

Büyük Dil Modelleri ile Tarihî Metinlerde Adlandırılmış Varlık Tanıma: Evliya Çelebi Seyahatnâmesi'nden Botanik Bilgi Çıkarımı

Identifying Named Entities in Historical Texts Using Major Language Models: Extracting Botanical Information from Evliya Çelebi's Seyahatname (Travelogue)

Doç. Dr. Gönül ERDEM NAS¹

Süreç/History

Geliş/Submitted: 16/12/2025
Kabul/Accepted: 07/03/2026
Yayın/Published: 25/03/2026

İntihal/Plagiarism

Bu makale, en az iki hakem tarafından incelendi ve intihal içermediği teyit edildi. / This article has been reviewed by at least two referees and scanned via a plagiarism software.

Etik Beyan/Ethical Statement

Bu çalışmanın hazırlanma sürecinde bilimsel ve etik ilkelere uyulduğu ve yararlanılan tüm çalışmaların kaynakçada belirtildiği beyan olunur. / It is declared that scientific and ethical principles have been followed while carrying out and writing this study and that all the sources used have been properly cited.

Çıkar Çatışması / Competing Interests

Yazar(lar), çıkar çatışması olmadığını beyan eder. / The author(s) declares that have no competing interests.

Telif Hakkı/Copyright

Submitted for possible open access publication under the terms and conditions of the [Creative Commons Attribution \(CC BY-NC 4.0\) license](https://creativecommons.org/licenses/by-nc/4.0/).



Öz

Bu çalışma, 17. yüzyıl Osmanlı seyahat literatürünün en önemli seyyahı olarak kabul edilen Evliya Çelebi Seyahatnamesi içerisindeki botanik verilerin, Büyük Dil Modelleri (BDM) destekli otomatik sistemler aracılığıyla çıkarılmasını ve sınıflandırılmasını konu edinmektedir. Araştırmanın temel motivasyonu, dilsel katmanları ve hacmi itibarıyla manuel analizin pratik açıdan sürdürülebilir olmadığı bu tür devasa tarihî metinlerde, Adlandırılmış Varlık Tanıma (AVT) süreçlerini optimize etmektir. Bu kapsamda, Claude Sonnet 4.5 mimarisi kullanılarak geliştirilen algoritmik yapı ile eserin on cildi taranmış; bitki taksonomisine ait veriler sistematik bir biçimde etiketlenmiştir.

Metodolojik süreçte, Osmanlı Türkçesi'nin çok katmanlı yapısı göz önünde bulundurularak, özellikle Arapça ve Farsça kökenli arkaik botanik terimlerin (örn. *engür*, *sanavber*, *gendüm*) modern terminoloji ile semantik eşleştirilmesi üzerinde durulmuştur. Elde edilen bulgular, yapay zekâ tabanlı modellerin tarihî derlemelerin işlenmesinde yüksek başarı sağladığını göstermekle birlikte; metaforik kullanımlar ve düşük frekanslı terimlerin tespitinde modelin bağlamsal çıkarsama yeteneğinin belirli sınırlar dâhilinde kaldığını ortaya koymuştur. Çalışma sonucunda oluşturulan bu geniş ölçekli veri seti, Osmanlı dönemi biyo-çeşitliliği ve bitki coğrafyası üzerine yapılacak araştırmalar için bir zemin sağlamaktadır. Ayrıca bu girişim, manuel iş yükünü minimize ederek Dijital Beşeri Bilimler alanında tarihî veri madenciliği yöntemlerinin geçerliliğini ve ölçeklenebilirliğini somut bir örneklem üzerinden kanıtlamakta, dilsel değişim süreçlerinin izlenmesine de katkı sunmaktadır.

Anahtar Kelimeler: Adlandırılmış Varlık Tanıma, Büyük Dil Modelleri, Dijital Beşeri Bilimler, Evliya Çelebi, Seyahatnâme, Tarihî Metin Analizi

Abstract

This study focuses on the extraction and classification of botanical data from the *Seyahatname* (Book of Travels) by Evliya Çelebi—a seminal work of 17th-century Ottoman travel literature—utilizing automated systems powered by Large Language Models (LLMs). The primary motivation is to optimize Named Entity Recognition (NER) processes for voluminous historical texts where manual analysis proves practically unsustainable due to the sheer size and linguistic complexity of the corpus. To this end, an algorithmic framework built upon the Claude Sonnet 4.5 architecture was employed to scan all ten volumes of the work, systematically tagging data related to plant taxonomy.

Given the multi-layered nature of Ottoman Turkish, the methodological approach prioritized the semantic mapping of archaic botanical terms—particularly those of Arabic and Persian origin (e.g., *engür*, *sanavber*, *gendüm*)—to their modern equivalents. While the findings demonstrate the high efficacy of AI-driven models in processing historical corpora, they also reveal certain limitations regarding the model's contextual inference capabilities, particularly when dealing with metaphorical usage and low-frequency terms. The resulting large-scale data set provides an empirical foundation for future research into Ottoman-era biodiversity and phytogeography. Furthermore, by minimizing manual workload, this initiative validates the scalability of historical data mining methods within the Digital Humanities and contributes to the broader understanding of linguistic evolution.

Keywords: Named Entity Recognition, Large Language Models, Digital Humanity Sciences, Evliya Çelebi, Seyahatnâme, Historical Text Analysis

Atf için (tocite): Erdem Nas, Gönül. (2026). Büyük Dil Modelleri ile Tarihî Metinlerde Adlandırılmış Varlık Tanıma: Evliya Çelebi Seyahatnâmesi'nden Botanik Bilgi Çıkarımı. *Afyon Kocatepe Üniversitesi Sosyal Bilimler Dergisi*, 28(1), 259-267. <https://doi.org/10.32709/akusosbil.1843063>

¹Afyon Kocatepe Üniversitesi, Fen-Edebiyat Fakültesi, Afyonkarahisar/Türkiye. gonulerdem2002@yahoo.com.

Giriş

Evliya Çelebi'nin 17. yüzyılda kaleme aldığı Seyahatname, Osmanlı İmparatorluğunun coğrafyası, kültürü ve doğal çevresi hakkında en kapsamlı kaynaklardan biridir. O dönem Osmanlı Devleti'nin sınırları içerisindeki birçok yer için birçok konuda ayrıntılı bilgiler sunan eser, günümüzde ulusal ve uluslararası birçok disiplin çalışması için bilimsel eserlerine temel kaynak niteliğindedir. Eserde yer alan botanik unsurlar, yalnızca doğal çevreyi değil; tarım, ticaret ve gündelik yaşam pratiklerini de yansıtmaktadır. On cilt ve günümüz Türkçesiyle yazılmış biçimi kapa, giriş ve indeksler dışında 3629 sayfadan oluşan bu eser, Orta Avrupa'dan Afrika'ya uzanan geniş bir coğrafyayı kapsamaktadır.

Eserde geçen bitki adlarının sistematik olarak çıkarılması, tarih, çevre tarihi ve botanik coğrafya açısından önemli bir potansiyel taşımaktadır. Ancak mevcut doğal dil işleme araçları, Osmanlı Türkçesi gibi tarihî dillerde sınırlı başarı göstermektedir. Bu durum, çalışmanın temel motivasyonunu oluşturmaktadır.

Örneğin "üzüm" kelimesi metinde "engür", "engûr" veya "ineb" biçimlerinde geçebilmekte; bu çeşitlilik, kural tabanlı veya istatistiksel yöntemlerle ele alınması güç bir yapı ortaya koymaktadır.

Son yıllarda geliştirilen büyük dil modelleri, çok dilli ve bağlamsal öğrenme kapasiteleri sayesinde bu tür sorunlara alternatif bir yaklaşım sunmaktadır. Bu çalışma, söz konusu potansiyelin Seyahatname özelinde ne ölçüde gerçekleştirilebileceğini sorgulamaktadır.

Bu çalışmanın amacı, Claude Uygulama Programlama Arabirimi'ni (Application Programming Interface API) kullanarak Seyahatname'den bitki isimlerini otomatik olarak etiketleyen, bir sistem geliştirmektir. Çalışma kapsamında şu araştırma soruları ele alınmaktadır:

(1) Büyük dil modelleri, 17. yüzyıl Osmanlı Türkçesi metinlerinden botanik varlıkları ne düzeyde doğrulukla tanıyabilir?

(2) Osmanlıca-modern Türkçe terim eşleştirmesi dil modeline eğitim örneği vermeden sadece görev tanımları yapılarak başarılabılır mı?

(3) Büyük dil modelleri ile büyük ölçekli tarihî metin analizi etkin bir şekilde gerçekleştirilebilir mi?

1. Literatür Taraması

1.1. Tarihî Metinlerde Adlandırılmış Varlık Tanıma

Tarihî metinlerde AVT, dijital beşerî bilimlerin alanının önemli bir araştırma konusunu oluşturmaktadır. Piotrowski (2012, s. 124), tarihî metinlerin doğal dil işleme açısından karşılaştığı zorlukları sistematik olarak ele almış ve yazım standardizasyonu eksikliği, morfolojik karmaşıklık ve bağlamsal belirsizlik gibi temel problemleri tanımlamıştır. Ehrmann ve arkadaşları (2016), tarihî gazete arşivlerinde AVT uygulamalarını incelemiş ve alan-spesifik eğitim verilerinin önemini vurgulamıştır.

Osmanlı Türkçesi metinleri üzerine yapılan çalışmalar görece sınırlıdır. Türkçe AVT sistemleri genellikle modern Türkçe için optimize edilmiş olup (Şeker ve Eryiğit, 2012, s. 2462; Yeniterzi, 2011, s. 109) Arapça-Farsça kökenli Osmanlıca terimlerin tanınmasında yetersiz kalmaktadır. Can ve arkadaşları (2008, s. 411), Osmanlıca belgelerin optik karakter tanıma (OKT) ve metin madenciliği için hazırlanması konusunda önemli bir çerçeve sunmuştur; ancak bu çalışma AVT'ye doğrudan odaklanmamıştır.

1.2. Büyük Dil Modelleri ve Sıfır-Atış Öğrenme

GPT-3'ün (Brown ve ark., 2020, s. 1879) tanıtılmasından bu yana, büyük dil modelleri doğal dil işleme görevlerinde devrim niteliğinde gelişmeler sağlamıştır. Bu modellerin en dikkat çekici özelliklerinden biri, açık eğitim verisi olmadan yalnızca doğal dil talimatlarıyla (prompt) görevleri gerçekleştirebilme kapasitesidir. Wei ve arkadaşları (2022, s. 2210), bu "sıfır-atış" ve "az-atış" öğrenme yeteneklerinin model ölçeğiyle birlikte arttığını göstermiştir.

AVT görevlerinde BDM'lerin kullanımı hızla yaygınlaşmaktadır. Wang ve ark. (2023, s. 2309), GPT-4'ün geleneksel AVT sistemleriyle karşılaştırılabilir performans gösterdiğini; Li ve ark. (2023, s. 57) ise BDM'lerin özellikle düşük kaynaklı dillerde transfer öğrenimi avantajı sağladığını ortaya koymuştur. Ancak bu çalışmaların büyük çoğunluğu modern metinlere odaklanmış tarihî metinlerde BDM performansı yeterince araştırılmamıştır.

1.3. Evliya Çelebi Seyahatnâmesi Üzerine Çalışmalar

Evliya Çelebi Seyahatnamesi, edebiyat, tarih ve coğrafya alanlarında geniş bir araştırma külliyyatının konusu olmuştur. Dankoff (2004, 2006), yaptığı İngilizce çeviriler ve akademik analizlerle Seyahatnâme'nin temel referans kaynaklarını oluşturmuştur. Yapı Kredi Yayınları tarafından gerçekleştirilen transkripsiyon projesi (1996-2007), on cildin tamamının Latin harflerine aktarılmasını sağlamış ve dijital çalışmalar için zemin hazırlamıştır.

Lisansüstü tez çalışmalarının birçok bilim dalında belki de en çok çalışılan konularından biri de Evliya Çelebi'nin Seyahatname eseridir. Yerli ve yabancı birçok araştırmacının ilgisini çeken Seyahatname üzerine YÖK Tez'de yüzün üzerinde tez girdisi bulunmaktadır. Doğrudan tematik çalışmalar içinde; meslek adları (Kırmit, 2025), kavramsal metaforlar (Sağ Yener, 2024), mimari ve mimari süsleme terimleri (Küçük Ak, 2023), organ adları (Bakır, 2022), bitki adları (Karacaoğlan, 2018), yiyecek adları (Kartal, 2016), hayvanlarla ilgili söz varlığı (Peker, 2015), Anadolu'da yer adları (Öz, 2007) tezleri sayılabilir.

Bu tezler arasında Karacaoğlan'ın tezi çalışmanın konusuyla doğrudan ilgilidir. Karacaoğlan, tez çalışmasında Seyahatnâme'yi etnobotanik açıdan incelemiş ve 309 bitki adı tespit etmiştir. Araştırmacı, tekrar eden bitki isimlerini bir kez saymak suretiyle bu sayıya ulaşmış ve tüm bitki adlarının geçtiği yerleri bağlamlarıyla birlikte bir indekste toplamıştır. Elde edilen veriler, bitki adlarını kullanım amaçlarına göre sınıflandırmaya imkân tanımıştır. Buna göre Seyahatnâme'de geçen bitkiler, beslenme ve şifa amacıyla kullanılanlar, kültürel malzeme veya nesne olarak değeri olanlar, doğal çevre ve mekân unsuru olanlar, gündelik hayat malzemesi veya simge işlevi görenler ve zirai/ekonomik amaçla kullanılanlar şeklinde beş ana kategori altında incelenmiştir. Bu beş ana başlık, toplamda 13 alt kategoriye ayrılarak her bir bitkinin işlevi ve kullanım bağlamı detaylandırılmıştır. Örneğin Evliyâ Çelebi, bazı bitkileri yiyecek, ilaç ya da baharat olarak ele alırken, bazı ağaçları *doğal anıt* veya *yer işareti* olarak kaydetmiştir. Yapılan karşılaştırmalar, Seyahatnâme'de anılan bitkilerin bir kısmının günümüzde de benzer amaçlarla kullanıldığını, bazılarının ise unutulmaya yüz tuttuğunu göstermiştir. Sonuç olarak, bu çalışmalar Evliyâ Çelebi'nin eserinin dönemin geleneksel bitki bilgisi (etnobotanik bilgi) açısından önemli bir derleme olduğunu vurgular (Karacaoğlan, 2018).

Bununla birlikte, Seyahatnâme'nin dijital beşerî bilimler perspektifinden sistematik analizi sınırlı kalmıştır. Metnin büyüklüğü ve dilsel karmaşıklığı, elle etiketleme yaklaşımlarını pratik olmaktan çıkarmaktadır. Bu çalışma, BDM teknolojisini kullanarak bu açığı kapatmayı ve eserin yapılandırılmış veri setlerine dönüştürülmesi için ölçeklenebilir bir yöntem sunmayı hedeflemektedir.

2. Yöntem

2.1. Veri Hazırlığı ve Dosya Biçimi Seçimi

Çalışmada kullanılan kaynak metin "DOCX" (Ms Word) dosya biçiminde elde edilmiş ancak metnin orijinal baskı düzenini koruması nedeniyle "PDF" biçimine dönüştürülmüştür. Böylelikle Büyük Dil Modelinin çıktılarının orijinal metindeki sayfa numaralarından elde edilmesi amaçlanmıştır. Oluşturulan kural tabanlı aramaya (regex) dayalı kontrol yazılımı ile metinde geçen bitki isimlerinin sayfa numaraları ve sayıları tespit edilmiş ve işlenmeye hazır hale getirilmiştir.

2.2. Model ve Yönerge Stratejisi

Büyük Dil Modelleri, kendilerine verilen rol ve yönergeler doğrultusunda cevap verirler. Bu sebeple yönerge (prompt) tasarlanırken öncelikle modele rolü "Sen 17. yy Osmanlı Türkçesi ve tarihi coğrafya uzmanısın. Görevin Evliya Çelebi Seyahatnamesi'nden bitki isimlerini çıkarmak" şeklinde tanımlandı. Yönerge yazılırken rol ile başlamak modelin verdiği görevi o role bürünerek yerine getirmesini sağlar. Örneğin bir projeyi pazarlama müdürü ile finans müdürü tamamen farklı bakış açılarıyla size sunacaktır. Bu durumda hiçbir rol tanımlanamamak modelin size genel-geçer cevaplar hazırlamasına neden olacaktır. Daha sonra bitki kategorisine girecek örnekleri ayrıntılı bir şekilde listeleyp Osmanlı Türkçesinde –varsa- farklı şekildeki kullanımları günümüz Türkçesindeki karşılığını da yanında belirtilerek listelenmesi istenildi. Böylelikle bir bitki bulunmak istendiğinde farklı yerlerde veya isimlerde kullanımlarının da elde edilme imkânına sahip olundu. Yönerge (promptu) geliştirilirken ilk cilt üzerinde çok sayıda deneme yapılarak bulunan sonuçlar titizlikle değerlendirildi ve yönerge tekrar tekrar düzenlenerek istenilen ölçüte gelindiğinden emin olundu. Ardından kullanılan

modelin (Claude 4.5 Sonnet) bağlam işleme kapasitesini göz önünde bulundurularak onar sayfalık parçalar şeklinde işlemenin hız ve güvenilirlik açısından iyi bir nokta olduğunu belirlendi ve **3629** sayfalık metin işletilip sonuçlar rapor haline dönüştürüldü.

2.3. Doğrulama ve Hata Analizi

Elde edilen kategorik bitki isimleri listesi öncelikle BDM'ye sunularak içerisinde bitki ismi olmayan sonuçları işaretlemesi istendi. Böylelikle "yiyecek" bağlamından bitki ile yakın görülen 'peynir, su, tuz' gibi yine bitkilerin işlenmesiyle elde edilen eşyalar gibi sonuçları listeden çıkararak tespit edilmesi sağlandı. Daha sonra elde edilen son liste uzman bakışıyla titizlikle kontrol edilip son kontrol aşamasına geçildi. Son kontrol aşamasında kural tabanlı aramaya (regex) dayalı algoritmayla ilgili bitki isimlerinin metinde tam olarak nerede kaç kere geçmiş olduğu tespit edildi. Bu yöntem sayesinde metinde geçmediği halde modelin uydurduğu (halüsinasyon) kelimeler işaretlemeyen elendi ve modelin bulunduğu sonuçların doğruluğu ortaya çıkarıldı. Son olarak tespit edilen bitki adları ses farklılıkları uzman kontrolüyle bir kez daha kontrol edilerek ses farklılığı içerenler güncel Türkçe adlandırmaları başlığında birleştirildi (ermud, emrud>armut; narenci,narenc>narenciye gibi) böylece ağız farklılıklarından kaynaklı ortak adlar teke düşürülerek son biçim tematik bir biçimde düzenlendi.

Tablo 1. Analiz sonuçları

Metrik	Değer
İşlenen sayfa sayısı	3629
Toplam karakter	15.6 milyon
Bilinen bitki çeşitliliği	802
Metinde bulunan	679
Metinde bulunamayan	123

Tablo 2. Eserin cilt düzeyinde analiz sonuçları

Cilt	İşlenen sayfa sayısı	Bulunan bitki adı sayısı (BDM tarafından)
I.Cilt	476	209
II.Cilt	240	146
III.Cilt	288	131
IV.Cilt	348	141
V.Cilt	316	120
VI.Cilt	321	166
VII.Cilt	336	142
VIII.Cilt	398	121
IX. Cilt	486	124
X. Cilt	520	246

Tablo 2'deki veriler BDM tarafından ilk analize dayalı sonuçları içermektedir. Bu tarama il olarak BDM'ye tanımlanarak yöntem bölümünde belirtilen yönergelerle geliştirildi. İlk olarak tüm ciltler tüm sayfalarıyla tarandığından buna indeks kısımları da dâhil edilmiştir. Bu sebeple sayıtlı fazla çıkmış bu da çalışmada sistemin, tekrarlı ya da farklı harflerle aynı anlamları taşıyan bitki adlarını tek tek farklı kodladığını göstermiştir. İkinci kontrolde eserlerdeki tüm indeksler çıkarılmış sayfa sayfa kontrol edip bağlama dayalı kendisinin bulması istenmiştir. Bu aşamadan sonra sayıtlının gerçekliğini kontrol amacıyla kural tabanlı (regex) aramaya dayalı bir python yazılımıyla aynı biçimde bulması istenmiştir

ardından her iki sistemin sonucu karşılaştırmalı olarak değerlendirilmiştir. Kural tabanlı arama (regex) metin içerisinde aranan bitki isimlerini eksiksiz bir biçimde yakalamaya / saymaya olanak tanır. Ancak bu avantajının yanında metin içindeki “Gül Teyze” ile bitki ismini ayırma yeteneğine sahip değildir. Ancak LLM doğası gereği bağlam içinde bu kelimelerin çiçek adı olarak mı yoksa başka bir anlamda kullanılıp kullanılmadığını değerlendirip listeye ekleme kararı verebilir. Bu yeteneklerinin yanı sıra aynı yönergeyle bile çalıştırılsa modelin vereceği yanıtlar her zaman birebir olmaz. Bazen bir sorguda tespit ettiği liste ile diğer bir sorguda belirledikleri farklı olabilir. LLM’ in doğası gereği olan bu zayıf tarafı çalışmanın farklı dil modelleriyle / farklı yönergelerle tekrar tekrar yapılarak bir modelin bulunduğu sonuçları diğer bir modele değerlendirilebilir. Bununla birlikte çalışmada yapıldığı üzere klasik doğal dil işleme yöntemleriyle yazılan yazılımlarla karşılaştırılarak saflaştırılabilir. Sonuçların güvenilirliği artırılabilir. Bu karar metnin büyüklüğüne (çok büyük metinlerde bir yerde kaçırılan bir bitki isminin diğer bir yerde tespit edilebilmesi), çalışmanın hedef kitlesine veya araştırmacının teknik yeterliliklerine, bütçesine göre değişiklik gösterebilir.

Tüm eser Claude Sonnet 4.5 büyük dil modelinin ve python ile geliştirilmiş kural tabanlı (regex) aramaya dayalı klasik doğal dil işleme (DDİ) programlarında birbirlerinin sağlaması biçiminde tekrarlı biçimde taranmış ve tespit edilen tüm bitki adları öncelikle rafine halde büyük dil modelinin kontrolünden geçirilmiş, daha sonra uzman kontrolünden manuel olarak geçirilmiştir. Bu tabloda da belirtilen bulunamayan kısımlar özel karakterli (tırnak işareti gibi) veya birleşik sözcüklerden oluşan adları kapsamış (za’ferân-safran, na’nâ-nane; adaçayı, çörek otu gibi) bunlar da hata kapsamına sistem anlayışıyla dâhil edildiği için sistem tarafından otomatik elenmiştir. Bir başka deyişle bu tip özel karakterler / yazım farklılıklarını klasik kural tabanlı (regex) sorgularla yakalamak çok zor olabilir bu sebeple elenmiş olsa da bu sistemin kısıtı sayılır.

3.2. Bitki Adları ve Tematik Sınıflandırma

Sistemleri sağlamalı ve kontrollü çoklu teyit sonrası toplamda 278 bitki adından en sık geçen 10 bitki aşağıda gösterilmektedir.

Tablo 3. En sık geçen bitkiler (İlk 20)

Sıra	Bitki adı	LLM (BDM) sayısı	Regex sayısı	Osmanlıca varyasyonlar
1	Gül	233	500	dehen-i gül': 1, 'gülâbdân': 24, 'gülâb': 56, 'gül': 388
2	Üzüm	176	197	engür': 20, 'çakır': 28, 'engür': 9, 'misket': 19, 'üzüm':...
3	Buğday	150	1061	'hınta': 21, 'ekmek': 195, 'buğday': 244, 'gendüm'
4	Servi	122	197	'serv': 45, 'servi': 118, 'servden': 4, 'servi': 30
5	İncir	116	84	cümmeyz': 26, 'incir': 56, 'encâlu': 1, 'encâs': 1
6	Zeytin	115	126	'zeytûn': 41, 'zeytun': 6, 'zeytin': 47, 'zeyt': 32
7	Limon	101	128	'limonu': 37, 'limon': 91
8	Çam	98	189	'çenâr': 1, 'çinâr': 33, 'çınar': 86,
9	Arpa	94	135	'arpa': 135

Sıra	Bitki adı	LLM (BDM) sayısı	Regex sayısı	Osmanlıca varyasyonlar
10	Nar	93	243	rummân': 5, 'nar': 44, 'nâr': 187,
11	Çınar	89	146	'çenâr': 1, 'çınâr': 33, 'çınar': 86...
12	Hurma	88	275	'hurmâ': 17, 'nâbıka': 37, 'hûrmâ': 119, 'hurma': 101,
13	Darı	72	157	'darı': 136, 'erzen': 21
14	Turunç	70	80	'turunç': 45, 'nârenc': 10, 'turunc': 25
15	Pirinç	69	229	'pirinci': 35, 'pirinc': 86, 'pirinç': 106, 'çeltik': 2
16	Kiraz	64	80	'kiras': 5, 'kiraz': 51, 'kiraza': 1...
17	Sümbül	63	85	'sümbül': 7, 'sünbül': 78
18	Elma	59	105	'sîb': 1, 'tüffâh': 12, 'elma': 68,...
19	Ceviz	57	82	'cevîz': 1, 'cevzî': 3, 'ceviz': 49, 'cevz': 27, 'cev': 2
20	Dut	51	150	'dûd': 73, 'dut': 31, 'dud': 46

Dikkat çekici bir bulgu, sistemin Osmanlıca-modern Türkçe eşleştirmelerini yüksek doğrulukla gerçekleştirmiş olmasıdır. Örneğin "engür" kelimesi birçok farklı bağlamda doğru şekilde "üzüm" ile eşleştirilmiştir. Benzer şekilde, Arapça kökenli "verd" (gül), Farsça "sanavber" (çam) ve "gendüm" (buğday) gibi terimler başarıyla tanınmıştır.

Veriler, tematik olarak sınıflandırmaya tabii tutulduğunda sistem aşağıdaki verileri sunmuştur (Bu veriler sistemin sunduğu ilk ve ham veriler değildir yine burada da uzman kontrolü sonrası talimatlar girilerek düzeltilmeler sisteme yaptırılmıştır. Bununla birlikte bu veriler her kategoride sıklık sırasına göre 278 bitki listelenmiştir):

3.2.1. Ağaçlar (42 bitki)

servi (*serv, servî, servden*), **çınar** (*çınâr, cümmeysz, çenâr*), **çam** (*santa, sanavber*), **dut** (*dud, dūd, tūt*), **kavak**, **ağaç** (*ūd, dirahî, şecere, ud*), **söğüt** (*bîd, söğüd*), **ardıç**, **kestane** (*kestâne*), **meşe** (*sindiyân, mefle, mişe*), **salkımsöğüt** (*söğüt*), **çimşir** (*çimşir*), **ihlamur** (*ohlamur, uhlamur*), **âbinûş** (*âbnûs*), **sanavber** (*çam*), **sindiyân**, **çenâr** (*çınar*), **santa**, **defne** (*tefne*), **palasanta**, **karaağaç** (*dirahî-karaağaç*), **köknar** (*koknar, koknâr, göyner*), **şimşâd**, **abanoz** (*âbinûş, abanos*), **bîd**, **gürgen**, **şimşir**, **pelîd**, **Hindistancevizi** (*cevz-ihindî, hindistâncevizi*), **dişbudak**, **dutağacı** (*şecere-itût*), **ılgın**, **muşmula**, **pirnâr**, **sakızağacı** (*sakız*), **sandaloz**, **sığla**, **üvez**, **kara çam**, **alıç**, **kızılağaç**, **mazi**

3.2.2. Tahıllar (13 bitki)

buğday (*hâs, ekmek, gendüm, hunta*), **arpa**, **darı** (*erzen*), **pirinç** (*pirinc, pirinci, çeltik*), **gendüm** (*buğday*), **yulaf**, **mâş**, **şa'ir** (*arpa*), **çavdar**, **çeltik** (*pirinç*), **hunta** (*buğday*), **sarıpirinç** (*sarıpirinc*), **erzen**

3.2.3. Baklagiller (7 Bitki)

mercimek (*ades*), bakla (*fûl, fûlü*), nohut (*nohud, leblebi, hummûs*), fûl, b r lce (*b gr lce*), f l, turmus

3.2.4. Meyveler (35 bitki)

 z m (* akır, eng r, misket, eng r*), hurma (*h rm , n bika, hurm , dom*), incir (*c mmeyz, inc r, enc lu, enc s*), zeytin (*zeyt n, zeyt, zeytun*), limon (*limonu*), nar (*n r, rumm n, g ln r, r mm n*), turun  (*turunc, n renc*), ceviz (*ceviz, cevz , cev, cev z*), karpuz (*hind v nesi*), Őeftali (*Őeft l , Őeftal , Őeft l , papa Őeft l su*), badem (*b dem, b d m*), armut (*emr d, armud, ahl d*), portakal (*turun , turuncu, turunc, n renc*), viŐne (*viŐn b*), fistic (*fisdık, botom, Ő mfisdıŐı*), ayva, c mmeyz (*c mmeyzi*), n bika, kızılıcık (*kızılıcıŐ*), eng r (* z m*), h nnap (*unn b, unn b *), narenciye (*n renc, narenci*),  ilek, mıŐmıŐ, muz,  l , g ln r (*g ln r*), koruk, kiŐmiŐ, zerdali, nahl-ihurma (*hurma*), b g rdlen, g gem, kaysı, yabanasması

3.2.5. Sebzeler (22 Bitki)

soŐan, kabak, lahana (*kelem, kapusta*), salatalık (*hıy r, hıyar*), havu , turp, pırasa (*purasa*), patlıcan, hıyar (*hıy r*), Őalgam, ispanak, bamya, kerefiŐ, tere, kereviz, karnabahar (*karneb t*), karnebit (*karneb t*), labada, hıy Őenbe, ispenet, pazı, acur

3.2.6.  i ekler (21 Bitki)

g l (*verd, g l b, g l bd n, dehen-ig l*), s mb l (*s nb l*), menekŐe (*benefŐe*), karanfil (*karanf l*), erguvan (*erguv n, erŐuvan*), nergis (*nerkis*), yasemin (*y semen, semen*), zambak (*zanbak*), Őakayık (*Őak yık*), zerrin (*zerr n*), nesrin, g l b, gelincik, nil fer, papatya, g lhatmi,  iŐdem, gon a, yerg l , yabang l ,   rg l

3.2.7. BaharatlarveTibbiBitkiler (31 Bitki)

tar ın (*dar ın*), safran, zencefil (*zenceb l*), nane (*n fe*), kına (*hunn *), meyank k  (*s s, meyank k ,  rk-ı s s, miyan*), sumak (*somm k *), karabiber (*f lf l, biber, d r-ı f lf l*), kakule, z ter, hardal, mahleb, haŐhaŐ, kimyon, afyon, demirhindi (*temirhind *), b dy n, biber, ravend (*r vend*), r vend-iR m  (*r vendi*), anison, rezene (*r zy ne*), haŐhaŐ,  erek, keb be, besb se,  p cini, cedv r, cedv r-iR m , havlic n-iR m , s nb l-iR m 

3.2.8. OtlarveYeŐillikler (29 Bitki)

reyhan (*reyh n*), yonca (*bers m*), t rfil (*turf l, terfil*), sarımsak, ayırık,  emen,  alı, s sen, eŐir,  im, g s l, ıŐkın, Ő htere, m rver, bers m, alef, evŐan, evŐesi, roka, kengeri, mantar, baldıran, berberis, geven, halfa, mugayl n, funda, laden, deveboynu

3.2.9. Tarım Ve Sanayi Bitkileri (13 Bitki)

kamiŐ, pamuk, keten (*kett n, ket n*), susam (*simsim, s s m*), ŐekerkamiŐı (*neb t-ı hamavi*), asma,  ay, kenevir, t t n, kendir,  ivit, kovasazı, bezir

3.2.10. DiŐer / Belirsiz (65 Bitki)

kebb t, avŐıla,  alaŐ,  emŐir, heldine, kurkas, n y, reybas (*reyb s*), rummani, tula (*tulŐa*), abd ll v , alakerde, b na, bedevine, belis n, belsem, c notu,  ıka,  iŐe, darfı, eŐ ke, fı ıda, germ germ, g l l, g neyik, Hırvat, h m l,  rkı'l-cen h, ıslama, ineb 'd-dibs, ıpsime, k f r , kakı, kameri, karafirik, kavza, kehle, kıjı, k lt r, kiss , kızılca, k fteri, lazudi, levad n, l zi'n-neb , mavuza, mele e, mersim, nardengi, nilva, pana, pırankop, r Ő, saparna, sarfirik, saya, ser med, s dl ce, s ngerı, t v s , tirkiz, yabr hu's-sanem, yebr hu's-sanem, yernebas, zerdeva

4. TartıŐma

4.1. Sistem Performansı

Elde edilen bulgular, büyük dil modellerinin tarihî Osmanlı Türkçesi metinlerinde eğitimsiz etiketleme için kullanılabilirliğini göstermektedir. Ancak bu başarının, her sayfa ve her bağlam için eşit düzeyde olmadığı da belirtilmelidir.

Model, Arapça ve Farsça kökenli botanik terimleri çoğu bağlamda doğru şekilde tanımıştır. Buna karşın, "gül yüzlü" gibi edebî ifadelerde yanlış pozitif etiketlemeler gözlemlenmiştir.

Varyasyon birleştirme: Aynı varlığın farklı yazımlarını (engür/engûr, çınar/çınâr) tek bir normalized forma eşleştirme son uzman kontrolü tespitiyle girilen yönerge (prompt) ile başarıyla gerçekleştirilmiştir.

Belirsizlik çözümü: "Gül" kelimesinin kişi adı mı yoksa çiçek mi olduğunun bağlamdan ayırt edilmesi, yönergedeki açık talimatlar sayesinde büyük ölçüde başarılmıştır.

4.2. Sınırlılıklar

Bu çalışmanın en önemli sınırlılığı, sistematik bir nicel doğrulama çalışmasının henüz gerçekleştirilmemiş olmasıdır. Manuel etiketleme sürecinin zaman alıcı olması, bu aşamanın çalışmanın kapsamı dışında bırakılmasına neden olmuştur.

4.3. Karşılaştırmalı Değerlendirme

Bu çalışma, bildiğimiz kadarıyla Osmanlı Türkçesi tarihî metinlerde BDM tabanlı AVT'nin ilk kapsamlı uygulamasıdır. Geleneksel AVT sistemleriyle doğrudan karşılaştırma yapılamamakla birlikte, Türkçe AVT literatüründeki en iyi sonuçlar modern metinlerde %90 civarında F1 skoru bildirmektedir (Şeker ve Eryiğit, 2012). Tarihî metinlerdeki zorluklar göz önüne alındığında, sistemin nitel değerlendirmesi umut verici görünmektedir.

5. Sonuç ve Öneriler

Bu çalışmada, Evliya Çelebi Seyahatnâmesi'nden bitki adlarını çıkarmaya yönelik BDM tabanlı bir AVT sistemi geliştirilmiş 10 cilt üzerinde test edilmiştir. Sistem, Osmanlıca-modern Türkçe terim eşleştirmelerinde yüksek oranda tutarlılık göstermiştir.

Sistem, 3630 sayfadan tekrarsız 278 bitki türü ismi çıkarmış, Osmanlıca-modern Türkçe terim eşleştirmelerini yüksek doğrulukla gerçekleştirmiştir. Bu sayı, Karacaoğlan (2018)'in manuel etiketlemeyle tespit ettiği ve tez çalışmasının temelini oluşturan 309 bitki adı sayısına yakındır.

Tespit edilen 278 bitki adı 10 farklı tema altında eserde geçtikleri sıklık sayısına göre sınıflandırılmıştır. Çalışma, tarihî metinlerin otomatik analizi konusunda BDM'lerin sunduğu imkânları göstermesi bakımından bir başlangıç niteliği taşımaktadır. Gelecek çalışmalarda, doğrulama analizleri ve yeni varlık türleriyle sistemin kapsamı genişletilecektir.

Çalışmanın temel katkıları şu şekilde özetlenebilir:

(1) Metodolojik katkı: Tarihî Osmanlı metinleri için ölçeklenebilir ve etkin bir AVT yaklaşımı önerilmiştir.

(2) Veri seti katkısı: Seyahatnâme'nin birinci cildi için yapılandırılmış botanik veri seti oluşturulmuştur. Bu veri seti, tarihî botanik ve dijital beşerî bilimler araştırmacıları için değerli bir kaynak oluşturmaktadır.

Gelecek çalışmalar için öneriler:

Uzman değerlendiricilerle manuel etiketleme gerçekleştirilerek performans hesaplanmalı böylece doğrulama çalışmaları desteklenmelidir. Çalışmanın kapsamı başka başlıklarla da genişletilebilir. Kişi adları, mimari yapılar, yiyecek-içecekler ve meslekler gibi ek varlık kategorileri sisteme eklenmelidir.

Bitki-yer co-occurrence analizleri ile tarihî botanik coğrafya haritaları üretilebilir ve elde edilen tüm veri setleri yöntem bakımından araştırmacılara açık erişimle sunulabilir.

Kaynakça:

- Bakır, M. (2022). *Evliya Çelebi Seyahatnamesi'nde organ adları* (Yayımlanmamış yüksek lisans tezi). Kocaeli Üniversitesi, Kocaeli.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... ve Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
- Can, F., Kocerberber, S., Balcık, E., Kaynak, C., Ocalan, H. C. ve Vursavas, O. M. (2008). Information retrieval on Turkish texts. *Journal of the American Society for Information Science and Technology*, 59(3), 407-421.
- Dankoff, R. (2004). *An Ottoman mentality: The world of Evliya Çelebi*. Brill.
- Dankoff, R. (2006). *Evliya Çelebi Seyahatnamesi okuma sözlüğü*. Yapı Kredi Yayınları.
- Ehrmann, M., Colavizza, G., Rochat, Y. ve Kaplan, F. (2016). Diachronic evaluation of NER systems on old newspapers. *Proceedings of the 13th Conference on Natural Language Processing (KONVENS)*, 97-107.
- Evliya Çelebi b. Derviş Mehmed Zillî. (2007). *Evliya Çelebi Seyahatnamesi* (Cilt 1-10) (R. Dankoff, S. A. Kahraman ve Y. Dağlı, Haz.). Yapı Kredi Yayınları.
- Karacaoğlan, Ç. (2018). *Evliya Çelebi'nin Seyahatname'sinde yer alan bitki adları* (Yayımlanmamış yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.
- Kartal, A. (2016). *Evliya Çelebi Seyahatnâmesi'nde geçen yiyecek adları üzerine bir dil incelemesi* (Yayımlanmamış yüksek lisans tezi). Erciyes Üniversitesi, Kayseri.
- Kırmit, D. (2025). *Evliya Çelebi Seyahatnamesi'nde meslek adları* (Yayımlanmamış doktora tezi). Ankara Yıldırım Beyazıt Üniversitesi, Ankara.
- Küçük Ak, F. (2023). *Evliyâ Çelebi Seyahatnâmesi'nde mimari ve mimari süsleme terimleri* (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Li, J., Sun, A., Han, J. ve Li, C. (2023). A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, 34(1), 50-70.
- Öz, S. (2007). *Evliya Çelebi Seyahatnamesi'ne göre Anadolu'da yer adları* (Yayımlanmamış yüksek lisans tezi). Kırıkkale Üniversitesi, Kırıkkale.
- Peker, S. D. (2015). *Evliya Çelebi Seyahatnamesi'nde hayvanlarla ilgili söz varlığı* (Yayımlanmamış yüksek lisans tezi). Gazi Üniversitesi, Ankara.
- Piotrowski, M. (2012). *Natural language processing for historical texts*. Morgan & Claypool.
- Sağ Yener, Z. (2024). *Evliya Çelebi'nin Seyahatnamesi'nde kavramsal metaforlar* (Yayımlanmamış doktora tezi). Fatih Sultan Mehmet Vakıf Üniversitesi, İstanbul.
- Şeker, G. A. ve Eryiğit, G. (2012). Initial explorations on using CRFs for Turkish named entity recognition. *Proceedings of COLING 2012*, 2459-2474.
- Wang, S., Sun, X., Li, X., Ouyang, R., Wu, F., Zhang, T., ... ve Wang, H. (2023). GPT-NER: Named entity recognition via large language models. *arXiv preprint arXiv:2304.10428*.
- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., ... ve Fedus, W. (2022). Emergent abilities of large language models. *Transactions on Machine Learning Research*.
- Yeniterzi, R. (2011). Exploiting morphology in Turkish named entity recognition system. *Proceedings of the ACL 2011 Student Session*, 105-110.