

## Türkiye’de Kullanılan Posta Adres Bilgilerinde Uyum Düzeylerinin Belirlenmesi

(Determination of Consistency Levels in Postal Address Information Used in Turkey)

Batuhan KILIÇ , Fatih GÜLGEN 

Yıldız Teknik Üniversitesi Harita Mühendisliği Bölümü Kartografya Anabilim Dalı 34220 Davutpaşa, İstanbul  
batuhank@yildiz.edu.tr, fgulgen@yildiz.edu.tr

Geliş Tarihi (Received): 10.04.2018

Kabul Tarihi (Accepted): 26.12.2018

### ÖZ

Adres bilgileri, coğrafi verinin kullanımındaki zenginleşmeye paralel olarak posta hizmetlerinin yanında mekânsal referanslı çeşitli coğrafi ve demografik analizlerin gerçekleştirilmesinde altlık olarak kullanılmaktadır. Adresi oluşturan bileşenlerin standart bir yapıda saklanması, adres odaklı gerçekleştirilen uygulamaları kolaylaştırır, ekonomik ve zamansal maliyetleri azaltır. Bu çalışmada, Türkiye’de farklı kaynaklardan elde edilen posta adres bilgileri arasındaki uyumlar incelenmektedir. Adresler arasındaki ilişkiyi sayısal olarak ortaya koyabilmek için Damerau-Levenshtein mesafe algoritmasından yararlanılmıştır. Damerau-Levenshtein söz dizileri arasındaki benzerlikleri hesaplamak için geliştirilmiş bir mesafe belirleme algoritmasıdır. Test işleminde kullanılan adresler İstanbul ili Fatih ilçe sınırlarında yer alan 74 konaklama tesisine aittir. Beş farklı web portalından bu adres verileri web kazıma yöntemiyle toplanmıştır. Ayrıca aynı konaklama tesisi adresleri dört farklı harita servisinden coğrafi kodlama işlemiyle elde edilmiştir. Karşılaştırma işlemlerine referans oluşturan veriler ise Nüfus Kayıt ve Vatandaşlık İşleri Genel Müdürlüğü’nün Adres Kayıt Sistemi’nden alınmıştır. Tüm karşılaştırma işlemleri Damerau-Levenshtein algoritmasıyla adres bileşeni odaklı olarak gerçekleştirilmiştir. Sonuçlar, adresleme sisteminin Türkiye’de yeterince anlaşılmadığını ve uygun şekilde yapılmadığını göstermektedir.

**Anahtar Kelimeler:** Adres Kayıt Sistemi (AKS), Adres Bileşeni, Dizi Benzerlik, Damerau-Levenshtein Algoritması, Coğrafi Kodlama, Web Kazıma

### ABSTRACT

Address information, in parallel with the enrichment of geographical data, is used as a base for various geographical and demographic analyzes beside postal services. Storing the address components in a standard structure facilitates address-driven implementations, reducing economic and time costs. In this study, the consistency between various postal addresses obtained from different sources in Turkey was examined. For quantifying the relationship between the addresses, the Damerau-Levenshtein distance algorithm was used. Damerau-Levenshtein is a distance-finding algorithm developed to calculate similarities between a set of strings. The addresses, used in the case study, belong to 74 accommodation facilities in the district of Fatih in Istanbul. These addresses are collected by using web scraping method from five different web portals. In

addition, the same accommodation facility addresses were obtained by geocoding from four different map services. The reference data for the comparison process is taken from the Address Registration System of the General Directorate of Population Registration and Citizenship Affairs. All comparison operations were performed using the Damerau-Levenshtein algorithm based on each of address components. The results show that the addressing system has not been comprehended adequately and done properly in Turkey.

**Keywords:** Address Registration System (ARS), Address Component, String Similarity, Damerau-Levenshtein Algorithm, Geocoding, Web Scraping

### 1. GİRİŞ

Adres, coğrafi mekânda bir konumu tanımlamak için kullanılan söz dizimidir. Geleneksel arazi kullanım türlerine bakılmaksızın herhangi bir toprak parçası için en temel tanımlayıcı unsur olan adres, bir coğrafi detayın konum bilgisine karşılık olarak kullanılır (Kellison, 2012). Dünya üzerinde yerel yönetimler ile kamu kurum ve kuruluşları ihtiyaçlar doğrultusunda birçok çalışmada adres verilerine ihtiyaç duyar. Kamu hizmetlerinin doğru planlanması ve eş güdümlü olarak yürütülmesi devletin tüm vatandaşlarının temel bilgilerine hızlı erişebilmesiyle mümkündür. Kişilerin ikamet adresleriyle kimlik numaralarını içeren güvenilir nüfus kayıt sistemlerinin kurulması, bu erişimi kolaylaştırır. Bireyler ile konumları arasında ilişki kurmada kullanılan adres verileri ulusal ya da uluslararası bilgi altyapılarının kurulmasında da oldukça önemlidir (Coetzee ve diğerleri, 2008; Geymen, Yomralıoğlu ve Baz, 2008; Lind, 2008; Zandbergen, 2008; Zandbergen, 2010; Li, Zhang ve Chen, 2010; Yıldırım, Yomralıoğlu, Nişancı ve İnan, 2014).

Kişiler, kurumlar ve kuruluşlar tarafından adres bilgilerinin verimli kullanılması adresi oluşturan bileşenlerin doğru ve tutarlı olarak ifade edilmesiyle sağlanır. Adres bileşenleri bir adresin tanımlanmasında kullanılan coğrafi unsurlardır. Mahalle, cadde, sokak, bulvar ve meydan isimleri, bina numarası, posta kodları, il ve ilçe isimleri

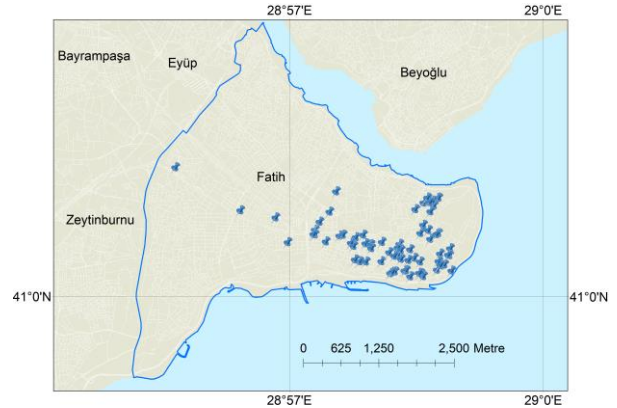
adresleri tanımlamak için kullanılan temel bileşenlerdir. Adres bileşenlerinin belirli bir sırada kullanılması standart adres kavramını ortaya çıkarır. Coğrafi analizlerin standart adres yapılarına bağlı olarak gerçekleştirilmesi, sonuçların kalitesini yükselterek ekonomik ve zamansal kayıpları ortadan kaldırır. Güney Afrika, Avustralya, Yeni Zelanda, Büyük Britanya, Danimarka, Amerika Birleşik Devletleri vb. birçok ülke adres yapılarının standartlaşması üzerine çalışmaktadır. Ayrıca, Uluslararası Posta Birliği (UPU), Uluslararası Standartlar Örgütü (ISO) ve Yapılandırılmış Bilgi Standartları Gelişim Örgütü (OASIS) gibi organizasyonlar adres standartlarının oluşturulmasına katkı sağlamaya devam etmektedir (Coetzee ve diğerleri, 2008). Avrupa ülkelerinin mekânsal veri altyapılarının oluşturulması sürecinde 2007 yılında *Infrastructure for Spatial Information in Europe (INSPIRE)* Direktifi yürürlüğe girmiştir (*INSPIRE*, 2007). Türkiye'de adres temasını da kapsayan mekânsal veri altyapılarının oluşturulması, kamu kurumları arasında veri paylaşımının sağlanması ve geleceğe yönelik stratejik planların oluşturulabilmesi için *INSPIRE* direktifine uygun olarak Türkiye Ulusal Coğrafi Bilgi Sistemleri (TUCBS) altyapı oluşturma faaliyetleri Coğrafi Bilgi Sistemi Genel Müdürlüğü öncülüğünde sürdürülmektedir.

Adres sisteminin standardizasyonuna yönelik çalışmalar ülkemizde büyük bir hızla devam etmesine karşın, kamu kurum ve kuruluşları tarafından farklı adres formatlarının kullanılması, adresi tanımlayan bileşenler arasında tutarsızlıklar olması, kısaltma kullanımlarındaki farklılıklar, değişen cadde, sokak, bulvar ve meydan isimlerinin güncellenme sıklığı, vatandaşların sokak isimlerinde ya da numarataj bilgilerinde yapılan değişikliklerin farkında olmaması vb. sorunlar, adres odaklı birçok uygulamayı olumsuz yönde etkilemektedir (Yıldırım ve diğerleri, 2014). Benzeri sorunlar, adres verilerini kullanan paydaşlar tarafından çeşitli platformlarda dile getirilmesine karşın, bunların etkileri henüz sayısal olarak analiz edilmemiş ve ortaya konulmamıştır. Bu çalışma kapsamında, adres tanımlamalarının standart kullanımına yönelik farkındalık oluşturmak için farklı kaynaklardan elde edilen adres verileri arasındaki uyum düzeylerini belirlemeyi hedefleyen bir yöntem tasarlanmıştır. Bu doğrultuda, bir test bölgesinden toplanan aynı konuma ait farklı kaynaklardaki adres içerikleri, metinler arasındaki benzerlikleri belirlemede kullanılan bir ölçüm yaklaşımı kullanılarak karşılaştırılmaktadır. Bu

yaklaşımın ayrıntıları çalışmanın üçüncü bölümünde açıklanmaktadır.

## 2. ÇALIŞMA BÖLGESİ VE VERİLERİN ELDE EDİLMESİ

Çok sayıda ülkede konaklama tesislerine ait bilgiler çeşitli web portalları aracılığıyla kullanıcılara sunulmaktadır. Bu portalların büyük kısmı global olarak hizmet veren şirketlere aittir ve her bir bağımsız portalda konaklama tesislerine ait adres içerikleri tutulmaktadır. Ülkemizde *Agoda.com*, *Booking.com*, *Hotels.com*, *Hotelscombined.com*, *Tripadvisor.com* gibi çeşitli siteler bu kapsamda hizmet vermektedir. Bu çalışmada, yukarıda bahsedilen beş farklı portaldan toplanan aynı otellere ait farklı adres içerikleri karşılaştırılmıştır. Test bölgesi, konaklama tesislerinin yoğun olduğu İstanbul'un Fatih ilçesi olarak seçilmiştir (Şekil 1). Bölgede bulunan 74 otelin yukarıda sayılan beş portal tarafından kullanıcılara konaklama tesisi olarak önerildiği gözlemlenmiştir.



Şekil 1. Fatih sınırlarında kalan ve test işlemlerinde kullanılan 74 otel

Çalışmaya altlık oluşturan adresler internet sitelerinden web kazıma olarak adlandırılan bir yöntemle elde edilmiştir. Web kazıma, bir internet sayfasının belirli konumunda yer alan standart bilgilerin toplanmasında kullanılan bir veri çekme işlemidir. Günümüzde web ortamında veri madenciliği çalışmalarının hız kazanmasıyla birlikte ekran kazıma ve web hasadı gibi kavramlar da benzer anlamda kullanılmaktadır (Mitchell, 2015). *Visual Web Ripper*, *Web Content Extractor*, *Web Scraper Plus+* gibi çeşitli ticari web kazıma yazılımları internetten veri toplama amacıyla kullanılmaktadır. Bu çalışmada otellerin beş ayrı sitede yazan adres içerikleri, Newprosoft şirketi tarafından geliştirilen *Web Content Extractor* yazılımıyla toplanmıştır.

Toplanan adres bilgilerine ek olarak her bir otelin, *ArcGIS Online*, *Bing*, *Google* ve *Yandex* hizmetlerinde tanımlı standart adres içerikleri coğrafi kodlama yöntemi kullanılarak da elde edilmiştir. Coğrafi kodlama, bir adres bilgisini enlem ve boylam açıları ile ifade edilen coğrafi koordinatlara çeviren ve bu noktaya ait standart adres format dönüşümünü sağlayan bir işlem aracıdır. Günümüzde mekânsal analizlere altlık oluşturan temel işlemlerden biri olan coğrafi kodlama, CBS ortamında klasik ve çevrimiçi olarak gerçekleştirilmektedir. Bu çalışmada, adres içeriklerinin toplanmasında çevrimiçi yöntem kullanılmıştır.

Son olarak her bir otelin adresi, Nüfus Kayıt ve Vatandaşlık İşleri Genel Müdürlüğü tarafından oluşturulan Adres Kayıt Sistemi (AKS) web sitesinden bulunmuştur. AKS’nin adres bileşenleri arasında posta kodu yer almadığından bu bileşene ait veriler T.C. Posta ve Telgraf Teşkilatı (PTT) resmi sayfasından alınmıştır. İki resmi kurumdan elde edilen adres içerikleri referans tanımlamalar olarak kabul edilmiştir.

### 3. YÖNTEM

#### a. Hazırlık Aşaması

Farklı portallardan toplanan otel vb. konaklama tesislerinin adres tanımlamalarında çeşitli farklılıklar rastlanmaktadır. Bir portalda kullanılan bir adres bileşenine ait içerik, bazı durumlarda diğer portalda hiç kullanılmayabilir. Çalışmanın hazırlık aşamasında, adres içeriklerinde görünen böylesi tutarsızlıkları saptayabilmek için toplanan tüm adres içerikleri incelenerek kullanılan her bir adres bileşeninin AKS uygulama yönergesindeki karşılığı tespit edilmiştir. AKS’de tanımlı adresler; il, ilçe, köy, mezra, belediye, mahalle, meydan, bulvar, cadde, sokak, küme ev, site, blok, mevki ismi, dış kapı ve iç kapı numarası ile posta kodu bileşenlerinden oluşmaktadır (“Adres Kayıt Sistemi”, 2007). Söz konusu standart uyarınca bu bileşenlerden il, ilçe ismi, dış kapı numarası ve posta kodunun her adreste yer alması zorunludur. Ayrıca, il merkezlerinde; il, ilçe, mahalle ismi ile meydan, bulvar, cadde, sokak ve küme evlerin isimlerinden sadece birinin ve dış kapı numarası ile posta kodunun bulunması gerekir. Bu kapsamda, portallar üzerinden elde edilen adreslerin bileşen içerikleri AKS’ye göre manuel olarak ayrıştırılmıştır. Servisler üzerinden elde edilen adreslerin bileşen içerikleri ise çevrimiçi coğrafi kodlama işlemi sayesinde otomatik olarak ayrıştırılabilmektedir. Çalışma bölgesi yerleşim içinde kaldığından ayrıştırılan adresler; il, ilçe mahalle,

yol adı, numarataj ve posta kodu olmak üzere toplam altı bileşenden oluşmuştur.

#### b. Adres Benzerlik Ölçümü

On farklı kaynaktan elde edilen adresler arasındaki uyum oranları, bileşen odaklı olarak metin benzerlik ölçüm yöntemiyle hesaplanmıştır. Metin benzerlik ölçümü, bilgi çıkarma, metin sınıflandırma, belge kümeleme, konu izleme, soru cevaplama, metin özetleme, intihal araştırma vb. işlemlerde kullanılmaktadır. Benzerlik ölçümü, sözcüksel ya da anlamsal olarak çalışan algoritmalar kullanılarak yapılır. Sözcüksel olarak çalışan dizi odaklı algoritmalar dizi serileri ve karakter bileşimleri üzerine yoğunlaşır (Gomaa ve Fahmy, 2013). Anlamsal çalışan algoritmalar, yapay zekâ, doğal dil işleme, bilişsel bilim gibi farklı uygulama alanlarında kullanılmaktadır. Kümeleme ve bilgi çıkarma teknikleri sayesinde kelimeler arasındaki anlamsal benzerlik, geliştirilen çeşitli algoritmalarla anlamsal olarak ölçülür. İki kelimenin anlamsal olarak benzer olup olmadığına karar vermek için bu iki kelime arasındaki semantik ilişkileri kavramak gerekir. Bu çalışmada adres bileşenlerinin içerikleri arasındaki benzerlik oranları sözcüksel algoritmalar arasında en çok bilinen Damerau-Levenshtein mesafe algoritması kullanılarak hesaplanmıştır (1).

$$DL_{benzerlik} = 1 - \frac{d(s_1, s_2)}{maks\_uzunluk(|s_1|, |s_2|)} \quad (1)$$

Burada,  $s_1$  ve  $s_2$  karşılaştırılan iki örnek diziyi göstermekte,  $d$  ise bu iki dizinin birbirlerine kaç hamle sonrasında eşit olacağını gösteren uzaklık ölçütünü temsil etmektedir.

1964’lü yıllardan günümüze kadar birçok çalışmaya altlık oluşturan Damerau-Levenshtein mesafe algoritması Levenshtein algoritmasının genişletilmiş bir versiyonu olarak tanımlanmaktadır (Damerau, 1964; Levenshtein, 1966; Hall ve Dowling, 1980; Peterson, 1980; Brill ve Moore, 2000; Bard, 2007; Peled, Fire, Rokach ve Elovici, 2013; van der Loo, 2014). Levenshtein algoritması iki diziyi karakter odaklı karşılaştırarak, bir dizinin diğerine dönüşebilmesi için gereken minimum işlem sayısını karakter odaklı olarak hesaplamaktadır.

Kılıç ve Gülgen (2017) posta adresleri arasındaki benzerlik ilişkilerini, adres ifadelerinin bütünü ele alarak, Levenshtein mesafe algoritması ile karşılaştırmıştır. Sonuçlar, portallarda kayıtlı adres ifadelerinde yapılmış olan

basit hataların benzerlik oranlarını düşürdüğünü göstermiştir.

Bu çalışmada önerilen Damerou-Levenshtein algoritması ise basit hataları göz ardı edebilmektedir. Algoritmanın en önemli özelliği komşu karakterin yer değişimine izin vermesidir (Damerou, 1964). Örneğin, “Mahalle” kelimesinin kısaltması olan “Mah.” yapılan bir basit hatayla “Mha.” şeklinde yazıldığında, Levenshtein algoritması iki kısaltma arasında iki birimlik mesafe tespit ederken; Damerou-Levenshtein algoritması mesafeyi bir birim olarak hesaplamaktadır. Mesafenin daha kısa hesaplanması ise benzerlik oranının daha yüksek çıkmasını sağlamaktadır.

Hesaplama işlemleri, çalışma bölgesinde bulunan toplam 74 otel için bileşen odaklı olarak gerçekleştirilmiştir. Ele alınan her bir otelin referans olarak kullanılan (AKS ve PTT veri tabanlarından elde edilen) adresleri, mahalle adı, yol adı, numarataj, ilçe adı, posta kodu ve il adı bileşenlerine ait bilgileri içermektedir. Buna karşın, servisler ve portallardan çıkarılan bileşenlere ait içeriklerde bazı eksiklikler mevcuttur. Örneğin Google ve Bing servisi bir otele ait adreste mahalle bilgisini vermeyebilir. Eğer ikili olarak karşılaştırılan içeriklerin her ikisi de eksikse bir uyum değeri hesaplanamaz. İçeriklerden biri mevcut diğeri değilse uyum %0 olarak hesaplanır. İki içeriğinde mevcut olduğu durumda ise uyum hesabı Damerou-Levenshtein algoritması kullanılarak yapılır. Bu algoritma karakter odaklı olarak çalıştığından dolayı karşılaştırma sonucu elde edilen uyum değerlerinin, bu çalışmada %80’den fazla olması beklenmektedir.

Test bölgesindeki 74 otele ait adreslerin mahalle adı, yol adı, numarataj, ilçe adı, posta kodu ve il adı bileşenleri için ikili karşılaştırmalardan hesaplanan ortalama uyumlar altı ayrı tabloda verilmiştir. Tablolar üzerinde her bir servisten elde edilen posta adreslerinin referans verilerle olan ortalama uyum değerleri sarı renkte görünmektedir. Portallardan elde edilen veriler ile referans veriler arasındaki ortalama uyumlar yeşil renktedir. Dört servisin birbirleriyle olan ortalama uyumları mavi renk; beş portalın birbiriyle uyumu ise turuncu renk ile gösterilmiştir. Portallar ve servislerden hesaplanan uyum değerleri ise gri renktedir. Tablolarda geçen kısaltmalar ise aşağıda verilmiştir:

- **S1** → ArcGIS Online servisi
- **S2** → Bing servisi
- **S3** → Google servisi

- **S4** → Yandex servisi
- **P1** → Agoda.com portalı
- **P2** → Booking.com portalı
- **P3** → Hotels.com portalı
- **P4** → Hotelscombined.com portalı
- **P5** → TripAdvisor.com portalı

Tablo 1 mahalle bileşenine ilişkin değerleri göstermektedir. Servislerin referans verileri olan uyumları (sarı) incelendiğinde Google’ın oldukça iyi sonuç verdiği, buna karşın Bing’in mahalle bilgisini hiç tutmadığı görülmektedir. ArcGIS Online ve Yandex’in ise mahalle bileşenini tutmasına karşın bu bilgiyi kullanıcılarına sağlıklı yansıttığı söylenemez. Servislerin kendi arasında da çok uyumlu olduğu söylenemez (mavi). Buna karşın, Google ve ArcGIS Online arasındaki uyumun daha yüksek olmasının nedeni, ArcGIS Online servisinin mahalle bileşeni içinde sadece mahalle adını standart olarak tutmasıdır. Örneğin; ‘Cankurtaran Mahallesi’ için ArcGIS Online mahalle bilgisini sadece ‘Cankurtaran’ olarak tutarken, Google ‘Cankurtaran Mahallesi’ şeklinde tutmaktadır. Bu durum sonuçların düşük çıkmasında etkilidir. Yandex ise mahalle bilgisini düzenli olarak yansıtmadığından diğer iki servis ile olan uyum değerleri düşük çıkmaktadır. Portallardan elde edilen mahalle bilgisinin referans veri tabanındaki mahalle bilgileriyle uyumları (yeşil) birbirine oldukça yakındır. Beş portal için ortalama %45 civarında olan bu değer de yetersizdir. Portalların kendi içindeki ortalama uyum değeri (turuncu) %63 civarındadır ve değerler birbirine oldukça yakındır. Bu değerler ışığında, portallardan elde edilen mahalle bilgilerinin genel olarak çok sağlıklı olmadığı söylenebilir. Bunun temel nedeni portallarda mahalle bilgisinin düzenli olarak yer almamasıdır. Bu nedenle, portallar ve servisler arasındaki uyumlar (gri) oldukça değişkendir ve düşük seviyelerdedir.

Tablo 2 yol adı bileşenine ilişkin değerleri göstermektedir. Servislerin referans verileri olan uyumları (sarı) oldukça yüksektir. Bing, referans verileriyle yaklaşık %99 uyumludur. Google ve Yandex için referans verileriyle uyum nispeten biraz daha düşük seviyededir. Servislerin kendi aralarındaki uyumlar (mavi) incelendiğinde ArcGIS Online ile Bing’in, Google ile de Yandex’in kendi içlerinde daha uyumlu olduğu söylenebilir. Uyumların iki servis arasında daha yüksek çıkmasının nedeni bunların benzer standartlarda kısaltmalar kullanmasıdır. Yol adı bileşeni için portallarının kendi aralarında (turuncu) genelde %80’in üzerinde yüksek uyum sağladığı, ancak AKS ile olan uyumlarında (yeşil) bu oranın %70 seviyelerine indiği görülmektedir. Hotels.com için

ise referans verilerle olan uyum %57 seviyesine kadar düşmektedir. Referans veriyle portalların uyumunda gözlemlenen bu azalmanın nedeni, portallardaki yol adlarının birden fazla yolu kapsamaması ve bunun AKS yönergesine aykırı bir durum olmasıdır. Örneğin; *Hotels.com*’da ‘Mimar Mehmet Ağa Cad. Amiral Tafdil Sok.’ olarak

tutulan bir sokak, AKS’de ‘Amiral Tafdil Sokak’ olarak yer aldığından uyum değeri düşmektedir. Servislerin portallar ile olan uyumlarında bazı yol adı bilgilerinde eksiklikler olması nedeniyle düşük değerler çıkmakta ve uyumsuzluklar görülmektedir.

Tablo 1. Mahalle adı bileşenine ilişkin ortalama uyum bilgileri

%	Ref.	S1	S2	S3	S4	P1	P2	P3	P4	P5
<b>Ref.</b>	100	66.78	0	89.73	45.23	45.73	45.05	44.76	47.06	41.41
<b>S1</b>		100	0	72.64	50.04	34.86	32.47	33.88	37.93	30.07
<b>S2</b>			-	0	0	0	0	0	0	0
<b>S3</b>				100	50.59	39.92	38.97	39.68	43.12	33.22
<b>S4</b>					100	28.29	31.63	37.51	34.49	30.55
<b>P1</b>						100	67.29	53.91	61.29	55.05
<b>P2</b>							100	63.39	68.08	64.43
<b>P3</b>								100	62.08	66.76
<b>P4</b>									100	63.81
<b>P5</b>										100

Tablo 2. Yol adı bileşenine ilişkin ortalama uyum bilgileri

%	Ref.	S1	S2	S3	S4	P1	P2	P3	P4	P5
<b>Ref.</b>	100	88.35	98.69	81.71	78.68	75.79	73.8	57.39	72.5	75.95
<b>S1</b>		100	88.63	73.18	70.04	61.02	57.97	61.13	55.24	59.9
<b>S2</b>			100	82.03	78.73	68.47	64.84	68.66	61.35	67.35
<b>S3</b>				100	92.15	65.14	64.7	72.79	71.91	68.78
<b>S4</b>					100	63.2	62.45	70.65	70.91	66.68
<b>P1</b>						100	86.36	79.1	80.34	81.5
<b>P2</b>							100	85.25	84.21	85.89
<b>P3</b>								100	87.26	86.68
<b>P4</b>									100	85.25
<b>P5</b>										100

Tablo 3. Numarataj bileşenine ilişkin ortalama uyum bilgileri

%	Ref.	S1	S2	S3	S4	P1	P2	P3	P4	P5
<b>Ref.</b>	100	88.03	98.65	92.53	95.5	72.11	85.6	56.61	79.24	79.76
<b>S1</b>		100	89.24	82.53	83.53	23.74	23.13	21.11	20.02	24.49
<b>S2</b>			100	91.18	94.14	26.26	25.47	22.7	22.36	27.13
<b>S3</b>				100	89.38	67.66	79.81	64.88	70.94	72.36
<b>S4</b>					100	25.73	24.32	22.05	21.55	25.91
<b>P1</b>						100	78.09	69.58	70.25	70.89
<b>P2</b>							100	79.79	82.56	83.83
<b>P3</b>								100	78.59	77.66
<b>P4</b>									100	82.12
<b>P5</b>										100

Tablo 4. İlçe adı bileşenine ilişkin ortalama uyum bilgileri

%	Ref.	S1	S2	S3	S4	P1	P2	P3	P4	P5
<b>Ref.</b>	100	97.88	0	99.21	98.65	27.38	51.88	29.6	44.89	33.55
<b>S1</b>		100	0	97.1	96.53	17.64	37.42	20.76	33.12	24.77
<b>S2</b>			-	0	0	0	0	0	0	0
<b>S3</b>				100	97.86	19.17	37.47	19.26	33.22	23.17
<b>S4</b>					100	18.46	36.9	20.24	32.65	24.14
<b>P1</b>						100	62.41	52.68	47.15	54.49
<b>P2</b>							100	58.62	68.14	61.23
<b>P3</b>								100	42.55	61.87
<b>P4</b>									100	46.6
<b>P5</b>										100

Tablo 3 numarataj bileşenine ilişkin değerleri göstermektedir. Referans veriler ve servisler arasındaki numarataj bilgisi uyumu (sarı) oldukça yüksek çıkarken, referans veriler ve portallar arasındaki uyumun (yeşil) nispeten daha düşük ve değişken olduğu görülmektedir. Özellikle *ArcGIS Online* dışında kalan üç servisin referans verilerle çok yüksek derecede uyumlu olduğu söylenebilir. *Bing* tarafından sağlanan numarataj bilgisi referans verilerle çok duyarlıdır. Bunlara paralel olarak, servislerin (mavi) portallara (turuncu) göre kendi içlerindeki uyum değerleri de daha yüksektir. Portallar ve servisler arasındaki uyumlar (gri) birkaç değer dışında oldukça düşüktür. Yaklaşık %20-%80 arasında değişen bu dalgalanmaların nedeni; portallarda tutulan numarataj bilgilerinin ülkemizde 'No' eki ile birlikte tanımlanmasıdır.

Tablo 4 ilçe adı bileşenine ilişkin değerleri göstermektedir. Referans veriler ve servisler arasındaki uyum (sarı) *Bing* haricinde oldukça yüksektir. *Bing* herhangi bir ilçe bilgisi vermemektedir. *Google* ile referans veriler arasındaki uyum ise çok yüksektir. Benzer olarak, diğer servislerin kendi içlerindeki uyumlarının da (mavi) yüksek olduğu söylenebilir. Referans veriler ve portallar arasındaki uyum değerleri (yeşil) genelde düşük ve oldukça değişkendir. Bunun nedeni; karşılaştıran portal verilerinin birden fazla (örneğin; 'Kadırga Sultanahmet' ve 'Eminönü Sultanahmet') ya da kendini yineleyen (örneğin; 'Sultanahmet Sultanahmet') gibi fazladan bilgiler içermesidir. Bu nedenden dolayı, portalların kendi içlerinde (turuncu) ve servislerle (gri) olan uyumu da beklenenden düşüktür ve değişkendir.

Tablo 5 posta kodu bileşenine ilişkin değerleri göstermektedir. *ArcGIS Online* ve *Bing* servisleri posta kodu bilgilerini referans verilerle (sarı) %100 uyumlu sağlamaktadır. *Google*'in uyumu ise biraz daha düşüktür. Buna karşın *Yandex* posta kodu bilgisi vermemektedir. Servislerin kendi içlerindeki uyumları da (mavi) buna paraleldir. *Hotelscombined.com* posta kodu bilgisi vermemesine karşın diğer portalların referans verilerle uyumu (yeşil) ortalama %75 civarındadır ve değerler birbirlerine oldukça yakındır. Portalların kendi içlerindeki uyumları da (turuncu) birbirlerine yakındır. Bu durum servisler ve portallar arasında da geçerlidir. *Yandex* ve *Hotelscombined.com* arasında veri olmadığından bir karşılaştırma yapılamamıştır.

Tablo 6 il adı bileşenine ilişkin değerleri göstermektedir. Genel olarak farklı kaynaklardan sağlanan il adı bileşenine ait bilgiler, referans verilerle, kendi aralarında ve birbirleriyle yüksek uyumludur. Fakat *Hotels.com*'un sağladığı il adları diğer tüm veri setleriyle düşük uyumludur. Bunun nedeni, *Hotels.com* adreslerinin il adı bileşeninin 'İstanbul İstanbul' şeklinde tekrarlayan bilgiler içermesidir.

Tablo 5. Posta kodu bileşenine ilişkin ortalama uyum bilgileri

%	Ref.	S1	S2	S3	S4	P1	P2	P3	P4	P5
<b>Ref.</b>	100	100	100	95.95	0	74.59	79.2	75.68	0	70.86
<b>S1</b>		100	100	95.95	0	73.43	78.11	75.14	0	69.34
<b>S2</b>			100	95.95	0	73.24	78.1	75.13	0	69.34
<b>S3</b>				100	0	71.62	76.49	72.7	0	66.64
<b>S4</b>					-	0	0	0	-	0
<b>P1</b>		S1 → ArcGIS Online servisi S2 → Bing servisi				100	62.41	70.54	0	66.35
<b>P2</b>		S3 → Google servisi S4 → Yandex servisi					100	79.46	0	73.96
<b>P3</b>		P1 → Agoda.com portalı P2 → Booking.com portalı						100	0	70.73
<b>P4</b>		P3 → Hotels.com portalı P4 → Hotelscombined.com portalı							-	0
<b>P5</b>		P5 → TripAdvisor.com portalı								100

Tablo 6. İl adı bileşenine ilişkin ortalama uyum bilgileri

%	Ref.	S1	S2	S3	S4	P1	P2	P3	P4	P5
<b>Ref.</b>	100	98.65	87.5	100	100	99.51	100	68.7	98.82	100
<b>S1</b>		100	86.32	98.65	98.65	97.93	98.65	53.41	97.47	98.65
<b>S2</b>			100	87.5	87.5	86.87	87.5	47.61	86.32	87.5
<b>S3</b>				100	100	99.28	100	54.04	98.82	100
<b>S4</b>					100	99.28	100	54.04	98.82	100
<b>P1</b>		S1 → ArcGIS Online servisi S2 → Bing servisi				100	99.51	69.1	98.33	99.51
<b>P2</b>		S3 → Google servisi S4 → Yandex servisi					100	68.7	98.82	100
<b>P3</b>		P1 → Agoda.com portalı P2 → Booking.com portalı						100	67.94	68.7
<b>P4</b>		P3 → Hotels.com portalı P4 → Hotelscombined.com portalı							100	98.82
<b>P5</b>		P5 → TripAdvisor.com portalı								100

#### 4. SONUÇ

Bu çalışmada, ülkemizde kullanılan adres tanımlamalarının uyumları, karakter odaklı çalışan bir algoritma kullanılarak, servisler ve web portalları üzerinden, bileşenlere bağlı olarak incelenmiştir. Hesaplanan ortalama uyum değerleri, servisler tarafından sağlanan adres bilgilerinin portal bilgilerine göre AKS ile daha uyumlu olduğunu göstermektedir. Mahalle ve ilçe adı bileşenleri için *Google*'in; yol adı ve numarataj bileşenleri için *Bing*'in sonuçları oldukça başarılıdır. Posta kodu bileşeni için *ArcGIS Online* ve *Bing* tam uyumlu sonuçlar vermektedir. İl adı bileşeni için ise *Google* ve *Yandex* tam başarılıdır. Buradan *Bing* servisinin yol adı, numarataj ve posta kodu bilgisine bağlı çalıştığı söylenebilir. *Google* ve *ArcGIS Online* bahsedilen tüm bileşen bilgilerini sağlamaktadır. Mahalle adı bileşenine


ilişkin bilgileri sadece *Google*'in başarılı olarak sağladığı görülmektedir. *Yandex*'in ise diğer servislerden daha az başarılı olduğu söylenebilir.


Portalların sonuçları genel olarak çok değişkendir. Yaklaşık %27 gibi çok başarısız kabul edilebilecek uyum değerleri içermektedir. Bunun temel nedeni adres bilgilerinin ülkemizde standart şekilde başarılı olarak ifade edilememesidir. Bu noktada topluma adres kullanımlarıyla ilgili gerekli bilincin okullarda kazandırılması ya da kamu spotu vb. uygulamalarla farkındalık yaratılması gerekir.

Karşılaştırmalar için kullanılan Damerau-Levenshtein algoritması karakter odaklı olarak çalıştığından farklı kısaltmalar kullanılması ve karakterler arasındaki boşluklar bırakılması benzerlik sonuçlarını düşürmektedir. Bu durum Damerau-Levenshtein algoritmasının bir eksiğidir.

Kısaltmalar ve boşluklar göz ardı edilerek mantıksal ilişkileri ön plana çıkaran bir algoritma bu sorunun üstesinden gelebilir. Ayrıca, sayısal ifadelerle tanımlanan numarataj ve posta kodu bilgilerinin herhangi bir karakterindeki farklılık, Damerau-Levenshtein algoritmasının ürettiği uyum değerini belli bir oranda düşürmektedir. Buna karşın karakter odaklı bu fark, adresler arasında uyumsuzluk olduğunun bir göstergesidir. Bu nedenle, algoritmanın sayısal değerlerin karşılaştırılmasında kullanılması anlamsız şekilde yüksek uyum değerleri üretebilir. Gelecekte bu sorunların üstesinden gelebilecek bir adres karşılaştırma algoritması tasarımı üzerinde çalışılacaktır.

### ORCID

Batuhan KILIÇ  <https://orcid.org/0000-0002-0529-8569>

Fatih GÜLGEN  <https://orcid.org/0000-0002-8754-9017>

### KAYNAKLAR

- Adres Kayıt Sistemi Uygulama Yönergesi. (2007). Erişim adresi: <https://www.nvi.gov.tr/PublishingImages/mevzuat/nufus-mevzuati/yonerge/AdresKayitSistemiUygulamayonerge.doc>
- Bard, G.V. (2007, Ocak). *Spelling-error tolerant, order-independent pass-phrases via the Damerau-Levenshtein string-edit distance metric*. In Proceedings of the fifth Australasian symposium on ACSW frontiers, Ballarat, Avustralya.
- Brill, E. ve Moore, R.C. (2000, Ekim). *An Improved error model for noisy channel spelling correction*. In Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, Hong Kong.
- Coetsee, S., Cooper, A.K., Lind, M., Wells, M.M., Yurman, S.W., Wells, E., ... Nicholson, M.J. (2008, Şubat). *Towards an international address standard*. 10th International Conference for Spatial Data Infrastructure, St. Augustine, Trinidad & Tobago.
- Damerau, F.J. (1964). A Technique for computer detection and correction of spelling errors. *Communications of the ACM*, 7(3), 171–176. doi:10.1145/363958.363994
- Geymen, A., Yomralioglu, T. ve Baz, I. (2008). Developing an urban information system for local governments. *Proceedings of the Institution of Civil Engineers-Municipal Engineer: Published for the Institution of Civil Engineers*, 161(3), 163-173. doi: 10.1680/muen.2008.161.3.163
- Gomaa, W.H. ve Fahmy, A.A. (2013). A Survey of text similarity approaches. *International Journal of Computer Applications*, 68(13), 13-18. doi: 10.5120/11638-7118
- Hall, P.A. ve Dowling, G.R. (1980). Approximate string matching. *ACM computing surveys (CSUR)*, 12(4), 381-402. doi:10.1145/356827.356830
- INSPIRE Directive. (2007). Directive 2007/2/EC of the EU Parliament and of the Council (14 March 2007) Establishing An Infrastructure for Spatial Information in the EU Community (INSPIRE). *Official Journal of the European Union*, 50.
- Kellison, M.T. (2012). *Address points and A Master address file: Improving efficiency in the city of Chino* (Doktora Tezi). ProQuest Dissertations and Theses veri tabanından erişildi. (UMI No. 1532831)
- Kılıç B. ve Gülgen F. (2017, Kasım). *A Research on Standard Address Usage in Turkey*. UCTEA International Geographical Information Systems Congress 2017, Adana, Türkiye.
- Levenshtein, V.I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *In Soviet physics doklady*, 10(8), 707-710.
- Li, B., Zhang, X. ve Chen, Y. (2010, Ağustos). *Automatic construction and visualization of address models*. 2010 Sixth International Conference on Natural Computation, Yantai, Shandong, Çin.
- Lind, M. (2008, Mayıs). *Addresses as an Infrastructure component: Danish experiences and perspectives*. Proceedings of the ISO Workshop on Address Standards: Considering the Issues related to an International Address Standard. Kopenhag, Danimarka.
- Mitchell, R. (2015). *Web scraping with Python: Collecting data from the modern web*. Sebastopol, California: O'Reilly Media, Inc.



- Peled, O., Fire, M., Rokach, L. ve Elovici, Y. (2013, Eylül). *Entity matching in online social networks*. 2013 International conference on social computing (SocialCom), Washington D.C., ABD.
- Peterson, J.L. (1980). Computer programs for detecting and correcting spelling errors. *Communications of the ACM*, 23(12), 676-687. doi:10.1145/359038.359041
- Van der Loo, M.P. (2014). The Stringdist package for approximate string matching. *The R Journal*, 6(1), 111-122.
- Yıldırım, V., Yomralıoğlu, T., Nişancı, R. ve İnan, H. (2014). Turkish street addressing system and geocoding challenges. *Proceedings of the Institution of Civil Engineers: Municipal Engineer*, 167, 99–107. doi: 10.1680/muen.13.00008
- Zandbergen, P.A. (2008). A Comparison of address point, parcel and street geocoding techniques. *Computers, Environment and Urban Systems*, 32, 214-232. doi: 10.1016/j.compenvurbsys.2007.11.006
- Zandbergen, P.A. (2010). Influence of street reference data on geocoding quality, *Geocarto International*, 26(1), 35–47. doi: 10.1080/10106049.2010.537374