

COVID-19 Hakkındaki Soruları Cevaplayan bir Akıllı Sanal Asistan

A Smart Virtual Assistant Answering Questions About COVID-19

Yusuf Uğurlu

Bilgisayar Mühendisliği Bölümü
İstanbul Esenyurt Üniversitesi
İstanbul, Türkiye
yusuf-ugurlu@outlook.com

Murat Karabulut

Bilgisayar Mühendisliği Bölümü
İstanbul Esenyurt Üniversitesi
İstanbul, Türkiye
mkarabulut44@hotmail.com

İslam Mayda

Bilgisayar Mühendisliği Bölümü
Yıldız Teknik Üniversitesi
İstanbul, Türkiye
islam.mayda@stu.khas.edu.tr

Özetçe—COVID-19 salgını sürecinde farklı kaynaklar tarafından ortaya atılan sayısız bilgi ve haber, insanların kafasında salgınla ilgili çok sayıda soru işaretinin oluşmasına neden olmuştur. Bu çalışmada, COVID-19 ile ilgili bu bilgi kirliliğini ortadan kaldırmak ve insanların sorularına güvenilir kaynaklardan elde edilen bilgilerle doğru cevaplar verebilmek için bir akıllı sanal asistan geliştirilmiştir. Akıllı sanal asistanlar, insanların merak ettikleri sorulara kolay ve hızlı bir şekilde cevap alabilmelerini sağlayan soru cevaplama sistemi yazılımlarıdır. Eğitim, e-ticaret, turizm, finans, sağlık başta olmak üzere birçok alanda tüm dünyada bu sanal asistanların kullanımı yaygınlaşmaktadır. Ancak, Türkçe üzerine yapılmış soru cevaplama sistemi çalışmalarının sayısı oldukça az ve yetersizdir. Hatta literatürde Türkçe için sunulmuş sağlık alanında bir soru cevaplama çalışmasına rastlanmamıştır. Çalışma kapsamında sunulan akıllı sanal asistan bu bakımdan bir ilktir. Bu akıllı sanal asistan iki farklı türde soruya cevap vermektedir. Bunlardan birincisi, COVID-19 ile ilgili halk arasında sıkça sorulan bilgi sorularıdır. İkincisi ise dünyadaki ülkelerin COVID-19 vaka sayısı, ölüm sayısı, test sayısı, vb. sayısal verilerine dair sorulardır. Geliştirilen sanal asistanın, kullanıcıların COVID-19 ile ilgili sorularının büyük çoğunluğuna cevap verebildiği görülmüştür. Bu akıllı sanal asistanın COVID-19 salgınıyla mücadelede faydalı olabileceği düşünülmektedir.

Anahtar Kelimeler—akıllı sanal asistan, soru cevaplama sistemi, doğal dil işleme

Abstract—Numerous information and news put forward by different sources during the COVID-19 pandemic have caused many questions about the epidemic in people's minds. In this study, a smart virtual assistant was developed in order to eliminate this information pollution about COVID-19 and to give correct answers to people's questions with information obtained from reliable sources. Intelligent virtual assistants are question answering system software that enable people to get answers easily and quickly to their questions curious about. The use of these virtual assistants is becoming widespread all over the world in many fields, especially in education, e-commerce, tourism, finance and health. However, the number of question-answering system studies on Turkish is very few and insufficient. In fact, no question-answering study was found in the field of health that is presented for Turkish in the literature. The smart virtual assistant presented within the scope of this study is a first in this respect. This smart virtual assistant answers two different types of questions. The first of these is frequently asked information questions about COVID-19 among the people. The second is the questions about numerical data such as number of

cases, number of deaths, number of tests, etc. of countries in the world. It has been observed that the developed virtual assistant can answer the vast majority of users' questions about COVID-19. It is thought that this smart virtual assistant could be useful in fighting the COVID-19 pandemic.

Keywords—smart virtual assistant, question answering system, natural language processing

I. GİRİŞ

2019 yılının son aylarında Çin'in Hubei eyaletinin Wuhan şehrinde ilk olarak görülen Yeni Koronavirüs Hastalığı (COVID-19) hızla yayılarak tüm dünyayı etkisi altına almış ve küresel bir salgına dönüşmüştür. İlk zamanlarda hakkında çok az şey bilinen bu hastalık insanlar arasında korku ve endişeye neden olmuştur. COVID-19 hakkında farklı medya platformlarında kısa süre içerisinde çok sayıda açıklama ve haber paylaşılmış, insanlar salgın sürecinde yoğun bir bilgi bombardımanına maruz kalmıştır. Bu durum aynı zamanda hastalık hakkında bilgi kirliliğine yol açmış, farklı kaynakların birbirleriyle çelişen açıklamaları veya kaynağı belli olmayan birtakım haberler insanların kafasında COVID-19 hakkında çok sayıda soru işaretinin doğmasına sebep olmuştur. Toplum, söz konusu hastalıkla ilgili olarak doğru kaynaklarla bilgilendirilmeye ve yönlendirilmeye ihtiyaç duymaktadır.

İnsanlara merak ettikleri bir konu hakkında kapsamlı ve detaylı bir dokümantasyon sunulsa bile insanlar uzun uzun doküman okumak yerine akıllarındaki sorulara daha hızlı ve net bir şekilde cevap bulmak istemektedir. Bu yüzden birçok web sitesi, kılavuz, duyuru, vb. yerlerde “Sıkça Sorulan Sorular” bölümüne yer verilmektedir. Otomatik soru cevaplama sistemleri de bu ihtiyaca yönelik olarak geliştirilen yazılımlardır. İnsanların belirli bir konuya yönelik veya genel olarak tüm sorularına otomatik yanıt vermeye çalışan bu programlar, *akıllı sanal asistan* şeklinde de adlandırılmaktadır. Bu yazılımlar, her geçen gün giderek artan insan bilgisayar etkileşiminin önemli birer ürünüdür. Doğal dil işleme kullanılarak geliştirilen akıllı sanal asistanların sayısı son yıllarda artış göstermekte ve birçok farklı alanda kullanımı yaygınlaşmaktadır. 2018 yılı itibarıyla küresel akıllı sanal asistan pazarı 2.2 milyar \$ değerine ulaşmıştır ve 2024 yılında bu pazarın 11.3 milyar \$ büyüklüğe ulaşması tahmin edilmektedir [1].

Halk sağlığı gibi hayati öneme sahip bir konuda insanların akıllarındaki sorulara kolay ve hızlı bir şekilde yanıt alabilmeleri için soru cevaplama sistemleri

kullanılabilir. Bu çalışmada da, COVID-19 ile ilgili mevcut olan bilgi kirliliğinin önüne geçmek ve hastalıkla ilgili insanları doğru kaynaklardaki veriler ışığında yönlendirmek amacıyla Türkçe dili için bir akıllı sanal asistan geliştirilmiştir. Geliştirilen bu sanal asistanın kullanımıyla birlikte, insanlar hastalıkla ilgili temel sorulara çok kısa süre içerisinde otomatik olarak yanıt alabileceğinden dolayı, başta Sağlık Bakanlığı olmak üzere salgınla mücadele veren kamu kurumlarının, hastanelerin, toplum sağlığı merkezlerinin, halkın hastalıkla ilgili bilgi talebinden doğan iş yükünün önemli ölçüde azalacağı düşünülmektedir.

Çalışmanın bundan sonraki bölümlerinde öncelikle, soru cevaplama sistemlerine yönelik daha önce yapılmış olan literatürdeki araştırmalar incelenecek, sonrasında bu çalışmada kullanılan yöntem detaylıca ortaya koyulacak, daha sonra çalışma kapsamında geliştirilen uygulamaya dair bilgiler sunulacak, son olarak ise çalışmanın sonuçları ve gelecek çalışmalara yönelik fikirler paylaşılacaktır.

II. LİTERATÜR TARAMASI

Soru cevaplama sistemleri, bilgi çekme (*information retrieval*) ve doğal dil işleme yöntemleri kullanılarak geliştirilen yazılımlardır. Akıllı sanal asistan olarak da adlandırılan bu sistemler özellikle son yıllarda popülerliğini arttırmış ve örnekleri hızla çoğalmıştır. Bu alanda yayınlanmış olan çalışmalar literatür taramalarında da incelenmiştir [2-4]. Bu bölümde soru cevaplama sistemlerine dair temel yöntemler özetlenecek ve çeşitli örnekler sunulacaktır.

Bir soru cevaplama sistemi genel olarak 3 ana aşamadan oluşmaktadır: soru işleme, doküman işleme, cevap çıkarımı. Öncelikle kullanıcı bir arayüz aracılığıyla sorusunu yazar. Birinci aşama olan soruyu işleme adımı sorunun türü belirlenir, yani soru sınıflandırılır ve anahtar kelimeleri çıkarılır. Sorular farklı sistemlerde farklı şekillerde sınıflandırılmaktadır. İkinci aşamada ise sorulan soruya dair muhtemel cevaplar dokümanlarda taranır. Burada doküman belirli bir klasördeki dosya kümesi, web kaynağı veya yapısal bir veritabanı olabilir. Bu tarama sonucunda aday cevaplar belirlenir. Son aşama olan cevap çıkarımı aşamasında ise, aday cevaplar değerlendirilir, puanlandırılır, kendi aralarında sıralanır ve doğru cevap olma olasılığı en yüksek olan cevap yine arayüz aracılığıyla kullanıcıya sunulur [2].

Soru cevaplama sistemleri cevap verdikleri alanların kapsamına göre, genel konulara yönelik ve sınırlı alandaki konulara yönelik olarak iki farklı grupta sınıflandırılabilir. Genel konulara yönelik olan soru cevaplama sistemleri genellikle web'den veya büyük veritabanlarından bilgi çekme yöntemiyle geliştirilmektedir. Burada temel amaç kullanıcının herhangi bir konuyla ilgili bilgiye kolay ve hızlı bir şekilde ulaşmasıdır. Genel amaçlı sohbet robotları farklı maksatlarla geliştiriliyor olsa da kullandıkları veri kaynağı bakımından bu gruptaki soru cevaplama sistemlerine dâhil edilebilir.

Sınırlı alana yönelik soru cevaplama sistemi çalışmaları ise, genel konulara yönelik yapılan çalışmalardan daha az sayıdadır. Çünkü sonuçları çok spesifik ve genelleştirmeye açık değildir [5]. Öte yandan, geliştirilen soru cevaplama sisteminin doğruluğunu arttırmak amacıyla genel konulara dair tüm soruları yanıtlamayı hedeflemek yerine sadece belirli alandaki sorulara yanıt veren bir soru cevaplama

sistemi geliştirmek etkili bir yaklaşımdır [2]. Bu tür sistemlerde kapsamın daraltılması başarı oranını yükselttiği için daha iyi bir kullanıcı deneyimi sağlanmaktadır. Bu çerçevede, özellikle sağlık, turizm, e-ticaret, eğitim başta olmak üzere birçok hizmete yönelik sınırlı alanda sorulara yanıt veren akıllı sanal asistanlar geliştirilmiştir [3].

Literatürdeki çalışmalara bakıldığında soru cevaplama sistemleri üzerine en çok çalışılan dilin İngilizce olduğu görülmüştür [3]. Türkçe ile İngilizce farklı dil ailelerine üye oldukları için İngilizce için geliştirilen yöntemler Türkçe için doğrudan uygulanabilir değildir. Türkçe'nin sondan eklemeli bir dil olması, Türkçe kelimelerin anlamının çıkarılabilmesi için biçimbilimsel analizi zorunlu kılmaktadır [6].

Sohbet robotlarının tarihinin başlangıcı, matematikçi ve bilgisayar bilimcisi Alan Turing'in 1950'de yayınladığı makaleye [7] dayandırılmaktadır. O tarihten bugüne kadar İngilizce için çok sayıda sohbet robotu geliştirilmiş olmasına rağmen, Türkçe için yapılan çalışmalar çok az sayıda ve yetersizdir. Türkçe üzerine geliştirilen ilk diyalog sistemlerinden biri selamlama ve tanışma ifadelerini içeren basit bir sohbet robotudur [8]. Geliştirilen sohbet robotu sohbetin sürekliliği bakımından yetersiz olsa da bu alandaki ilk çalışmalardan biri olması bakımından önemlidir.

Referans [9]'da genel sohbet amaçlı Türkçe konuşan daha kapsamlı ve kullanışlı bir sohbet robotu sunulmuştur. Çalışmada AIML dili ¹ (Artificial Intelligence Markup Language) kullanılmıştır. AIML dili ile soru-cevap grupları XML formatında dosyalara kaydedilmektedir. Geliştirilen sohbet robotu da bu dosyaları kullanarak kullanıcıya uygun bir cevap vermektedir. Program kullanıcıya sunabileceği uygun bir cevap bulamadığı durumda kaçamak cevaplar verme, kullanıcıyı cevap verebileceği konulara çekme gibi özelliklere sahiptir. Burada AIML dosyaları elle oluşturulduğu için dosyalardaki soru-cevap gruplarının sayısı arttırıldıkça sistemin başarısı da artmaktadır.

Sohbet robotları dışındaki ilk Türkçe soru cevaplama sistemi çalışmalarından biri, *BayBilmiş* adı verilen, arka planda arama motoru sonuçlarını kullanan internet tabanlı ve konu bağımsız bir soru cevaplama sistemidir [10]. Öncelikle, kullanıcının sorusu arama motorunda sorgulanıp, arama motoru sonuçlarında listelenen sayfalardan aday cevaplar belirlenmektedir. Sonrasında ise bu aday cevaplar çeşitli kriterlere göre puanlandırılıp, en yüksek puanı alan ilk 5 aday cümle kullanıcıya cevap olarak sunulmaktadır. Geliştirilen sistem, “neden”, “nasıl”, “nelerdir” soruları haricinde “kim”, “nedir”, “ne zaman”, “nerede” gibi sorular için başarılı çalışmaktadır.

Türkçe için geliştirilen başka bir soru cevaplama sistemi çalışmasında [11] sistemin kapsamı ve cevap kümesinin nasıl oluşturulduğuna dair detaylar verilmemiştir. Çalışmada kullanılan örnek soru ve aday cevap metinleri kelimelerine ayrılıp, Türkçe'de sık kullanılan edat, bağlaç, zamir ve fiiller metinlerden çıkartıldıktan sonra kalan kelimelerin gövde kısımları ile doküman-terim matrisi elde edilmiştir. Vektör uzay modeliyle her bir doküman çok boyutlu uzayda birer nokta olarak ifade edilmiş ve soruya en yakın olan aday cevap kosinüs benzerliği formülü ile hesaplanmıştır.

Referans [12]'de ise Türkçe soru cevaplama sistemleri için soru analizi kısmında kullanılmak üzere sorunun odağını, yani cevabın türü ve niteliğini belirten kelimeleri

¹ www.aiml.foundation

belirleyen kural tabanlı bir yöntem önerilmiştir. Önerilen yöntem elle oluşturulmuş bir soru kümesi ile test edilmiştir.

Türkçe üzerine yayınlanmış kısıtlı alanlara yönelik birkaç soru cevaplama sistemi örneği de mevcuttur. Örneğin, [13]'te kullanıcıların çeşitli vakıflara bağış yapmalarına yardımcı olmak üzere bir rehber robot geliştirilmiştir. Diğer bir çalışma olan [6]'da ise bir e-ticaret sitesinde kullanılmak üzere kullanıcılara alışverişlerinde yardımcı olmayı amaçlayan bilgi tabanlı bir soru cevaplama sistemi önerilmiştir. Literatürde sağlık alanına özel olarak Türkçe için geliştirilmiş bir soru cevaplama sistemine rastlanmamıştır. Bu bildiride sunulan çalışma bu bakımdan bir ilk olma özelliği taşımaktadır. Öte yandan İngilizce üzerine sağlık alanında çok sayıda soru cevaplama sistemi araştırması yayınlanmıştır [14]. Soru cevaplama sistemlerinin sağlık alanındaki uygulamalarına dair literatür taraması niteliğinde olan [14]'te, doğal dil işleme yaklaşımının, kullanıcıya sunulan cevabın kesinlik (*precision*) değerinin çok önemli olduğu durumlar için uygun olduğu ifade edilmiştir.

Sağlık alanında İngilizce dili için geliştirilen soru cevaplama sistemlerinin yanı sıra farklı diller için geliştirilmiş sistemler de yayınlanmıştır. Örneğin, [15]'te İtalyanca dili için bir kişisel sağlık asistanı geliştirilmiştir. Telegram üzerinden çalışan akıllı bot, doğal dil işleme yaparak kullanıcıların semptomlarını anlayıp, hastalığını tahmin edip, ilgili hastalığı için hastaya en yakın doktoru önermektedir. Aynı zamanda, sanal asistan kullanıcının tedavisini ve sağlık değerlerini izlemesine de yardımcı olmaktadır.

COVID-19 salgınıyla mücadele sürecinde de bilgi yayma, kişisel semptomları izleme, davranış değişikliği ve akıl sağlığı konularında destek olma gibi farklı görevlerde akıllı sanal asistanlardan faydalanılmıştır [16]. Bu bildiride sunulan, Türkçe dilindeki COVID-19 ile ilgili sorular için geliştirilen akıllı sanal asistanın bir benzerinin [17] daha önce İngilizce için geliştirildiği de görülmüştür. *Jennifer*² adı verilen bu soru cevaplama sistemi vaka istatistiklerinden, COVID-19 ile ilgili gıda güvenliğine ve hastalığın önlenmesine dair tedbirlere kadar geniş konularda bilgi sunmaktadır. Sistemin veritabanı, dünya çapında yüzlerce gönüllü bilim insanı ve tıp uzmanı tarafından oluşturulmuştur. Araştırmacılar, veritabanındaki soru-cevap gruplarının güvenilir olmasına son derece önem vermişler ve girilen tüm veriler denetim mekanizmasından geçirilip onaylandıktan sonra veritabanına dâhil edilmiştir. Elle oluşturulan bu veritabanının yanı sıra sistem güvenilir kaynaklardan otomatik olarak çektiği sayısal verileri kullanarak bir ülkedeki veya bir eyaletteki vaka sayısı veya ölüm oranı gibi sorulara da yanıt verebilmektedir. Akıllı sanal asistan kullanıma sunulduktan sonra kullanıcılar tarafından 1,480 soru sorulduğu ve *Jennifer*'in bu soruların %71'ine cevap verebildiği tespit edilmiştir.

III. METODOLOJİ

Bu çalışma kapsamında geliştirilen akıllı sanal asistan iki farklı türde soruya yanıt verebilmektedir. Bunlardan birincisi COVID-19 ile ilgili sıkça sorulan sorulardır. İkincisi ise ülkelere dair sayısal verilere yönelik istatistik sorulardır. Öncelikle bu iki soru türüne dair detaylar verilip, daha sonra geliştirilen sanal asistanın sistem mimarisi sunulacaktır.

² www.newvoicesnasem.org/jennifer-ai-chatbot

A. COVID-19 ile İlgili Sıkça Sorulan Sorular

Bu tür sorular, kullanıcıların hastalıkla ilgili merak ettiği bilgi sorulardır. Örneğin, “COVID-19’un kuluçka süresi ne kadardır?”, “Evcil hayvanlar koronavirüsü taşıyıp, bulaştırabilir mi?”, “Koronavirüs açık yaradan bulaşabilir mi?” gibi kullanıcıların sıkça sorduğu sorular bu türe ait veri kümesinde yer almaktadır. İngilizce için kamuya açık olarak paylaşılmış COVID-19 soru cevaplama veri kümesi [18] olsa da literatürde Türkçe için benzer bir veri kümesi henüz sunulmamıştır. O yüzden, bu çalışma için sıfırdan bir veri kümesi üretilmiştir. İngilizce veri kümesinde olduğu gibi bu çalışmada kullanılan veri kümesi de elle oluşturulmuştur.

Sağlık alanında kullanılan otomatik soru cevaplama sistemleri insanlara yardımcı olma amacı taşıyor olsa da içerisinde ciddi riskler de barındırmaktadır. Kullanıcıların sağlık alanındaki konularla ilgili bilgi almak için *Siri*, *Alexa*, and *Google Assistant* kullanımını inceleyen bir çalışmada, uygulamaların verdiği cevapların %29'unun insanlara zararlı olabileceği, verilen cevapların hayata geçirilmesinin ise %16 oranında ölümlerle sonuçlanabileceği ifade edilmiştir [19]. Söz konusu çalışma sağlık alanındaki sanal asistanların kullanıcıları doğru bilgi verme hususundaki zorunluluğunu açıkça ortaya koymaktadır. Çünkü sanal asistan tarafından verilen yanlış bir cevabın bedeli kullanıcı için sadece zaman veya para kaybı olmayıp, insan hayatının kaybı da söz konusu olabilmektedir.

Bu çalışmada kullanılan soru cevaplama veri kümesi oluşturulurken kullanıcıya kesin ve doğru bilgilerle cevap verebilmek için sadece T.C. Sağlık Bakanlığı'nın web sitesi³ üzerinden kamuoyuna paylaştığı rehberler ve diğer bilgilendirme dokümanları ile Sağlık Bakanlığı Koronavirüs Bilim Kurulu üyesi uzman doktorların açıklamalarından faydalanılmıştır. Bu şekilde oluşturulan 54 soru-cevap grubu PHP ile geliştirilen bir web arayüzü kullanılarak yazarlar tarafından elle veritabanına girilmiştir. Veri girişi sırasında her bir soru-cevap grubu için etiketler de tanımlanmıştır. Bu etiketler, soru veya cevap metninde geçen anahtar kelime niteliğindeki isim, sıfat ve fiillerin çekim eki almamış halleridir. Ayrıca, bir soru-cevap grubu için tanımlanan etiketlerin, eğer varsa eş anlamlıları da veritabanına girilmiştir. Örneğin, semptom için belirti, salgın için pandemi etiketleri de ilgili soru-cevap grubunun etiket listesine dâhil edilmiştir.

B. Ülkelere dair Sayısal Verilere Yönelik İstatistik Soruları

Küresel salgından etkilenen ülkeler COVID-19 ile ilgili günlük verilerini kamuoyu ile paylaşmaktadır. Ülkelerin sağlık bakanlıklarının veya ilgili diğer kurumlarının paylaştığı bu veriler toplu halde Worldometer⁴ sitesinde yayınlanmaktadır. Worldometer'da yayınlanan veriler çok sayıda devlet ve uluslararası kuruluş tarafından güvenilir kabul edilmekte ve kullanılmaktadır. Worldometer koronavirüs sayfasında ülkelere dair paylaşılan veriler şöyledir: toplam vaka sayısı, yeni vaka sayısı, toplam ölüm sayısı, yeni ölüm sayısı, toplam iyileşen sayısı, yeni iyileşen sayısı, aktif vaka sayısı, durumu kritik olan vaka sayısı, 1 milyondaki toplam vaka sayısı, 1 milyondaki ölüm sayısı, toplam test sayısı, 1 milyondaki test sayısı, ülkenin nüfusu.

³ <https://covid19.saglik.gov.tr>

⁴ www.worldometers.info/coronavirus

Worldometer koronavirüs sayfasındaki ülkelere dair tüm veriler, yazılan bir kod parçasıyla HTML çözümleme yapılarak çekilmekte ve veritabanına kaydedilmektedir. Worldometer sayfasındaki verilere ek olarak ülkelerdeki ölüm oranları da hesaplanıp, bu veri de veritabanına girilmektedir. Veritabanındaki bu veriler Cron Job⁵ yöntemi kullanılarak dakikada bir otomatik olarak güncellenmektedir.

C. Sistem Mimarisi

Geliştirilen akıllı sanal asistan bir web sayfasında kullanıma sunulmuştur. Web sayfasına girildiğinde sanal asistan kullanıcıyı selamlayıp, kısaca kendini tanıtmaktadır. Kullanıcı sanal asistana sorusunu sorduğunda, öncelikle soru metnindeki tüm karakterler küçük harfe çevrilmekte ve noktalama işaretleri temizlenmektedir. Daha sonra ise sistem sorunun türünü tespit etmektedir. Eğer soru metninde bir ülke ismi ve sayısal verilerle ilgili bir anahtar kelime (*toplam vaka, ölüm oranı, test sayısı*, vb. gibi) varsa, bu sorunun ülkelere dair istatistik sorusu olduğu tespit edilmektedir. Eğer soruda bir ülke ismi veya sayısal verilere dair anahtar kelime yoksa o zaman bunun COVID-19 ile ilgili bir bilgi sorusu olduğu anlaşılmaktadır.

Kullanıcının sorusu ülkelere dair istatistik sorusu ise ülkenin adıyla birlikte tespit edilen anahtar kelime, yazılan web servis aracılığıyla ülke verilerinin tutulduğu veritabanındaki tablodan sorgulanmaktadır. Aynı anda birden fazla anahtar kelime verisi de sorulabilmektedir. Kullanıcı aşağıdaki örnek sorulara sanal asistan aracılığıyla cevap alabilmektedir:

- *Türkiye'deki bugünkü vaka sayısı nedir?*
- *Rusya'daki ölüm oranını söyler misin?*
- *Brezilya'nın test sayısını öğrenebilir miyim?*
- *Hindistan'daki bugün iyileşen sayısı nedir?*
- *İspanya'nın 1 milyondaki ölüm sayısı nedir?*
- *Peru'daki aktif vaka sayısı ve ölüm oranı kaç?*
- *Tayland'daki toplam test sayısı ve toplam vaka sayısı kaçtır?*

Soru metninde sadece ülke ismi tespit edilip, sayısal verilere dair bir anahtar kelime bulunamamışsa, kullanıcıya o ülkeye dair veritabanındaki tabloda bulunan tüm sayısal veriler listelenmektedir. Ayrıca, ülkelerin verilerinden yola çıkarak dünyadaki en fazla, en az ve toplam gibi istatistikler de sorgulanabilmektedir. Eğer soru metninde, sayısal verilere dair bir anahtar kelime ve *en az, en yüksek, en düşük, en fazla, en çok, toplam* vb. ifadelerden biri geçiyorsa kullanıcıya bu sorunun cevabı da sunulabilmektedir. Örneğin, kullanıcı aşağıdaki gibi sorulara yanıt alabilmektedir:

- *Dünyadaki toplam vaka sayısı nedir?*
- *Dünyadaki en fazla test sayısını öğrenebilir miyim?*
- *En düşük ölüm oranını söyler misin?*
- *Dünyada en yüksek yeni ölüm sayısı kaçtır?*

Ülke verilerine yönelik istatistik sorularını yanıtlarken kullanılan tüm anahtar kelimeler (*ölüm sayısı, vaka sayısı*, vb.), bunların alternatif yazımları (*ölüm yerine vefat* gibi) ve diğer ifadeler (*en fazla, en düşük*, vb.), sistem tarafından sorunun cevaplanması sırasında harici bir metin dosyasından okunmaktadır. Yani, ilerleyen zamanlarda sanal asistanın koduna müdahale etmeden bu harici metin dosyaları üzerinde yapılacak düzenlemelerle sistemin iyileştirilebilmesi oldukça kolaydır.

Kullanıcının sorusu COVID-19 ile ilgili bilgi sorusu ise soru metni ön işleme adımlarından geçirildikten sonra sorunun önemli kelimeleri elde edilmektedir. Sorudan elde edilen bu kelimeler yine web servis aracılığıyla veritabanından sorgulanmaktadır. Bilgi sorusu olduğu anlaşılan soru metinleri için sırasıyla aşağıdaki ön işleme adımları uygulanmaktadır:

1. Soru metni boşluklarına göre kelimelere ayrılmaktadır.
2. Soru metni içinde geçen, tek başına bir anlamı olmayan kelimeler (*stop words*) temizlenmektedir. Burada web tabanlı bir depolama servisinde olan Github'da açık kaynak olarak paylaşılan Türkçe dolgu sözcükleri⁶ listesi kullanılmıştır.
3. Açık kaynak kodlu Türkçe doğal dil işleme kütüphanesi olan NZemberek⁷ ile kelimelerin gereksiz olan çekim ekleri atılarak kelime gövdeleri elde edilmektedir. Java için geliştirilen Zemberek kütüphanesinin [20] .NET platformuna taşınması projesi kapsamında geliştirilen NZemberek'in içinde kelimelerin gövdesini bulmak için hazır bir metot olmadığı için, kelimeler önce *kelimeCozumle* metodu ile kök ve eklerine ayrılıp, daha sonra kelimenin kökü ve yapım ekleri *kelimeUret* metodu ile tekrar birleştirilmiş ve kelime gövdeleri bu şekilde üretilmiştir.

Ön işleme adımlarının ardından elde edilen kelime gövdeleri, veritabanındaki soru-cevap gruplarının etiketleri ile karşılaştırılarak, en az bir eşleşme sağlanan soru-cevap gruplarından aday cevaplar belirlenmektedir. Aday cevaplar arasından doğru cevabı bulabilmek için kelime örtüşüm (*word overlap*) [21] yönteminden yararlanılmıştır. Buna göre, en yüksek oranda eşleşme sağlayan aday cevap seçilerek, kullanıcının sorusuna cevap olarak döndürülmektedir. Eşleşme oranı, eşleşen etiket sayısının soru-cevap grubunun toplam etiket sayısına bölünmesiyle hesaplanmaktadır.

Elbette ki soru cevaplama sistemlerinde kullanıcıların bir soruyu farklı şekillerde sorması söz konusudur. Geliştirilen sistemde soru metni üzerinde morfolojik analiz yapıldığı için aynı soru farklı ifadelerle sorulsa da, soru içerisinde geçen ve cevabın niteliğini belirten odak kelimelerin tespit edilebilmesi mümkün olmaktadır. Örneğin, klimanın COVID-19 salgını üzerindeki etkisini öğrenmek isteyen kullanıcılar aşağıdaki sorulardan herhangi birini veya benzerlerini sorarak doğru cevabı alabilmektedir:

- *Kapalı ortamlarda klimayı kullanabilir miyiz?*
- *Salgın döneminde klima riskli midir?*
- *Klimanın virüsü bulaştırma durumu var mı?*

⁵ www.cron-job.org

⁶ <https://github.com/ahmetax/trstop>

⁷ <https://code.google.com/archive/p/nzemberek>

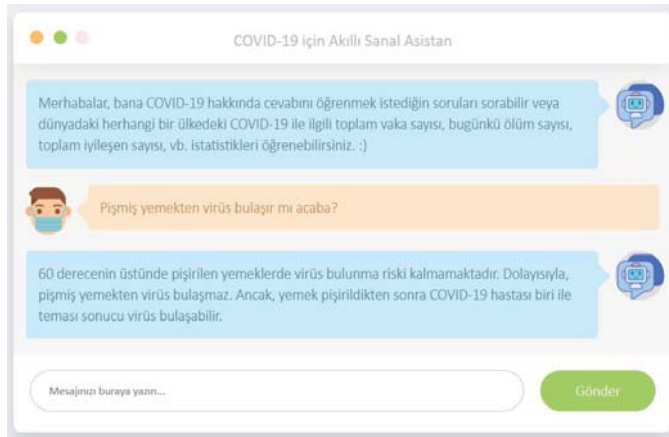
- *Klimalı ortalarda bulunmak sakıncalı mıdır?*
- *Klimaların salgında zararlı olduğu doğru mu?*

IV. UYGULAMA

Akıllı sanal asistanın önyüzü .NET platformunda C# dili ve AJAX tekniği ile, sonucu tarafı ise PHP ile geliştirilmiştir. Uygulamanın veritabanı, MySQL veritabanı yönetim sistemi kullanılarak oluşturulmuştur. Kullanıcı arayüzü için Bootsnipp sitesinden açık kaynak olarak paylaşılan bir sohbet robotu tasarımı⁸ alınmış ve özelleştirilerek kullanılmıştır. Geliştirilen sanal asistana ait örnek ekran görüntüleri Şekil 1 ve 2’de sunulmuştur.

Akıllı sanal asistanın soru cevaplama performansını izleyebilmek ve daha sonra iyileştirme çalışmalarında kullanmak amacıyla kullanıcıların sormuş olduğu her soru ve sanal asistanın o kullanıcıya verdiği cevap veritabanına kaydedilmektedir. Bu kayıtlar bir arayüz üzerinden listelenmekte ve sanal asistanın kullanımı bu kayıtlar sayesinde yazarlar tarafından gözlemlenmektedir.

Log kayıtlarına bakıldığında, akıllı sanal asistan yayınlandıktan sonra çok sayıda kişi tarafından kullanıldığı ve yüzlerce soru sorulduğu görülmüştür. Sanal asistan bu soruların büyük çoğunluğuna doğru cevap verirken bazı sorulara cevap verememiş veya başka bir sorunun cevabını vermiştir.



Şekil 1. Akıllı sanal asistanın örnek ekran görüntüsü



Şekil 2. Akıllı sanal asistanın örnek ekran görüntüsü

Log kayıtlarına göre sanal asistanın cevap veremediği veya başka bir sorunun cevabını verdiği kullanıcı sorularından bazıları aşağıdaki gibidir:

- *Ne tip maske takmalıyım?*
- *Virüs kaparsam ölür müyüm?*
- *Çok yanyorum, hasta mıyım?*
- *Karantina nedir?*
- *Bir kişinin ateş aralığı kaç olmalıdır?*
- *Aşı nasıl bulunur?*

Bu soruların haricinde sanal asistana “sen kimsin”, “ne biliyorsun”, “nasılsın” gibi soruların yanı sıra “covid19 muyum”, “hasta mıyım”, “ölecek miyim” gibi soruların da sorulduğu ve sistemin bu soruları yanıtlayamadığı gözlemlenmiştir. Bu kayıtlar sayesinde akıllı sanal asistanın cevap veremediği bilgi soruları görülebilmekte ve sistemin soru cevaplama performansını iyileştirmek için yine güvenilir kaynaklardan elde edilen bilgilerle yeni soru-cevap grupları sisteme girilebilmektedir.

V. SONUÇ VE GELECEK ÇALIŞMALAR

Bu çalışmada, COVID-19 salgını sürecinde ortaya çıkan bilgi kirliliği sorununa karşı, insanların hastalıkla ilgili merak ettikleri sorulara hızlı ve kolay bir şekilde doğru cevaplar alabileceği bir soru cevaplama sistemi ile çözüm önerilmiştir. Bu doğrultuda, COVID-19 salgını hakkında halk arasında sıkça sorulan sorulara T.C. Sağlık Bakanlığı ve konunun uzmanlarının bilgilendirmeleri ışığında cevap veren ve ayrıca dünyadaki ülkelerin COVID-19 ile ilgili toplam vaka sayısı, bugünkü ölüm sayısı, toplam iyileşen sayısı, test sayısı, vb. gibi sayısal verilerini kullanıcılara sunan bir akıllı sanal asistan geliştirilmiştir.

Kullanıcı kayıtlarına bakıldığında, akıllı sanal asistanın kullanıcıların sorularını büyük ölçüde doğru cevaplandığı görülmüştür. Sistemin cevap veremediği sorular için de yeni soru-cevap grupları veritabanına girilerek düşünülmektedir. Ayrıca, tanışma ve günlük sohbet soruları ile sistemin kullanıcı ile etkileşiminin artırılması da planlanmaktadır. Sanal asistanın, sadece kullanıcının sorularına cevap vermek yerine, kullanıcıya sağlık durumuyla ilgili çeşitli sorular sorarak kullanıcının cevaplarına göre yönlendirmeler yapabilmesi de önemli bir özellik olarak sisteme dâhil edilebilir. Bu sayede, çeşitli semptomlar gösterdiğini belirten ve COVID-19 virüsünü kapmış olması muhtemel olan kullanıcıların en yakın sağlık kuruluşuna yönlendirilebilmesi mümkün olabilecektir.

Çalışma kapsamında sunulan akıllı sanal asistan Türkçe için sağlık alanında geliştirilen literatürdeki ilk soru cevaplama sistemi olma niteliği taşımaktadır. Günümüzde soru cevaplama sistemlerinin sayısının hızla arttığı ve kullanımının yaygınlaştığı göz önüne alındığında Türkçe için bu konuda daha fazla çalışma yapılması gerektiği açıktır. Türkçe soru cevaplama sistemleri geliştirilebilmesi için her şeyden önce Türkçe soru cevaplama veri kümelerinin oluşturulup araştırmacılara sunulmasına ihtiyaç vardır. Büyük ve kapsamlı veri kümelerinin araştırmacılarla paylaşılması halinde sadece COVID-19 için değil, diğer hastalıklar için de bu tür akıllı sanal asistanlar geliştirilebilecektir. Bu sanal asistanlar, özellikle

⁸ www.bootsnipp.com/snippets/ZlkBn

bugünlerde yaşadığımız salgın süreci gibi kritik dönemlerde sağlık sisteminin yükünü azaltmada önemli bir rol üstelenebilir.

KAYNAKÇA

- [1] IMARC Group, "Intelligent Virtual Assistant Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2019-2024", Rapor No: 4775648, Haziran 2019, [Online]. Available: www.researchandmarkets.com/reports/4775648/intelligent-virtual-assistant-market-global
- [2] P. Gupta ve V. Gupta, "A survey of text question answering techniques", International Journal of Computer Applications, vol. 53, no. 4, pp. 1-8, Eylül 2012.
- [3] M. A. C. Soares ve F. S. Parreiras, "A literature review on question answering techniques, paradigms and systems", Journal of King Saud University – Computer and Information Sciences, vol. 32, pp. 635-646, 2020.
- [4] M. Sayıt ve B. Karaoğlu, "Soru cevaplama üzerine bir araştırma", Akademik Bilişim, Kütahya, Türkiye, 31 Ocak - 2 Şubat, 2007.
- [5] D. Molla ve J. L. Vicedo, "Question answering in restricted domains: An overview", Computational Linguistics, vol. 33, no. 1, pp. 41-61, Mart 2007.
- [6] P. Yaşar, İ. Şahin ve E. Adalı, "Türkçe bilgi kaynaklı soru yanıtlama dizgesi", Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, vol. 12, no. 2, pp. 40-49, 2019.
- [7] A. Turing, "Computing machines and intelligence", Mind, vol. LIX, no. 236, pp. 433-460, Ekim 1950.
- [8] Ö. Özyurt ve C. Köse, "Türkçe tabanlı diyalog sistemi tasarımı ve kodlanması", Elektrik, Elektronik ve Bilgisayar Mühendisliği Sempozyumu ve Fuarı (ELECO), Bursa, Türkiye, 6-10 Aralık, 2006.
- [9] S. Topçu, C. Şen ve M. F. Amasyalı, "Türkçe sohbet robotu", Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu (ASYU), Trabzon, Türkiye, 3-4 Temmuz, 2012.
- [10] M. F. Amasyalı ve B. Diri, "Bir soru cevaplama sistemi: BayBilmiş", Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, vol. 1, no. 1, pp. 37-51, 2005.
- [11] S. İlhan, N. Duru, Ş. Karagöz ve M. Sağır, "Metin madenciliği ile soru cevaplama sistemi", Elektrik, Elektronik ve Bilgisayar Mühendisliği Sempozyumu ve Fuarı (ELECO), Bursa, Türkiye, 26-30 Kasım, 2008.
- [12] C. Derici ve diğerleri, "Türkçe soru cevaplama sistemlerinde kural tabanlı odak çıkarımı", 22. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SİU), Trabzon, Türkiye, 23-25 Nisan, 2014.
- [13] N. Albayrak, A. Özdemir ve E. Zeydan, "Yapay zeka tabanlı rehber robotlara genel bir bakış ve örnek bir rehber robot uygulaması", 26. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SİU), İzmir, Türkiye, 2-5 Mayıs, 2018.
- [14] A. Andrenucci, "Automated question-answering techniques and the medical domain", 1st International Conference on Health Informatics (HEALTHINF), Funchal, Madeira, Portekiz, 28-31 Ocak, 2008.
- [15] M. Polignano ve diğerleri, "HealthAssistantBot: A personal health assistant for the Italian language", IEEE Access, vol. 8, pp. 107479-107497, Haziran 2020.
- [16] A. S. Miner, L. Laranjo ve A. B. Kocaballi, "Chatbots in the fight against the COVID-19 pandemic", npj Digital Medicine, vol. 3, no. 65, 2020.
- [17] Y. Li ve diğerleri, "Jennifer for COVID-19: An NLP-powered chatbot built for the people and by the people to combat misinformation", ACL 2020 Workshop on Natural Language Processing for COVID-19 (NLP-COVID), Seattle, Washington, ABD, 9 Temmuz, 2020.
- [18] R. Tang ve diğerleri, "Rapidly bootstrapping a question answering dataset for COVID-19", 2020, *arXiv: 2004.11339 [cs.CL]*.
- [19] T. W. Bickmore ve diğerleri, "Patient and consumer safety risks when using conversational assistants for medical information: An observational study of Siri, Alexa, and Google Assistant", Journal of Medical Internet Research, vol. 20, no. 9, 2018.
- [20] M. D. Akın ve A. A. Akın, "Türk dilleri için açık kaynaklı doğal dil işleme kütüphanesi: Zemberek", Elektrik Mühendisliği, no. 431, pp. 38-44, 2007.
- [21] M. Light, G. S. Mann, E. Riloff ve E. Breck, "Analyses for elucidating current question answering technology", Natural Language Engineering, vol. 7, no. 4, pp. 325-342, Aralık 2001.