşende
1.2	îki cânibde durur serv-i hırâmân saf saf
2.1	Mescid içre göre tâ kimlere hemzânûsın
2.2	Şekl-i sakkada gezer dide-i giryân saf saf
3.1	Gökde efgân iderek sanma geçer hayl-i küleğ
3.2	Çekilür kûyune mürgân-ı dil ü cân saf saf

1.1	پسنى سیر ایتکون ره گز گلشنده
1.2	ایکی جانیده طوزره سپرو خرامان
2.1	مسجد اچره گوره تا کیلره همزانوسین
2.2	شکل سقاده دگز دین کریمان
3.1	گوکدن افغان ایدرک صنمه کچرخل
3.2	چکیلور کویکه مرغان دل و جان

Redif Extraction in Handwritten Ottoman Literary Texts. Ethem F. Can¹, Pınar Duygulu, Fazlı Can, Mehmet Kalpaklı, Int. Conf. on Pattern Recognition (ICPR) 2010

Redif extraction

کم ای دوست
 نیست بتو اشتباهم ای دوست
 یغوشود نکاهم ای دوست
 ابروس تو فدا کام ای دوست
 در جلد جهان شام ای دوست
 خندم و چشم و سباهم ای دوست

نه بگنزر
 کبر عمل اولدم اندر کانه بگنزر
 وطن جبر همان ایمانه بگنزر
 خیالی کعبه ده حست نه بگنزر
 رکاب خسر و دوانه بگنزر

ت بولمز
 دعوای حسن و جمال ایلیه صورت بولمز
 یوست ایگز حمله سزمت بولمز
 قصه درد و غم یار نمانست بولمز
 قنق عاشق که اگر او غم نیست بولمز

لا اول
 کل کبسی کوله صورت عالم بهار اول
 عکسه رخکله سخن حین لاله زار اول
 کربنزم در یسه قتی سیر اول
 حنک قشده اکاذخی نه اعتبار اول

باف سید
 شول اقی زیان پلن زبان سید
 ارم سرای ماهن پوستای سید
 بو صم فانیه ابروغ نه حاضر سید
 چمنده سیر ز بتس پرو پوستای سید

ادن کج
 گرم ایله دل شیدا دن کج
 زور قیله یوزی در پلون کج
 چنبره چرخ مطلق دن کج
 ای امید می حمر دن کج

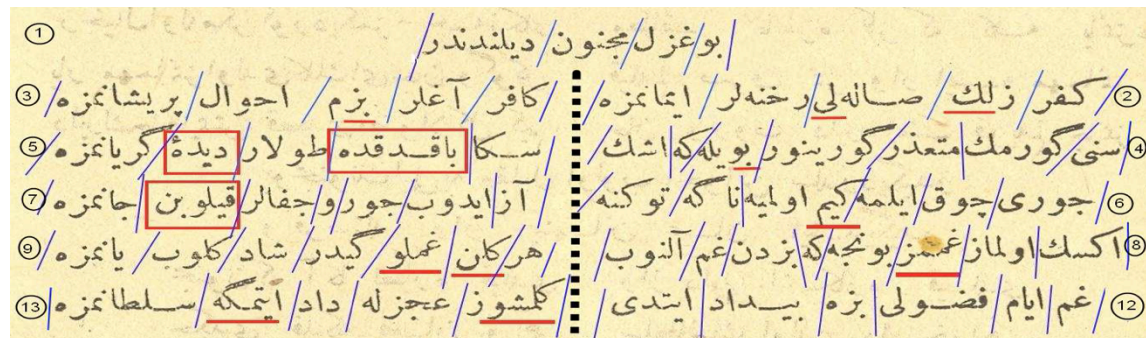
ان صفت
 گویا چنگله دور ریزه گذار صفت
 ایکی جاننده دور رسد خفا صفت
 کوزور مو حله برین لبه آغ صفت
 شکل ستاده کوز دیره کرب صفت
 حکم کوز کوز فاعل اول صفت

اگور
 کله دی تیشدی آخر افندی بلای کور
 مزات جایی آل نظر ایله صفای کور
 بر مقله کوز ستراله قبسه نمایی کور
 رنگ اندی که معنیه زید و ریایی کور
 او کوب کوز که چقره دلای دل جانی کور

Redif Extraction in Handwritten Ottoman Literary Texts. Ethem F. Can1, Pinar Duygulu, Fazlı Can, Mehmet Kalpaklı, Int. Conf. on Pattern Recognition (ICPR) 2010

Cross-document word matching

- ① بو/غزل/مجنون/دیلندندر/
- ② کفر/زلفک/صالحه/لی/رخنه لری/ایمانمزه/
- ③ کافر/آغلر/بزم/احوال/پرشانمزه/
- ④ سنی/کورمک/متعذر/کورینور/بویله/که اشک/
- ⑤ سکا/باخذقده/طولار/دیده/کریانمزه/
- ⑥ جورى/چوق/ایلمه/کیم/اولیه/ناکه/توکنه/
- ⑦ آز/ایدوب/جور/و/جفالر/قبیلا/سن/جانمزه/
- ⑧ اکسک/اولما/زغممز/بوغجه/که/بزدن/غم/آلوب/
- ⑨ هر/کلن/غملو/کیدر/شاد/کلوب/یانمزه/
- ⑩ وار هر حلقه زنجیر مزک بر آغزی
- ⑪ متصل ویرمکه افشا غم پنهانمزه
- ⑫ غم/ایام/فضولی/بزه/بیداد/ایتدی/
- ⑬ کلمشوز/عجزله/داد/ایتمکه/سلطانمزه/



حشمت ایله	یجیله	سرشته سن	یتزمی که
حشمتیله	یجیله	سررشته سن	یتزمیکه

منی	بتر	ایتمز	کیم	غم	قیلم
بی	بتر	ایتمز	کیم	غم	قیلم

Duygulu, Pinar; Arifoglu, Damla; Kalpakli, Mehmet; "Cross-document word matching for segmentation and retrieval of Ottoman divans", Pattern Analysis and Applications, 1-17, 2014, Springer London

Word-segmentation with cross-document word matching

لطف اید و بسن مکر ای باد بو گوندن بویله / ویره سن بر خیر اول سرو خراما تمدن
 لطف اید و بسن مکر ای باد بو گوندن بویله / ویره سن بر خیر اول سرو خراما تمدن

فضای عشق چون کوردم صلاح عقلدن دورم / نبی رسوا گوروب عیب اتمه ای ناصح که معذورم
 فضای عشق چون کوردم صلاح عقلدن دورم / نبی رسوا گوروب عیب اتمه ای ناصح که معذورم

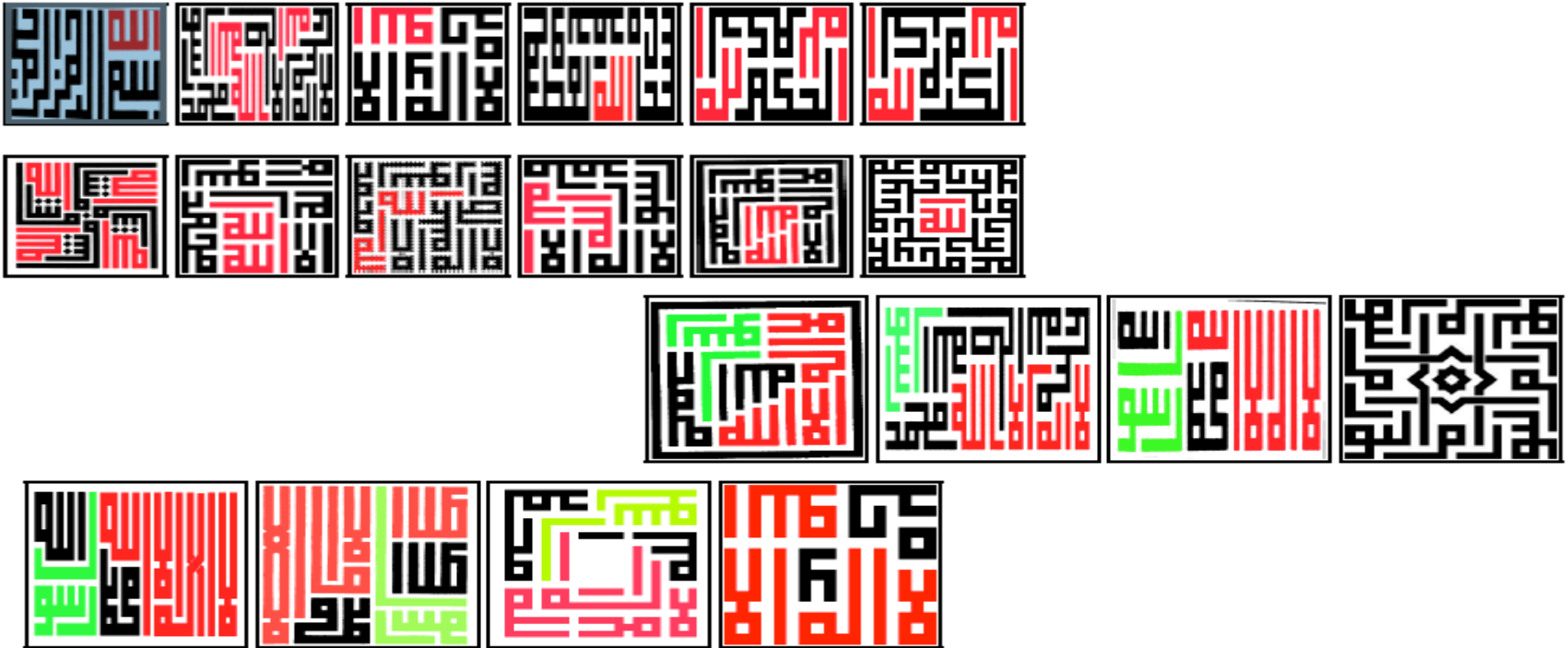
عشق دردی ای معالج قابل درمان دکل / جوهر نندن ایلمک جسمی جدا آسان دکل
 عشق دردی ای معالج قابل درمان دکل / جوهر نندن ایلمک جسمی جدا آسان دکل

شهر دن بحر ایه بر فرق اولدوغن هر کیم بیلور / بیلمش اول کیم عشق صحر اسنده سر گردان دکل
 شهر دن بحر ایه بر فرق اولدوغن هر کیم بیلور / بیلمش اول کیم عشق صحر اسنده سر گردان دکل

	Proposed Approach		Baseline
	Manual queries	Automatic queries	Vertical projection based segmentation
Recall	0.70	0.65	0.43
Precision	0.74	0.68	0.60

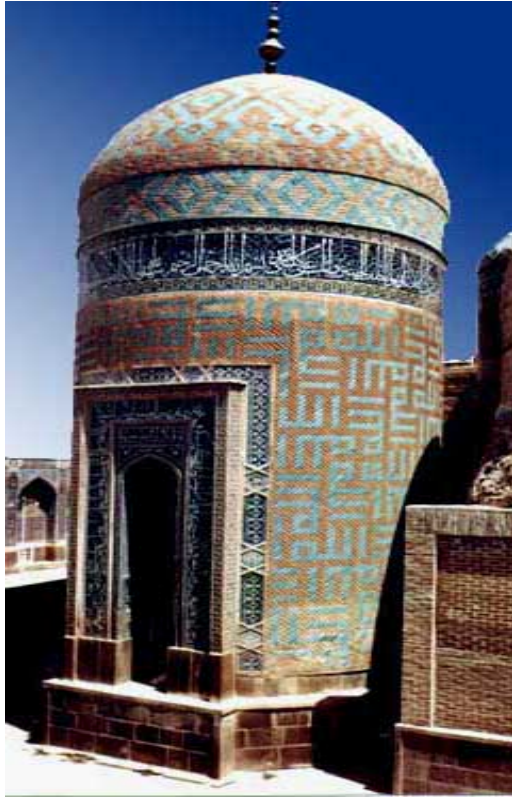
Duygulu, Pinar; Arifoglu, Damla; Kalpakli, Mehmet; "Cross-document word matching for segmentation and retrieval of Ottoman divans", Pattern Analysis and Applications, 1-17, 2014, Springer London

Kufic



D Arifoglu, E Sahin, H Adiguzel, P Duygulu, M Kalpakli, "Matching Islamic Patterns in Kufic Images", Pattern Analysis and Applications, 2015

Kufic



Ottoman keyword search system



Query word
دوست



دوست
ایچون ندر
ندر دوست

Ottoman keyword search system

پنهانند



بو غزل مجنون دلندندر

وفا هر کیمسه دن کیم ایستدم آندن جفا کوردم

کیمی کیم بی وفا دنیا ده کوردم بیوفا کوردم

کیمه کیم دردی اظهاری قیلدم ایستیوب درمان

اوز یعدن میک بتر درد و بلا یه مبتلا کوردم

مکدر خاطرمدن قیلمدی بر کیمسه دفع غم

صفادن دم اورن همدملری اهل ریا کوردم

اگر صودا من طوتدم روان دوندردی یوز مندن

و کرکز کوزدن اومدم صدق عکس مدعا کوردم

ایاق باصدم درامیده سرکر دانلق ال ویردی

أمل سر شته سن طوتدم الیمده اژدها کوردم

مکا کوستردی کردون تیره بختیم کوز کو سن یوز کز

من بدبخت آکا هرگاه کیم باقدم قرا کوردم

نیشوری عیب قیلمه یوز چورسم اهل عالمندن

ندن کیم هر کیمه یوز طوتدم آندن یوز بلا کوردم

Manual labeling tool

Word Labeling

(غزل مقتضای حال)

Word	Label	Verify	Operations
حال	<input type="text" value="h3a1l"/> Help	<input type="text" value="حال"/>	Edit Delete

OTAP transliteration alphabet

Tr	Ottoman	Description	Tr	Ottoman	Description	Tr	Ottoman	Description	Tr	Ottoman	Description
a5		initial a (short)	a2	آ	initial ā (Arabic)	a3	ا	initial a (Turkish)	e2	اي	initial é (double)
e5		initial e (short)	i2	اي	initial i (double)	i4	ا	initial i (short)	i3	اي	initial i (double)
i5		initial i (short)	o2	او	initial o (double)	o4	او	initial o (short)	o3	او	initial ö (double)
o4		initial ö (short)	u2	او	initial u (double)	u4	او	initial u (short)	u3	او	initial ü (double)
u5		initial ü (short)	a1	ا	ā [long]	a7	ا	a [A/O script character]	a8	او	ā [as in خواب]
a0	ه	vocalic he	a6	و	dagger elif	a9	ي	elif maksur	e2	ي	è
e0	ه	vocalic he	i1	ي	i1	i6	ي	i [long]	i6	ي	i [Turkish]
i8	ه	["ki" he]	i9	وي	ī [as in خویش]	i7	ي	i [Turkish]	o1	و	o [long]
o6	و	o [Turkish]	o7	و	ô [Turkish]	u1	و	u [long]	u6	و	u [Turkish]
u8	ي	u [final as yā]	u7	و	ü [Turkish]	u9	ي	ü [final as yā]	b	ب	b
b7	ب	b [stops to "p"]	c	چ	c	c7	چ	c [stops to "ç"]	c2	چ	ç
d	د	d	d7	د	d [stops to "t"]	d6	ط	d [reverse to t]	d3	ظ	d
f	ف	f	g	گ	g	g7	گ	g [stops to "k"]	g6	گ	ğ / {g=}
g4	غ	ğ / {g.}	h	ه	h	h3	ح	h	h5	ح	h
j	ج	j	k	ک	k	k3	ق	k	l	ل	l
m	م	m	n	ن	n	n6	ک	ñ	p	پ	p
r	ر	r	s	س	s	s2	ش	ş	s3	ص	s
s5	ت	s [under-bar]	t	ت	t	t7	ة	[tā marbūta]	t3	ط	t
v	و	v	y	ي	y	z	ز	z	z3	ظ	z
z4	ض	z	z5	ذ	z	'2	ع	['ayin]	'3	ء	? [hamza]
'4	و	hamza with "he/yā"	w2	و	shedda						

Automatic labeling

(غزل مقتضای حال)

فلك آیردی بنی جور ایله جانانمدن * حذر ایتزمی عجب ناله و افغانمدن
اوده یاندرسم اکر شعله ایله نه فلکی * نه بتر آتش آه دل سوزانمدن
غم بنهان بنی اولدردی بوهم برغم کیم * کله خیم اولمدی آگه غم پنهانمدن
آه ایدی هم نفسم آه که اول هم آخر * چیقدی اکر اه قیلوب کلبه احزانمدن

Text Output

<p> gazel mukteza y hal </p> <p> felek ayırd beny cevr ile cananmdn - hazer etmez mi acep nale efganmdn </p> <p> odh yandırsım eđer şule ile ne feleki - ne [biter/beter] ateş ah dil suzanmdn
bnhn beny oldrdy bu hem bir gam kım - gelir khm olmdy agah gam pinhanmdn </p> <p> ah idi hem nefesm ah ki ol hem ahır - çkdy ikrah kıl p kulb y ahzanmdn </p> <p> ben ne hacet ki kıl m dađ y
akıbet zahır olur çak garıbanmdn </p>

Transcription File

grl mze y xel
m flb r d y by x wr e blh xebbrā b x d r e g r mğ xb beblhwe fg d b
e wd bebd e kr s glk bBh bnfl ky g br el s e d l sb wr e brā b
gm e bbb y e w ldd d y bw m r g k h klr x m e r l d ye kə7 g m s h e brā b
e h e bd y bf e k h e w l hm e x r x brā ye e kr flw b k e n a x r e b m d
ko7 sm

Acknowledgements

Contributors

- Emre Şahin
- Esra Ataer Cansızođlu
- Ethem Fatih Can
- Damla Arifođlu
- Hande Adıgüzel
- Ceyhun Karbeyaz
- Bilge Körođlu

Contact info

Pinar Duygulu

Associate Professor

Hacettepe University, Department of Computer Engineering

Ankara, Turkey

pinar.duygulu@gmail.com

Damla Arifoglu

PhD student

Computer Engineering Department

Sabanci University, Istanbul, Turkey

damla.arifoglu@gmail.com

Thank you!

ANY QUESTIONS?