

PAPER DETAILS

TITLE: Stackoverflow gönderilerinde tartisilan trend konularin kelime frekans analizi ile belirlenmesi

AUTHORS: Fatih GÜRCAN,Özcan ÖZYURT

PAGES: 357-368

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/1346883>

Stackoverflow gönderilerinde tartışılan trend konuların kelime frekans analizi ile belirlenmesi

Identification of trend topics discussed in stackoverflow posts by word frequency analysis

Fatih GÜRCAN^{*1,a}, Özcan ÖZYURT^{2,b}

¹ Karedeniz Teknik Üniversitesi, Uzaktan Eğitim Uygulama ve Araştırma Merkezi, 61080, Trabzon

²Karadeniz Teknik Üniversitesi, Of Teknoloji Fakültesi, Yazılım Mühendisliği Bölümü, 61830, Trabzon

• Geliş tarihi / Received: 15.10.2020

• Düzeltilerek geliş tarihi / Received in revised form: 02.02.2021

• Kabul tarihi / Accepted: 09.02.2021

Öz

Günümüzde yazılım geliştiriciler ve bilgisayar bilimciler sıklıkla çevrimiçi bilgi paylaşım platformlarını kullanmaktadır. StackOverflow bu platformların başında gelmektedir. Bu ortamda paylaşılan bilgilerin analizi güncel konu ve eğilimlerin belirlenmesinde önemli çıkarımlar sunabilir. Bu bağlamda yürütülen çalışmada 2019 yılı içerisinde StackOverflow platformu üzerinde paylaşılan gönderilere ait etiketlerin incelenmesi amaçlanmıştır. Kelime frekans analizine dayalı metinsel içerik analizinin yapıldığı bu çalışmada, StackOverflow gönderilerinde kullanılan etiketlerle bir veri seti oluşturulmuş ve bu etiketler içerisinde en çok kullanılanlar, ivmesi artan ve azalan etiketler analiz edilmiştir. Bu analiz sonucunda elde edilen etiketlerden en çok kullanılan elli etiket üzerinden detaylı sonuçlar verilmiştir. En çok kullanılan bu 50 etiketin altı sınıf altında kümelendiği görülmüştür. Çalışmanın sonuçlarına dayanarak başta yenilikçi web, mobil ve ilişkisel olmayan veri tabanı teknolojileri olmak üzere güncel teknolojilerin genel olarak sık kullanıldığı ve ivmelerinin arttığı söylenebilir. Bunun yanında programlama dilleri, araçları ve kütüphanelerinde Python odaklı güçlü bir eğilimin olduğu açıktır. Çalışmanın sonuçlarının başta yazılım geliştiriciler ve bu alanda eğitim alan ve kariyerini bu alanda planlayan bireyler olmak üzere, müfredat yapımcılar ve karar vericilere önemli bilgiler sunması beklenmektedir.

Anahtar kelimeler: Kelime frekans analizi, Soru ve cevap forumları, StackOverflow gönderileri, Trend konular.

Abstract

Today, software developers and computer scientists frequently use online information sharing platforms. StackOverflow is one of these important platforms. The analysis of the information shared in this environment can offer important implications in revealing current issues and trends. In this context, it is aimed to examine the tags of the posts shared on the StackOverflow platform in 2019. In this study, in which textual content analysis based on word frequency analysis was made, a data set was created with the tags used in StackOverflow posts, and the most used tags among these tags, whose acceleration increased and decreased, were analyzed. Detailed results were given on the fifty most used tags among the tags obtained as a result of this analysis. It has been observed that these 50 most used tags are clustered under six classes. Based on the results of the study, it can be said that current technologies, especially innovative web, mobile and non-relational database technologies, are generally used frequently and their acceleration has increased. In addition, it is clear that there is a strong Python-oriented trend in programming languages, tools and libraries. The results of the study are expected to provide important information to curriculum makers and decision-makers, especially software developers and individuals who have received training in this field and plan their career in this field.

Keywords: Word frequency analysis, Question and answer forums, StackOverflow posts, Trend topics.

^{*a} Fatih GÜRCAN; fgurcan@ktu.edu.tr, Tel: (0505) 536 55 62, orcid.org/0000-0001-9915-6686

^b orcid.org/0000-0002-0047-6813

1. Giriş

Bilişim teknolojilerinin hızla geliştiği ve değiştiği günümüzde, bilgisayar bilimcileri ve yazılım geliştiricileri de bu değişime ayak uydurma adına sürekli öğrenme ve bilgi paylaşımı içerisine girmektedir. Son yıllarda bilgiye erişimin ve bilgi paylaşımının arttığı göz önünde bulundurulduğunda (Bakır vd., 2020), sosyal medya başta olmak üzere, özel amaçlı geliştirilen ortak çalışma siteleri bu amaçla etkin bir biçimde kullanılmaktadır (Johri ve Bansal, 2018). Bu platformlar, giderek artan veri ve bilgi kaynağı olması bakımından aynı zamanda birer öğrenme ortamı olarak da değerlendirilebilir (Ahmed ve Srivastava, 2017; Özyurt ve Özyurt, 2010). Araştırmacılar ya da geliştiriciler yeni bir konu hakkında bilgi edinmeye veya karşılaştıkları bir problemin çözümünde yeni bilgilere ihtiyaç duyduklarında bu platformları sıklıkla tercih etmektedirler (Joorabchi vd., 2016). Diğer yandan, sürekli değişen ve gelişen teknolojiye ayak uydurmak için geliştiriciler bu platformları kullanarak bilgi alanlarını paylaşmakta ve daha az bilgi sahibi oldukları alanlarda diğer geliştiricilerden yardım istemektedirler (Johri ve Bansal, 2018; Özyurt ve Özyurt, 2010).

Son yıllarda topluluk temelli Soru ve Cevap (SvC) forumlarının popülaritesi gittikçe artmıştır (Ahmed ve Srivastava, 2017; Barua vd., 2016; Joorabchi vd., 2016). Bu forumlar bilgiyi paylaşmak için mükemmel bir fırsat sunmanın yanı sıra öğrenme için de ideal bir ortam sağlarlar (Ahmed ve Srivastava, 2017). Bilişim alanında farklı SvC forumları olmakla birlikte StackOverflow bu alanda önde gelen platformlar arasında yer almakta ve teknik konularda fikir paylaşımı yapmak için en iyi forumlardan biri olarak kabul edilmektedir (Ahmed ve Srivastava, 2017; Barua vd., 2016; Johri ve Bansal, 2018; Yang vd., 2016; Zhang ve Zhu, 2018). Günümüzün dinamik, yenilikçi ve tamamen yetkin işgücüne dayanan yazılım endüstrisinde güncel bilgi, beceri ve teknolojik araçlara olan eğilimler büyük bir hızla değişmekte ve gelişmektedir (Gürcan, 2019; Gürcan ve Köse, 2017). Bu nedenle, dinamik ve rekabetçi iş ortamlarına sahip yazılım endüstrisi içerisinde önemli bir role sahip olan bilgisayar bilimciler ve yazılım geliştiriciler bilgi, beceri ve farklı uzmanlık alanlarına ilişkin yetkinliklerini sürekli olarak güncel tutmak zorundadırlar (Gürcan ve Şevik, 2019). Yazılım endüstrisindeki güncel temalar ve teknolojik trendlere ilişkin önemli bir bilgi kaynağı olarak görülen StackOverflow son zamanlarda yazılım geliştiriciler tarafından oldukça yoğun ilgi gören bir bilgi paylaşımı ve çevrimiçi öğrenme platformuna dönüşmüştür

(Barua vd., 2016). Yazılım geliştiriciler StackOverflow üzerinde kendi aralarında çok çeşitli teknik konuları tartışabilir, bilgi paylaşabilir ve böylelikle alana özgü güncel trendleri yakından takip edebilirler. Bu kişilerin, üzerinde tartıştıkları konuların anlaşılması, programlama dilleri ve yazılım araçlarına ilişkin olarak yazılım geliştiricilerin tercihlerinin ve eğilimlerinin belirlemesi ve bilgi içeriklerine dair kullanım alışkanlıklarının ortaya koyulabilmesi amacıyla bu platformda paylaşılan bilgilerin analiz edilmesi önemli çıkarımlar sağlayabilir. Nitekim bu ortamdaki paylaşılan verilerin analiz edilmesi, gerek öğrenci ve öğretmenler, gerek başlangıç düzeyindeki geliştiriciler, gerekse de deneyimli geliştiriciler gibi farklı bilgi düzeyine sahip kullanıcıların karşılaştıkları teknik problemlerle ilgili de ilginç bilgiler ortaya çıkarabilmektedir.

SvC platformları üzerindeki verilerle yapılacak olan analizlerin önemi ortada olmakla birlikte, bu platform üzerindeki verilerin büyüklüğü, manuel analiz yapmayı zorlaştırmaktadır (Johri ve Bansal, 2018; Yang vd., 2016). Bu bağlamda bu tür analizler için doğal dil işleme ve metinsel veri madenciliği tekniklerinden faydalanılabilir. Nitekim, bir çok uygulama alanında başarıyla uygulanmış olan bir dizi metin madenciliği ve görselleştirme yöntemleri, bu tür SvC platformlarının içerdiği verileri analiz etme sürecinde de etkin olarak kullanıldığı rahatlıkla söylenebilir (Aggarwal ve Zhai, 2012; Joorabchi vd., 2016). Çevrimiçi etkileşime dayalı bu tür platformlardaki kullanıcı paylaşımlarından oluşan büyük hacimli verilerin analizinde, kelime frekans analizi, metin sınıflandırma yaklaşımları, konu modelleme analizi, kümeleme analizi, metin özetleme, duygu analizi, n-gram analizi gibi metin analizi yöntemleri yoğun olarak kullanılmaktadır (Aggarwal ve Zhai, 2012). Özellikle kelime frekans analizine dayalı metinsel içerik analizleri metin madenciliği ve doğal dil işleme uygulamalarında temel bir yöntem bilim olarak görülmektedir (Lijffijt, Papapetrou, Puolamäki ve Mannila, 2011).

1.1. İlişkili çalışmalar

Literatürde StackOverflow üzerinde paylaşılan verilerin analizini konu alan farklı kapsamlardaki çalışmalardan söz etmek mümkündür. Nitekim StackOverflow verilerini kullanarak uzman davranışlarını tanımlamak, geliştiricilerde davranışsal temaları tespit etmek, etiketleri analiz etmek ve tavsiye etmek, katılımı oyunlaştırmak ve soruların durumunu tahmin etmek gibi farklı çalışmaların yapıldığı bilinmektedir (Ahmed ve

Srivastava, 2017; Cavusoglu ve Huang, 2015; Correa ve Sureka, 2013; Murgia vd., 2014; Nasehi vd., 2012; Pal vd., 2012; Xia vd., 2013). Bunların yanında doğrudan yazılım geliştiriciler ve programcılar için teknik tartışmalar üzerine yapılan metinsel içerik analizlerine dayalı çalışmalardan da söz edilebilir.

Barua vd. (2014) çalışmalarında, yazılım geliştiricilerin teknik paylaşımları ve tartışmalarına dair ana temaları ve eğilimleri otomatik olarak keşfetmek için istatistiksel konu modelleme tekniğine dayalı bir yöntem bilim kullanmışlardır. Çalışmanın sonuçları göstermiştir ki, geliştiricilerin ilgilendiği konular, işlerden sürüm kontrol sistemlerine ve C # sözdizimine kadar geniş bir yelpazede yer almaktadır; bazı konulardaki sorular diğer konulardaki tartışmalara yol açar; ve zaman içinde en popüler hale gelen konular web geliştirme (özellikle jQuery), mobil uygulamalar (özellikle Android), Git ve MySQL'dir (Barua vd., 2014). Johri ve Bansal (2018) yaptıkları çalışmada, StackOverflow platformu üzerinde paylaşılan metin verileri üzerinde konu modelleme yaklaşımını kullanılarak bir analiz gerçekleştirmişlerdir. Çalışmanın sonuçları, bu bağlamda gerçekleştirilen içerik analizlerinin geliştirici tartışmalarında trend konuların keşfedilmesine yardımcı olduğunu göstermiştir. Sonuçlar, web sitesi tasarımı/css'nin en etkili konu olduğunu göstermiştir. Veri analizi/görselleştirme ve mobil uygulama geliştirme gündemdeki konulardır ve popülerliği artmaktadır, nesneye dayalı programlama ve kodlama stili/uygulaması zamanla azalmıştır. Öte yandan, kimlik doğrulama/güvenlik ve UI geliştirme gibi konular zaman içinde istikrarlı eğilimler göstermiştir. Ayrıca, veri analizi/görselleştirme konusunda R ve Python hâkim, Oracle ve MySQL en popüler veri tabanı platformu, Python en etkili kodlama dili olarak ortaya çıkmıştır (Johri ve Bansal, 2018). Gürcan ve Özyurt (2019) çalışmalarında, Türk yazılım endüstrisindeki iş ilanlarını analiz ederek programlama dilleri ve araçları için ortaya çıkan eğilimleri tespit etmeyi amaçlamışlardır. Çalışmanın sonuçlarına göre, Türk yazılım endüstrisinde en çok talep gören ilk beş programlama dili sırasıyla Javascript, Html, Css, Java ve Php olarak ortaya çıkmıştır. En çok talep gören veri tabanları ise SQL, MySQL, Oracle, MsSQL ve MongoDB olarak elde edilmiştir. Benzer bir çalışmada ise Gürcan ve Şevik (2019) farklı ülkeleri kapsayan iş ilanlarını analiz ederek yazılım endüstrisindeki yurtdışı odaklı eğilimleri ve teknik becerileri ortaya koyan bir çalışma yürütmüşlerdir. Bu çalışmanın sonuçları yazılım endüstrisinde ihtiyaç duyulan güncel iş unvanlarını

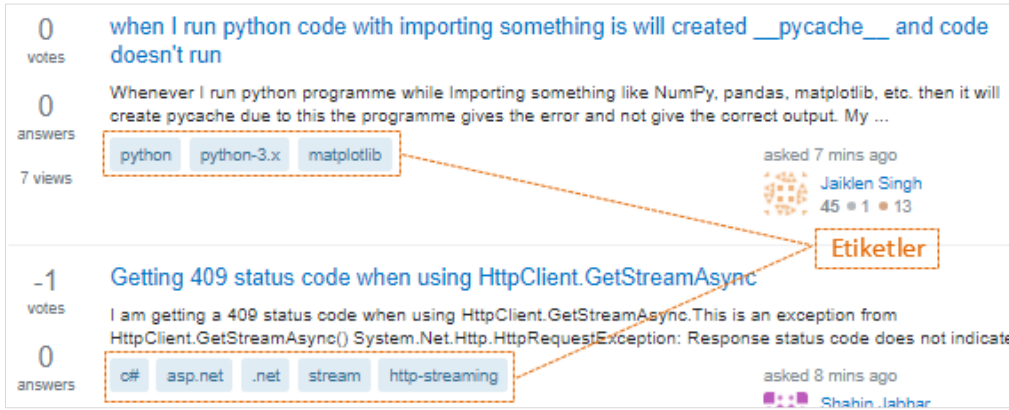
ve bunlara ait bilgi ve becerileri ortaya koymuştur. Çalışmanın sonuçlarına göre, yazılım geliştirme araçları içerisinde Python, Java, Javascript, C# ve C++ programlama dillerinin baskın bir şekilde ortaya çıktığı söylenebilir (Gürcan ve Şevik, 2019). Yang vd. (2016), StackOverflow üzerindeki güvenlikle ilgili sorularla büyük ölçekli bir çalışma yürütmüşlerdir. Metinlerine dayalı olarak güvenlikle ilgili farklı soruları kümelemek için konu modelleme yöntemini kullanmışlardır. Çalışmanın sonuçlarına göre güvenlikle ilgili şu beş ana kategori ortaya çıkmıştır: web güvenliği, mobil güvenlik, şifreleme, yazılım güvenliği ve sistem güvenliği. Rosen ve Shihab (2016) çalışmalarında mobil geliştiricilerin StackOverflow üzerindeki mobil uygulamalar ile ilgili paylaşımlarını özetlemek için konu modelleme yöntemine dayalı bir içerik analizi yapmışlardır. Çalışmanın sonuçları göstermiştir ki; geliştiriciler yoğun olarak uygulama dağıtımı, mobil API'ler, veri yönetimi, sensörler ve bağlam, mobil araçlar ve kullanıcı ara yüzü geliştirme konularında gönderi paylaşmışlardır. Başka bir çalışmada ise, Zhang ve Zhu (2018) konu modelleme yaklaşımını kullanarak StackOverflow üzerinde bir içerik analizi yürütmüşlerdir. Çalışmanın sonuçlarına göre en yoğun paylaşımın yapıldığı konu olarak Java, kullanıcılardan gelen sorularda en sık geçen favori kelimeler olarak da JavaScript ve C programlama dilleri ön plana çıkmıştır.

Sürekli değişen ve gelişen yazılım teknolojilerine ilişkin güncel eğilimlerin takip edilmesi ve bu bağlamda farklı çalışmaların yapılması alana özgü değişim ve gelişimin betimlenmesi açısından önemlidir. Bu açıdan StackOverflow üzerinde yapılan tartışmalardaki konular yazılım teknolojileri ve programlama dillerine ilişkin güncel teknolojileri ve eğilimleri belirlemede yardımcı olabilir. Nitekim StackOverflow bilişim alanında uluslararası platformda en çok tercih edilen SvC platformu olarak kabul edilmektedir (Zhang ve Zhu, 2018; Gürcan ve Köse, 2017). Bu platform üzerinde yürütülen paylaşımlar ve bu paylaşımlara ilişkin gönderilerde kullanılan etiketlerin analizi, yazılım teknolojileriyle ilgili güncel eğilimler hakkında önemli çıkarımlar ve öngörüler sağlayabilir. Benzer bir bakış açısıyla yürütülen bu çalışmada StackOverflow üzerinde 2019 yılını kapsayan paylaşımlar dikkate alınarak, bu gönderiler üzerinde kelime frekanslarının dağılımına dayalı metinsel içerik analizi ile yazılım endüstrisindeki güncel tema ve eğilimlerin belirlenmesi amaçlanmıştır.

2. Yöntem

Bu çalışma, StackOverflow'daki gönderilere kullanıcı tarafından atanan ve gönderinin ilgili olduğu konuları yansıtan etiketleri kelime frekans analizi yöntemiyle analiz ederek yazılım geliştiricilerin en çok yoğunlaştığı tartışma konularını ortaya koymayı amaçlamaktadır. StackOverflow, kullanıcıların yeni sorular sorma ve mevcut soruları yanıtlama, ayrıca gönderinin algılanan değerine göre soruları ve yanıtları yukarı veya aşağı "oylama" becerisine sahiptir. StackOverflow gönderilerinde soruyu soran kullanıcı sorduğu soruyla ilgili etiketleri listeden

seçerek soruya ekler ve soru bu etiketlerle ilgili kategorilere atanır. Bu şekilde soruyu paylaşan kullanıcı bir soruya en fazla beş etiket ekleyebilir. Böylelikle sistemde önceden tanımlanmış ve kullanıcı tarafından soruya eklenen bu etiketler sayesinde sisteme eklenen her bir sorunun ilgili olduğu ana tartışma kategorileri belirlenir. Örnek olarak, Şekil 1'de iki adet gönderinin ekran görüntüsü verilmiştir. Şekil 1'de görüldüğü gibi ilk sorunun etiketleri soruyu soran kullanıcı tarafından "Python", "Python-3.x", "Matplotlib" olarak atanmıştır. Şekildeki ikinci sorunun etiketleri ise "C#", "Asp.net", ".Net", "Stream", "Http-streaming" olarak atanmıştır.



Şekil 1. StackOverflow web sitesinde paylaşılan gönderilere ilişkin örnek görünüm

Bu çalışmada StackOverflow üzerinde 1 Ocak 2019 – 31 Aralık 2019 tarihleri arasında paylaşılan gönderilerin oluşturduğu bir yıllık veri çekilmiş ve bu verilerle deneysel veri seti oluşturulmuştur. Veri çekim işlemi Mayıs 2020'de StackExchange tarafından sağlanan API (Stack Exchange API, 2020) yardımıyla gerçekleştirilmiştir. Bu kapsamda, ilgili API her bir ay için ayrı ayrı çalıştırılmış ve bunun sonucunda 2019 yılına ait gönderileri içeren 12 aylık bir veri seti oluşturulmuştur. Veri seti oluşturulduktan sonra veriler içerisinde sadece her bir gönderiye ait etiketlerin yer aldığı ayrı bir veri seti oluşturulmuştur. Oluşturulan bu veri seti 42727 tekil etiket içermekte olup toplamda ise 5545951 etiket içermektedir. Daha sonra, etiketlerle oluşturulan bu veri seti üzerinde etiketlerin aylık frekansları hesaplanarak kelime frekans analizi yöntemine dayalı bir metinsel içerik analizi gerçekleştirilmiştir (Gürcan, 2009). Kelime frekans analizi veri madenciliğinde kullanılan, önemli yöntemlerden birisidir (Gürcan, 2018; Lijffijt vd., 2011). Herhangi bir belgedeki kelime frekanslarının belirlenmesi, metnin ana temalarının ve metnin anlamsal olarak içerdiği kelime öbeklerinin ortaya konulması bakımından çoğu

zaman güçlü çıkarımların elde edilmesine olanak sağlar (Gürcan, 2009; Rajput, Ahuja, ve Riyal, 2019).

3. Bulgular

Bu bölümde mevcut çalışma kapsamında elde edilen bulgulara yer verilmiştir. Kelime frekans analizine göre, StackOverflow üzerinde 2019 yılındaki gönderilerden elde edilen verilere ilişkin detaylı bilgiler Tablo 1'de verilmiştir. Tablo 1'de görüldüğü gibi StackOverflow üzerinde 2019 yılındaki gönderilerde toplam 5545951 etiket kullanılmış olup bu etiketler içerisindeki tekil (tekrar etmeyen) etiket sayısı 42727 olarak hesaplanmıştır. En çok kullanılan ilk 20 etiketin toplam etiket sayısına oranı %28.22 olarak elde edilmişken en çok kullanılan ilk 50 etiketin toplam etiket sayısına oranı ise %39.24 olarak elde edilmiştir. Diğer bir ifade ile 42727 tekil etiketten ilk 20 tanesi toplam etiket sayısının yaklaşık %30'unu temsil ederken, ilk 50 tanesi toplam etiket sayısının yaklaşık %40'ını ve ilk 100 tanesi ise toplam etiket sayısının yaklaşık %50'sini temsil etmektedir.

Tablo 1. StackOverflow gönderilerinden elde edilen 2019 yılı etiket verilerine ilişkin bilgiler

Kategoriler	Veriler
Tekil etiket sayısı	42727
Toplam etiket sayısı	5545951
Aylık ortalama etiket sayısı	462163
En çok kullanılan ilk 20 etiketin toplam sayısı	1564823
En çok kullanılan ilk 20 etiketin toplam etiket sayısına oranı	%28.22
En çok kullanılan ilk 50 etiketin toplam sayısı	2175966
En çok kullanılan ilk 50 etiketin toplam etiket sayısına oranı	%39.24
En çok kullanılan ilk 100 etiketin toplam sayısı	2628544
En çok kullanılan ilk 100 etiketin toplam etiket sayısına oranı	%47.40
Sadece 1 defa kullanılmış tekil etiket sayısı	6739
En fazla 5 defa kullanılmış tekil etiket sayısı	18852
İvmesi artan tekil etiket sayısı	2837
İvmesi azalan tekil etiket sayısı	1513
En fazla kullanılan ilk üç etiket (Sıralı)	Python, Javascript, Java
İvmesi en çok artan ilk üç etiket (Sıralı)	Python, Reactjs, Flutter
İvmesi en çok azalan ilk üç etiket (Sıralı)	Php, C#, JQuery

Tablo 2 ve tablo 3'te sırasıyla toplam etiket ve tekil etiket sayılarının aylara göre dağılımları verilmiştir. Tablo 2'de görüldüğü gibi aylık etiket sayıları genel olarak dengeli ve ortalama değere (462163) yakın olmakla birlikte Ekim ve Kasım aylarında bir miktar artış olduğu söylenebilir.

Tablo 2. Toplam etiket sayısının aylara göre dağılımı

Aylar	N
Ocak	422795
Şubat	409909
Mart	465928
Nisan	430527
Mayıs	478853
Haziran	455790
Temmuz	495907
Ağustos	454064
Eylül	451992
Ekim	515057
Kasım	501580
Aralık	463549
Toplam	5545951

Tablo 3'te görüldüğü gibi toplam 42727 olan farklı etiket sayısının her bir aydaki dağılımı birbirine yakın değerler olarak elde edilmiştir. Örneğin 42727 tekil etiketten Ocak ayında kullanılan tekil etiket sayısı 20510 iken Şubat ayında bu sayı 20358, Aralık ayında ise 22190 olarak elde edilmiştir.

Tablo 4'te, StackOverflow üzerinde 2019 yılı boyunca paylaşılan gönderilerde en çok kullanılan ilk 50 etiketin aylara göre dağılımı ve bu etiketlerin toplam sayısı sıralı bir biçimde verilmiştir.

Tablo 3. Tekil etiket sayısının aylara göre dağılımı

Aylar	N
Ocak	20510
Şubat	20358
Mart	21446
Nisan	20421
Mayıs	21899
Haziran	21749
Temmuz	22439
Ağustos	21771
Eylül	21710
Ekim	22810
Kasım	22811
Aralık	22190

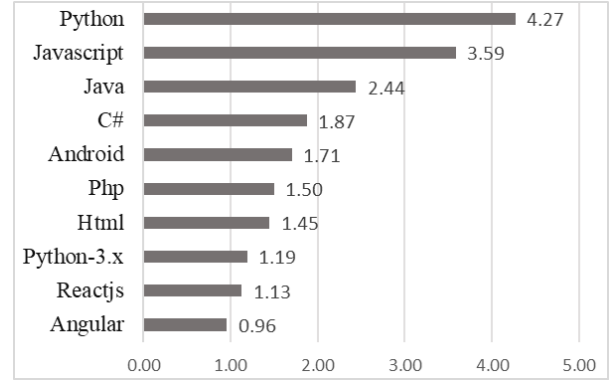
Tablo 4'te görüldüğü gibi, gönderilerde en çok geçen etiket Python (n=236569) iken bunu Javascript ile Java takip etmektedir. Gönderilerde en çok kullanılan etiket olan Python'un (n=236569) toplam etiket sayısına (n=5545951) oranı da %4.27 olarak elde edilmiştir. Her bir etiketin artış/azalış oranı (ivmesi) ise bu tablonun son sütununda verilmiştir. İvme hesaplanırken aylık artış/azalış miktarı dikkate alınarak 12 aylık ortalama değeri hesaplanmıştır. Her bir etiket için ilgili ayın toplam sayısından bir önceki ayın toplam sayısı çıkarılarak toplamda 11 adet ivme değeri hesaplanmıştır. Daha sonra, bu 11 değer ortalama alınarak her bir etiket için ortalama bir ivme değeri hesaplanmıştır. Pozitif ivme değerleri o etikete olan ilginin arttığını, negatif ivme değerleri ise o etikete olan ilginin zamanla azaldığını gösterir. Buna göre, gönderilerde en çok kullanılan Python etiketi için ivme %4.53 olarak elde edilmiştir.

Tablo 4. Gönderilerde en çok geçen ilk 50 etikete ilişkin veriler

Etiketler	Ocak	Şubat	Mart	Nisan	Mayıs	Haziran	Temmuz	Ağustos	Eylül	Ekim	Kasım	Aralık	Toplam	Toplam (%)	İvme (%)
Python	16436	16826	19686	19292	20206	19355	21388	19706	18441	22624	22478	20131	236569	4.27	4.53
Javascript	15553	15040	17126	15773	17172	16494	17663	16702	16233	17806	17327	16043	198932	3.59	0.6
Java	10761	10649	11958	10943	12089	10999	12055	10256	10899	12716	11590	10387	135302	2.44	-0.46
C#	8519	8036	8894	8914	9577	8433	9163	8334	8500	9145	8634	7809	103958	1.87	-0.87
Android	7338	7027	7822	7063	8063	8292	9437	7929	7812	8503	7939	7396	94621	1.71	0.07
Php	7060	7125	7787	7147	7652	6816	7349	6368	6213	6816	6717	5961	83011	1.50	-1.35
Html	6404	6183	7270	6556	6945	6556	7339	6434	6222	6978	6820	6490	80197	1.45	0.11
Python-3.x	4562	4368	5099	5524	5644	5433	6387	5963	5532	6422	5810	5219	65963	1.19	0.81
Reactjs	4327	4314	4979	4525	5321	5042	5629	5397	5422	6060	5983	5780	62779	1.13	1.78
Angular	4229	4200	4699	4155	4653	4532	4694	4237	4363	4721	4700	4119	53302	0.96	-0.13
Node.js	3974	3742	4240	4121	4748	4613	4947	4512	4415	4790	4748	4389	53239	0.96	0.51
R	3802	4060	4456	4486	4647	4473	4631	4377	4292	4968	4628	4159	52979	0.96	0.44
Css	4328	4153	4709	4174	4421	4110	4607	4331	4119	4509	4225	3995	51681	0.93	-0.41
C++	3782	3619	4369	4324	4281	3883	4452	3411	3979	5138	4933	4505	50676	0.91	0.89
Sql	3897	4011	4575	4151	4130	3727	4179	3675	3996	4500	4438	4042	49321	0.89	0.18
Mysql	3378	3354	3886	3455	3552	3079	3443	3014	2971	3483	3553	3059	40227	0.73	-0.39
Jquery	3488	3276	3687	3448	3570	3273	3717	3186	3169	3317	3136	2927	40194	0.72	-0.69
Pandas	2712	2822	3217	3116	3256	3152	3586	3561	3161	3942	3703	3527	39755	0.72	1.00
Swift	2844	2605	2905	3027	3321	3463	3400	2818	2760	3337	3124	2940	36544	0.66	0.12
IOS	2855	2608	2867	2758	3115	3021	3148	2725	2896	3544	3167	2869	35573	0.64	0.02
Arrays	2376	2483	2874	2495	2660	2334	2477	2191	2427	2934	2972	2686	30909	0.56	0.38
Django	2176	2186	2516	2458	2597	2426	2743	2624	2578	2686	2821	2628	30439	0.55	0.55
Typescript	2245	2299	2400	2370	2565	2532	2711	2482	2499	2710	2844	2612	30269	0.55	0.45
Json	2138	2213	2615	2380	2530	2457	2619	2333	2233	2538	2576	2292	28924	0.52	0.19
Laravel	2213	2051	2333	2250	2402	2273	2590	2431	2564	2672	2621	2422	28822	0.52	0.26
Excel	2405	2184	2559	2132	2638	2586	2757	2423	2305	2404	2374	1996	28763	0.52	-0.5
Sql-server	2231	2264	2350	2290	2278	2095	2222	2119	2042	2219	2222	1898	26230	0.47	-0.41
C	2220	1999	2344	2411	2185	1804	1787	1575	2056	2572	2552	2258	25763	0.46	0.05
Spring-boot	1682	1628	1863	1679	1971	1881	2117	1974	1938	2250	2103	1875	22961	0.41	0.24
React-native	1692	1628	1724	1620	1755	1889	2120	2110	1955	2194	2087	1897	22671	0.41	0.25
Vba	1869	1672	2030	1741	2036	2055	2188	1940	1763	1803	1701	1398	22196	0.40	-0.58
Flutter	1292	1235	1391	1369	1692	1615	1932	2050	2053	2095	2016	2359	21099	0.38	1.31
Vue.js	1389	1319	1595	1438	1548	1658	1731	1777	1749	1921	1708	1787	19620	0.35	0.49
Firebase	1636	1549	1692	1541	1670	1599	1703	1657	1533	1696	1687	1538	19501	0.35	-0.12
Docker	1521	1310	1608	1412	1733	1595	1655	1597	1694	1840	1669	1532	19166	0.35	0.01
Spring	1440	1414	1619	1369	1585	1498	1708	1516	1533	1692	1650	1501	18525	0.33	0.07
Regex	1471	1403	1697	1579	1687	1544	1598	1524	1454	1646	1557	1346	18506	0.33	-0.15
AWS	1405	1309	1510	1276	1572	1476	1680	1620	1480	1736	1649	1612	18325	0.33	0.25
Wordpress	1298	1254	1410	1161	1597	1523	1626	1502	1234	1464	1484	1404	16957	0.31	0.13
Postgresql	1172	1109	1395	1320	1393	1256	1500	1325	1357	1557	1492	1431	16307	0.29	0.32
Mongodb	1180	1163	1212	1141	1416	1447	1474	1395	1382	1474	1448	1480	16212	0.29	0.37
Dataframe	1141	1142	1283	1210	1272	1274	1464	1336	1266	1698	1591	1452	16129	0.29	0.38
Azure	1170	1200	1275	1201	1339	1368	1388	1224	1306	1444	1333	1189	15437	0.28	0.02
Tensorflow	1018	962	1258	1220	1344	1330	1473	1298	1177	1416	1426	1399	15321	0.28	0.47
.Net	1191	1212	1195	1101	1193	1146	1269	1088	1239	1291	1290	1327	14542	0.26	0.17
Kotlin	1027	939	1076	1015	1160	1237	1394	1237	1318	1408	1378	1251	14440	0.26	0.27
Ruby-on-rails	1323	1145	1252	1056	1127	1084	1106	1082	1082	1124	1058	1049	13488	0.24	-0.34
Asp.net-core	975	828	1046	1001	1213	1062	1203	1126	1211	1292	1256	1193	13406	0.24	0.27
Asp.net	1076	1026	1151	1042	1247	1144	1160	1041	1054	1120	1098	1094	13253	0.24	0.02
Numpy	882	1031	1186	1100	1047	968	1080	1078	980	1257	1233	1120	12962	0.23	0.29

Tablo 4'e ek olarak, etiketlerin en çok kullanılması sayısına göre ilk 10 sıraya yerleşen etiketler yüzdelik oranları ile birlikte Şekil 2'de grafik olarak verilmiştir. Şekil 2'ye göre ilk dört sırayı Python, Javascript, Java ve C# programlama dilleri paylaşmaktadır. Beşinci sırada ise bir mobil işletim sistemi ve mobil yazılım geliştirme platformu olan Android yer almaktadır. Bunu altıncı sırada bir web programlama dili olan Php ve yedinci sırada betik bir dil olan Html takip etmektedir. Sekizinci sırada Python programlama dilinin son sürümü olan Python-3.x yer almaktadır. İlk 10'un son sıralarında ise Javascript kütüphaneleri olarak yaygın bir kullanıma sahip olan Reactjs ve Angular yer almaktadır.

En çok kullanılan ilk 50 etiketin hangi alanlarla ilişkili olduğuna dair bilgiler ile bu etiketlerin en çok kullanılan ilk 50 ve toplam etiket sayısına oranı Tablo 5'te verilmiştir. Tablo 5'te görüldüğü gibi en çok kullanılan ilk 50 etiket altı kategori altında sınıflandırılmıştır. En fazla etiket "Web programlama dili/teknolojiler/araçlar" başlığı altında toplanmışken (n=19), en az etiket ise "Bulut bilişim" kategorisi altında toplanmıştır (n=2).



Şekil 2. Gönderilerde en çok geçen ilk 10 etiketin grafiksel gösterimi

Bunlara ek olarak en çok kullanılan ilk 50 etiketin toplam etiket sayısına oranı %39.24 olarak hesaplanmıştır. Bu oranın alt kategorilere dağılımı da % (Toplam) sütunu olarak verilmiş olup ilk sırayı %15.43 ile "Web programlama dili/teknolojiler/araçlar" kategorisi altındaki etiketlerin oranlarının toplamının oluşturduğu görülmüştür.

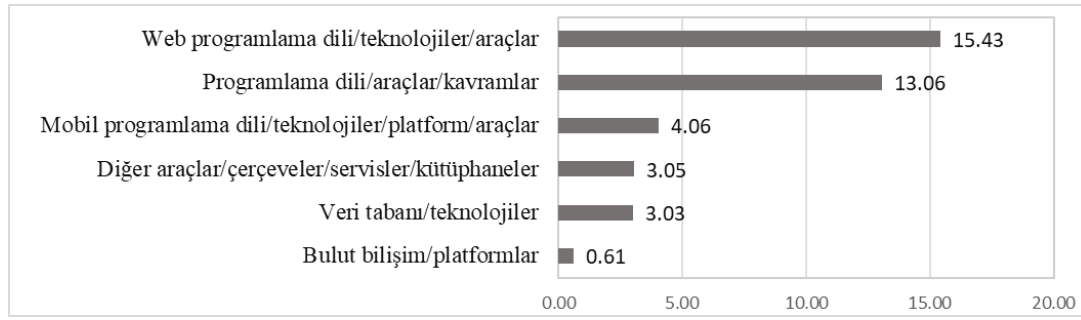
Tablo 5. Gönderilerde en çok kullanılan 50 etiketin kategorizasyonu

Kategoriler	Etiketler	N	%(Toplam)
Programlama dili/araçlar/kavramlar	Python, Java, C#, Python-3x, R, C++, Arrays, C, Vba	9	13.06
Web programlama dili/teknolojiler/araçlar	Javascript, Php, Html, Reactjs, Angular, Node.js, Css, Jquery, Django, Typescript, Json, Laravel, Spring-boot, Vue.js, Wordpress, .Net, Ruby-on-rails, Asp.net-core, Asp.net	19	15.43
Mobil programlama dili/teknolojiler/platform/araçlar	Android, Swift, IOS, React-native, Flutter, Kotlin	6	4.06
Veri tabanı/teknolojiler	Sql, MySql, Sql-server, Firebase, Postgresql, MongoDB	6	3.03
Bulut bilişim/platformlar	AWS, Azure	2	0.61
Diğer araçlar/çerçeveler/servisler/kütüphaneler	Excel, Docker, Spring, Pandas, Dataframe, Tensorflow, Numpy, Regex	8	3.05

Şekil 3'te de, gönderilerde en çok kullanılan ilk 50 etiketin dahil olduğu sınıfların yüzdelikleri sıralı bir biçimde verilmiştir. Şekil 3'te görüldüğü gibi, ilk sırayı %15.43 ile Web programlama dili/teknolojiler/araçlar sınıfı oluşturmaktadır. Bunu %13.06 ile Programlama dili/araçlar/kavramlar sınıfı takip ederken, son sırada ise Bulut bilişim/platformlar sınıfı yer almaktadır.

Veri seti üzerinde en fazla kullanılan etiketlerin analizinden sonra, 2019 yılı içerisinde etiketlere

ilişkin eğilimlerdeki değişimlerin belirlenmesi amacıyla etiketlerin ivmeleri incelenmiştir. Bu kapsamda, tüm veri setinde yer alan 42727 tekil etiket içerisinde ivmesi en çok artan ve azalan ilk 50 etiket Tablo 6'da sıralı bir biçimde verilmiştir. Tablo 6'da görüldüğü gibi ivmesi en çok artan ilk üç etiket sırasıyla Python, Reactjs ve Flutter olarak ortaya çıkmıştır. Benzer şekilde, ivmesi en çok azalan ilk üç etiket ise sırasıyla Php, C# ve Jquery etiketleridir.



Şekil 3. Gönderilerde en çok geçen 50 etiket kategorizasyonun grafiksel gösterimi

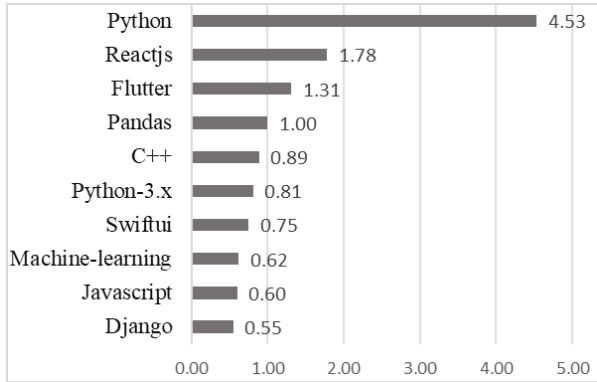
Tablo 6. İvmeleri en çok artan ve azalan ilk 50 etikete ilişkin bilgiler

Etiketler	İvme (%)	Etiketler	İvme(%)
Python	4.53	Php	-1.35
Reactjs	1.78	C#	-0.87
Flutter	1.31	Jquery	-0.69
Pandas	1	Vba	-0.58
C++	0.89	Excel	-0.5
Python-3.x	0.81	Java	-0.46
Swiftui	0.75	Angular6	-0.44
Machine-learning	0.62	Css	-0.41
Javascript	0.6	Sql-server	-0.41
Django	0.55	Angularjs	-0.4
Node.js	0.51	Mysql	-0.39
Vue.js	0.49	Python-2.7	-0.36
Tensorflow	0.47	Ruby-on-rails	-0.34
List	0.46	Ionic3	-0.27
React-hooks	0.45	Angular7	-0.24
Typescript	0.45	Android-studio	-0.23
Angular8	0.45	Laravel-5.7	-0.22
R	0.44	Visual-studio-2017	-0.22
Dart	0.39	Laravel-5	-0.2
Dataframe	0.38	Ruby	-0.19
Arrays	0.38	Ecmascript-6	-0.17
Mongodb	0.37	Twitter-bootstrap	-0.16
Xcode	0.36	Scala	-0.15
Flask	0.35	Regex	-0.15
Visual-studio-code	0.32	Mysqli	-0.15
Postgresql	0.32	Git	-0.14
.net-core	0.31	Login	-0.14
Visual-studio-2019	0.29	Xamarin	-0.13
Numpy	0.29	Bootstrap-4	-0.13
Api	0.28	Android-gradle-plugin	-0.13
Kotlin	0.27	Tsql	-0.13
Asp.net-core	0.27	Angular	-0.13
Tensorflow2.0	0.27	Firebase-realtime-database	-0.13
Authentication	0.26	Symfony	-0.12
Laravel	0.26	Firebase	-0.12
Amazon-web-services	0.25	Java-8	-0.12
React-native	0.25	Perl	-0.12
Deep-learning	0.24	Wpf	-0.11
Function	0.24	Vuejs2	-0.11
Spring-boot	0.24	Xaml	-0.11
Google-sheets-formula	0.22	Ionic-framework	-0.11
Google-sheets	0.22	Exception-handling	-0.11
Oracle	0.21	Ajax	-0.10
Keras	0.21	Asp.net-mvc	-0.09

Tablo 6'nın devamı

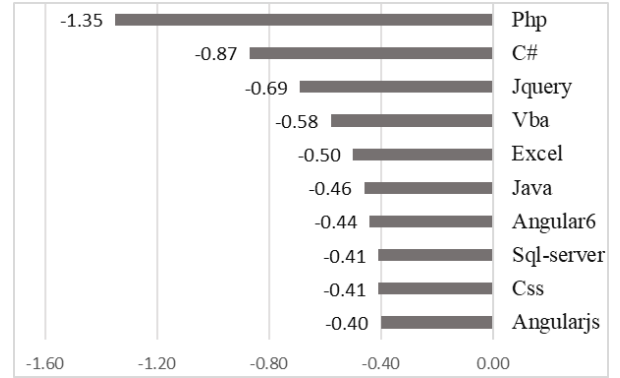
Pyspark	0.21	Xamarin.forms	-0.09
Loops	0.21	Delphi	-0.09
Express	0.19	Asp.net-core-2.1	-0.09
Pytorch	0.19	Angular5	-0.09
Scikit-learn	0.19	Bash	-0.09
Json	0.19	Ms-access	-0.08

Tablo 6'ya ek olarak, ivmesi en çok artan ilk 10 etiketın grafiksel gösterimi Şekil 4'te verilmiştir. Buna göre ivmesi en çok artan etiketler içerisinde Python programlama dili birinci, bir Javascript kütüphanesi olan Reactjs ikinci ve bir mobil uygulama geliştirme ara yüzü olan Flutter üçüncü sırada yer almaktadır. Üst düzey veri işleme aracı olarak kullanılan ve bir Python kütüphanesi olan Pandas ise dördüncü sırada görülmektedir. Beşinci sırada C++ programlama dili, altıncı sırada Python programlama dilinin son sürümü Python-3.x ve yedinci sırada ise Apple platformları için kullanıcı ara yüzleri oluşturmaya yönelik yeni bir araç olan Swiftui yer almaktadır. Bunu bir veri madenciliği tekniği olan ve makine öğrenmesi olarak adlandırılan Machine-learning etiketi takip etmektedir. Listenin dokuzuncu sırasında Javascript programlama dili ve son sırada ise Python için bir web uygulama geliştirme aracı olarak kullanılan Django yer almaktadır.



Şekil 4. İvmesi en çok artan ilk 10 etiketın grafiksel gösterimi

Benzer bir şekilde, ivmesi en çok azalan ilk 10 etiketın grafiksel gösterimi ise Şekil 5'te verilmiştir. Şekil 5'e göre ivmesi en çok azalan etiketler sıralamasında Php birinci, C# programlama dili ikinci, bir Javascript kütüphanesi olan JQuery üçüncü, Visual Basic programlama dilini temsil eden Vba dördüncü ve ofis paketinin bir bileşenin olan Excel ise beşinci sırada yer almaktadır.



Şekil 5. İvmesi en çok azalan ilk 10 etiketın grafiksel gösterimi

4. Sonuç ve tartışma

Bu çalışmada, StackOverflow üzerindeki 2019 yılına ait gönderilerdeki etiketlerden oluşturulan veri seti üzerinde kelime frekanslarına dayalı metinsel içerik analizi yapılmış ve bu analizle yazılım endüstrisine ilişkin güncel tema ve eğilimlerin belirlenmesi amaçlanmıştır. Veri seti içerisinde en çok kullanılan ilk 50 etiketın toplam tekil etiketlerin (n=42727) yaklaşık %40'ını temsil ettiği görülmüştür. Bu sonuca göre ilk 50 etiketın veri setini örneklediği söylenebilir. Dolayısıyla ilk 50 etiket üzerinden yürütülecek sonuç ve tartışmalar tüm veri setini betimlediği düşünülebilir. Çalışmanın sonuçlarının başta yazılım geliştiriciler ve bu alanda eğitim alan ve kariyerini bu alanda planlayan bireyler olmak üzere, müfredat yapımcılar ve karar vericilere önemli çıkarımlar ve öngörüler sağlaması beklenmektedir.

StackOverflow'daki 2019 yılına ait gönderilerde kullanılan etiketler üzerinde yapılan bu içerik analizinde önemli örüntüler elde edilmiştir. Şöyle ki, son bir yılda en fazla kullanılan ve ivmesi en fazla artan etiketın Python olarak çıkmış olması, Python programlama dilinin geliştiriciler ve araştırmacılar tarafından en çok konuşulan konu olduğunu göstermektedir. Özellikle büyük verinin yazılım mimarilerine etkin bir şekilde katılmasıyla veri odaklı bir programlama dili olan Python, programlama dilleri içerisinde oldukça baskın bir konuma gelmiştir. Bu durum [Johri ve Bansal \(2018\)](#)'ın sonucu ile paralellik göstermiş olup

Python'un son yıllarda popülaritesini artırdığını gösterir niteliktedir. Nitekim ilk 10 etiket içerisinde yer alan Python, Python-3.x ve ile Python kütüphaneleri olan Pandas, Numpy ve Tensorflow gibi kütüphaneler de dikkate alındığında Python'un yazılım teknolojileri ve programlara dilleri içerisinde önemli bir yer tuttuğu söylenebilir. Nitekim Pandas, Numpy ve Tensorflow kütüphaneleri Python ile ilişkili olup ivmesi artan etiketler arasında yer almaktadırlar.

İlk 50 etiketin kategorik olarak analizi altı sınıfa ortaya çıkarmıştır. Bu sınıflar büyük çoğunlukla güncel programlama dilleri, araçlar ve teknolojilerden oluşmaktadır. Bu sınıflar içerisinde en fazla etiket içeren sınıfın "Web programlama dili/teknolojiler/araçlar" olduğu görülmüştür. Bu durum web teknolojilerinin popülaritesini koruduğu ve farklı teknolojilerin bu alanda kullanıldığının bir göstergesi olarak kabul edilebilir. Bu sonuçlar Barua vd. (2014) tarafından web teknolojilerinin öneminin vurgulandığı çalışma ile paralellik gösterse de ivmesi artan ve azalan etiketler dikkate alındığında önemli bir fark ortaya çıkmaktadır. Nitekim ivmesi artan etiketlere bakıldığında Reactjs, Javascript, Django, Node.js, Vue.js gibi yenilikçi web teknolojisi araçlarının popülaritesini artırdığı; Php, JQuery ve Css gibi nispeten daha eski teknolojilerin popülaritesinin düştüğü söylenebilir. Çalışmanın bu sonuçları, Johri ve Bansal (2018), çalışmalarında web tasarımı/Css'in popüler olduğu sonucuyla ters düşmekte, Css' nin popülaritesinin azaldığını ortaya koymaktadır.

Ortaya çıkan kategorilerden bir diğeri olan "Programlama dili/araçlar/kavramlar" sınıfı "Web programlama dili/teknolojiler/araçlar" sınıfının ardından içerdiği etiket bakımından ikinci sırada yer almıştır. Bu sınıf altında yer alan etiketlere ve ivmesi artan/azalan etiketlere bakıldığında daha önce de söylendiği gibi Python dilinin önemli bir yere sahip olduğu açık ve nettir. Her ne kadar ivmesi azalan etiketler içerisinde Python-2.7 olsa da bu durumun Python-2.7'nin eski sürüm olmasından kaynaklandığı düşünülmektedir. Nitekim ilk Python, Python-3.x etiketlerine ilişkin ivmelenmenin pozitif olduğu net bir şekilde görülmektedir. Diğer yandan bu sınıftaki etiketlerin ivmelerine bakıldığında Python'un yanı sıra C++'ın ivmesi artmışken, C#, Vba, Java gibi dillerin ivmesinin azaldığı söylenebilir. C++'ın hem en çok konuşulan konular içerisinde önde gelmesi hem de ivmesinin artması birçok alanda temel programlama dili olarak kabul edilmesinden kaynaklı olduğu düşünülmektedir. Bu sonuçlar, Gürcan ve Özyurt (2019)'un sonuçları ile kısmen

paralellik kısmen zıtlık taşımaktadır. Nitekim Gürcan ve Özyurt (2019) Java'nın en popüler dil olduğunu ortaya koymuşken, bu çalışmada Java en çok konuşulan konular içinde yer almakla birlikte ivmesinin azaldığı görülmüştür. Ayrıca elde edilen sonuçlara göre C#, Vba gibi geleneksel dillere olan ilginin düştüğü söylenebilir.

Diğer bir kategori olarak "Mobil programlama dili/teknolojiler/platform/araçlar" kategorisi dikkat çekmektedir. Bu kategori içerisinde "Android" etiketi çok kullanılan etiketlerin başında gelmektedir. Bunun yanı sıra bu kategori altında yer alan altı etiketten tamamının ivmesi pozitif olmakla birlikte "Flutter" en çok ivmelenen etiketler içinde üçüncü sırada gelmektedir. Bunlara ek olarak, en hızlı ivmelenen etiketlere bakıldığında içerisinde mobil teknolojilere ait çok sayıda etiketin olduğu görülmektedir. Bu durum genel olarak mobil programlara teknolojilerine olan ilginin arttığının bir göstergesi olarak kabul edilebilir. Barua vd. (2014) ile Johri ve Bansal (2018) çalışmalarında başta Android olmak üzere mobil ortamlarla ilgili popülaritenin arttığından bahsetmişlerdir. Bu sonuçlar desteklenmekle birlikte IOS ve hibrit ortamlarında popülaritesini artırdığı söylenebilir.

Veri tabanı teknolojileri açısından bakıldığında da önemli değişim göze çarpmaktadır. Şöyle ki, en çok artan etiketler içerisinde yer alan veri tabanı kavramlarından beş tanesi ilişkisel veri tabanı yönetim sistemleri ile ilgili sadece Mongoddb yenilikçi ilişkisel olmayan veri tabanı yönetim sistemleri türündendir. İvmelenmelere bakıldığında da Mongoddb'nin en yüksek ivmeye sahip olduğu görülmektedir. Bu durum yazılım teknolojilerindeki veri tabanı gelişimi ile paralellik göstermekte olup önemli bir husustur. Çalışmanın sonuçları bu bağlamda Gürcan ve Özyurt (2019)'un çalışması ile genel olarak paralellik taşımaktadır. Sadece Gürcan ve Özyurt (2019) çalışmasında Oracle öne çıkmışken bu çalışmada açık kaynak kodlu yeni nesil veri tabanı yönetim sistemleri olan Firebase ve Postgresql dikkat çekicidir. Bunun yanında Sql-server, Mysql gibi eski nesil veritabanı araçlarının negatif ivmeye sahip olması, veritabanı sistemlerinde yeni jenerasyon veri tabanlarına doğru önemli bir yönelime ve teknik dönüşüme vurgu yapmaktadır.

Çalışmanın kısıtları dikkate alındığında, bu çalışma StackOverflow üzerindeki 2019 yılına ait gönderilerde yer alan kullanıcı tanımlı etiketlere ilişkin veriler ile sınırlıdır. Gelecek çalışmalarda gerek yıl bazında gerekse etiketler bazında periyotlar oluşturulup yıl bazında ve etiket bazında daha derinlemesine analizler yapılması

planlanmaktadır. Ayrıca uygulanan yöntem açısından, kelime frekans analizine ek olarak olasılıksal konu modelleme yöntemleriyle soru ve cevaplar üzerinde anlamsal bir kümeleme analizi de uygulanarak farklı çalışmalar yürütülebilir. Bu çalışmalarla yazılım ve bilişim teknolojileri alanındaki trendlere ilişkin farklı perspektifler ortaya konulabilir.

Kaynaklar

- Aggarwal, C.C. and Zhai, C. (2012). Mining Text Data. New York: Springer Science & Business Media.
- Ahmed, T. and Srivastava, A. (2017). Understanding and evaluating the behavior of technical users. A study of developer interaction at StackOverflow. Human-centric Computing and Information Sciences, 7(1), 1-18. <https://doi.org/10.1186/s13673-017-0091-8>
- Bakir, C., Hakkoymaz, V, Diri, B. ve Güçlü, M. (2020). Dağıtık veritabanlarında saldırı önleme metotları. Gümüşhane Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 10(2), 425-441.
- Barua, A., Thomas, S.W. and Hassan, A. E. (2014). What are developers talking about? An analysis of topics and trends in stack overflow. Empirical Software Engineering, 19(3), 619-654. <https://doi.org/10.1007/s10664-012-9231-y>
- Cavusoglu, H., Li, Z. and Huang, K.W. (2015). Can gamification motivate voluntary contributions? The case of StackOverflow Q&A community. In Proceedings of the 18th ACM conference companion on computer supported cooperative work & social computing (pp. 171-174). New York.
- Correa, D. and Sureka, A. (2013). Fit or unfit: analysis and prediction of closed questions on stack overflow. In Proceedings of the first ACM conference on Online social networks (pp. 201-212). Boston.
- Gurcan, F. (2019). Extraction of core competencies for big data: Implications for competency-based engineering education. International Journal of Engineering Education, 35(4), 1110-1115.
- Gurcan, F. and Kose, C. (2017). Analysis of software engineering industry needs and trends: Implications for education. International Journal of Engineering Education, 33(4), 1361-1368.
- Gürcan, F. and Özyurt, Ö. (2019). Analysis of requirements for programming languages and tools in Turkish software industry. 2nd Turkish World Engineering and Science Congress (pp. 307-311). Antalya.
- Gürcan, F. and Şevik, S. (2019). Expertise roles and skills required by the software development industry. In 2019 1st International Informatics and Software Engineering Conference (UBMYK) (pp. 1-4). Ankara.
- Gürcan, F. (2009). Web içerik madenciliği ve konu sınıflandırılması. Yüksek Lisans Tezi, Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü, Trabzon.
- Gürcan, F. (2018). Multi-class classification of Turkish texts with machine learning algorithms. In 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-5). Ankara.
- Johri, V. and Bansal, S. (2018). Identifying trends in technologies and programming languages using topic modeling. 12th International Conference on Semantic Computing (ICSC2018) (pp. 391-396). California.
- Joorabchi, A., English, M. and Mahdi, A.E. (2016). Text mining stackoverflow: An insight into challenges and subject-related difficulties faced by computer science learners. Journal of Enterprise Information Management, 29(2), 255-275. <https://doi.org/10.1108/JEIM-11-2014-0109>
- Lijffijt, J., Papapetrou, P., Puolamäki, K. and Mannila, H. (2011). Analyzing word frequencies in large text corpora using inter-arrival times and bootstrapping. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 341-357). Berlin.
- Murgia, A., Tourani, P., Adams, B. and Ortu, M. (2014). Do developers feel emotions? An exploratory analysis of emotions in software artifacts. In Proceedings of the 11th working conference on mining software repositories (pp. 262-271). Hyderabad.
- Nasehi, S.M., Sillito, J., Maurer, F. and Burns, C. (2012). What makes a good code example?: A study of programming Q&A in stackoverflow. In 28th IEEE International Conference on Software Maintenance (ICSM) (pp. 25-34). Trento.
- Özyurt, Ö. and Özyurt, H. (2010). Forum based learning: Content analysis of asynchronous discussion forums for the teaching of computer programming languages. International Educational Technology Conference (pp. 618-622). İstanbul.
- Pal, A., Harper, F.M. and Konstan, J.A. (2012). Exploring question selection bias to identify experts and potential experts in community question answering. ACM Transactions on Information Systems, 30(2), 1-28. <https://doi.org/10.1145/2180868.2180872>

- Rajput, N.K., Ahuja, B. and Riyal, M.K. (2019). A statistical probe into the word frequency and length distributions prevalent in the translations of Bhagavad Gita. *Pramana*, 92(4), 1-6. <https://doi.org/10.1007/s12043-018-1709-8>
- Rosen, C. and Shihab, E. (2016). What are mobile developers asking about? A large scale study using stack overflow. *Empirical Software Engineering*, 21(3), 1192-1223. <https://doi.org/10.1007/s10664-015-9379-3>
- Stack Exchange API. (2020, May 29). Retrieved from <https://api.stackexchange.com/>.
- Xia, X., Lo, D., Wang, X. and Zhou, B. (2013). Tag recommendation in software information sites. In 10th Working Conference on Mining Software Repositories (MSR) (pp. 287-296). San Francisco.
- Yang, X.L., Lo, D., Xia, X., Wan, Z.Y. and Sun, J.L. (2016). What security questions do developers ask? A large-scale study of stack overflow posts. *Journal of Computer Science and Technology*, 31(5), 910-924. <https://doi.org/10.1007/s11390-016-1672-0>
- Zhang, J. and Zhu, G.B. (2018). Hot topic discovery research of StackOverflow programming website based on CBOW LDA topic model. *Computer Science*, 45(4), 208-214.