

**TOBB EKONOMİ VE TEKNOLOJİ ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**

**METİN TABANLI İNSAN ETKİLEŞİM İSPATI SİSTEMLERİ İÇİN  
İNSAN HESAPLAMA KULLANIMI**



**DOKTORA TEZİ**

**Hakan Ezgi KIZILÖZ**

**Bilgisayar Mühendisliği Anabilim Dalı**

**Tez Danışmanı: Prof. Dr. Kemal BIÇAKCI**

**Ağustos 2016**



Fen Bilimleri Enstitüsü Onayı

.....  
**Prof. Dr. Osman EROĞUL**  
Müdür

Bu tezin Doktora derecesinin tüm gereksinimlerini sağladığını onaylarım.

.....  
**Doç. Dr. Oğuz ERGİN**  
Anabilim Dalı Başkan Vekili

TOBB ETÜ, Fen Bilimleri Enstitüsü'nün **101111041** numaralı Doktora öğrencisi **Hakan Ezgi KIZILÖZ**'ün ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı "**METİN TABANLI İNSAN ETKİLEŞİM İSPATI SİSTEMLERİ İÇİN İNSAN HESAPLAMA KULLANIMI**" başlıklı tezi **10.08.2016** tarihinde aşağıda imzaları olan jüri tarafından kabul edilmiştir.

**Tez Danışmanı: Prof. Dr. Kemal BIÇAKCI** .....  
TOBB Ekonomi ve Teknoloji Üniversitesi

**Jüri Üyeleri: Doç. Dr. M. Fatih DEMİRCİ**  
(Başkan) .....  
TOBB Ekonomi ve Teknoloji Üniversitesi

**Prof. Dr. Bülent TAVLI** .....  
TOBB Ekonomi ve Teknoloji Üniversitesi

**Doç. Dr. Tolga Kurtuluş ÇAPIN** .....  
TED Üniversitesi

**Yrd. Doç. Dr. Tayfun KÜÇÜKYILMAZ** .....  
TED Üniversitesi



## TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, alıntı yapılan kaynaklara eksiksiz atıf yapıldığını, referansların tam olarak belirtildiğini ve ayrıca bu tezin TOBB ETÜ Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırlandığını bildiririm.

Hakan Ezgi KIZILÖZ



## ÖZET

Doktora Tezi

### METİN TABANLI İNSAN ETKİLEŞİM İSPATI SİSTEMLERİ İÇİN İNSAN HESAPLAMA KULLANIMI

Hakan Ezgi KIZILÖZ

TOBB Ekonomi ve Teknoloji Üniversitesi  
Fen Bilimleri Enstitüsü  
Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı: Prof. Dr. Kemal BIÇAKCI

Tarih: Ağustos 2016

İnsan Etkileşim İspatı (İEİ) testleri, otomatik saldırıları azaltmak için kullanılır. Bu testlerin insanlar tarafından kolaylıkla geçilebildiği fakat bilgisayar veya otomatik programlar tarafından geçilemeyecek şekilde tasarlandığı düşünülmektedir. Güvenlik ve kullanılabilirlik İEİ'ler için her zaman kritik bir problem olmuştur. Özellikle “erişilebilirlik” bir sistem gereksinimi olduğunda bu ihtiyaçlar daha ön plana çıkmaktadır. İşitsel İEİ'ler ya otomatik olarak çözülebilecek şekilde çok kolay, ya da insanların bile çözemeyeceği şekilde çok zor olup, genellikle meşru kullanıcılarla saldırganları güvenilir bir şekilde ayırt edemezler. Metin tabanlı İEİ'lerin kullanılabilirlik ve erişilebilirlik özellikleri tam istenilen şekilde olsa bile güvenlik sorunu henüz çözülememiştir. Metin tabanlı İEİ'leri üretmek için tamamen otomatik teknikleri kullanan güvenli bir sistem olmadığı göz önünde bulundurularak, bu tez çalışmasında güvenli şekilde metin tabanlı İEİ üretebilmek için insan hesaplamasının kullanılması önerilmiştir. Bu çalışmada halihazırda kullanımda olan metin tabanlı bir İEİ sisteminin, textCAPTCHA'nın, kullanılabilirliği Google'ın reCAPTCHA servisi ile karşılaştırılmıştır. Sonuçlar incelendikten sonra, SMARTCHA adında yeni bir sistem tasarlanmış ve uygulamaya geçirilmiştir. SMARTCHA, insanlar tarafından üretilen İEİ'lerin güvenli olduğunu anlayabilmek için otomatik olarak proaktif kontroller gerçekleştiren bir güvenlik motoru ve İEİ sorularının sayısını artırmak için insan hesaplama ile otomatik yöntemleri birleştiren bir modül içermektedir. Üç adet insan hesaplama çalışması yapılmış ve bu çalışmalarda İEİ operatörlerinden yararlanılmıştır. İEİ operatörleri SMARTCHA sistemi için toplamda yaklaşık

22.000 soru üretmişlerdir. Bu insan hesaplama çalışmalarının yöntemi, verimliliği ve sonuçları detaylı olarak analiz edilmiştir. SMARTCHA sisteminin kullanışlılığı 372 kullanıcının katıldığı geniş bir kullanıcı çalışmasıyla değerlendirilmiştir. Kullanıcılar SMARTCHA'nın metin tabanlı İEİ'lerini çözmeyi görsel reCAPTCHA İEİ'lerini çözmeye göre istatistiksel anlamlı bir şekilde daha eğlenceli bulmuşlardır. Bu çalışmada ayrıca, soru ön filtreleme ve otomasyon teknikleri kullanımının etkileri de incelenmiştir. Sonuçlara göre soru ön filtreleme SMARTCHA testlerinin çözülme süresini kısaltırken, otomasyon teknikleri uygulamak bu süreyi artırmaktadır. 31 adet görme engelli kullanıcı arasında yapılan başka bir kullanıcı çalışması da SMARTCHA'nın erişilebilirlik yönünden değerlendirilebilmesini sağlamıştır. Çalışma sonucuna göre SMARTCHA İEİ'lerini çözmek, yeni sürüm işitsel reCAPTCHA İEİ'lerini çözmeye göre daha eğlencelidir ve daha kısa zaman almaktadır. Her iki çalışma da metin tabanlı İEİ'lerin güvenli, kullanışlı ve erişilebilir İEİ'ler elde edebilmek için gelecek vaat ettiğini göstermektedir.

**Anahtar Kelimeler:** İnsan etkileşim ispatı, İnsan hesaplama, Kullanıcı çalışması, Güvenlik, Kullanışlılık, Erişilebilirlik



## ABSTRACT

Doctor of Philosophy

### LEVERAGING HUMAN COMPUTATION FOR PURE-TEXT HUMAN INTERACTION PROOFS

Hakan Ezgi KIZILÖZ

TOBB University of Economics and Technology  
Institute of Natural and Applied Sciences  
Department of Computer Engineering

Supervisor: Prof. Dr. Kemal BIÇAKCI

Date: August 2016

Human-Interaction Proofs (HIPs) are used to mitigate automated attacks. They are assumed to be easily passed by humans but not by computers or automated programs. Security and usability have always been a critical problem for HIPs, especially when “accessibility” is a system requirement. Audio HIPs usually cannot reliably distinguish attacks from legitimate use; they are either easy, and can be automatically solved, or hard even for humans. Even though purely text-based HIPs have desirable usability and accessibility attributes; they could not overcome the security problems yet. Given the fact that fully automated techniques to generate pure-text HIPs securely do not exist, leveraging human computation for this purpose is proposed in the thesis study. In the study, the usability of a currently used pure-text HIP service, textCAPTCHA, is compared against Google’s reCAPTCHA. After analyzing the results, a system called SMARTCHA is designed and implemented. SMARTCHA involves a security engine to perform automated proactive checks on the security of human-generated HIPs and also a module for combining human computation with automation to increase the number of HIP questions. HIP operators were employed in three human computation studies, in which they generated around 22,000 questions in total for SMARTCHA system. The methodology, efficiency and results of these human computation studies are analyzed in detail. The usability of SMARTCHA system is evaluated with a large user study of 372 participants. Users found solving pure-text HIPs of SMARTCHA system significantly more enjoyable than solving

reCAPTCHA visual HIPs. The effects of question pre-filtering and use of automation techniques are also evaluated in the study. Results suggest that question pre-filtering reduces solving time of SMARTCHA, whereas applying automation techniques increase it. Another user study among 31 visually impaired users helped evaluation of accessibility. The study results show that SMARTCHA takes less time and is more enjoyable to solve than the new reCAPTCHA audio HIPs. Both studies suggest that pure-text HIPs could be a promising solution for secure, usable and accessible HIPs.

**Keywords:** Human interaction proof, Human computation, User study, Security, Usability, Accessibility



## TEŞEKKÜR

Öncelikle doktora öğrenimim süresince bilgi birikimi ve deneyimleri ile bana yol gösteren, anlayış ve sabırla çalışmalarına yardımcı olan danışmanım sayın Prof. Dr. Kemal Bıçakcı'ya teşekkürü bir borç bilirim. Ayrıca, defalarca kez değerli vakitlerini ayıran ve yaptıkları yorumlarla tezime katkı sağlayan tez izleme komitesi üyeleri sayın Doç. Dr. M. Fatih Demirci ve sayın Prof. Dr. Bülent Tavlı'ya; tez savunmama katılarak yaptıkları yorumlarla bu çalışmaya değer katan sayın Doç. Dr. Tolga Kurtuluş Çapın ve sayın Yrd. Doç. Dr. Tayfun Küçükylmaz'a, son olarak eski bölüm başkanımız sayın Prof. Dr. Erdoğan Doğdu başta olmak üzere ders ve ders dışında aktardıkları bilgilerle bakış açımı genişleten TOBB Ekonomi ve Teknoloji Üniversitesi Bilgisayar Mühendisliği Bölümü'nün değerli öğretim üyelerine teşekkür ederim.

Tez çalışmasının insan hesaplama ve kullanıcı çalışması adımlarında yer alan ismini bildiğim ve bilmediğim bütün katılımcılara çok teşekkür ederim.

Doktora eğitimim boyunca sağladığı burs ile maddi destek sağlayan TOBB Ekonomi ve Teknoloji Üniversitesi'ne teşekkür ederim. 111E148 nolu TÜBİTAK 1002 projesi kapsamında bu tez çalışmasının maddi kaynağını sağladığı için TÜBİTAK'a ayrıca teşekkürlerimi sunarım.

Projenin her aşamasında yardımını esirgemeyen iş ve oyun arkadaşım Cumhuriyet Kılıç'a ve her an sağladığı manevi destekle hep yanımda olan Umur Özhan Şengül'e ayrıca teşekkür ederim. B11 ve 214 nolu laboratuvarlarda birlikte çalıştığım Fahri Aydos, Salih Atılay Oto, Tashtanbek Satiev, Ömer Selçuk Demirci'ye, üniversiteden yakın arkadaşlarım Uğur Şahin, Kazım Sarıkaya, Murat Karaöz'e ve adını sayamadığım diğer arkadaşlarıma ayrıca teşekkür ederim.

Beni büyütüp bugünlere getiren sevgili annem Kamuran Şansal ile beni öz evlatları gibi karşılıksız seven anne ve babam Emine ve Hamdi Deniz başta olmak üzere bütün aile üyelerime verdikleri destekten dolayı teşekkür ederim.

Son olarak, hayatıma girdiği ilk günden beri her gün daha mutlu uyanmamı sağlayıp, hastalıkta ve sağlıkta yanımda olan; sadece bu tez çalışmasının değil, kendisi olmasa benim için dünyanın yürümesinin imkanı olmayan sevgili eşim, diğer yarım, mutluluk kaynağım Firdevsi Ayça Deniz-Kızılöz'e sonsuz kez teşekkür ederim. Seni çok seviyorum ve kendimi seninle tamamlanmış hissediyorum. İyi ki varsın.



## İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET .....	iv
ABSTRACT .....	vi
TEŞEKKÜR .....	viii
İÇİNDEKİLER .....	ix
ŞEKİL LİSTESİ.....	xi
ÇİZELGE LİSTESİ.....	xiii
<b>1. GİRİŞ .....</b>	<b>1</b>
<b>2. GENEL BİLGİLER .....</b>	<b>5</b>
2.1 Literatür Taraması .....	5
2.2 Kullanımda Olan İEİ Sistemleri .....	7
2.2.1 textCAPTCHA .....	7
2.2.1.1 Güvenlik .....	8
2.2.1.2 Kullanışlılık .....	11
2.2.1.3 Ölçeklenebilirlik .....	11
2.2.2 reCAPTCHA .....	11
2.3 Kullanıcı Çalışması Adımları .....	14
2.3.1 Çalışmanın planlanması.....	15
2.3.2 Katılımcıların bulunması .....	15
2.3.3 Çalışmanın yapılması .....	16
2.3.4 Verilerin yorumlanması .....	16
2.3.4.1 Kolmogorov-Smirnov testi .....	17
2.3.4.2 Wilcoxon işaretli sıra testi .....	17
2.3.4.3 Mann-Whitney U testi.....	18
2.3.4.4 Friedman testi .....	18
2.3.4.5 Kendall uyum katsayısı (W) .....	18
<b>3. PROBLEM TANIMI VE ÖNERİLEN SİSTEM .....</b>	<b>21</b>
3.1 SMARTCHA.....	22
3.1.1 Güvenlik .....	22
3.1.2 Kullanışlılık .....	27
3.1.3 Ölçeklenebilirlik .....	28
<b>4. İNSAN HESAPLAMA ÇALIŞMALARI .....</b>	<b>33</b>
4.1 Sistem Mimarisi .....	33
4.1.1 Veritabanı .....	33
4.1.2 Kodlama .....	34
4.1.3 Sunucu yazılımı.....	34
4.1.4 Sunucu ve alan adı kiralanması .....	35
4.2 Sistemin Yapısı .....	35
4.3 Soru Hazırlama Süreci .....	38
4.3.1 Birinci çalışma.....	39
4.3.2 İkinci çalışma .....	39
4.3.3 Üçüncü çalışma.....	41

4.4 Sonular .....	41
4.4.1 Birinci alıřma .....	41
4.4.2 İkinci alıřma .....	43
4.4.3 Üüncü alıřma .....	44
4.5 Sonuların Analizi .....	45
4.5.1 Birinci alıřma .....	46
4.5.2 İkinci alıřma .....	48
4.5.3 Üüncü alıřma .....	48
4.5.4 Karřılařtırmalı istatistikler .....	49
4.6 ıkarılan Dersler .....	53
<b>5. KULLANICI ALIřMALARI .....</b>	<b>57</b>
5.1 textCAPTCHA Kullanıcı alıřması .....	57
5.1.1 Yöntem .....	57
5.1.2 Hipotezler .....	60
5.1.3 Sonular .....	61
5.1.4 Kısıtlar .....	64
5.1.5 Tartıřma .....	64
5.2 SMARTCHA Kullanıcı alıřması .....	65
5.2.1 Yöntem .....	66
5.2.2 Hipotezler .....	67
5.2.3 Sonular .....	68
5.2.4 Kısıtlar .....	73
5.2.5 Tartıřma .....	73
5.3 Görme Engelli Kullanıcılarla Yapılan Kullanıcı alıřması .....	74
5.3.1 Yöntem .....	74
5.3.2 Hipotezler .....	75
5.3.3 Sonular .....	76
5.3.4 Tartıřma .....	78
5.4 Sonuların Tartıřması .....	82
<b>6. SONU VE GELECEKTEKİ ALIřMALAR .....</b>	<b>89</b>
<b>KAYNAKLAR .....</b>	<b>93</b>
<b>EKLER .....</b>	<b>97</b>
<b>ÖZGEMİř .....</b>	<b>101</b>

## ŞEKİL LİSTESİ

	<u>Sayfa</u>
Şekil 1.1: Örnek bir İEİ testi.....	1
Şekil 1.2: slashdot.org sitesinde kullanılmakta olan ve otomatik olarak çözülebilen bir İEİ örneği. ....	2
Şekil 1.3: random.irb.hr sitesinde kullanılmakta olan ve otomatik olarak çözülebilen bir Math CAPTCHA örneği. ....	3
Şekil 2.1: textCAPTCHA servisinin sunduğu testlerden renk soru tipindeki soruları çözebilen Python fonksiyonu [22].....	10
Şekil 2.2: Eski kitapların sayısallaştırılmasında kullanılan görsel reCAPTCHA İEİ'lere bir örnek. ....	12
Şekil 2.3: reCAPTCHA servisinin şu anda kullandığı görsel İEİ testlerine bir örnek. ....	13
Şekil 3.1: SMARTCHA güvenlik motorunun akış şeması. ....	26
Şekil 3.2: Baz sorulardan otomatik yöntemlerle soru üretme işlemi. ....	29
Şekil 4.1: Operatör giriş sayfası.....	36
Şekil 4.2: Operatörlerin kendisine verilen cevaba soru yazdığı sayfa .....	37
Şekil 4.3: Soru ve cevabı operatörlerin belirlediği sayfa.....	37
Şekil 4.4: Operatörlerin aldıkları ücret tipine göre çalışma verimlilikleri. ....	42
Şekil 4.5: İkinci insan hesaplama çalışmasında operatörlerin hazırladıkları soru ve güvenli soru sayıları. ....	44
Şekil 4.6: İkinci insan hesaplama çalışmasında günlük hazırlanan soru sayısı. ....	45
Şekil 4.7: Üçüncü insan hesaplama çalışmasında operatörlerin hazırladıkları toplam soru ve güvenli soru sayıları.....	46
Şekil 4.8: Her bir çalışmada hazırlanan güvenli soru sayısının toplam soru sayısına oranı (%). ....	50
Şekil 4.9: Her bir çalışmada operatörlerin ortalama soru hazırlama süreleri..	50
Şekil 4.10: Her bir çalışmada güvenlik motorunun soru başına ortalama çalışma süresi. ....	51
Şekil 4.11: Her bir çalışmada güvenlik motorunun soru başına ürettiği muhtemel cevap sayısı.....	51
Şekil 4.12: Her bir çalışma için Bing arama motorunun soru başına bulduğu kelime sayısı.....	52
Şekil 4.13: Her bir çalışma için güvenli olmayan soruların hangi aşamada güvenli olmadığına karar verildiğinin incelenmesi. ....	53
Şekil 5.1: Kullanıcı çalışmasında kullanılan görsel İEİ arayüzü.....	59
Şekil 5.2: Kullanıcı çalışmasında kullanılan metin tabanlı (solda) ve işitsel (sağda) İEİ arayüzleri. ....	60
Şekil 5.3: Kullanıcı çalışmasında kullanılan anket sayfası arayüzü. ....	61

Şekil 5.4: SMARTCHA ve görsel İEİ'lerin çözülme süreleri.....	70
Şekil 5.5: SMARTCHA ve görsel İEİ'lerin anket sonuçları (Ankette verilen '5' cevabı tamamen katılıyorum, '1' cevabı ise tamamen katılmıyorum anlamına gelmektedir).....	71
Şekil 5.6: Kullanıcıların SMARTCHA ve görsel İEİ'leri "Birinci" olarak seçme oranları.....	72
Şekil 5.7: SMARTCHA ve görsel İEİ'lerin başarılı çözüm oranları.....	73
Şekil 5.8: SMARTCHA ve işitsel İEİ'lerin çözme süreleri.....	78
Şekil 5.9: Kullanıcıların SMARTCHA ve işitsel İEİ sistemleriyle üçüncü karşılaşmalarındaki çözme süreleri. ....	79
Şekil 5.10: SMARTCHA ve işitsel İEİ'lerin anket sonuçları (Ankette verilen '5' cevabı tamamen katılıyorum, '1' cevabı ise tamamen katılmıyorum anlamına gelmektedir).....	80
Şekil 5.11: Kullanıcıların SMARTCHA ve işitsel İEİ'leri "Birinci" olarak seçme oranları.....	81
Şekil Ek.1: Birinci insan hesaplama çalışmasının duyurusu için kullanılan afiş.....	100



## ÇİZELGE LİSTESİ

	<u>Sayfa</u>
Çizelge 2.1: textCAPTCHA servisinde yer alan İEİ'lerden örnekler. ....	7
Çizelge 4.1: Birinci insan hesaplama çalışmasında üretilen İEİ testlerine örnekler. ....	42
Çizelge 4.2: İkinci insan hesaplama çalışmasında üretilen İEİ testlerine örnekler. ....	43
Çizelge 4.3: Üçüncü insan hesaplama çalışmasında üretilen İEİ testlerine örnekler. ....	44
Çizelge 4.4: Yapılmış olan çalışmaların grafiklerde kullanılmak üzere numaralandırmaları. ....	49
Çizelge 5.1: textCAPTCHA üzerine yapılan kullanıcı çalışmasında toplanan demografik bilgiler. ....	58
Çizelge 5.2: İlk kullanışlılık çalışmasında toplanan verilerin ortalama ve standart sapma değerleri (parantez içinde). ....	63
Çizelge 5.3: Kullanıcı çalışmasında test edilmiş SMARTCHA sistemindeki durumlar. ....	66
Çizelge 5.4: İkinci kullanışlılık çalışmasında toplanan verilerin ortalama ve standart sapma değerleri (parantez içinde). ....	69
Çizelge 5.5: SMARTCHA ve işitsel İEİ'lerin saniye cinsinden ortalama çözme süreleri. ....	77
Çizelge 5.6: Erişilebilir İEİ sistemlerindeki soru sayıları hakkında bilgi. ....	86
Çizelge 5.7: Kullanıcı çalışmalarında kullanılan İEİ servislerinin karşılaştırması. ....	87
Çizelge Ek.1: SMARTCHA veritabanı tabloları. ....	98
Çizelge Ek.2: SMARTCHA sistemindeki İngilizce soruların sayısını artırmak için kullanılabilecek ek örüntüler. ....	99



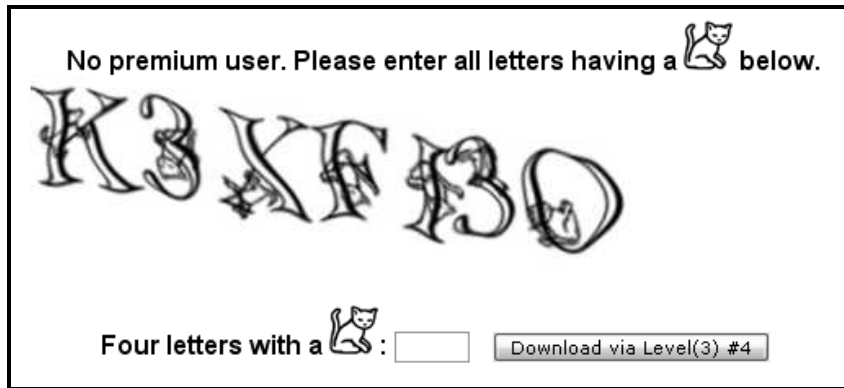
## 1. GİRİŞ

İnsan Etkileşim İspatı (İEİ) sistemleri, insanların kolayca geçebileceği fakat bilgisayarların ya da programların geçemeyeceği testlerdir. İEİ sistemlerinin kullanılması ile otomatik olarak yapılan servis engelleyici saldırılar, internet sitelerinde otomatik kullanıcı hesabı almak veya internet sitelerine otomatik olarak yorum yazmak gibi istenmeyen aktiviteler engellenmeye çalışılmaktadır. İEİ sistemleri bu amaçlar doğrultusunda günümüzde oldukça yaygın olarak kullanılmaktadır.

İEİ testleri her ne kadar teknik olarak işe yarıyor gözükse de, Şekil 1.1'de de bir örneği görülebileceği üzere, pratikte insanların kolay kullanımını merkeze alan bir yapıda değildir. En az kullanılabilirlik problemi kadar kötü olan erişilebilirlik problemlerine karşı bir çözüm olmaktan da uzaktır. Erişilebilirlik problemlerine, göz sağlığı yerinde olmayan kullanıcıların bu testleri geçebilmesinin mümkün olmayışı örnek olarak verilebilir. Oysa tanımı gereği İEİ testlerini bütün insanların geçebilmesi gerekmektedir.

Yukarıda bahsedilen erişilebilirlik problemine bir çözüm olarak grafik öğeler içermeyen ve tamamen metin olarak sunulan İEİ'ler düşünülebilir. Fakat tamamen otomatik yöntemlerle üretilen metin İEİ testleri bilgisayarlar tarafından otomatik olarak çözülmeye daha elverişli olup, istenilen güvenliği sağlayamamaktadır.

Günümüzde çoğu pratik uygulamada tercih edilen karakter veya sözcük tanıma tabanlı İEİ'ler, Şekil 1.2'deki gibi karmaşık olsalar dahi, gelişen optik karakter



Şekil 1.1: Örnek bir İEİ testi.

tanıma ve kesimleme teknolojisi karşısında güvensiz hale gelmişlerdir [1, 2, 3]. Yapılan bir çalışmada [3] bilgisayarların, aralarında Msn, Yahoo ve Google gibi sitelerin de bulunduğu yedi farklı sistemin kullandığı İEİ testlerini en az normal görebilme yetisine sahip insanlar kadar iyi çözebildiği, hatta sorular karmaşıklaştıkça insanlardan daha başarılı oldukları tespit edilmiştir. Alternatif olarak önerilen ve pek çok farklı türü olan görsel İEİ sistemleri [1] İEİ'lerin güvenlik problemlerine çözüm potansiyeline sahip olmakla birlikte, kullanışlılık ve bilhassa erişilebilirlik problemlerine çare olamamaktadırlar. Görsel İEİ'ler gözleri iyi görmeyen kullanıcılarca yine çözümü imkansız testlerdir.

İnsan Etkileşim İspatı kavramı ilk kez 1996 yılında Moni Naor tarafından ortaya atılmıştır [4]. Bu kavram pratikte ilk olarak Altavista şirketi tarafından 2001 yılında kullanılmaya başlanmıştır [5]. İEİ kavramını nitelemek için kullanılan CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) terimi ise ilk defa 2003 yılında kullanılmıştır [6].

İEİ'lerin erişilebilirlik problemi uzun zamandır bilinmektedir. Bu problemin çözümü için W3C<sup>1</sup> tarafından önerilen alternatiflerden [7] en yaygın olarak tercih edilen işitsel İEİ'lerdir. İşitsel İEİ'ler gürültülü bir ortamda ve/veya kesik kesik seslendirilen harf veya kelimelerin doğru bir şekilde anlaşılmasını ve klavye ile girilmesini gerektirmektedir. Yahoo, Google, Ebay gibi çoğu büyük şirket tarafından tercih edilen işitsel İEİ'ler hem kullanışlılık problemleri içermekte [8] hem de başarılı güvenlik saldırılarına hedef olmaktadır [9, 10]. Yapılan bir çalışmada [8] işitsel İEİ'lerin çözümünün ortalama 25 saniyelere varan sürelerde gerçekleştiği, doğru çözme oranının ise Google örneğinde %35'e kadar düştüğü gözlemlenmiştir.

Grafik unsurları içermeyen tamamen metin tabanlı İEİ'ler ekran yazılarını sesli olarak okuyan sistemler tarafından algılanabildiklerinden görme engelli İnternet kullanıcılarının çözebildikleri testler arasında yer almaktadır [11]. Bu tür test-

---

<sup>1</sup>World Wide Web Consortium, <http://www.w3.org>.

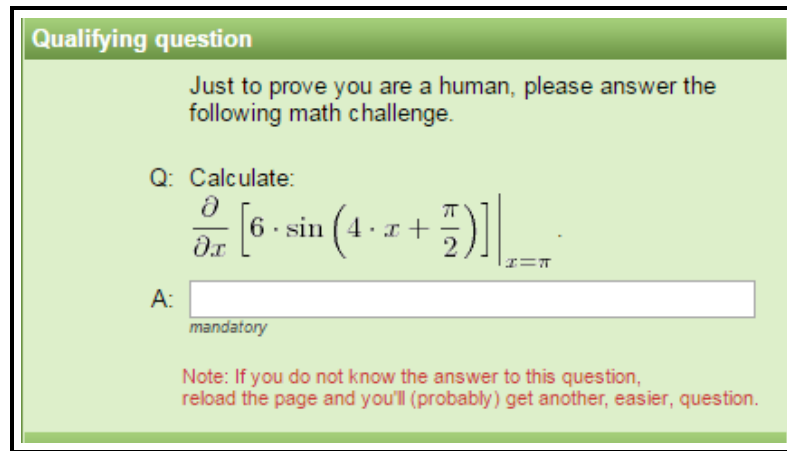


Şekil 1.2: slashdot.org sitesinde kullanılmakta olan ve otomatik olarak çözülebilen bir İEİ örneği.

lerin tamamen otomatik yöntemlerle “güvenli” bir şekilde üretilebilmesi henüz çözümlenmemiş bir araştırma problemidir [12, 13]. Grafik unsurları içermeyen İEİ'lere verilecek enteresan bir örnek Hırvatistan'daki bir enstitü tarafından geliştirilen “Math CAPTCHA” yöntemidir. Şekil 1.3'te bir örneği görülebilecek olan bu yöntemde basit aritmetik sorular yerine logaritma, trigonometri vb. ileri düzey matematik bilgisi gerektiren sorular İEİ testi olarak kullanılmaktadır. Ne yazık ki bu sistem dahi başarılı saldırılara hedef olmuştur [14].

İEİ'lere otomatik olarak doğru cevap verilme olasılığının hangi seviyelerde olmasının kabul edilebilir olduğu konusunda uzmanlar hemfikir değillerdir. Fakat, bütün sistemlerin cevap kümesi sonlu sayıda elemana sahip olduğundan, bu oranın %0 olamayacağı bilinen bir gerçektir. Bununla birlikte, diğer bazı güvenlik araştırmacıları İEİ'lerin pratik ve teorideki sağladığı güvenliğin karıştırılmaması gerektiğini ve düşünülebilen en basit İEİ'nin bile pratikte işe yarayabileceğini belirtmişlerdir [15]. Bir web sayfasında kullanıcıya metin olarak sunulan bir kelimenin (örneğin “portakal” kelimesinin) bir metin kutusuna klavyeden girilmesini gerektiren bir sistem bu duruma bir örnek olabilir.

Bu tez çalışmasının devamı şu şekilde düzenlenmiştir: Bölüm 2'de metin tabanlı İEİ sistemleri ile ilgili literatür taraması verilmiş, bu çalışmada kullanılan İEİ sistemleri tanıtılmış ve kullanıcı çalışmalarında izlenmesi gereken adımlar anlatılmıştır. Bölüm 3'te problem tanımı yapılmış ve bu problemi çözebilmek için önerilen yöntemden bahsedilmiştir. Önerilen yöntemin yapılan üç insan hesaplama çalışması ile gerçekleştirilmesi ve bu çalışmaların analizleri Bölüm 4'te detaylıca anlatılmıştır. Önerilen sistemin kullanılabilirliğini ölçmek için yapılan üç kullanıcı çalışmasının yöntemi, hipotezleri, sonuçların analizi ve tartışmaları Bölüm 5'te



Şekil 1.3: random.irb.hr sitesinde kullanılmakta olan ve otomatik olarak çözülebilen bir Math CAPTCHA örneği.

verilmiştir. Tez çalışması, Bölüm 6'da yer alan son yorumlar ve gelecek çalışmalarla bitirilmiştir.



## 2. GENEL BİLGİLER

Bu bölümde öncelikle metin tabanlı İEİ'ler hakkında detaylı bir literatür taraması, sonrasında halihazırda kullanımda olan bazı İEİ sistemleri, son olarak da kullanıcı çalışmalarında izlenen adımlar hakkında kısaca bilgiler verilmiştir.

### 2.1 Literatür Taraması

Metin tabanlı İEİ konsepti Godfrey tarafından ortaya atılmıştır [12]. Godfrey, bir paragraf içerisindeki rastgele seçtiği bir kelimeyi değiştirerek yerine yeni bir kelime yazmış ve kullanıcılara yeni paragrafı göstererek onlardan değiştirilmiş kelimeyi bulmalarını istemiştir. Önermiş olduğu bu sisteme 3-gram modeli kullanılarak tasarladığı bir saldırıda %39 başarı elde etmiş ve bu sebeple sistemin kullanılamayacağını belirtmiştir. Bunun üzerine Godfrey aynı çalışmada ikinci bir sistem önerisinde bulunmuştur. Bu öneriye göre biri gerçek ve kurallı, birisi rastgele oluşturulmuş iki cümleye bir 3-gram dönüşümü uygulamıştır. Kullanıcılardan, uygulanan 3-gram dönüşümü sonrasında değişen iki cümleden ilk baştaki gerçek ve kurallı cümle ile rastgele oluşturulmuş cümleyi ayrıştırmasını istemiştir. Ancak sonuçlara göre kullanıcıların başarı oranı oldukça düşük seviyelerde kalmıştır.

Bir başka çalışmada Bergmair *ve diğerleri* [16] metin tabanlı İEİ'ler oluşturabilmek için anlam kayması kullanmayı önermişlerdir. Bunun için bir cümle içerisinde birden fazla anlamı bulunan (eş sesli) bir kelimeyi seçmiş ve bu kelimenin her bir farklı anlamı için o anlama gelen başka kelimeleri (eş anlamlılarını) bulmuşlardır. Sonrasında seçmiş oldukları eş sesli kelimenin bütün anlamlarındaki eş anlamlılarını o cümle içerisinde onun yerine yazarak yeni cümleler türetmişlerdir. Kullanıcılardan gerçek cümlede bulunan kelimenin gerçek anlamına gelen bütün anlamdaşlarının bulunduğu cümleleri seçmelerini istemişlerdir. Örneğin Türkçe'de *yüz* kelimesinin bir anlamındaki eş anlamlıları sima, çehre, surat iken, başka bir anlamındaki eş anlamlıları yan ve taraf kelimeleridir. Bir cümle içerisinde *yüz* kelimesi sima/çehre/surat anlamında kullanılmışken bu eş sesli kelimenin yerine diğer eş anlamlıları olan yan/taraf kelimeleri konulduğunda cümlede anlam kayması olacaktır. Kullanıcılardan anlam kayması olmayan cümleleri seçmeleri istenmiştir.

Ximenes *ve diğerlerinin* yaptığı başka bir çalışmada [17] metin tabanlı İEİ'leri oluşturmak için Tak-Tak şakalarından faydalanılmıştır. Bu çalışmada kullanıcılara üç adet Tak-Tak şakası gösterilmiştir. Bu şakalardan sadece bir tanesi gerçekten mizah ögesi barındırırken, diğer iki tanesi sadece format olarak Tak-Tak şakasına benzetilmiş fakat mizah ögesi içermeyen metinlerdir. Bu çalışmada kullanıcıların gerçekten mizah ögesi barındıran metni bulmaları istenmiştir. Bu şakaların Türkçe'deki en yakın karşılığı "Sana X'in selamı var." ile başlayan şakalardır. Örneğin;

–Sana Hasan'ın selamı var.  
+Hangi Hasan?  
–Floresan.

metni mizahi bir değer taşıırken,

–Sana Hasan'ın selamı var.  
+Hangi Hasan?  
–Bizim sınıftaki Hasan.

metni düzgün ve kurallı bir yapıda olmasına rağmen mizah ögesi bulundurmamaktadır. İEİ testini geçebilmek için arka arkaya iki farklı test çözmeleri istenen kullanıcıların sadece %30'u testi başarıyla geçebilmiştir.

Yamamoto *ve diğerleri* [18] metin tabanlı İEİ testleri üretmek için otomatik tercüme araçlarını kullanmayı önermiştir. Bu sistemde cümleler otomatik tercüme araçları kullanılarak önce başka bir dile çevrilmekte, sonrasında elde edilen tercüme metin tekrar ana diline çevrilmektedir. Bu araçlar günümüzde mükemmel çeviri yapamadıklarından diller arası tercüme yapılırken bozulmalar olmakta ve dolayısıyla gerçeğe yakın fakat gramer kurallarına uymayan cümleler elde edilmektedir. Bu sistemde kullanıcılara sadece 5 tanesi düzgün ve kurallı olan toplam 15 cümle gösterilmiş ve kullanıcılardan düzgün ve kurallı olan 5 tanesini bulmaları istenmiştir. Çalışmanın sonucunda yazarlar bu sistemin kullanılabilirliğinin yeterli olmadığını ve iyileştirilmesi gerektiğini belirtmişlerdir. Bu çalışmada yazarlar tarafından vurgulanan önemli bir nokta, güvenlik sebebiyle geçerli cümlelerin internette ulaşılabilecek kaynaklardan alınmaması gerekliliğidir. Ancak yazarlar geçerli cümlelerin otomatik yöntemlerle nasıl elde edileceği hakkında geçerli bir yöntem sunmamışlardır.

Chew *ve diğerleri* [19] işbirlikçi filtreleme kullanarak bir İEİ sistemi tasarlamışlardır. Önerdikleri sistemde İEİ testlerinin net bir cevabı yoktur. Bu sistem ilk



başta gerçek kullanıcıların cevapları ile eğitilir ve daha sonra yeni gelen her bir kullanıcının verdiği cevaplar ilk cevaplar ile karşılaştırılır. Yazarlar önerdikleri sistemde farklı zevklere sahip meşru kullanıcıların testi geçemeyebileceğini belirtmiştir. Ayrıca bu sistemin bir İEİ sistemi olarak kullanılabilmesi için yeterli olmadığını da eklemiştir.

Bir diğer çalışmada Johansson ve Östlund [20] alt alta iki cümle göstererek kullanıcılardan bu cümleleri birbiriyle karşılaştırmalarını istemiştir. Bu sistemde cümleleri inceleyen kullanıcılar bu cümlelerin ikisinin de kuralsız olduğu, sadece ilkinin ya da sadece ikincisinin kurallı olduğu, ya da ikisinin de kurallı olup cümlelerin birbirleriyle anlamsal olarak benzeşip benzeşmediği yönünde bir geri bildirim vermelidir. Güvenlik sebebiyle bir kullanıcının meşruiyetini kanıtlayabilmesi için arka arkaya 16 test çözmesi gerektiği belirtilmiştir.

## 2.2 Kullanımda Olan İEİ Sistemleri

Bu bölümde halihazırda kullanımda olan textCAPTCHA ve reCAPTCHA servislerine dair bilgiler verilecektir.

### 2.2.1 textCAPTCHA

textCAPTCHA uygulaması [21] aktif olarak kullanılmakta olan metin tabanlı bir İEİ sistemidir. Uygulamanın web sayfasında 180 milyonun üzerinde soruya sahip bir veritabanından bahsedilmektedir. Ücretsiz bir web servisi olan textCAPTCHA, WordPress ve bunun gibi pek çok web sitesine günlük ortalama 271.000 civarında soru sunmaktadır. Soruların nasıl üretildiği konusunda bir bilgiye rastlanmamakla birlikte, Çizelge 2.1’de örnekleri görülen soru türlerinden ve veritabanının büyüklüğünden otomatik yöntemlerin kullanıldığı sonucu çıkarılmaktadır.

Çizelge 2.1: textCAPTCHA servisinde yer alan İEİ’lerden örnekler.

---

Arm, bee or elephant: the body part is?

Nine + 8 is what?

What is the 1st colour in the list stomach, pink, lion, brown, tracksuit and green?

The 2nd colour in green, red and house is?

---

textCAPTCHA servisinin yapısı, bir İEİ sistemin sağlaması gereken özellikler kapsamında üç alt başlıkta incelenebilir:

### 2.2.1.1 Güvenlik

Bu uygulamanın en büyük dezavantajı güvenlik konusundadır. Soru üretmek için birkaç soru tipi belirlenerek buna göre otomatik soruların üretildiği gözlemlenmektedir. Dolayısıyla bu soru tiplerini tespit ederek buna göre doğru cevabı otomatik olarak bulacak kodu yazmak oldukça kolaydır. Nitekim internette açık kaynak kodlu geliştirilen bir projeye [22] söz konusu İEİ sorularına otomatik olarak %99,5 başarı oranıyla doğru cevap vermek mümkün olmaktadır.

Python programlama dili ile geliştirilmiş bu projede textCAPTCHA servisinin İEİ testi olarak kullandığı soru tipleri belirlenmiş ve bu soru tiplerini çözülebilen örüntüler oluşturulmuştur. textCAPTCHA servisinin sağladığı soru sırasıyla renk, isim, vücut kısımları, kelime veya sayı içerisindeki yeri bulma, basit toplama veya çıkarma işlemi yapma, bir liste sayı arasında seçim yapma ve günler ile ilgili örüntülerin bulunduğu fonksiyonlara gönderilerek doğru cevap bulunmaya çalışılır. Fonksiyonlardan bir tanesi cevap döndürdüğü anda diğer fonksiyonların çalışmasına gerek kalmaz.

Örneğin renk soru tipi için hazırlanmış olan fonksiyonun çalışma algoritması aşağıdaki gibidir:

- Renk soru tipine gelen soru metni küçük harflere çevrilir ve kelimelerine ayrılır.
- Bu kelimeler arasında “colour” ya da “colours” kelimelerinin geçip geçmediği incelenir. Eğer bu kelimeler soru içerisinde geçmiyorsa sorunun renk soru tipine ait olmadığına karar verilerek bu fonksiyondan çıkılır.
- Eğer bu kelimeler arasında “colour” kelimesi varsa, cevabın tekil bir renk ile ilgili olduğu sonucuna varılır. Bu durumda soru metninde yer alan bütün renk adları tespit edilir.
  - Soru metninde yalnızca bir adet renk adı geçmişse, bulunan renk adı cevap olarak kaydedilir ve fonksiyondan çıkılır.
  - Tekil bir renk aranmasına rağmen soru metninde birden fazla renk adı geçmişse soru metninde sıralama belirten ifadeler aranır (örneğin

“first”, “2nd”, vb.). Sıralama belirtilen ifadede hangi sıradaki rengin arandığı bilgisi çözümlendikten sonra bütün renk adları sıralamasında ilgili sırada geçen renk adı cevap olarak kaydedilir ve fonksiyondan çıkarılır.

- Eğer bu kelimeler arasında “colours” kelimesi varsa, soruda kaç tane renk adı geçtiği bilgisinin sorulduğu anlaşılır. Soru metninde geçen renk adı sayısı hesaplanarak cevap olarak kaydedilir ve fonksiyondan çıkarılır.

Renk soru tipi ile ilgili bu çözümlenmeler, Şekil 2.1’de de görülebildiği üzere, yorumlarla birlikte toplam 39 satırlık bir kod ile yapılabilmektedir.

Diğer soru tiplerinden kısaca bahsetmek gerekirse;

- İsim soru tipini çözen fonksiyon “What is Susan’s name?” gibi insan ismi soran soruları çözebilmektedir.
- Vücut kısımları soru tipini çözen fonksiyon renk soru tipine benzer şekilde çalışmakta ve “Cat, apple, finger, elephant or hospital: the body part is?” gibi vücut parçası soran soruları çözebildiği gibi, “The list chin, cat, head, toe, T-shirt and hair contains how many body parts?” gibi sorunun içerisinde kaç tane vücut parçasının geçtiğini soran soruları da çözebilmektedir.
- Kelime veya sayı içerisinde yer bulma soru tipini çözen fonksiyon “What is the 6th digit in 7911863?” gibi soruları çözebilmektedir.
- Basit toplama veya çıkarma işlemi yapan soru tipini çözen fonksiyon ise “What is ten - 3?” gibi basit aritmetik işlemleri çözebilmektedir. Bunu yaparken en fazla iki basamaklı sayıların tespiti yapılmış ve yapılacak işlemin seçimi için soru metni içerisinde “plus”, “+”, “add” ya da “minus”, “-”, “subtract” ifadeleri aranmıştır.
- Bir liste arasından seçim yapma soru tipini çözen fonksiyon, bir liste sayı arasından en büyüğünün, en küçüğünün veya belirli bir sıradakinin arandığı soru tipidir. Bu soru tipi “Sixty two, 14, 75, fifty three, fifty seven or seventy: the largest is?” gibi soruları çözebilmektedir.
- Günler soru tipini çözen fonksiyonda ise önce soru metni içerisinde “weekend” kelimesinin var olup olmadığı incelenir ve bu kelimenin bulunması durumunda “Saturday” ya da “Sunday” kelimelerinden ilk geçen kelime cevap olarak döndürülür. Eğer soru metninde “weekend” kelimesi bulunmuyorsa soru metni içerisinde “today” kelimesi aranır ve soru metninde geçen

```

def solve(question):

    tokens = re.sub(r'[^\w\d\s]', '', question).lower().split(' ')
    found = []
    which = -1

    # Singular colour:
    # 'The pink church is what colour?' - pink
    # 'What is the 1st colour in the list coat, purple, yellow, black, pink and white?' - purple
    if 'colour' in tokens:

        for i in range(len(tokens)):

            # Go through each token and save those that are colors and those that are ordinal.
            if tokens[i] in colors:
                found.append(tokens[i])
            elif tokens[i] in ordinals:
                which = ordinals[tokens[i]]

        if len(found) == 1 or which == -1:
            return found[0]
        else:
            return found[which - 1]

    # Plural colours:
    # 'The list blue, brown, T-shirt, white and purple contains how many colours?' - four
    # 'Red, cheese and blue: how many colours in the list?' - two
    elif 'colours' in tokens:

        for i in range(len(tokens)):

            # Go through each token and save those that are colors.
            if tokens[i] in colors:
                found.append(tokens[i])

        return str(len(found))

    else:
        return None

```

Şekil 2.1: textCAPTCHA servisinin sunduğu testlerden renk soru tipindeki soruları çözebilen Python fonksiyonu [22].

gün değeri bulunur (örneğin Thursday). Daha sonra “today”, “yesterday” ya da “tomorrow” gibi bulunmuş olan gün değeriyle ilişkilendirilecek bir kelime aranır. Bulunan ilişkiye göre cevap hesaplanarak kaydedilir. Bu soru tipi “Which day from Sunday, Thursday, Tuesday or Monday is part of the weekend?” ve “What day is today, if yesterday was Wednesday?” sorularına cevap verebilmektedir.

Çizelge 2.1’de sunulan örnek sorular incelendiğinde %99,5 başarı oranında yapılan saldırıların başarısı şaşırtmamaktadır. Örnek sorulardan çıkarılabilecek bir başka sonuç ise, cevabın soru içerisinde yer alması durumunda soruda yer alan kelimelerden rastgele birinin cevap olarak seçildiği saldırılarda dahi başarı oranının hayli yüksek seviyelerde olacağıdır.

### 2.2.1.2 Kullanışlılık

Günlük ortalama 271.000 soru servis etmesine rağmen textCAPTCHA'nın kullanılışlılığını ölçen bir çalışmaya rastlanmamıştır. Bu servisin pratikte tercih ediliyor olması kullanılışlılık konusunda başarılı olduğuna dair ipuçları olarak algılanabilir. Yine de metin tabanlı bir servis olan textCAPTCHA'nın kullanılışlılığından somut olarak bahsedebilmek için bir kullanıcı çalışması yapılmıştır. Bu kullanıcı çalışmasının yöntemi ve sonuçları Bölüm 5.1'de ayrıntılı şekilde anlatılmıştır. Kısaca bahsetmek gerekirse textCAPTCHA testleri, reCAPTCHA testlerinden hem daha hızlı çözülebilmektedir, hem de kullanıcılar textCAPTCHA testlerini çözmeyi daha kolay ve eğlenceli bulmaktadır.

### 2.2.1.3 Ölçeklenebilirlik

Veritabanında 180 milyona yakın İEİ testi bulunması sebebiyle textCAPTCHA servisinde bulunan soruların tamamı bir veritabanına kaydedilip sonradan insan gücüyle teker teker çözülemez<sup>2</sup>. Bu sebeple textCAPTCHA servisi ilk bakışta ölçeklenebilir bir servis gibi gözükmemektedir. Fakat textCAPTCHA servisinin soruları küçük bir soru tipleri havuzunda yer alan sorular üzerine otomatik yöntemler uygulanarak üretildiğinden bu sorular otomatik olarak çözülebilmektedir. Bu sebeple, textCAPTCHA servisinin soru tipleri havuzunun pratik anlamda ölçeklenebilir olmadığı söylenebilir.

### 2.2.2 reCAPTCHA

reCAPTCHA [23] Google'a ait ücretsiz bir İEİ servisi olup, görsel ve işitsel İEİ'ler sağlamaktadır. Günümüzde birçok web sitesinde bu servis kullanılmaktadır.

Bölüm 5.1 ve 5.2'de anlatılan kullanıcı çalışmalarında görsel reCAPTCHA'nın o anki güncel sürümü kullanılmıştır. Bu sürüm, kullanıcılardan gürültülü bir resimdeki iki bozulmuş metni doğru bir şekilde tespit edip yazmalarını istemekteydi. Bu metinler eski bir kitabın ya da derginin bir kopyasından taranarak elde ediliyordu. Bu metinlerden birisi önceden optik karakter okuyucular tarafından başarıyla okunabilmiş bir metindi ve gerçek İEİ testi olarak kullanılıyordu. Di-

---

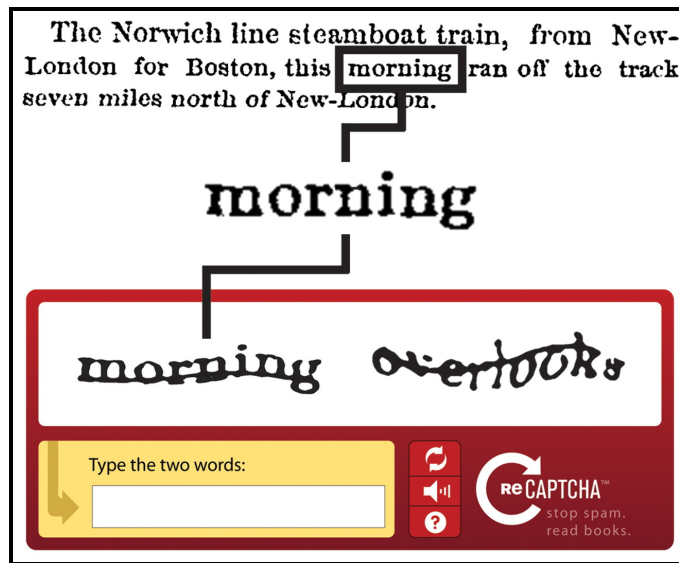
<sup>2</sup>Bölüm 5.1.3'te gösterildiği üzere, bir textCAPTCHA İEİ'nin çözümü için yaklaşık 15 saniye harcanmaktadır. Bu süre baz alındığında veritabanında bulunan soruların tamamını çözebilmek için 85 yıldan fazla iş gücü gerekmektedir.

ğer metin ise optik karakter okuyucu tarafından okunamadığı halde kullanıcılara birinci metin ile birlikte gösteriliyordu. Kullanıcılar İEİ testini başarılı şekilde geçebilmek için ekranda gördükleri her iki metni de doğru şekilde yazmaya çalıştıklarından kullanıcıların verdikleri cevaplar üzerine yapılan istatistiksel analizlerle daha önce optik karakter tanıma yöntemiyle otomatik olarak okunamamış kelimeler belirlenebilmekteydi. Bu sayede birçok eski kitap ve dergi sayısal ortama geçirilebilmiştir. Sayısallaştırma işleminin nasıl yapıldığı Şekil 2.2’de gösterilmektedir.

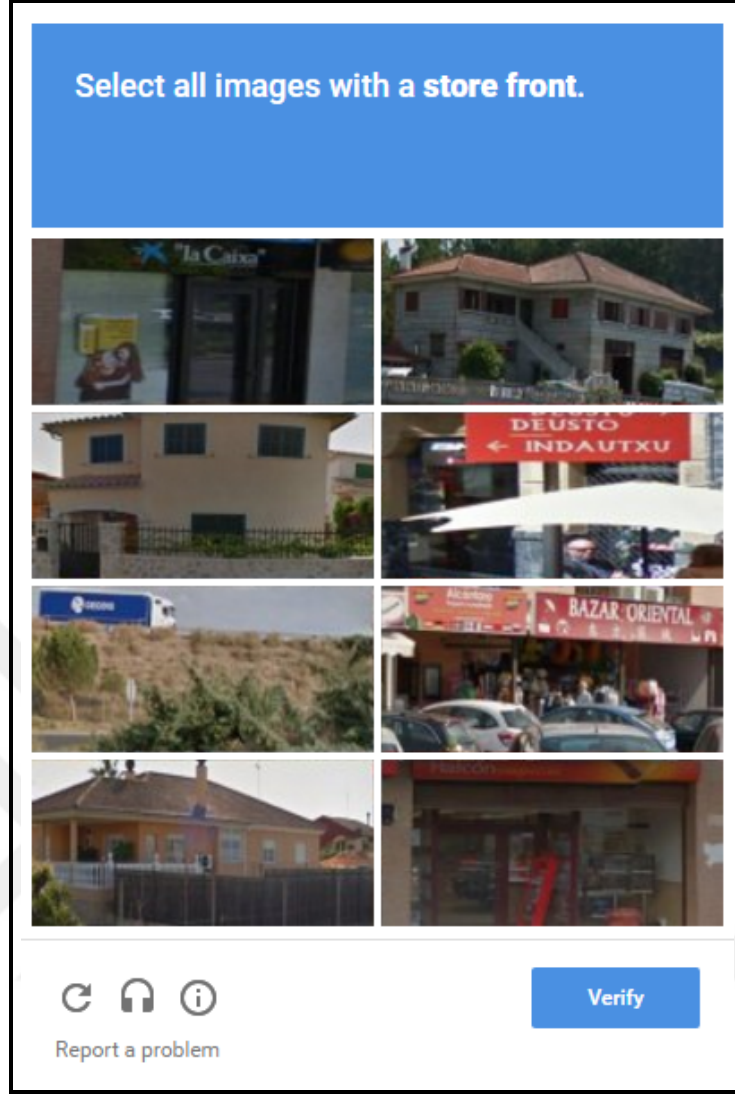
Görsel reCAPTCHA İEİ’lerin bu sürümünün güvenliği birçok araştırmacı tarafından incelenmiştir [2, 24, 25]. Bursztein *ve diğerleri* [2], 2011 yılındaki çalışmalarında görsel reCAPTCHA İEİ’leri otomatik bir şekilde çözememiştir. Sano *ve diğerleri* [25], 2013’teki çalışmalarında Hidden Markov modeli kullanarak testlerin %31,75’ini çözmeyi başarmıştır. Goodfellow *ve diğerlerinin* [24] yakın zamanda yaptıkları bir çalışmada, görsel reCAPTCHA İEİ’lerin en zor kategorisindeki testlerin derin katmanlı sinir ağı kullanılarak %99,8’lik bir başarı oranıyla çözülebildiği görülmüştür.

Görsel reCAPTCHA’nın güncel sürümüne ait bir örnek Şekil 2.3’te verilmiştir. Görüldüğü üzere, eski sürümde kelime tanıma ve yazma testleri kullanılmaktayken, yeni sürümde görseller içerisinde nesne tanıma tabanlı testler yer almaktadır.

Görsel reCAPTCHA İEİ’ler görme engelli kullanıcılar tarafından çözülemeyeceğinden reCAPTCHA bu kullanıcılara işitsel İEİ’ler sunmaktadır. Bölüm 5.1’de



Şekil 2.2: Eski kitapların sayısallaştırılmasında kullanılan görsel reCAPTCHA İEİ’lere bir örnek.



Şekil 2.3: reCAPTCHA servisinin şu anda kullandığı görsel İEİ testlerine bir örnek.

anlatılan kullanıcı çalışması yapıldığı sıralarda işitsel reCAPTCHA testleri kullanıcılarından gürültülü bir arka planda üçlü bloklar halinde seslendirilen 9 ila 12 tane rakamı doğru şekilde tespit ederek cevap kutusuna yazmaları istenmekteydi. Bu testlerin seslendirilmeleri 15 saniye civarında sürmekteydi. Bu sürümde kesimleme saldırılarını [26] engelleyebilmek adına bazı rakamların seslendirilmesi üst üste bindirilmekteydi ve bu uygulama testlerin kullanışsız olmalarına sebep olmaktaydı [27, 28].

Google bu duruma daha yeni ve daha kolay bir işitsel reCAPTCHA sürümünü kullanıma sunarak cevap vermiştir [29] <sup>3</sup>. Yeni sürümün duyurulmasından yaklaşık iki ay sonra bir saldırgan, ironik bir şekilde Google Speech API'yi kullanarak, bu yeni sürüm üzerine başarılı saldırılar yapılabildiğini kendi blog sayfasında açık-

<sup>3</sup>Bölüm 5.3'te anlatılan kullanıcı çalışmasında bu yeni sürüm kullanılmıştır.

lamıştır [30]. Hem saldırının tekrarlanabilirliğini ölçmek, hem de saldırının başarı oranı hakkında somut veriler sunabilmek adına saldırının aşağıda özetlenen saldırı yöntemi tekrarlanmıştır:

- reCAPTCHA servisinden bir işitsel İEİ talep edilir ve servisin sağladığı bağlantı yardımıyla işitsel İEİ'nin ses dosyası *.mp3* formatında indirilir.
- Ses dosyası eşit uzunluktaki parçalara bölünür.
- Her bir parçanın ses seviyesi iki katına çıkarılır.
- Google Speech API'nin her bir parçayı doğru tanımlama şansını artırmak için her bir parçadan ikişer kopya olacak şekilde parçalar arka arkaya eklenerek bir *.flac* dosyası oluşturulur.
- Oluşturulan bu dosya Google Speech API'ye gönderilir.
- Google Speech API'den gelen cevaplar reCAPTCHA servisine testin cevabı olarak yollanır.
- reCAPTCHA servisinin cevabına göre saldırının başarılı veya başarısız olduğu not edilir.

Yukarıda belirtilen yöntem 40 tane rastgele üretilmiş işitsel reCAPTCHA İEİ'leri üzerinde denenmiş ve 25 tanesinde (%62,5) başarılı sonuç alınmıştır. Normalde gerçek kullanıcıların konuşmalarını tanımlamaya çalışan Google Speech API, analizlerinde kelime ya da rakam özelleştirmesi yapmamaktadır. Bu durum birkaç kez bazı seslerin kelime olarak tanımlanmasına ve dolayısıyla saldırının başarısız olmasına sebep olmuştur. Sadece rakamları tanımlamak üzerine özelleşmiş bir konuşma tanımlama aracı kullanılarak yapılacak bir saldırı ile bu çalışmada elde edilen orandan daha yüksek başarı oranları elde edilebileceği düşünülmektedir.

### 2.3 Kullanıcı Çalışması Adımları

Kullanıcı çalışmaları, bir ürün ya da servisin muhtemel kullanıcılarıyla test edilmesidir. Bu çalışmalar farklı amaçlarla yapılabilmektedir. Sistemin hatalarını bulmak, kullanıcıların ürünü ne kadar sevdiğini tespit edebilmek, katılımcıların verilen bir görevi ne kadar sürede tamamlayabildiğini görmek bu amaçlara örnek olarak verilebilir. Kullanıcı çalışmalarında katılımcılar kendilerine verilen görevi



yaparken bazı bilgiler toplanır ve elde edilen bu bilgiler ışığında ürünün kullanılabilirliği hakkında fikir elde edebilmek amaçlanır.

Etkili bir kullanıcı çalışması dört adımdan oluşmaktadır: çalışmanın planlanması, katılımcıların bulunması, planlanan çalışmanın yapılması ve elde edilen verinin yorumlanması. Aşağıda yer alan alt bölümlerde bu adımlardan bahsedilmiştir.

### **2.3.1 Çalışmanın planlanması**

İlk aşama olan planlama aşaması, çalışmanın gidişatına yön vermek açısından oldukça önemlidir. Çalışma ile ilgili birçok karara bu aşamada varılır. Bu aşamada verilen eksik ya da hatalı kararlar bütün çalışmanın verimini düşürebilmektedir. Bu aşamanın sonunda çalışmanın kapsamı ve amacı, çalışmanın nerede ve ne zaman yapılacağı, oturumların ne kadar süreceği, çalışmada kullanılacak aletlerin neler oldukları, katılımcıların sayısı ve bu katılımcılara nasıl ulaşılabileceği, katılımcıların hangi görevleri yapmaları isteneceği, bu esnada hangi verilerin toplanacağı gibi bilgiler belirlenmelidir.

Kullanıcılardan toplanabilecek veriler nitel ve nicel olmak üzere ikiye ayrılabilir. Katılımcıların sistem ile ilgili sorulara verdikleri yanıtlar nitel verilere örnektir. Örneğin kullanıcıların sistem ile önceden tanışıklıklarının olup olmadığı, tamamladıkları her bir görevin zorluğu hakkında neler hissettikleri, sistemi sevip sevmedikleri, ya da açık uçlu sorulara verdikleri cevaplar nitel verilere örneklerdir. Diğer taraftan, kullanıcının hislerinden bağımsız şekilde elde edilen verilere nicel veriler denmektedir. Kullanıcıların verilen görevleri tamamlama süreleri, görevi başarılı şekilde tamamlayıp tamamlayamadıkları bilgisi, görevi tamamlarken hata yapıp yapmadıkları veya kaç kere hata yaptıkları bilgisi gibi veriler nicel verilerdir. Kullanıcı çalışmalarının yorumlanmasında hem nitel hem de nicel veriler önem arz etmektedir.

### **2.3.2 Katılımcıların bulunması**

Planlanması yapılan kullanıcı çalışmasında yer alacak katılımcıların sayısı, seçimi ve bu katılımcılara nasıl ulaşılabileceği da oldukça önemlidir. Yeterli sayıda katılımcı ile yapılmayan çalışmalar anlamlı sonuç vermezken, gereğinden çok sayıda katılımcı ile yapılan çalışmalar ek maliyete sebep olabilmektedir. Her bir katılımcıya belli miktarda ücret ödemek maddi maliyete, her bir katılımcı ile ayrı oturum

yapmak zamansal maliyete örnektir. Bazı arařtırmacılara göre [31] bir sistemin hataları aranıyorsa 5 katılımcı yeterli iken, istatistiksel analizler yapılacaksa en az 20 katılımcı gerekmektedir. İstatistiksel analiz hedeflenen çalıřmalarda katılımcı sayısı arttıka sonuçların güvenilirliđi artmaktadır.

Katılımcıların sayısı kadar nasıl seğıldikleri de oldukça önemlidir. Sistemin tasarımcıları sistemi zaten tanıdıklarından bu kişilerin çalıřmaya katılmaları sađlıklı sonuçlar vermeyecektir. Benzer şekilde, sistemi hiç kullanmayacak olan katılımcıların çalıřmada yer alması da dođru deđildir. Katılımcılar sistemin muhtemel kullanıcıları arasından seğılmeli ve mümkün olduđunca gerçek kullanıcı profilini yansıtmalıdır. Bu sayede toplanan verilerin yorumlanmasıyla birlikte daha verimli sonuçlar elde edilebilir.

Son olarak, katılımcılara dođru şekilde ulařabilmek gerekmektedir. Özenle planlanan çalıřmanın duyurusu etkili yapılamadıđı takdirde gerekli katılımcılara ulařmakta zorluk çekilebilir ve dolayısıyla çalıřma başarısız olabilir. Etkili hazırlanmış bir duyuru metninde çalıřmanın kapsamı, amacı, yeri ve zamanı, ne kadar süreceđi, katılımcıların ne kadar ücret alacakları, hangi özelliklere sahip katılımcıların arandıđı gibi bilgiler yer almalıdır.

### **2.3.3 Çalıřmanın yapılması**

Önceki aşamalarda planlanan ve katılımcıları bulunan çalıřmanın dođru şekilde uygulanması gerekmektedir. Bunu yaparken mümkün olduđunca plana sadık kalınmalıdır. Olası plandan sapma durumlarında sapmanın sebepleri ve sonuçları dikkatlice incelenmelidir. Gerekmesi halinde plan güncellenmeli ve çalıřmaya baştan başlanmalıdır.

### **2.3.4 Verilerin yorumlanması**

Verilerin elde edilmesi kadar dođru şekilde yorumlanmaları ve raporlanmaları da önemlidir. Elde edilen nitel ve nicel veriler çalıřmanın amacına göre farklı şekillerde incelenmelidirler. Örneđin yapılan kullanıcı çalıřmasıyla bir sistemin hataları tespit edilmeye çalıřılıyorsa bulunan hataların önem düzeyi belirlenmeye çalıřılmalıdır. Yapılan çalıřmayla iki sistemin kullanılabilirlikleri karşılaştırılmaya çalıřılıyorsa, elde edilen veriler istatistiksel analizlerle test edilmelidir.

Sonuçları doğru yorumlayabilmek için hem amaca, hem de veriye uygun istatistiksel testlerin kullanılması oldukça önem arz etmektedir. Çalışmada yapılan her bir değişiklik, uygulanması gereken istatistiksel testi değiştirebilmektedir. Örneğin çalışmanın amacı sistemler arasındaki ilişkiyi bulmaksa korelasyon katsayısı hesaplanmalıdır. Çalışmanın amacı sistemlerin farkını bulmaksa karşılaştırma testleri uygulanmalıdır. Benzer şekilde, verilerin sayısal ya da sözel oluşu, normal dağılıma uyup uymamaları, çalışmada yer alan grupların sayısı gibi değişkenler de uygulanacak istatistiksel testin belirlenmesi için kullanılan bilgilerdir. Sayısal veriler normal dağılıma uyuyorlarsa parametrik testler, uymuyorlarsa parametrik olmayan testler uygulanmalıdır. Eğer veriler sadece bir grubun verilen görevleri her sistem için yapmasıyla elde edildiyse grup içi testler, her sistemi farklı kullanıcıların bulunduğu gruplar test ettiyse gruplar arası testler kullanılmalıdır.

Bu tez çalışmasında önerilen sistemin kullanılabilirliğini ölçebilmek amacıyla üç kullanıcı çalışması yapılmıştır. Elde edilen sonuçların raslantısal olup olmadıklarını anlayabilmek adına sonuçlar istatistiksel testlerle incelenmişlerdir. Bu çalışmada kullanılan istatistiksel testler aşağıda kısaca anlatılmaktadır.

#### **2.3.4.1 Kolmogorov-Smirnov testi**

Kolmogorov-Smirnov testi, verilerin normal dağılıma uyup uymadığını kontrol etmek için kullanılan bir uyum iyiliği testidir. Uyum iyiliği testi olarak kullanılabilmesi için veriler üzerinde ön işlem uygulayarak onları standartlaştırır ve sonrasında sonucu standart normal dağılım ile karşılaştırır. Eğer karşılaştırma sonucuna göre anlamlı bir fark bulunursa verinin normal dağılıma uymadığına, anlamlı bir fark bulunamazsa verinin normal dağılıma uyduğuna karar verilir.

#### **2.3.4.2 Wilcoxon işaretli sıra testi**

Grup içi ölçümler arasındaki farkı test etmek için kullanılan ve parametrik olmayan bir testtir. Başka bir deyişle, aynı kullanıcıların farklı sistemleri test etmeleri sonucu elde edilen verilerin aynı dağılımı gösterip göstermediğini belirleyebilmek için kullanılır.

Bu testte aynı kullanıcıların farklı sistemlerde elde ettikleri veriler karşılaştırılır. Aynı olan değerler atıldıktan sonra geriye kalanların farkı hesaplanır ve bu fark değerleri küçükten büyüğe sıralanır. Gerçek değerler, sıralarını belirten değerlerle

değiştirilir ve bu yeni değerlerin toplamı hesaplanır. Elde edilen toplam belli bir eşik değerinden büyükse verilerin aynı dağılımı göstermediğine karar verilir.

#### **2.3.4.3 Mann-Whitney U testi**

Gruplar arası ölçümler arasındaki farkı test etmek için kullanılan ve parametrik olmayan bir testtir. Başka bir deyişle, farklı kullanıcıların farklı sistemleri test etmeleri sonucu elde edilen verilerin aynı dağılımı gösterip göstermediğini belirleyebilmek için kullanılır.

Bu testte farklı kullanıcıların farklı sistemlerde elde ettikleri veriler bir araya getirilerek küçükten büyüğe sıralanırlar. Gerçek değerler, sıralarını belirten değerlerle değiştirilir ve bu yeni değerler kullanılarak her bir grubun toplamı ayrı ayrı hesaplanır. Elde edilen toplam değerleri kullanılarak her iki grubun  $U$  değerleri hesaplanır ve büyük olan  $U$  değeri belli bir eşik değerinden büyükse verilerin aynı dağılımı göstermediğine karar verilir.

#### **2.3.4.4 Friedman testi**

Friedman testi, grup içi tekrarlı ölçümlerde elde edilen veriler arasındaki farkları test etmek için kullanılabileceği gibi, ikiden fazla sistem olduğu durumlarda herhangi bir sistemin diğerlerinden farklı olup olmadığını anlamak için de kullanılabilir.

Friedman testinde de verilerin sıra değerleri kullanılmaktadır. Her bir kullanıcının farklı ölçümlerde elde ettiği değerler kendi aralarında küçükten büyüğe sıralanır ve gerçek değerler sıra belirten değerlerle değiştirilir. Her bir farklı ölçüm için bu yeni değerler toplanır. Elde edilen toplam değerleri ile yeni değerlerin ortalaması bilgileri kullanılarak yeni hesaplamalar yapılır ve bu hesaplamalara göre farklı ölçümler arasında anlamlı fark olup olmadığına karar verilir.

#### **2.3.4.5 Kendall uyumu katsayısı (W)**

Kendall uyumu katsayısı (W), Friedman testinin normalleştirilmiş halidir. Friedman testinde olduğu gibi burada da katılımcıların verdikleri oyların birbiriyle ne kadar uyumlu olduğu belirlenmeye çalışılır. Normalleştirilmiş olduğundan so-

nuç 0 ile 1 arasında çıkmaktadır. 0 değeri tamamen uyumsuzluğu, 1 değeri ise tamamen uyumluluğu belirtmektedir.





### 3. PROBLEM TANIMI VE ÖNERİLEN SİSTEM

Önceki bölümlerde görsel İEİ'lerin erişilebilirlik, kullanılabilirlik ve hatta güvenlik sorunlarından; işitsel İEİ'lerin kullanılabilirlik ve güvenlik sorunlarından, ve şu ana kadar sunulmuş olan metin tabanlı İEİ'lerin ise güvenlik sorunlarından bahsedilmiştir. Güvenli şekilde üretilebildikleri takdirde metin tabanlı İEİ'ler, İEİ'lerin erişilebilirlik ve kullanılabilirlik problemlerini çözmeye adaydır.

Bu bilgiler ışığında, bu tez çalışmasında metin tabanlı İEİ testlerinin güvenlik problemine ilişkin bir çözüm aranmıştır. Bölüm 2.2.1.1'de ayrıntılı bahsedildiği üzere, textCAPTCHA servisi birkaç belli başlı soru tipi seçerek bunlardan yeni sorular türettiği için güvenlik bakımından yetersiz kalmıştır. Bunun yerine bu çalışmada, insan hesaplama yardımıyla soru tipi sayısı artırılarak her soru tipini çözen bir kodun yazılması engellenmeye çalışılmıştır.

İnsan hesaplama kavramının felsefe ve psikoloji literatüründe kullanımı 1838 yılında başlamıştır [32]. Bilgisayar bilimleri literatürüne girişi 1950 yılında Alan Turing ile olmuştur [33]. İnsan hesaplama kavramının modern anlamda kullanımına ise 2005 yılında Louis von Ahn'ın "İnsan Hesaplama" başlıklı tezi [34] ve sonrasındaki çalışmaları ivme kazandırmıştır. İlgili tez çalışmasında insan hesaplama kavramı "insan işlem gücü kullanılarak bilgisayarların henüz çözemediği problemlerin çözülmesi" olarak tanımlanmıştır.

Bu tez çalışmasında bilgisayarların henüz çözemediği problem, otomatik olarak çözülmesi zor metin İEİ'lerin oluşturulması işlemidir. SMARTCHA (SeMi Automated Reverse Turing test to tell Computers and Humans Apart) ismi verilen çözümde İEİ operatörleri adı verilen çalışanlardan ücret karşılığında soru üretmeleri istenmiştir. Fakat operatörlerin oluşturduğu bu soruların da güvenlik ve kullanılabilirlik kriterlerini sağlaması gerekmektedir. Soruların kullanılabilir olabilmesi için operatörlere nasıl kullanılabilir sorular hazırlayacakları yönünde talimatlar verilmiş, bu talimatlar doğrultusunda soru hazırlamadıkları takdirde kendilerine ücret ödemesi yapılmayacağı bildirilmiştir. Soruların güvenli olabilmesi için ise operatörlerin hazırladığı sorular doğrudan sisteme kaydedilmemiş, ilk önce güvenlik motorundan geçirilerek otomatik olarak çözülebilirliği test edilmiştir. Operatörlerin ürettiği soru-cevap ikilileri "baz sorular" olarak kullanıl-

mış ve bu baz sorular daha sonra otomatik yöntemlerle sayıları artırılarak sistem ölçeklenebilir hale getirilmeye çalışılmıştır.

### 3.1 SMARTCHA

Önerilen SMARTCHA sisteminin güvenlik, kullanılabilirlik ve ölçeklenebilirliğine dair bilgiler aşağıda yer alan bölümlerde ayrı ayrı incelenmektedir.

#### 3.1.1 Güvenlik

textCAPTCHA servisinin en büyük dezavantajının güvenlik konusunda olduğu ve bu açığa küçük bir soru havuzundan yeni sorular türetmenin neden olduğu önceki bölümlerde belirtilmişti. Sözkonusu güvenlik açığını kapatabilmek için insan hesaplama kullanarak soru havuzunu genişletmek önerilmektedir. Ancak bu noktada soru havuzunun genişlemesi ne kadar önemliyse, soru havuzundaki soruların otomatik olarak çözülüyor olmaları da o kadar önemlidir. Soru havuzunda yer alan bir baz soru otomatik olarak çözülebiliyorsa, bu baz sorudan otomatik olarak üretilen bütün sorular da büyük ihtimalle otomatik olarak çözülebilir olacaktır. Otomatik olarak çözülebilen soruların sistemin güvenliğini tehlikeye atmasını engellemek için bir güvenlik motoru oluşturulmuştur. Hazırlanan bütün soru-cevap ikilileri güvenlik motorundan geçtikten sonra veritabanına güvenlik motorunun cevabıyla birlikte kaydedilmektedir.

Hazırlanan güvenlik motoru katmanlı bir yapıya sahiptir. Aşağıda ayrıntılı olarak verilen bu katmanların akış şeması Şekil 3.1’de görülebilmektedir.

- Aşamalardan ilki, sorunun doğru cevabının soru metni içerisinde yer alıp almadığının kontrolüdür. Cevabı sorunun içinde yer alan soru-cevap ikilileri güvenlik motoru tarafından “güvensiz” olarak işaretlenirler.
- Bir sonraki aşamada soru metni, textCAPTCHA servisi sorularını otomatik olarak çözebilen TextCaptchaBreaker projesinden [22] ilham alınarak Java programlama dili ile kodlanmış bir katman ile test edilir. Hazırlanan bu katman TextCaptchaBreaker projesinde belirlenmiş olan soru tiplerinin yanı sıra, textCAPTCHA servisi incelenerek tespit edilen yeni soru tiplerini de çözebilecek şekilde güncellenmiştir. Temel olarak TextCaptchaBreaker servisine çok benzeyen bu katman, ondan farklı olarak bir tek cevap yerine



olası cevaplar kümesi bulmayı amaçlamaktadır. Bunun sebebi, TextCaptchaBreaker projesinin amacı sorunun var olan bir tek doğru cevabını bulmakken, güvenlik motoruna eklenen bu katmanın amacı sorunun otomatik olarak çözülme olasılığını düşürmeye çalışmaktır.

- Bu kontrolden sonra sorunun basit bir aritmetik işlem sorusu olmadığı doğrulanmaktadır. Bu doğrulama için Dijkstra tarafından geliştirilmiş olan Shunting Yard algoritması [35] kullanılmıştır. Shunting Yard algoritması ardışık olarak yapılan toplama, çarpma, çıkarma, bölme ve üs alma gibi basit aritmetik işlemleri parantez ve işlem önceliklerini göz önünde bulundurarak çözebilmektedir. Shunting Yard algoritması normalde sadece sayılar ile çalışan bir algoritma olmasına rağmen güvenlik motoru için güncellenmiş ve yazıyla yazılmış sayılarla da işlem yapabilecek hale getirilmiştir. Örneğin “What is the result of (three hundred and three + five) / (four) + twelve?” sorusunun cevabını 89 olarak verebilmektedir.
- Sorunun basit bir aritmetik işlem sorusu olmadığı doğrulandıktan sonra soru Bing arama motoruna [36] sorgu olarak gönderilmekte ve ilk 50 sonuç incelenmektedir. Bing arama motorunun soru için döndürmüş olduğu ilk 50 cevap içerisinden; sorunun içerisinde bulunan kelimeler ile İngilizce bir anlamı olmadığı halde cümle bütünlüğünü sağlamaya yarayan ve adına Stop Words denen *a, the, there, are* gibi kelimeler çıkarılmaktadır. Kalan cevaplar üzerinde kelimelerin yer alma sıklıkları (frekans analizi) yapılmakta ve en çok geçen 10 kelime tespit edilmektedir. Daha sonra sorunun gerçek cevabı ile elde edilmiş olan en sık geçen 10 kelime karşılaştırılmakta ve sorunun Bing arama motoru tarafından otomatik olarak çözülebiliyor çözülemediğine karar verilmektedir. Bu karşılaştırma işlemi, cevap tek kelimedenden oluşuyorsa cevabın en sık geçen 10 kelime arasında yer alıp almaması; cevap birden fazla kelimedenden oluşuyorsa da cevapta yer alan en az bir kelimenin en sık geçen 10 kelime arasında ve geriye kalan kelimelerin bütün cevaplar arasında olup olmaması şeklindedir.
- Bir sonraki aşamada soru WolframAlpha servisinde [37] sorgulanmakta ve oradan gelen cevaplar ile doğru cevap karşılaştırılmaktadır. Ticari olmayan kullanımda ücretsiz olarak aylık 2000 sorguya kadar izin veren WolframAlpha web servisi, Wolfram Araştırma Şirketi tarafından geliştirilen bir hesaba dayalı bilgi motorudur.
- Son olarak İEİ operatörünün yazmış olduğu sorunun, aynı operatörün yazmış olduğu diğer sorularla benzerliği ölçülmektedir. Bölüm 4.3.1’de anlatılan ilk insan hesaplama çalışmasında bu kontrol yapılmamıştır. Daha sonra ya-

pılan incelemelerde bazı operatörlerin güvenli bir soru hazırladıktan sonra bu soru üzerinde küçük değişiklikler yaparak yeni sorular hazırlamış olmalarının fark edilmesi üzerine bu katman güvenlik motoruna eklenmiştir.

Güvenlik motoru bir soruyu ancak ve ancak soru yukarıda anlatılan bütün aşamaları çözülmeden geçerse güvenli olarak işaretlemektedir. Bir sorunun İEİ testi olarak kullanılabilmesi için güvenlik motoru tarafından güvenli olarak işaretlenmesi gerekli ancak yeterli değildir. Sorunun aktif ya da pasif soru olduğunun anlaşılabilmesi için ek kontroller yapılmaktadır. Ancak ve ancak aktif ve güvenli olarak işaretlenmiş sorular İEİ testi olarak kullanılabilir. Soruların aktif ya da pasif özelliği aşağıdaki şekilde açıklanabilir.

Soruların çoğunluğunun cevabının aynı olması durumunda sistem kaba kuvvet saldırısına açık hale gelecektir. Başka bir deyişle, saldırganların sistemin güvenliğini aşabilmeleri için herhangi bir işlem yapmadan doğru cevap olarak sadece en çok kullanılan cevabı vermeleri yeterli olacaktır. Bu sebeple otomatik olarak çözülemeyen bir sorunun cevabının kullanım sayısı, güvenli olarak işaretlenmiş bütün cevapların toplam kullanım sayısına oranlanmaktadır. Eğer bu oran %0,1'den büyükse soru "pasif güvenli soru", %0,1'den küçükse "aktif güvenli soru" olarak sisteme kaydedilmektedir.

SMARTCHA yaygın olarak kullanılmaya başlanırsa farklı saldırı tipleri ortaya çıkabilir. Katmanlı yapısı sayesinde güvenlik motoru yeni saldırılara karşı güncellenebilir<sup>4</sup>. Aslına bakılırsa diğer İEİ çeşitlerinde tecrübe edildiği gibi, İEİ testlerinin güvenliği saldırgan ve araştırmacı arasında geçen bir kedi-fare oyunu gibidir. Saldırganlar sistemin açıklarını yakaladıkça araştırmacılar bu saldırılara karşı yeni önlemler geliştirmeye çalışırlar. SMARTCHA da bu konuda bir istisna değildir.

Literatürdeki güvenlik saldırılarına karşı SMARTCHA'nın durumu kısaca aşağıdaki şekilde özetlenebilir. %0,1'den daha fazla kullanılmış cevaplar güvenlik motoru tarafından pasif olarak işaretlendiği için SMARTCHA kaba kuvvet saldırılarına karşı dayanıklıdır. SMARTCHA'nın güvenlik motoru, arama motorlarını kullanan saldırılara karşı da koruma sağlamaktadır. SMARTCHA'daki bazı sorular operatörün soruyu üretmesi aşamasında güvenlik motoru tarafından çözülemezken sonradan çözülebilir hale gelebilirler. Böyle bir duruma karşı periyodik olarak uygulanacak çevrimdışı güvenlik kontrolleri yararlı olabilir. SMARTCHA'nın güvenliği tamamen veritabanındaki soru-cevap ikililerinin güvenliğine

<sup>4</sup>Nitekim yapılan çalışmalar süresince bazı güvenlik kontrolleri eklemek üzere birkaç defa güncellenmiştir.

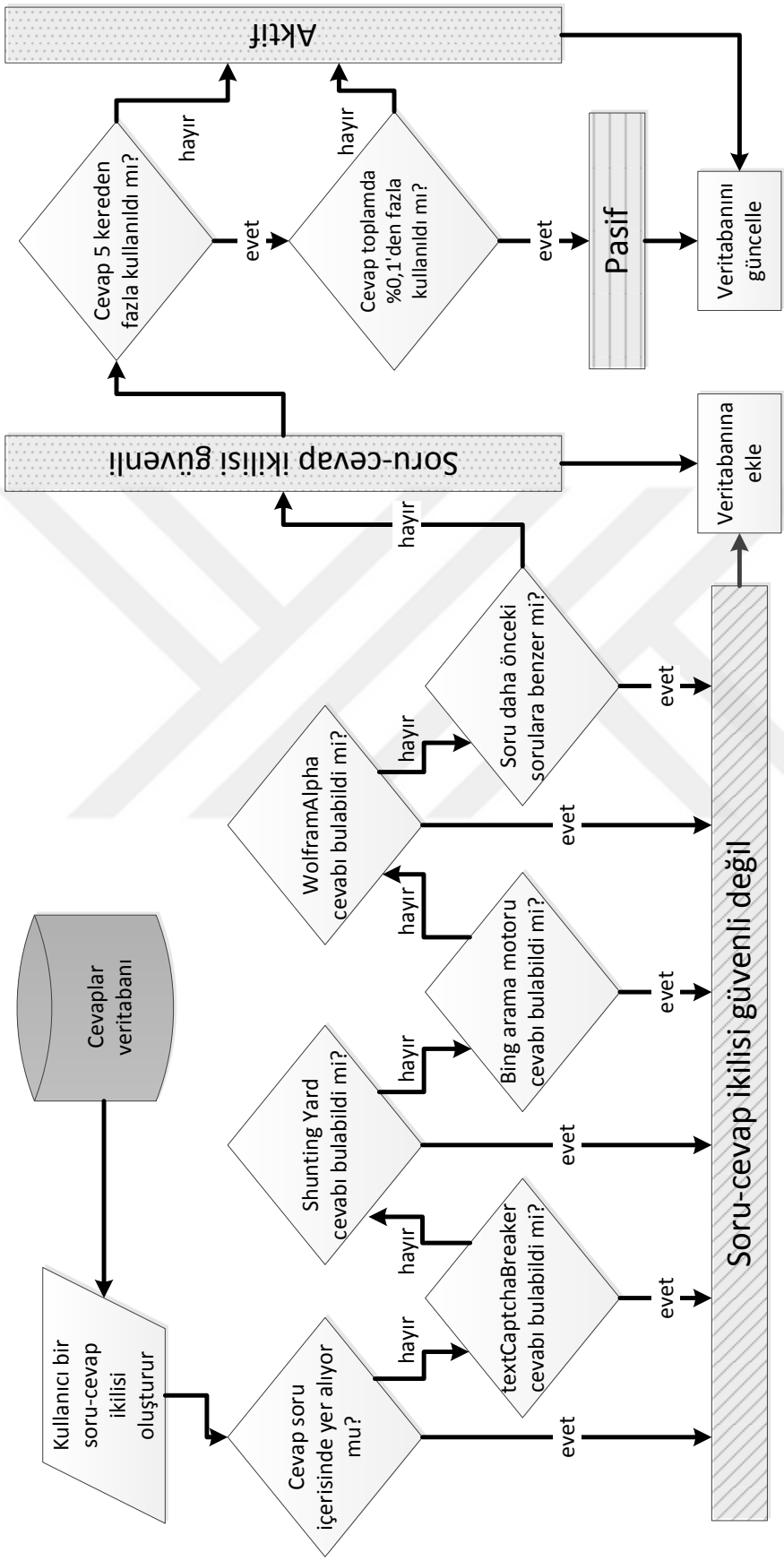
bağlıdır. Herhangi bir yolla veritabanına sızılması durumunda sistemin güvenlik yapısı bozulmuş olur. Zayıf bir önlem olsa da, bu duruma karşı soru-cevap ikilileri veritabanında şifrelenmiş olarak saklanabilirler. Saldırganlar SMARTCHA veritabanındaki bütün soruları elde etmek için art arda istek gönderebilirler. Bu tarz saldırılar İEİ isteğinde bulunan her IP adresini saklayarak ve art arda yapılan istek sayısının belli bir eşik değerini aşması durumunda ilgili IP'deki istemciye cevap vermeyerek hafifletilebilir. Bu durumda saldırganlar bu saldırıyı farklı IP adresine sahip bilgisayarlara dağıtarak gerçekleştirebilirler. Ancak saldırganlar bu şekilde bütün soruları toplasalar dahi, bütün testleri ya elle ya da insan hesaplama kullanarak çözmeleri gerekir. SMARTCHA, İEİ testlerinin çözmesi için bir insana yönlendirildiği aktarma saldırılarına (relay attacks) karşı zayıftır. Öte yandan aktarma saldırıları, var olan çoğu İEİ sistemini tehdit etmektedir [38].

Bir başka saldırı yöntemi olarak saldırganlar operatör gibi davranabilir ve veritabanını cevabını zaten bildikleri sorularla doldurabilirler. Bu saldırı yöntemi operatör başına düşen soru sayısı düşük tutularak engellenebilir. Bir kullanıcının kendisini birden fazla kullanıcıymış gibi gösterdiği Sybil saldırılarını [39] engellemek için operatörlerin banka hesap numarası, TC kimlik numarası ya da vergi numarası gibi bazı kişisel bilgilerle kayıt olması şart koşulabilir. Amazon Mechanical Turk sistemi [40] de bu yöntemi kullanmaktadır. Bunun yanı sıra, güvenli olarak işaretlenmiş soruları sistemde hemen aktifleştirmek yerine rastgele bir miktar zaman süresince beklemek de bu saldırının etkilerini azaltabilir.

Güvenlik motorunun amacı, günümüz teknolojisi ile soruların otomatik olarak çözülememesini sağlamaktır. Gelecekte anlamsal bilişimin gelişmesi SMARTCHA için bir tehdit olabilir. Soruların otomatik olarak aritmetik, mantıksal, genel kültür gibi kategorilere ayrılması bu tarz saldırılara karşı yararlı olabilir. Daha iyi bir önlem alınıncaya kadar ilgili kategorinin testleri devre dışı bırakılarak saldırılar hafifletilebilir. IBM'in Watson isimli doğal dil işleme yaparak soru cevaplayan bilgisayar sistemi anlamsal bilişimin gelişmesinin gerçekleşebileceğinin göstergesidir. Watson, *Jeopardy!*<sup>5</sup> isimli bir genel bilgi yarışmasında iki insan rakibini yenmeyi başarmıştır. Eğer böyle bir sistem genel kullanıma açık hale gelirse SMARTCHA da dahil olmak üzere birçok İEİ sistemi güvenli olmaktan çıkacaktır. Diğer yandan bu sistem genel kullanıma açık hale geldiğinde SMARTCHA'nın güvenlik motoruna bir katman olarak eklenebilir. Bu durumun gerçekleşmesi halinde hazırlanacak olan yeni SMARTCHA İEİ testlerinin anlamsal yapısında zorunlu olarak değişikliğe gidilecektir.

---

<sup>5</sup>Bu yarışma formatı Türkiye'de *Riziko!* ve *Büyük Risk* isimleriyle yayınlanmıştır.



Şekil 3.1: SMARTCHA güvenlik motorunun akış şeması.

### 3.1.2 Kullanışlılık

SMARTCHA sisteminde İEİ operatörleri tarafından üretilen soruların güvenlik kontrolü otomatik olarak yapılabilirse bile, kullanışlı olup olmadığının kontrolü otomatik olarak yapılamamaktadır. Bu sebeple operatörlere uymaları gereken bazı kurallar verilmiş ve eğer bu kurallara uymadıkları tespit edilirse kendilerine ücret ödenmeyeceği belirtilmiştir. Operatörlerin uymaları istenen kurallar sisteme giriş yaptıkları sayfada da yer almaktadır. Bölüm 4.3.3'te anlatılan üçüncü insan hesaplama çalışmasına ait kurallar örnek olarak aşağıda verilmiştir:

- Sorunuzun dili Türkçe olmalıdır.
- Sorularınız, en azından Türkçe dilini konuşanlar arasında, evrensel olarak çözülebilir olmalıdır. (örn. Ankara'ya daha önce gelmemiş birisi, Batıkent metro durağından sonra hangi durağın geldiğini vb. bilemeyecektir.)
- Sorularınız bütün bilgi düzeyindeki insanlar tarafından çözülebilmelidir. (örn. Bir avukat teknik mühendislik terimlerini bilemeyecektir, vb.)
- Cevap, sorunuz için TEKİL CEVAP olmalıdır. (örn. "Türkiye'nin en iyi futbol klübü hangisidir?" sorusuna başkasının cevabı sizin cevabınızdan farklı olabilir, vb.)
- Lütfen kişisel sorular sormayınız. (örn. Bir başkasının en sevdiği sayı / mevsim / gün sizinkinden farklı olabilir, vb.)
- Lütfen argo kelimeler kullanmayınız.

Fakat bunlara rağmen gözden kaçan durumlar veya soruların olabileceği öngörülmektedir. Sistemin kullanışlılığını düşüren bu tarz soruları ayıklamak üzere bir veya birkaç kişinin her soruyu incelemesi oldukça zahmetlidir. Bu tarz soruları sisteme koymadan tespit edip ayıklamak yerine, güvenli olarak işaretlenmiş her soruyu sisteme eklemek ve daha sonra soruların çözülebilmeye istatistiklerini tutarak belirli bir orandan daha düşük miktarda çözülebilmeye veya cevap verilmesi tercih edilmemiş soruları kullanışlı olmayan soru olarak etiketleyerek ayırma yöntemi düşünülmüştür. Taranmış kitapları sayısal ortama aktarmaya çalışan reCAPTCHA servisinde optik karakter tanıma teknikleriyle okunamayan kelimelerin hangi kelimeye karşılık geldiğine karar verilme yöntemi de benzer şekilde çalışmaktadır. reCAPTCHA servisinde kullanıcılara, bir tanesi önceden hangi kelimeye karşılık geldiği tespit edilmiş ve bir tanesi de tespit edilememiş olmak

üzere iki farklı taranmış metin gösterilir ve kullanıcılardan iki metni de doğru biçimde yazmaları istenir. Kullanıcıların vermiş oldukları cevaplardan daha önceden hangi kelimeye karşılık geldiği tespit edilmiş olan metin, kullanıcılar için gerçek İEİ testidir. Hangi kelimeye karşılık geldiği bilinmeyen metin ise, çok sayıda kullanıcıya gösterilmekte ve kullanıcıların belirli bir çoğunluğu (örneğin %90'ı) bu metin için aynı kelimeyi cevap olarak vermişse, o zaman kelimenin değerini bu cevapla eşleştirmektedir.

Hazırlanmış olan soruların kullanılabilirliğini ölçmeye yönelik çalışmalar Bölüm 5.2 ve 5.3'te detaylı olarak anlatılmaktadır. Kısaca bahsetmek gerekirse, yapılan iki kullanıcı çalışmasında da kullanıcılar SMARTCHA testlerini çözmeyi reCAPTCHA testlerini çözmekten daha eğlenceli bulmuşlardır. Ayrıca bazı durumlarda SMARTCHA testlerini reCAPTCHA testlerinden daha hızlı çözebilmişlerdir. Kullanıcılar SMARTCHA testlerini çözmekte en çok zorlandıkları durumda dahi reCAPTCHA testlerine kıyasla SMARTCHA'yı tercih etmişlerdir.

### 3.1.3 Ölçeklenebilirlik

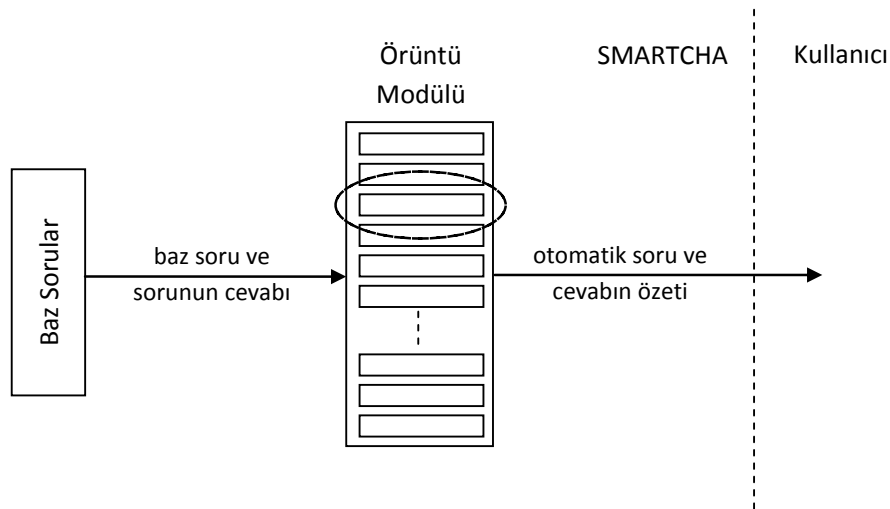
Bölüm 4.3'te anlatılan insan hesaplama çalışmalarına göre operatörlerin ortalama soru yazma süreleri özel durumlar haricinde 40 saniyeden kısa sürmemekte, hatta 56 saniyeye kadar çıkabilmektedir (Şekil 4.9). Oysa Bölüm 5'te anlatılan kullanıcı çalışmalarına göre kullanıcıların ortalama SMARTCHA sorusu çözme süresi 12 ila 24 saniye arasında değişmektedir. Ortalama soru üretme süreleri (40 - 50 saniye civarı), ortalama soru cevaplama süresinden (12 - 24 saniye civarı) daha fazla olduğundan, bir saldırgan yine "insan hesaplama" kullanarak sisteme bir saldırı düzenlediği durumda daha avantajlı konuma geçecektir. Üstelik daha az süreli insan hesaplama kullanacağından daha az para harcayacaktır. Fakat İEİ operatörlerince girilen her bir sorudan en az 40 - 50 yeni soru üretilebilirse operatörlerin efektif soru üretme süreleri 1 saniye mertebelerine inecek ve bahsedilen saldırganın avantajlı durumu ortadan kalkacaktır.

Operatörlerin üretmiş olduğu güvenli soruların "baz sorular" olarak kullanılabilirliği ve otomatik yöntemlerle sayılarının artırılabilirliği belirtilmiştir. Soruların otomatik yöntemlerle artırılabilmesi için bazı örüntülerin kullanılması gerekmektedir. Soru artırımı için örüntülerin kullanılması, aynı textCAPTCHA servisinde olduğu gibi, örüntülerin otomatik olarak tespit edilebilmesini beraberinde getirir. Fakat SMARTCHA soruları otomatik olarak çözülemediğinden örüntüler tespit edilerek baz sorular ayrıştırılabilirse bile, saldırganlar hala baz soruları iş gücü ya da

insan hesaplama kullanarak çözmelidir. Bunun yanı sıra saldırganların örüntüleri tanımlayıp baz soruları ayırıştırmak için ekstra zaman ve efor harcamaları gerekecektir. Kullanılan örüntü sayısı arttıkça yeni örüntüleri tespit edip baz soruları otomatik ayırıştırma süreci de uzayacaktır. Bütün bunlar da sistemin güvenliğini artırmaktadır.

Soruların otomatik artırımı için tasarlanan yapı oldukça basit ve verimlidir. Veritabanındaki İEİ sayısı bu yapı kullanılarak çokça artırılabilir olsa bile Şekil 3.2’de görülebilecek örüntü modülü sayesinde veritabanında herhangi bir şişkinlik olmayacak, sadece baz sorular saklanacaktır.

Örüntü modülü şu şekilde çalışmaktadır. Bir istemci SMARTCHA servisinden İEİ testi istediği zaman, o ana kadar en az miktarda kullanılmış olan baz sorular arasından rastgele bir tanesi seçilerek örüntü modülüne iletilmektedir. Bu modül içerisinde birbirinden bağımsız örüntüler hazırda beklemektedir. Seçilen baz soru üzerine örüntülerden bir tanesi rastgele seçilerek uygulanır ve bu sayede otomatik olarak yeni bir soru-cevap ikilisi üretilir. Bütün bu işlemler sunucu tarafında yapıldıktan sonra istemciye sadece otomatik üretilmiş soru ve üretilen cevabın özet bilgisi yollanacaktır. Eğer cevabın özeti değil de kendisi istemciye yollansaydı saldırganlar sunucu ve istemci arasındaki haberleşmeyi dinleyerek doğru cevabı tespit edebilecekti. Cevabın özetinin gönderildiği durumlarda saldırganlar sunucu ile istemci arasındaki haberleşmeyi dinleyerek cevabın özetini ele geçirseler bile bu bilgiyi cevap olarak kullanamayacaklardır. Oysa istemci için kullanıcının vermiş olduğu cevabın özetini hesaplayarak sunucudan gönderilmiş olan özet bilgisiyle



Şekil 3.2: Baz sorulardan otomatik yöntemlerle soru üretme işlemi.

karşılaştırmak oldukça kolaydır. Bu durumda kullanıcılar ancak soruyu doğru çözebilmeleri durumunda İEİ testini geçmiş olacaklardır.

Bahsedilen örüntü modülü içerisinde şu anda sekiz farklı örüntü hazır bulunmaktadır. Bu örüntüler kısaca aşağıdaki gibi özetlenebilir:

**Örüntü 1:** Bu örüntü, eğer sorunun cevabı bir sayıysa çalışmaktadır. Bu örüntü, baz soru tırnak içerisinde yazıldıktan sonra kullanıcıya cevabın  $[1, 10]$  eksiği, fazlası, katı sorulması şeklindedir.

*Örnek soru:* “baz soru” sorusuna vereceğiniz cevabın 3 katı kaçtır?

Bu örüntü tekniği ile cevabı sayı olan soru sayısının 30 katı kadar otomatik soru üretilebilmektedir. Bu tekniğin dezavantajı, baz sorular arasında cevabı sayı olan soru sayısının az olması ve bu soruların saldırganlar tarafından tespit edilmesi durumunda bu örüntünün güvenliğinin ortadan kalkmasıdır.

**Örüntü 2:** Bu örüntü, baz soru tırnak içerisinde yazıldıktan sonra kullanıcıdan cevabın sonuna, başına veya hem sonuna hem de başına otomatik olarak üretilmiş bir kelimeyi yazmasını istemek şeklindedir.

*Örnek soru:* “baz soru” sorusuna vereceğiniz cevabın başına “fkld” harflerini ekleyiniz?

Bu örüntü tekniği ile baz soru sayısının  $3m$  katı kadar yeni otomatik soru üretilebilmektedir. Burada  $m$  değeri, başa, sona veya hem başa hem de sona eklenecek olan kelimelerin sayısıdır. Eğer bu kelimeler anlamlı kelimeler olarak seçilmez de, rastgele uzunlukta rastgele harfler ve rakamlar olarak seçilirlerse,  $m$  değeri çok artacağından otomatik olarak üretilebilecek soru sayısı da çok fazla artacaktır. Bu tekniğin bir başka avantajı da, eklenecek kelime de tırnak içerisinde yazılacağından otomatik olarak yapılacak saldırıları yanıltabilecektir. Bu yöntemin bariz bir dezavantajı bulunmamaktadır.

**Örüntü 3:** Bu örüntü, baz soru tırnak içerisinde yazıldıktan sonra kullanıcıdan cevabın başına veya sonuna otomatik olarak üretilmiş basit bir aritmetik işlemin cevabını yazmasını istemek şeklindedir.

*Örnek soru:* “baz soru” sorusuna vereceğiniz cevabın sonuna “ $4 \times 3$ ” işleminin cevabını ekleyiniz?



Bu örüntü tekniği ile baz soru sayısının  $2m$  katı kadar yeni otomatik soru üretilebilmektedir. Burada  $m$  değeri, başa veya sona eklenecek olan aritmetik işlem sayısı kadardır. Aritmetik işlemin basit olması gerektiğinden  $m$  değeri bir önceki örüntüde olduğu kadar fazla büyüyemez. Bu yöntemin dezavantajı ise kullanıcılara bir yerine iki İEİ testi çözdürmesidir.

**Örüntü 4:** Kullanıcıdan, baz soruya vereceği cevabı iki kere yazması istenir.

*Örnek soru:* “baz soru” sorusuna vereceğiniz cevabı iki kere yazınız?

Bu örüntü tekniği ile baz soru sayısı kadar yeni otomatik soru üretilebilmektedir.

**Örüntü 5:** Kullanıcıdan, baz soruya vereceği cevaptan rastgele seçilmiş iki harfi yazması istenir. Burada seçilen iki harfin de boşluk karaktere denk gelmeyecek şekilde seçilmesi gerekmektedir.

*Örnek soru:* “baz soru” sorusuna vereceğiniz cevabın 3. ve 9. harflerini yazınız?

Bu örüntü tekniği ile baz sorulara verilen cevapların ortalama uzunluğunun karesi kadar yeni soru üretilebilmektedir. Bu tekniğin avantajları, günlük hayatta bankacılık işlemlerinde de kullanılması sebebiyle (“Anne kızlık soyadınızın 3. ve 5. harfleri nelerdir?” gibi) kullanıcıların alışkın olduğu tarzda bir soru olması ve çok fazla soru üretmeye olanak sağlamasıdır. Olası bir dezavantajı ise saldırganların rastgele iki harf deneyerek (belli bir olasılık ile) testi geçebilecek olmasıdır.

**Örüntü 6:** Bir önceki örüntüye çok benzeyen bu örüntüde, kullanıcıdan baz soruya vereceği cevapta rastgele seçilmiş iki harfin başına veya sonuna otomatik olarak üretilmiş bir kelimeyi yazması istenir. Burada seçilen iki harfin de boşluk karaktere denk gelmeyecek şekilde seçilmesi gerekmektedir.

*Örnek soru:* “baz soru” sorusuna vereceğiniz cevabın 4. ve 7. harflerini yazdıktan sonra “şalsdk” kelimesini de yazınız?

Bu örüntü tekniği ile baz sorulara verilen cevapların ortalama uzunluğunun karesi ile  $2m$  çarpımı kadar yeni soru üretilebilmektedir. Burada  $m$  ifadesi, otomatik olarak üretilen kelime sayısını göstermektedir. Bu örüntünün avantajları, bir önceki örüntünün avantajlarına benzemekle birlikte, üretilebilecek soru sayısı gerçekten çok fazladır. Diğer taraftan kullanıcılar sorulan soruyu algılamakta zorluk çekebilirler.

**Örüntü 7:** Bu örüntüde, kullanıcıya iki adet baz soru gösterilerek bu iki baz sorunun cevabını birbirine eklemesi istenmektedir.

*Örnek soru:* “baz soru 1” sorusu ile “baz soru 2” sorusuna vereceğiniz cevapları sırasıyla yazınız?

Bu örüntü tekniği ile otomatik olarak üretilebilecek soru sayısı, baz soru sayısının karesi kadardır. Bu örüntünün avantajı, reCAPTCHA servisinde de iki farklı test olduğundan kullanıcıların iki test çözmeye aşına olmalarıdır. Örüntünün dezavantajlarından bahsetmek gerekirse, kullanıcılara iki test çözdürüldüğü için kullanışlılık düşmektedir. Fakat daha önemli bir sorun, baz sorular açık olarak sorulacağından dolayı bu soruların bir veritabanına saklanması ve daha sonra veritabanından soru-cevap ikilisi eşlenerek hemen cevabın verilebilmesine olanak sağlamasıdır.

**Örüntü 8:** Bu örüntüde, kullanıcıya cevabı sayı olan iki adet baz soru gösterilir ve kullanıcıdan bu iki baz sorunun cevaplarını toplaması, çıkarması ya da çarpması istenir.

*Örnek soru:* “baz soru 1” sorusuna vereceğiniz cevabın “baz soru 2” sorusuna vereceğiniz cevap ile çarpımı kaçtır?

Bu örüntü tekniği ile cevabı sayı olan soru sayısının karesinin iki katı kadar daha otomatik soru üretilebilmektedir. Tekniğin bariz bir avantajı bulunmamakla birlikte bir önceki örüntüde bahsedilen dezavantajlar aynen korunmaktadır. Ayrıca kullanıcılardan üç adet test çözmeleri istenmektedir.

Bahsedilen bu örüntüler örüntü modülü içerisinde bulunmaktadır. Sistemin modüler yapısı gereği yeni örüntü teknikleri eklenebileceği gibi, önceden hazırlanmış olan bazı teknikler isteğe göre çıkarılabilir. Çizelge Ek.2’de İngilizce testler için kullanılacak 26 örüntü daha verilmiştir.

Soru yazma süresinin soru cevaplama süresinden daha fazla olması sebebiyle, otomatik yöntemler kullanmanın önemi açıktır. Her ne kadar otomatik yöntemler kullanılarak soru sayısı artırılabilir, insan hesaplama ve otomatik yöntemleri beraber kullanan bir saldırgan başarılı olabilir. Fakat insan hesaplama kullanarak yapılan bir saldırıya herhangi bir İEİ sisteminin koruma sağlaması zaten mümkün değildir. Örüntü modülünün amacı, olası saldırılarda saldırganların toplam harcadıkları emek ve/veya iş gücünün artırılmasını ve bu sayede sistemin bir miktar daha güvenli olmasını sağlamaktır.

## 4. İNSAN HESAPLAMA ÇALIŞMALARI

Bu bölümde ilk önce insan hesaplama çalışmalarında kullanılan sistemin mimarisi ve yapısı anlatılmıştır. Daha sonra ise sırasıyla çalışmaların süreci, sonuçları ve sonuçların analizi verilmiştir. Bu bölümde en son olarak, bu çalışmalardan çıkarılan derslerden bahsedilmiştir.

### 4.1 Sistem Mimarisi

Çalışmalarda kullanılan sistemin mimarisi; veritabanı, kodlama, sunucu yazılımı ve sunucu ile alan adı kiralanması olmak üzere dört kısımda incelenmiştir.

#### 4.1.1 Veritabanı

MySQL, günlük 65.000'den fazla indirilme miktarı ile dünyanın en popüler açık kaynak veritabanıdır [41]. MySQL çoklu iş parçacıklı, çok kullanıcı, hızlı ve gürbüz bir veritabanı yönetim sistemidir.

Bu avantajları sebebiyle veritabanı olarak MySQL kullanılmıştır. Oluşturulan veritabanı tabloları ve açıklamaları Çizelge Ek.1'de verilmiştir.

Veritabanı beş farklı tablodan oluşmaktadır. Bu tablolardan *Answers*, her bir cevap için tekil olan cevap tanımlayıcısını, cevap metnini ve de bu cevabı kaç sorunun kullanmakta olduğu bilgisini tutmaktadır. Bir başka tablo olan *Questions* ise, yine her soru için tekil olan soru tanımlayıcısını, soru metnini, bu soruya karşılık gelen cevabın tanımlayıcısını ve de sorunun güvenli olup olmadığı bilgisini tutmaktadır. İEİ operatörlerinin hazırlamış oldukları sorular ile ilgili analizi yapabilmek için oluşturulmuş olan *OperatorQuestions* tablosunda hangi operatörün hangi soruyu kaç saniyede yazdığı, soruyu hangi soru hazırlama sayfasında yazdığı, güvenlik motorunun kaç milisaniye çalıştığı ve otomatik olarak kaç tane muhtemel cevap ürettiği, Bing arama motorunun kaç tane cevap döndürdüğü ve de sorunun güvenlik motorunda hangi aşamaya kadar gelebildiği bilgileri tutulmaktadır. *OperatorsPersonalInfo* tablosu sadece birinci çalışmada kullanılmış

olup, diğ er ç alıřmalarda operatörlerin kiřisel bilgileri toplanmadıđı için kullanılmamıřtır. Operatörlerin adı, soyadı, yařı, cinsiyeti, üniversitede hangi bölümde öğrenim gördüđü ve öğrenci numarası, TC kimlik numarası ve iki farklı ücret ödeme tipinden hangisini seçmiř olduđu bilgileri tutulmaktadır. Son tablo olan *OperatorLoginInfo* tablosunda ise operatörlerin sisteme giriř yapabilmeleri için gerekli olan kullanıcı adı ve parola bilgileri, sisteme kayıt oldukları tarih ve en son giriř yaptıkları tarih ile birlikte sistemde toplam çalıřma süreleri bilgileri tutulmaktadır.

#### 4.1.2 Kodlama

JavaServer Pages (JSPler) dinamik web sayfaları ve uygulamalar oluşturulmasını sađlayan bir sunucu tarafı programlama teknolojisidir. HTML, XML, DHTML veya diğ er belge türlerinin içine Java kodu gömerek gerçekleştirilir. JSP'ler Java servlet teknolojisinin genişletilmiř versiyonudur. Java kodunun HTML ile bir dosyada birleřtirilmesini sađlar.

Oluřturulan internet sitesinin JSP dili ile kodlanmasında daha önceden JSP dili ile olan tanıřıklıđın yanı sıra, güvenlik motorunun Java programlama dili ile kodlanmış olması da büyük rol oynamaktadır. Eclipse IDE'nin kullanım kolaylıđı sađlaması ile birlikte önceden oluşturulmuş güvenlik motoru JSP sayfası üzerinde kolayca çağırılarak kullanılabilir.

#### 4.1.3 Sunucu yazılımı

İnternet sitelerini yayınlatabilmek için sunucu yazılımlarına ihtiyaç vardır. Tomcat de bir sunucu yazılımıdır.

Tomcat'in en önemli özelliđi esnek olmasıdır. Örneđin; Tomcat servisi ve JSP ile servletler bařka bir makinede iken Apache bařka bir fiziksel sunucuda çalıřtırılabilir. Bazı řirketlerin özellikle bu metodu seçmesinin sebebi de bu şekilde ek bir güvenlik sađlanmış olmasıdır. Tomcat'in bařka bir avantajı da kararlı bir yapıya sahip olmasıdır. Tomcat içinde önemli derecede bir hata durumu oluşursa sadece servletler ve JSP sayfaları işleneceđinden bu bütün Apache servisini kullanıřsız hale getirmez.

#### 4.1.4 Sunucu ve alan adı kiralınması

Bir internet sitesini yayınlatabilmek için sürekli açık bulunan bir sunucu gerekmektedir. İnternet sitesi sahipleri kendi bilgisayarlarını sunucu olarak kullanabilecekleri gibi, sunucu hizmeti veren firmalardan kendilerine özel bir alan kiralararak sayfalarını buradan yayınlatabilmektedirler. Sistemin her an erişilebilir olması için yüksek erişilebilirlik garantisi veren bir firmadan sunucu hizmeti kiralanmasına karar verilmiştir.

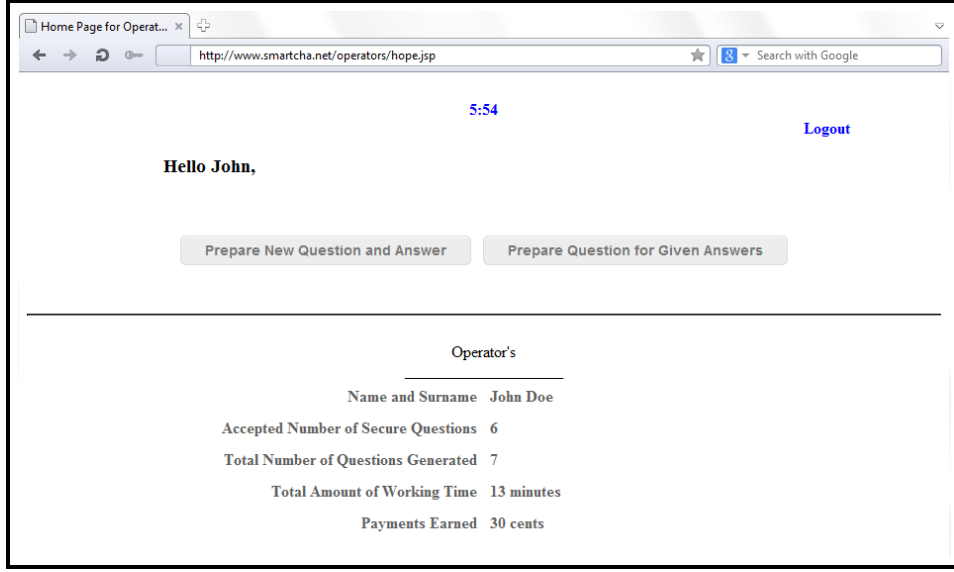
Hazırlanmış olan sistem gereği sunucunun MySQL veritabanı, JSP kodlama desteği vermesi ve de sunucu yazılımı olarak Tomcat kullanması gerekmektedir. Bu gereksinimlerle birlikte, sistemin aynı anda birden fazla kullanıcı tarafından kullanılabilir olması gerektiğinden yüksek miktarda bellek ihtiyacı vardır. Bu kriterler göz önüne alınarak bir araştırma yapıldığında [www.mochahost.com](http://www.mochahost.com) adresinde hizmet vermekte olan MochaHost firmasının aranılan özelliklerde sunucu kiralamakta olduğu görülmüştür.

Alan adı tescili de yapan MochaHost firmasından, SMARTCHA adı verilen sistem için [www.smartcha.net](http://www.smartcha.net) alan adı tescili yapılmış ve de yukarıda bahsedilen kriterleri sağlayan sunucu hizmeti kiralanmıştır.

## 4.2 Sistemin Yapısı

Sistem tasarlanırken soruların üretilmesi işinin insan hesaplama kullanılarak yapılacağı ve aynı anda birden fazla operatörün soru hazırlayabilmesi gerekliliği göz önünde bulundurulmuştur. Operatörlerin çalışmaları ortam, zaman, mekan, platform açısından kısıtlamamaya çalışılmıştır. Onlara istedikleri zamanlarda ve mekanlarda çalışabilme özgürlüğünü verebilmek için sistem bir internet sitesi olarak tasarlanmıştır. Bu sayede internet erişimi olan herhangi bir yerde çalışabilmeleri sağlanmıştır. [www.smartcha.net](http://www.smartcha.net) adresi üzerinden erişim sağlayabildikleri sitede operatörler sisteme kaydolduktan sonra kendi kullanıcı adları ve parolalarıyla giriş yapabilmektedir.

Operatörler web sitesinde kendi kullanıcı adları ve parolaları ile giriş yaparak Şekil 4.1'de görülen kendilerine özel giriş sayfasını açarlar. Bu sayfa her bir operatör için özelleşmiş olup operatörün adını ve soyadını, toplamda kaç soru hazırladığını, bu sorulardan kaç tanesinin güvenlik motoru tarafından güvenli olarak işaretlendiğini, operatörün saat karşılığı ücreti mi yoksa soru karşılığı ücreti mi tercih

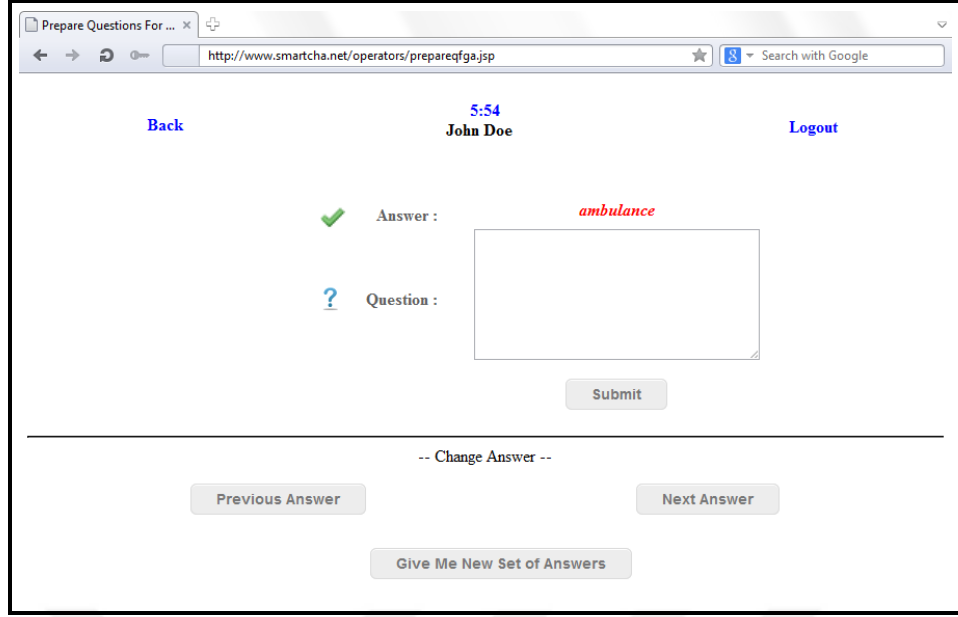


Şekil 4.1: Operatör giriş sayfası

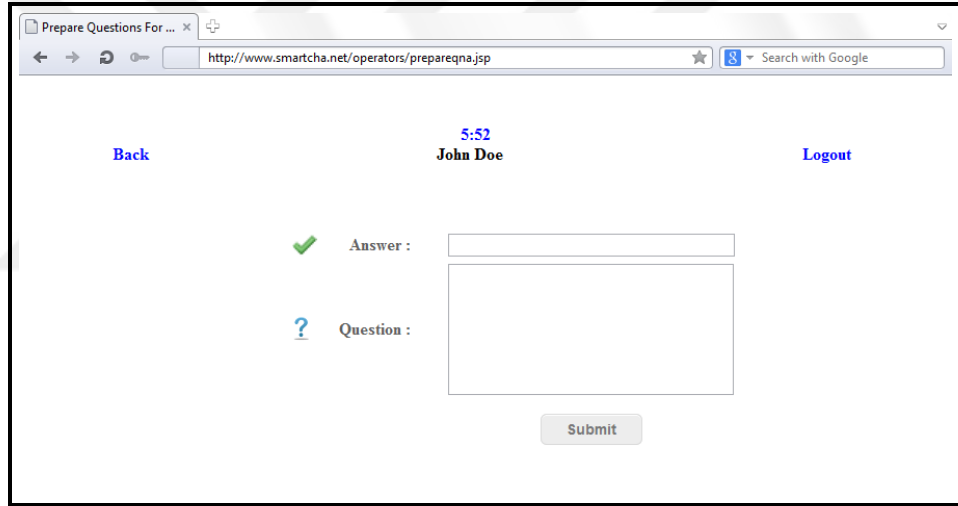
ettiğini ve o anda ne kadarlık ücret kazanmayı hak ettiğini görebildiği bir sayfa-  
dır. Operatörler ana sayfada bulunan “Prepare New Question & Answer” ya da  
“Prepare Question to Given Answers” düğmelerinden istedikleri birisine tıklayarak  
soru girme sayfasına girebilirler.

Ana sayfada bulunan operatör eğer “Prepare Question to Given Answers” düğ-  
mesine tıklarsa Şekil 4.2’de gösterilen 1. tip soru girme sayfasına yönlendirilir  
ve kendisine gösterilecek cevap kümesinin belirlenmesi işlemi başlar. Bu işlem,  
veritabanında bulunan cevapların kullanılma sayılarına göre azdan çoğa doğru  
sıralanmaları ve sonrasında daha az kullanılmış olanlara öncelik verecek şekilde  
rastgele 20 tanesinin seçilerek bir küme oluşturulmasıdır. 1. tip soru girme sayfa-  
sında operatöre cevaplar kümesinin ilk elemanı gösterilir ve operatörden sadece bu  
cevabın doğru olduğu bir soru üretmesi istenir. Operatör sadece bu cevabın doğru  
olduğu bir soruyu yazı alanına yazdıktan sonra “Submit Question For The Given  
Answer” düğmesine tıklamalı ve güvenlik motorunun sorunun otomatik olarak  
çözülüp çözülemediği hakkında vereceği kararı beklemelidir. Operatör herhangi  
bir cevap hakkında bir soru üretmek istemiyorsa “Next Answer” düğmesine tıkla-  
yarak kümedeki bir sonraki cevaba geçebilir veya “Previous Answer” düğmesine  
tıklayarak kümedeki bir önceki cevap ile ilgili soru üretebilir. Operatör talep et-  
tiği anda “Next Answer” veya “Previous Answer” düğmelerine tıklayarak cevap  
değiştirme hakkına sahip olduğu gibi, “Give Me New Set of Answers” düğmesine  
tıklayarak 20 soruluk yeni bir cevap kümesi seçim işlemini başlatabilir.

Eğer operatör ana sayfada bulunan “Prepare New Question & Answer” düğme-  
sine tıklarsa Şekil 4.3’te gösterilen 2. tip soru girme sayfasına yönlendirilir. 2.



Şekil 4.2: Operatörlerin kendisine verilen cevaba soru yazdığı sayfa



Şekil 4.3: Soru ve cevabı operatörlerin belirlediği sayfa

tip soru sayfasında operatörün hem bir soru, hem de bu sorunun tek cevabı olan uygun bir cevap yazması gerekmektedir. Operatörler bu sayfada altta yer alan yazı alanına yazmak istedikleri soruyu ve üstte yer alan yazı alanına ise bu sorunun cevabını yazdıktan sonra “Submit Question & Answer” düğmesine tıklayarak güvenlik motorunun soru ve cevabın otomatik olarak çözümlüğü hakkındaki kararını beklemelidir.

Her iki tip soru girme sayfasında da bulunan “Back” düğmesi yardımıyla operatörler giriş sayfasına dönüş yapma ve diğer soru girme sayfasına geçiş yapma hakkına sahiptir.

Operatörlerin soru oluşturma hızları veritabanına saniye olarak kaydedilmiştir. Bunun amacı operatörlerin hangi tipte daha hızlı biçimde soru hazırlayabildiklerini ölçebilmektir. Operatörlerin soru yazma hızlarını hesaplayabilmek için her iki tip soru hazırlama sayfasında benzer yöntemler uygulanmıştır. Eğer operatör 1. tip soru girme sayfasındaysa, operatör yazı alanını seçtikten sonra “Submit” düğmesine basana kadar geçen süre hesaplanarak kaydedilir. Eğer operatör soruyu 2. tip soru girme sayfasında hazırlıyorsa, operatör soru veya cevap alanından herhangi birisini seçtiği andan itibaren “Submit” düğmesine basana kadar geçen süre operatörün soru hazırlama süresi olarak veritabanına kaydedilir.

Bölüm 3.1.1’de çalışma mantığı anlatılmış olan güvenlik motorunun cevabı her iki tip soru hazırlama sayfasında da çok önemlidir. Eğer güvenlik kontrol motoru sorunun güvenli olduğuna karar vermişse hazırlanan soru-cevap ikilisi veritabanına güvenli soru, güvenli olmadığına karar vermişse soru-cevap ikilisi veritabanına güvenilir olmayan soru olarak eklenir. Her iki durumda da operatöre soru eklemenin başarılı veya başarısız olduğuna dair bilgilendirme yazısı gösterilir. Bilgilendirme yazısından sonra eğer operatör 1. tip soru girme sayfasındaysa operatörün aynı cevap için başka bir soru ekleyebilmesi için yazı alanı boşaltılır. Eğer operatör 2. tip soru girme sayfasında bulunuyorsa her iki yazı alanı da boşaltılarak operatörün yeni soru girebilmesi için hazırlanır.

Operatörler istedikleri anda giriş sayfasına dönmekte ya da sistemden çıkış yapmakta özgür bırakılmışlardır. Operatörlerin toplam çalışma zamanlarının hesabının doğru olarak yapılabilmesi için operatörlerin “Log Out” düğmesine basarak çıkış yapmaları gerekmektedir. “Log Out” düğmesine basmamış olmasına rağmen 6 dakika boyunca hiç bir işlem yapmayan operatörler de, çalışma süresi veritabanında güncellendikten sonra, sistemden otomatik olarak çıkarılırlar.

### **4.3 Soru Hazırlama Süreci**

Tasarlanan sistemin verimliliğini görebilmek için üç insan hesaplama çalışması yapılmıştır. Çalışmalar esnasında çıkarılan dersler bir sonraki çalışmada mümkün olduğunca kullanılmıştır. Birbiri arkasından gelen bu çalışmalar üç alt başlıkta toplanabilir:



### 4.3.1 Birinci çalışma

Bu çalışmada TOBB Ekonomi ve Teknoloji Üniversitesi lisans ve lisansüstü öğrencileri arasından ücret karşılığı çalışmaya gönüllü olmuş kişiler İEİ operatörleri olarak seçilmiştir. Bu ilk çalışma süresince operatörlerden İngilizce sorular hazırlamaları istenmiştir.

Bu çalışmada ücret karşılığı iş gücünden yararlanan operatörler hakkında sadece yaşı, cinsiyeti, hangi bölümde okuduğu gibi kişisel olmayan temel bilgiler toplanmıştır. Operatörlerin basit İngilizce sorular hazırlamaları gerektiğinden, öğrenciler seçilirken üniversitenin hazırlık programını geçmiş olmaları şartı aranmıştır.

Üniversite genelinde yapılmış olan duyuru (Şekil Ek.1) sonrası çalışmaya katılmayı planladığını belirten 78 operatörden 49 tanesi sisteme kayıt yaptırmış olsa da, bu operatörlerin sadece 29 tanesi 20 sorudan fazla üreterek sisteme katkıda bulunmuştur. 19 ila 29 yaş aralığında değişen bu 29 operatörün 20 tanesi (yaklaşık %69'u) erkek öğrencilerken, 9 tanesi (yaklaşık %31'i) kadın öğrencilerdir. Bu 29 öğrencinin 9 tanesi bilgisayar mühendisliği bölümü öğrencileriyken, diğer öğrenciler 10 farklı bölüm arasında, her bir bölümden en fazla 3 öğrenci olmak üzere dağılım göstermişlerdir. Sisteme kayıt olmuş 49 operatörün seçimleri incelendiğinde, bu operatörlerin 20 tanesi (yaklaşık %41'i) 5 saatlik çalışma karşılığında 15 TL ücret almayı tercih etmiş, geri kalan 29 operatör (yaklaşık %59'u) 100 adet güvenli soru hazırlama karşılığında 15 TL ücret almayı istemiştir.

Yaşları 19 ila 29 aralığında değişen 29 operatör, bir kereye mahsus olmak üzere bir sınıfta toplanılarak 1 saatlik oryantasyon süresince problem tanımı yapılmış, probleme karşı sunulan çözüm detaylıca anlatılmış, sunulan çözümün daha iyi nasıl olabileceği tartışılmış, var olan iki ödeme kriterinin neler olduğu belirtilmiş ve çalışma süresince sistemi nasıl kullanmaları gerektiği anlatılmıştır. Operatörlerin bütün soruları ayrıntılı biçimde cevaplanarak kafalarda soru işareti bırakılmaya özen gösterilmiştir.

Bu çalışma Mayıs 2012 sonlarına doğru başlayarak yaklaşık 1 ay sürmüştür.

### 4.3.2 İkinci çalışma

Bu çalışmada soruları oluşturacak operatörlerin bulunması için Amazon'un Mechanical Turk servisi [40] kullanılmıştır. Mechanical Turk, basit sayılabilecek işler

için ucuz iş gücüne ya da insan çeşitliliğine ihtiyaç duyan iş verenler ile bu işler arasında seçim yaparak istediği işlerde çalışan işçileri bir araya getiren bir platformdur. İş verenler kendi işlerinde çalışmasını istedikleri işçiler konusunda istedikleri gibi bir kısıtlama yapabildikleri gibi, herhangi bir kısıtlama yapmaksızın, işlerini isteyen herkesin yapabilmesine de imkan tanıyabilirler.

Bu çalışmada, ilk çalışmada elde edilen sonuçlara göre operatörlere ödenecek ücretin belirlenmesinde küçük değişikliklere gidilmiştir. Bu değişiklikler gereğince saat karşılığı ücret ödemesi uygulamadan kaldırılmıştır. Her 100 güvenli soru karşılığı olan ücret de her güvenli soru karşılığı ücret olarak güncellenmiştir. Bir başka değişiklik ise soru hazırlama yöntemlerinde yapılmıştır. Buna göre, operatörlerin hem soru hem de cevabı kendilerinin belirledikleri sayfa yayından kaldırılmış ve operatörlerin sadece kendilerine verilen cevaba soru yazabilecekleri bir sistem hazırlanmıştır.

Yapılmış olan bu çalışmada sistem yaklaşık olarak üç hafta boyunca açık kalmıştır. Mechanical Turk'ün evrensel bir sistem olması sebebiyle operatörlerden İngilizce sorular hazırlamaları istenmiştir. Bu üç haftalık süre içerisinde, dördüncü günün sonunda sistem beş gün için kapatılmış ve o ana kadar hazırlanmış olan sorular gözden geçirilmiştir. Sistemin kapalı olduğu beş gün içerisinde, belirlenen kullanılabilirlik kriterlerine uygun yapıda soru hazırlayan operatörler tespit edilmeye çalışılmış ve daha sonra sistem sadece bu operatörlerin kullanımına açılmıştır. İlk dört gün içerisinde toplamda 316 operatör çalışmada yer alırken, yapılan incelemelerden sonra bu operatörler arasından 44 tanesinin soru hazırlamaya devam etmesi uygun görülmüştür. Bu 44 operatörden 27 tanesi 20 taneden fazla soru hazırlayarak sisteme katkıda bulunmuştur.

Mechanical Turk servisinde işler HIT (Human Intelligence Task) olarak adlandırılır. Bu sistemde iş verenler işlerini küçük parçalara bölerler ve her parça için bir HIT açarak işlerinin yapılmasını sağlarlar. Bu çalışma kapsamında bir HIT, “bir güvenli sorunun sisteme girilmesi” olarak tanımlanmıştır. Mechanical Turk servisi iş verenleri, servisin sağlamış olduğu hazır şablonları kullanabileceği gibi, eğer uygun şekilde tasarlanmışlarsa, dış kaynaklı internet sitelerini de kullanabilmektedir.

Birinci çalışmada hazırlanan SMARTCHA sistemi, bu çalışma için Mechanical Turk ile uyumlu hale getirilmiştir. Bunun için her bir HIT sonrasında Mechanical Turk sistemine HIT'in tamamlandığına dair geri bildirim yapılmıştır. Sistem uyumlu hale getirildikten sonra HIT'in adını, tanımını, dış kaynaklı bir internet sitesi kullanılacağını, bu şekilde kaç tane HIT açılacağını, her bir HIT'in

ücretini ve bir operatörün en fazla kaç tane HIT yapabileceğini belirtmek için farklı parametrelerin olduğu bir dosya düzenlenmiştir. Sonrasında Java dili ile hazırlanmış olan bir program çalıştırılarak bu dosyada geçerli olan parametreler okunmuş ve HIT'ler oluşturulmuştur. Operatörlerin HIT başına hak ettikleri ücretlerin ödenmesi ise, Mechanical Turk servisinin web üzerinde sağladığı arayüz ile sağlanmıştır.

### 4.3.3 Üçüncü çalışma

Üçüncü ve son olarak yapılan bu çalışmada altı operatör bir hafta boyunca toplam 23 saat çalışmıştır. Bu çalışmada operatörlerden Türkçe sorular hazırlamaları talep edilmiştir.

Daha önceki çalışmaların aksine, sadece İngilizce dilinde servis veren Wolfram Alpha servisi güvenlik motorundan çıkarılmıştır. Buna karşın, soru içerisinde geçen bütün sayıları tespit ederek bu sayıların çarpımları, toplamları ile birlikte her birinin dokuz eksiğine ve dokuz fazlasına kadarını kontrol edip cevabın bu sayılar arasında olmadığından emin olan bir algoritma güvenlik motoruna yeni bir katman olarak eklenmiştir.

## 4.4 Sonuçlar

Yapılan her çalışma sayesinde bir sonraki çalışmada düzeltmeler yapılmış olması sebebiyle her çalışmada elde edilen sonuçlar ayrı ayrı incelenmiştir:

### 4.4.1 Birinci çalışma

Operatörler çalıştıkları bir ay süresince sisteme toplamda 10.776 soru üretmişlerdir. Sisteme girilmiş olan bu 10.776 sorunun 8.842 tanesi (yaklaşık %82'si) güvenlik motoru tarafından “güvenli” olarak işaretlenmiş ve geriye kalan 1.934 tanesi ise (yaklaşık %18'i) “güvensiz” olarak işaretlenerek sisteme kaydedilmiştir. Bu çalışmada üretilen sorulardan bazı örnekler Çizelge 4.1'de görülebilir. Örneklerdeki gramer hataları orijinal sorularda da mevcuttur, bu sebeple düzeltilmemişlerdir.

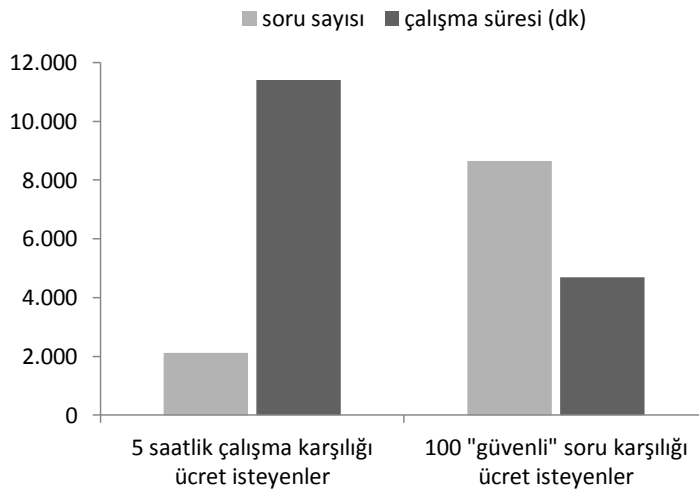
Projede süre olarak en çok çalışan operatörün 7.426 dakika çalışarak 200'ü güvensiz ve 503'ü güvenli olmak üzere toplamda 703 soru ürettiği gözlemlenmiştir.

Çizelge 4.1: Birinci insan hesaplama çalışmasında üretilen İEİ testlerine örnekler.

Soru	Cevap
How many eyes does a dog has?	2
What colour do you get when you mix blue and red?	purple
Increase each digit in 718 by 1. Result ?	829
1=A, 2=B, 3=C, 4=D, 5=?	E
What is the result of this, one hundred times ten plus one thousand plus sixty five?	2065
What is the result of this, one hundred times ten plus one thousand plus sixty six?	2066

Yaklaşık olarak 10 dakikada 0,9 soru üreten bu operatör, her 5 saatlik çalışma karşılığı ücret almayı talep etmiştir. Diğer taraftan, 203'ü güvensiz ve 5.729'u güvenli olmak üzere toplamda 5.932 soru ile en çok soru üreten operatörün toplam 1.990 dakika çalışmış olması dikkat çekmiştir. Yaklaşık olarak dakikada 2,98 soru üreten bu operatör ise, diğer operatörün aksine, yazmış olduğu her 100 güvenli soru karşılığı ücret almayı talep etmiştir.

Diğer bir taraftan, 5 saatlik çalışma karşılığı ücret almayı seçen operatörlerin toplamda 11.405 dakikada 2.118 soru ürettikleri (yaklaşık olarak dakikada 0,18 soru) gözlemlenmişken; 100 güvenli soru karşılığı ücret almayı tercih eden operatörlerin ise toplamda 4.697 dakikada 8.658 soru ürettikleri (yaklaşık olarak dakikada 1,85 soru) gözlemlenmiştir (Şekil 4.4).



Şekil 4.4: Operatörlerin aldıkları ücret tipine göre çalışma verimlilikleri.

#### 4.4.2 İkinci çalışma

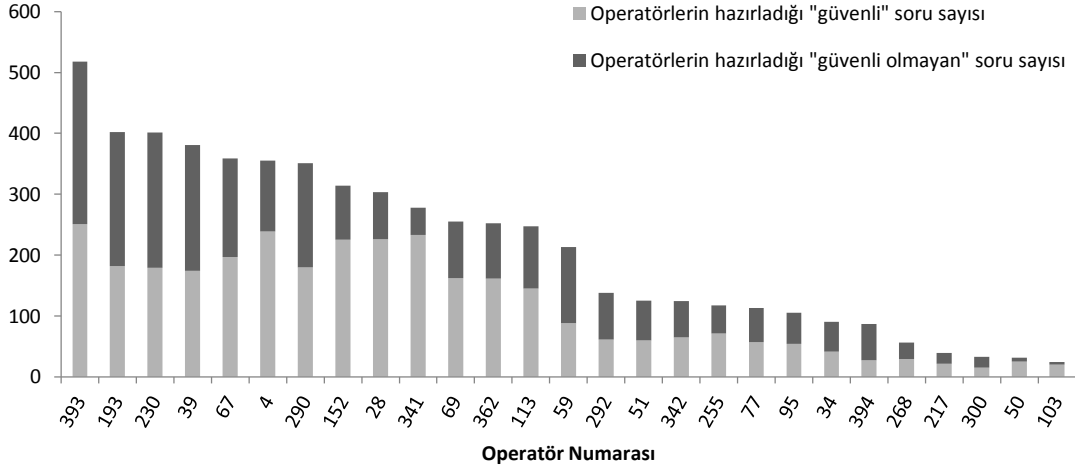
Yaklaşık olarak üç hafta süren bu çalışma süresince 5.460 tanesi güvenli olmak üzere toplamda 10.412 soru üretilmiştir. Bu sorulara örnekler Çizelge 4.2'de görülebilir. Bir günde 1.964 soru üretimine kadar çıkılan ilk dört günün sonunda 4.565 soru üretilmiştir. Bu noktada sistem beş gün için kapatılmış ve hazırlanmış olan sorular gözden geçirilmiştir. Bu beş gün içerisinde, belirlenen kullanılabilirlik kriterlerine uygun yapıda soru hazırlayan operatörler tespit edilmeye çalışılmış ve daha sonra sistem sadece bu operatörler için açılmıştır. İlk dört gün içerisinde toplamda 316 operatör soru üretirken, yapılan incelemeden sonra bu operatörler arasından 44 tanesinin soru hazırlamaya devam etmesi uygun görülmüştür.

Bu 44 operatörün çalışmaları sonucu 15 günde 3.220 tanesi güvenli olmak üzere toplam 5.847 soru üretilmiştir. Bu 44 operatör arasında en fazla soru üreten operatörün hazırladığı soru sayısı 518 iken, sadece bir soru hazırlayan operatörler de vardır. Seçilmiş olan 44 operatör arasından 27'si, 20 taneden fazla soru üreterek sisteme katkı sağlamıştır. Bu operatörlerin ürettiği soru sayıları Şekil 4.5'te görülebilir.

Bu çalışmada dikkat çeken bir başka önemli nokta ise, operatörlerin hazırladıkları soru sayısının günde 100 sorunun altına düşmemesi olmuştur. Şekil 4.6'da da görülebileceği üzere, en az soru hazırlanan günde bile 42 tanesi güvenli olmak üzere 113 tane soru üretilmiştir. Bu gözlem, aynı operatörler kullanıldığında bile günlük üretilebilen soru miktarının belirli bir seviyenin üzerinde tutulabilir olduğunun göstergesidir.

Çizelge 4.2: İkinci insan hesaplama çalışmasında üretilen İEİ testlerine örnekler.

Soru	Cevap
What vehicle takes you to the hospital?	ambulance
HOW MANY COLOURS DOES THE RAINBOW HAVE?	7
Type the letters "ITN" in lower case and in the following order: second letter first, first letter second, third letter third.	tin
Electronics Device that can spread the input voice.	microphone



Şekil 4.5: İkinci insan hesaplama çalışmasında operatörlerin hazırladıkları soru ve güvenli soru sayıları.

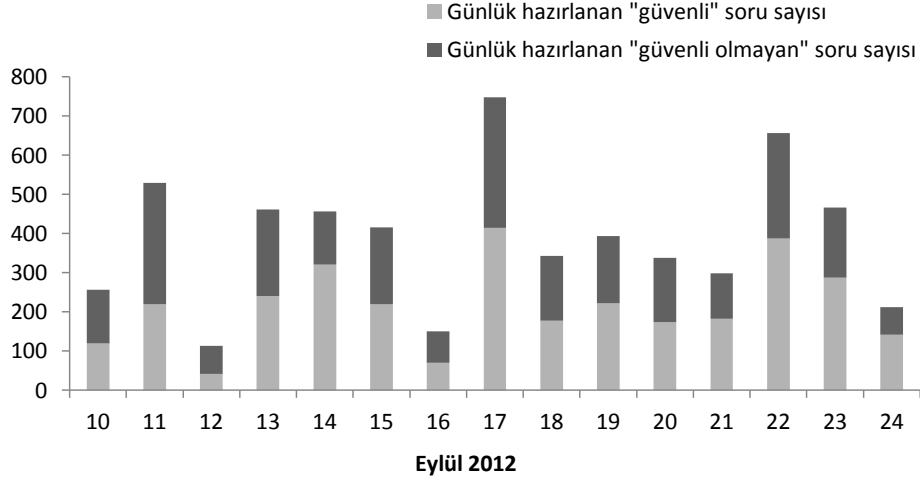
#### 4.4.3 Üçüncü çalışma

Yapılan bu son çalışmada amaç hem Türkçe soruların hazırlanması ile ilgili verileri elde edebilmek, hem de çalışma amacı üretmiş olduğu güvenli soru başına ücret almak değil de sistemi işlevsel hale getirebilmek olan operatörler çalıştığında nasıl bir sonuçla karşılaşılacağını görmektir. Bu çalışmada üretilmiş olan sorulara örnekler Çizelge 4.3'te görülebilir.

Bir hafta boyunca açık kalan sistemde altı operatör toplamda 23 saat çalışarak 739 soru üretmişlerdir. Bu soruların 692 tanesi güvenli soru olarak sisteme kaydedilmiştir. Altı operatörün soru hazırlama sayıları Şekil 4.7'deki gibidir.

Çizelge 4.3: Üçüncü insan hesaplama çalışmasında üretilen İEİ testlerine örnekler.

Soru	Cevap
1000 santimetre kaç metredir?	10
Adım atarak ilerleme şekline ne ad verilir?	yürümek
Miyavlaması ile bilinen ve evcil hayvan olarak da beslenen hayvanın adı nedir?	kedi
ana karaların çevresin kaplayan tuzu suya ne denir?	deniz
sihir kelimesinde s ve h harflerini f ve k harfleri ile değiştiriniz.	fikir



Şekil 4.6: İkinci insan hesaplama çalışmasında günlük hazırlanan soru sayısı.

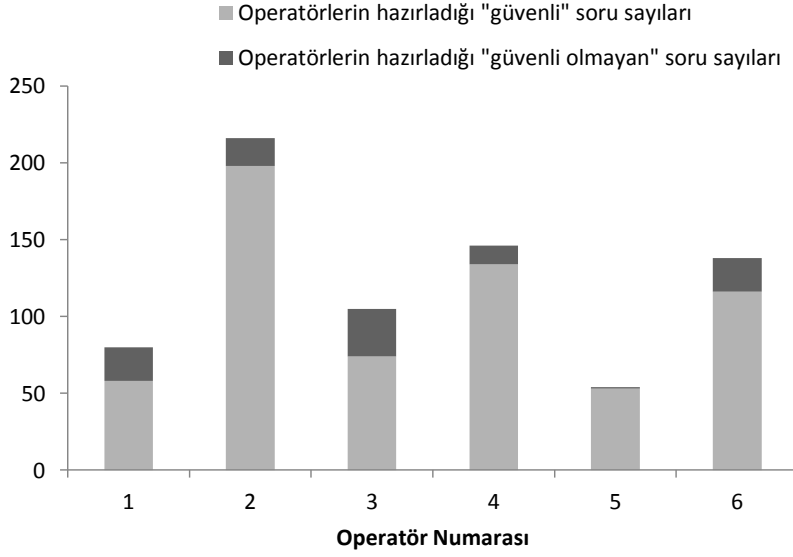
#### 4.5 Sonuçların Analizi

Çalışmalar süresince operatörlerin çalışmalarını ve hangi çalışmanın daha etkin olduğunu inceleyebilmek amacıyla bazı istatistiksel veriler tutulmuştur. Bu incelemeleri yapabilmek adına, her bir soru için,

- Operatörlerin soruyu yazma süreleri (saniye),
- Güvenlik motorunun çalışma süresi (milisaniye),
- Sorunun güvenlik motorunda hangi aşamaya kadar gelebildiği,
- Güvenlik motorunun otomatik olarak üretmiş olduğu muhtemel cevap sayısı,
- Bing arama motorunun kaç kelime bulabildiği

verileri saklanmıştır.

Burada operatörlerin soruyu yazma süreleri, soruyu yazacakları metin alanına tıkladıkları andan itibaren soruyu gönderdikleri ana kadar geçen süredir. Bu süre, bir sorunun ortalama olarak ne kadar sürede üretildiğini göstermektedir. Güvenlik motorunun çalışma süresi bilgisi, operatör soruyu sisteme gönderdikten sonra güvenlik motorunun sorunun güvenli olup olmadığına karar vermesine kadar geçen süredir. Eğer her soru için güvenlik motoru uzun süreler boyunca çalışırsa, operatörlerin soru üretme süreleri artacağı gibi çalışma şevkleri de kırılacağından incelenmesi gereken bir parametredir. Sorunun güvenlik motorunda hangi



Şekil 4.7: Üçüncü insan hesaplama çalışmasında operatörlerin hazırladıkları toplam soru ve güvenli soru sayıları.

aşamaya kadar gelebildiği bilgisi, güvenli olmayan soruların güvenli olmadığına hangi aşamada karar verildiğini göstermektedir. Bu tespit ile gereksiz olduğu düşünülen fakat fazla zaman alan aşamalar kaldırılabilir. Güvenlik motorunun otomatik olarak üretmiş olduğu muhtemel cevap sayısı, bütün aşamalarda üretilmiş olan muhtemel cevapların toplamıdır. Bu sayı, soruların hangi oranda güvenli olduğunun tespiti açısından önemlidir. Örneğin güvenlik motoru güvenli bir soru için otomatik olarak 20 farklı muhtemel cevap üretmişse, sorunun doğru cevabı otomatik olarak üretilen muhtemel ilk 20 cevap arasında yer almadığından o sorunun güvenliği en az %5'tir denilebilir. Bing arama motorunun güvenlik motorunda kullanımı, doğru cevabın Bing arama motorundan dönen cevaplar arasında en sık geçen 10 kelimedenden birisi olup olmadığı kontrolü yönündedir. Bing arama motorundan dönen cevaplar arasında en sık geçen 10 kelime yerine cevapların tamamı kullanıldığında daha etkin bir güvenlik motoru sağlanabilecektir. Bu bilgi ile soruların ortalama güvenliğinin hangi oranlara çekilebileceği hesaplanabilmektedir.

Her bir çalışmada elde edilen veriler çalışmaların kendi başlığında incelendikten sonra bu verilerin karşılaştırmalı grafikleri bölümün sonunda topluca verilecektir.

#### 4.5.1 Birinci çalışma

Birinci çalışma, operatörlerin hem soruyu hem de cevabı yazdıkları çalışma ile operatörlerin kendilerine verilen cevaba soru hazırladıkları çalışma olmak üzere iki farklı kısımdan oluşmaktadır.



Operatörlerin kendilerine verilen cevaba soru hazırladıkları sayfa aracılığıyla 1.656 tanesi güvenli olmak üzere toplamda 2.402 soru hazırlanmıştır. Hazırlanmış olan 2.402 soru için soru yazma hızı ortalaması 39,995 saniyedir. Bu sorular için güvenlik motorunun çalışma süresi ise 3.155 milisaniyedir.

2.402 sorunun 1.656 tanesi (%68,94) güvenli olup, hiçbir aşamada otomatik olarak çözülememiştir. Geriye kalan 746 sorunun bir tanesi (%0,04) daha önceki sorulardan biriyle birebir aynı olduğu için, 15 tanesinde (%0,63) cevap sorunun içerisinde yer aldığı için, 30 tanesi (%1,25) TextCaptchaBreaker algoritmaları ile çözülebildiği için, 27 tanesi (%1,12) aritmetik işlem sorusu olduğu için, 543 tanesi (%22,60) Bing arama motoru tarafından çözülebildiği için ve geriye kalan 130 tanesi de (%5,41) Wolfram Alpha servisi tarafından çözülebildiği için güvenli olmayan soru olarak işaretlenmiştir.

Bu sayfa aracılığıyla üretilen her bir soru için güvenlik motoru otomatik olarak ortalama 17 cevap üretmiştir. Yine bu sayfada üretilmiş her bir soru için Bing arama motoru ortalama olarak 538 farklı kelime döndürmüştür. Güvenlik motoru, Bing arama motorundan geriye dönen cevaplar arasında en çok kullanılmış olan 10 tanesini muhtemel cevap olarak kabul etmektedir. Bing arama motoru her soru için ortalama 538 farklı kelime döndürmüş olmasına rağmen soru başına otomatik olarak üretilen cevap sayısı ortalamasının 17 olmasının sebebi budur.

Operatörlerin hem soruyu hem de cevabı kendilerinin yazdıkları sayfa aracılığıyla 7.208 tanesi güvenli olmak üzere toplamda 8.368 soru hazırlanmıştır. Bu soruların yazım hızı ortalaması 20,8754 saniye iken, bu sorular için güvenlik motorunun çalışma süresi ortalama olarak 4.329 milisaniyedir.

Bu sayfa aracılığıyla hazırlanmış olan 8.368 sorunun 7.208 tanesi (%86,14) güvenlidir. Geriye kalan 1.160 sorunun 18 tanesi (%0,22) daha önceki sorulardan biriyle birebir aynı olduğu için, 16 tanesinde (%0,19) cevap sorunun içerisinde yer aldığı için, 32 tanesi (%0,38) TextCaptchaBreaker algoritmaları ile çözülebildiği için, 19 tanesi (%0,23) aritmetik işlem sorusu olduğu için, 926 tanesi (%11,07) Bing arama motoru tarafından çözülebildiği için ve kalan 149 tanesi de (%1,78) Wolfram Alpha servisi tarafından çözülebildiği için güvenli olmayan soru olarak işaretlenmiştir.

Bu sayfa aracılığıyla üretilen her bir soru için güvenlik motoru otomatik olarak ortalama 22 cevap üretmiştir. Yine bu sayfada üretilmiş her bir soru için Bing arama motoru ortalama olarak 491 farklı kelime döndürmüştür.

### 4.5.2 İkinci çalışma

Bu çalışmada 5.460 tanesi güvenli olmak üzere toplamda 10.412 soru hazırlanmıştır. Hazırlanmış olan 10.412 soru için operatörlerin ortalama soru yazma süresi 50,5972 saniyedir. Bu çalışmada üretilen sorular için güvenlik motorunun çalışma süresi ise 3.879 milisaniyedir.

Üretilmiş olan 10.412 sorunun 5.460 tanesi (%52,44) güvenli olup, güvenlik motoru tarafından otomatik olarak çözülememiştir. Geriye kalan 4.952 sorunun 610 tanesi (%5,86) daha önceki sorulardan en az biriyle çok benzer olduğu için, 410 tanesinde (%3,94) cevap sorunun içerisinde yer aldığı için, 202 tanesi (%1,94) TextCaptchaBreaker algoritmaları ile çözülebildiği için, 134 tanesi (%1,29) aritmetik işlem sorusu olduğu için, 3.323 tanesi (%31,92) Bing arama motoru tarafından çözülebildiği için ve geriye kalan 279 tanesi de (%2,68) Wolfram Alpha servisi tarafından çözülebildiği için güvenli olmayan soru olarak işaretlenmiştir.

Bu sayfa aracılığıyla üretilen her bir soru için güvenlik motoru otomatik olarak ortalama 19 cevap üretmiştir. Yine bu sayfada üretilmiş her bir soru için Bing arama motoru ortalama olarak 450 farklı kelime döndürmüştür.

### 4.5.3 Üçüncü çalışma

Bu çalışmada toplam 739 soru hazırlanmıştır ve bu soruların 692 tanesi güvenli olarak işaretlenmiştir. Bu çalışmada operatörlerin ortalama soru yazma süresi 55,9323 saniye iken, üretilen sorular için güvenlik motorunun çalışma süresi ise 1.524 milisaniyedir.

Üretilmiş olan 739 sorunun 692 tanesi (%93,64) güvenli olup, güvenlik motoru tarafından otomatik olarak çözülememiştir. Geriye kalan 47 sorunun üç tanesinde (%0,41) cevap sorunun içerisinde yer aldığı için ve 44 tanesinde ise soru (%5,95) Bing arama motoru tarafından otomatik olarak çözülebildiği için güvenli olmayan soru olarak işaretlenmiştir.

Bu sayfa aracılığıyla üretilen her bir soru için güvenlik motoru otomatik olarak ortalama 32 cevap üretmiştir. Yine bu sayfada üretilmiş her bir soru için Bing arama motoru ortalama olarak 283 farklı kelime döndürmüştür.

#### 4.5.4 Karşılaştırmalı istatistikler

Rahat okunabilmeleri açısından, bu kısımda yer alan grafiklerde çalışma isimleri numaralarla gösterilmiştir. Birinci çalışmada operatörlerin kendilerine verilen cevaba soru hazırladıkları sayfadan elde edilen istatistikler “1.1” ifadesi ile, yine birinci çalışmada operatörlerin hem soru hem de cevap hazırladıkları sayfadan elde edilen istatistikler “1.2” ifadesi ile, ikinci çalışmadan elde edilen istatistikler “2” ifadesi ile ve son olarak da üçüncü çalışmadan elde edilen istatistikler “3” ifadesi ile gösterilmiştir. Numaralandırmalar Çizelge 4.4’te görülebilir.

Her üç çalışmada da operatörlerin hazırlamış oldukları güvenli soru sayısının toplam soru sayısına oranı Şekil 4.8’de gösterilmektedir. Görüldüğü üzere en çok güvenli soru hazırlanabileme oranı Çalışma 3’te gerçekleşmiştir. Bunun sebebi, operatörlerin sistemin nasıl çalıştığını iyi biliyor olması ve mümkün olduğunca fazla hem güvenli hem de kullanışlı soru girmeye çalışmalarıdır. Çalışma 3’ten sonra en yüksek güvenli soru hazırlanma oranı Çalışma 1.2’de gözükmektedir. Fakat bu durum maalesef aldatıcıdır. Daha sonra daha ayrıntılı bahsedileceği üzere, Çalışma 1.2’de operatörlerin her seferinde yeni bir soru hazırlamaya çalışmaktansa bulmuş oldukları güvenli bir sorunun üzerinde küçük değişiklikler yaparak hemen hemen aynısını göndermiş oldukları tespit edilmiştir. Bir güvenli soru tespit edildikten sonra bu sorunun benzeri birçok soru da sisteme girilmiş olduğundan bu yanlısına doğaldır.

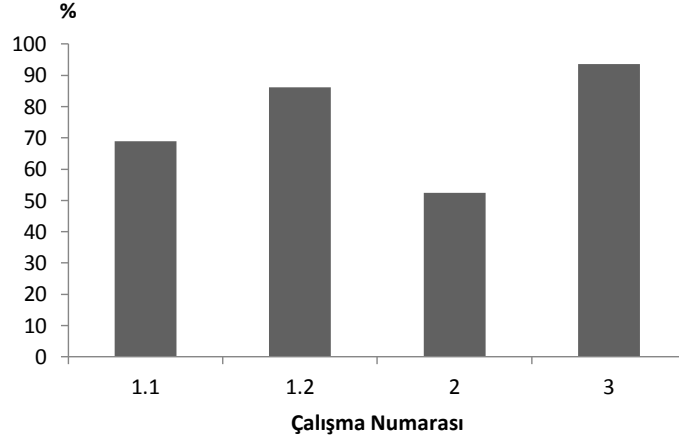
Benzer bir durum operatörlerin ortalama soru yazma süreleri karşılaştırmasında da gözlemlenmiştir. Şekil 4.9’da da görüldüğü üzere en kısa soru yazma süresi Çalışma 1.2’de gerçekleşmiş gibi gözükse de, bu durum operatörlerin güvenli olarak işaretlenen soruları “kopyala, yapıştır, değiştir ve gönder” şeklinde çalışmalarından dolayı kaynaklanmaktadır. Çalışma 3’te soru yazım hızlarının en fazla olmasının sebebinin ise, ücret yerine sistemin menfaati için çalışan operatörlerin

Çizelge 4.4: Yapılmış olan çalışmaların grafiklerde kullanılmak üzere numaralandırmaları.

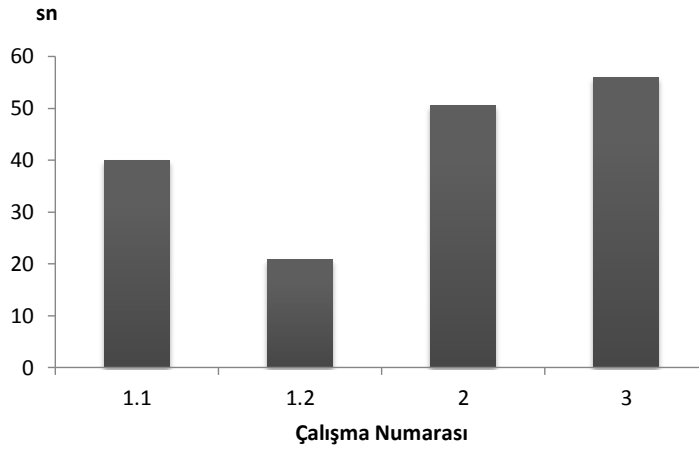
---

1.1	Birinci çalışmada operatörlerin kendilerine verilen cevaba soru hazırlaması ile elde edilen istatistikler
1.2	Birinci çalışmada operatörlerin hem soruyu hem de cevabı hazırlaması ile elde edilen istatistikler
2	İkinci çalışmadan elde edilen istatistikler
3	Üçüncü çalışmadan elde edilen istatistikler

---



Şekil 4.8: Her bir çalışmada hazırlanan güvenli soru sayısının toplam soru sayısına oranı (%).



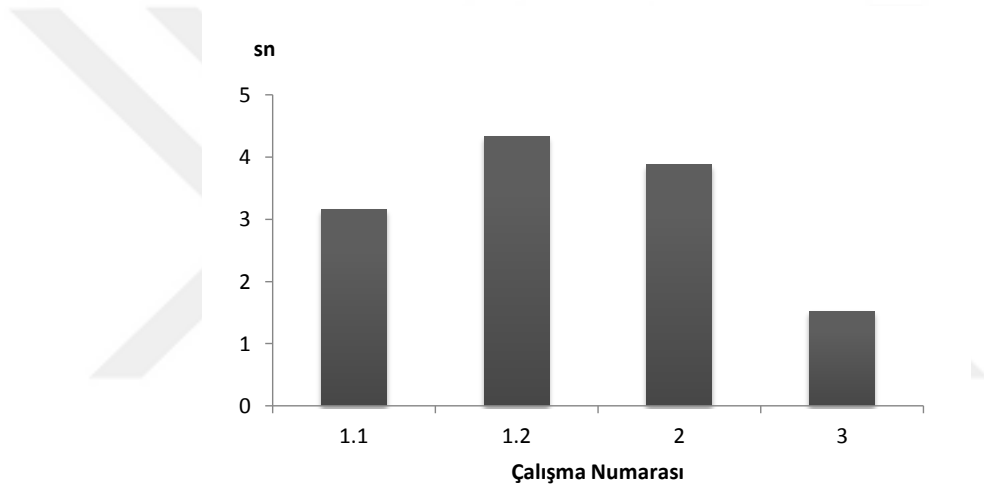
Şekil 4.9: Her bir çalışmada operatörlerin ortalama soru hazırlama süreleri.

yazmış oldukları soruları sisteme kaydetmeden önce tekrar okuyarak yazım hatasına karşı kontrol ve objektif gözle bakarak çözülebilirliğini teyit etmeleri olduğu düşünülmektedir.

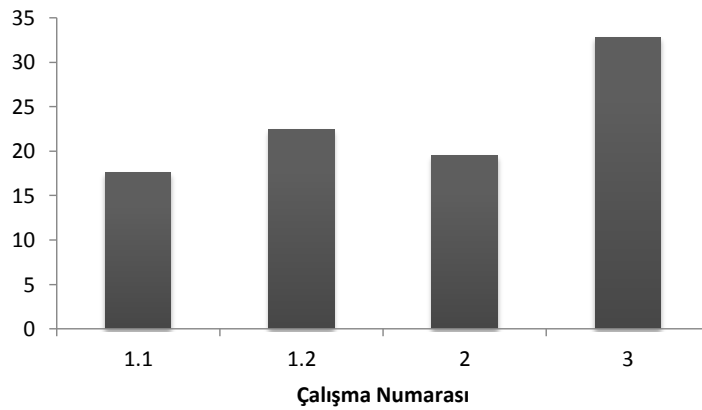
Her bir çalışmada güvenlik motorunun sorunun güvenli veya güvensiz olduğuna karar vermek için soru başına ortalama kaç milisaniye çalıştığı Şekil 4.10'da görülebilir. Güvenlik motorunun bir sorunun güvenli olup olmadığına en kısa sürede cevap verebildiği çalışmanın Çalışma 3 olduğu görülmektedir. Bunun sebebi, yalnızca İngilizce dilinde servis vermekte olan Wolfram Alpha servisinin diğer çalışmalarda kullanılmış olmasına rağmen Çalışma 3'te kullanılmamış olmasıdır. Diğer bir taraftan, güvenlik motorunun en uzun süreyle çalışması Çalışma 1.2'de gerçekleşmiştir.

Operatörlerin hazırladıkları sorulara güvenlik motorunun otomatik olarak ürettiği muhtemel cevap sayısı ortalaması Şekil 4.11’de görülebilir. Bu ortalama, üretilen sorunun hangi hassasiyette bir güvenlik testine tabii tutulduğuna dair bir fikir vermek için konulmuştur. Üretilen en çok muhtemel cevabın Çalışma 3’te elde edilmesinin sebebi, güvenlik motorunun soruda yer alan bütün sayıların çarpımları ve toplamları ile birlikte bu sayıların dokuz eksiği ve dokuz fazlasına kadar olan sayıları da muhtemel cevap olarak görmesidir. Daha önceki çalışmalarda böyle bir algoritma olmadığından o çalışmalarda üretilen muhtemel cevap sayıları birbirlerine yakındır.

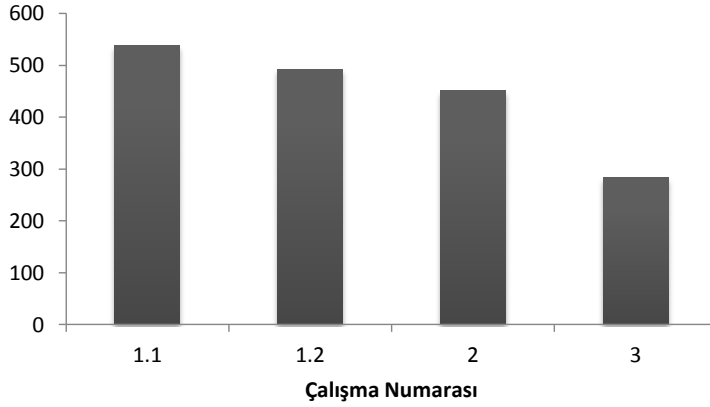
Şekil 4.12’de ise Bing arama motorunun her çalışma için sorulara döndürmüş olduğu ortalama kelime sayısı gözükmektedir. Çalışma 3’ün Türkçe yapılmış olması



Şekil 4.10: Her bir çalışmada güvenlik motorunun soru başına ortalama çalışma süresi.



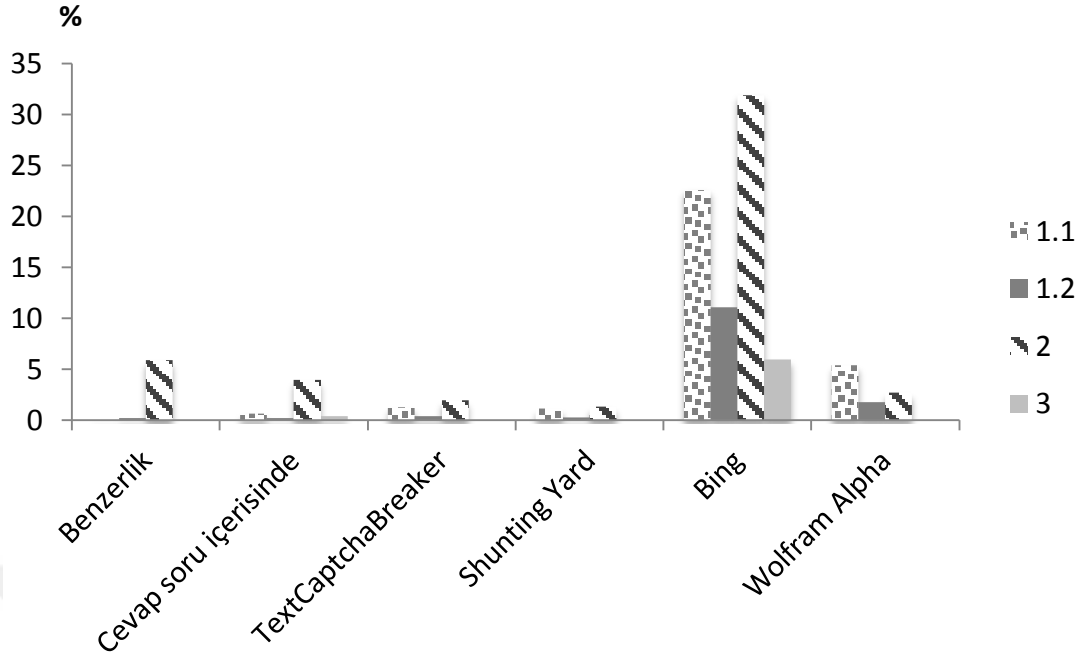
Şekil 4.11: Her bir çalışmada güvenlik motorunun soru başına ürettiği muhtemel cevap sayısı.



Şekil 4.12: Her bir çalışma için Bing arama motorunun soru başına bulduğu kelime sayısı.

sebebiyle Bing arama motorunun geriye daha az sonuç döndürdüğü düşünülmektedir. Bing arama motorundan geri dönen cevaplar arasında en çok kullanılan 10 kelime tespit edilerek cevabın bu 10 kelime arasında olup olmadığı incelendiğinden, güvenlik motorunun üretmiş olduğu muhtemel cevaplar arasına sadece ilk 10 kelime girmektedir. Bu sebeple Çalışma 3, Bing arama motorunun sorgu başına en az sonuç döndürdüğü çalışma olmasına rağmen, en çok muhtemel cevap üreten çalışma olma özelliğini kazanmıştır.

Yapılmış olan her çalışmada güvenli olmayan soruların hangi aşamada güvenli olmadığına karar verildiği tespit edilerek Şekil 4.13'te verilmiştir. Buna göre güvensiz sorular en çok Bing arama motorundan dönen sonuçlara göre tespit edilebilmektedir. Bing arama motorundan sonra soruların güvenilir olmadığını en çok tespit edebilen kısım Çalışma 1.1 ve 1.2 için Wolfram Alpha servisi, Çalışma 2 için ise benzerlik testidir. Çalışma 3'te Bing arama motoru dışında soruların güvenilir olmadığını tespit edebilen bir katman neredeyse yoktur. Sonuçların bu şekilde çıkması doğaldır, çünkü Çalışma 1.1, 1.2 ve 2'ye bakıldığında Wolfram Alpha servisinin soruları otomatik olarak çözebilme oranı birbirlerine yakındır. Buna karşın Çalışma 1.1 ve 1.2 süresince benzerlik ölçümü fazla etkin olmayıp yalnızca sorunun birebir aynısının daha önce veritabanında olup olmadığı kontrol edilmiştir. Çalışma 2'de operatörler ürettikleri güvenli soru başına ücret aldıklarından benzer sorular hazırlama eğilimleri artmış ve benzerlik ölçümü Wolfram Alpha servisine göre daha ön plana çıkmıştır.



Şekil 4.13: Her bir çalışma için güvenli olmayan soruların hangi aşamada güvenli olmadığına karar verildiğinin incelenmesi.

#### 4.6 Çıkarılan Dersler

Birinci çalışma süresince, her ne kadar bir ay gibi kısa sürede %82'lik güvenli soru oranıyla yaklaşık 11.000 soru üretilmiş olsa da, üretilmiş olan sorular incelendiğinde soruların tamamının orijinal olmadığı; bazı soruların güvenli olarak işaretlenmiş sorular üzerinde küçük değişiklikler yapılarak türetilmiş oldukları fark edilmiştir. Bu çalışmanın en büyük amaçlarından bir tanesi böyle bir sistemin sürdürülebilir olup olmadığını tespit etmeye çalışmak, başka bir deyişle, operatörlere yatırım yapmaya devam ettikçe yeni tür soruların bulunabilip bulunamadığını test etmektir. Bu sebeple operatörlerin güvenli olarak işaretlenmiş sorulardan yeni sorular üretmeleri sistemde otomatik olarak çözülemeyen soru sayısının artması açısından önemli olsa da, sürdürülebilirlik araştırmasına bir katkı sağlamamaktadır.

Bunun üzerine operatörlerin yazmış oldukları her soruyu kendisinin daha önceden yazmış olduğu diğer sorularla karşılaştırarak soruların benzerlik hesabını yapan bir algoritma güvenlik motoruna yeni bir katman olarak eklenmiştir. Bu algoritma gereğince her soru, aynı operatör tarafından yazılmış olan bütün güvenli sorularla karşılaştırılarak içlerinde en benzer olduğu soru bulunmuştur. Benzerlik

karşılaştırması için Levenshtein algoritması [42] kullanılmıştır. Eğer sorunun en benzer olduğu soru ile olan benzerliği belli bir orandan (%69 olarak seçilmiştir) fazla ise, bu sorunun başka bir sorunun değiştirilmesi yoluyla oluşturulduğuna karar verilmiştir. %69 değerinin seçilmesi işlemi için, kendisine en çok benzeyen soruya olan benzerliği %30 - %85 aralığında olan rastgele 200 soru seçilmiş ve en benzer oldukları sorularla karşılaştırılarak orijinal soru olup olmadıkları incelenmiştir. Burada elde edilen sonuçlara göre %69 değeri seçildiğinde, 200 soruda toplam üç tane hatalı tasnif işlemi yapıldığı ve bu oranın diğer oranlardan oldukça küçük olduğu gözlemlenmiştir. Bu algoritma ikinci ve üçüncü çalışmalarda güvenlik motoruna eklenmiş, eğer soru aynı operatörün bir başka sorusuna çok benziyorsa o soru güvensiz olarak işaretlenmiştir.

Bu algoritma sonucunda, sistemde bulunan 8.842 güvenli sorunun 1.322 tanesinin (yaklaşık %15'i) orijinal, kalan 7.520 tanesinin (yaklaşık %85'i) orijinal sorulardan türetme oldukları anlaşılmıştır. Operatörlerin bu şekilde soru türetmeyi tercih etmelerinin sebebinin daha fazla soru üreterek daha fazla ücrete hak kazanma güdüsü olduğu düşünülmektedir. Her ne kadar oluşturulmuş olan soruların büyük bir kısmı tekrar sorulardan oluşsa da, bu şekilde üretilen soruların hiç birisi güvenlik motoru tarafından otomatik olarak çözülemediğinden aslında hepsi İEİ olarak kullanılabilecek sorulardır.

Bu çalışmada dikkat çeken bir başka husus ise, operatörlerin hem soruyu hem de cevabı kendilerinin yazdıkları soru hazırlama sayfasını istenmeyen şekilde kullanmış olmalarıdır. Operatörler arasında soru yazılacak alana cevap, cevap yazılacak alana soru yazanlar olmuştur. Bu hata, bu sayfada yazılmış olan cevaplar otomatik olarak cevap veritabanına eklendiğinden, veritabanından otomatik olarak cevap isteyip ona soru yazmak isteyen diğer operatörleri de zor durumda bırakmıştır. Bundan daha tehlikeli bir başka hata ise, bazı operatörlerin cevabı yazarken yazım hatası yapmalarıdır. Bu sayfanın bir başka dezavantajı daha vardır. Bazı operatörler kendi yazdıkları sorular için çok spesifik cevaplar vermişlerdir, öyle ki, başka hiçbir operatör daha sonra bu cevaba soru hazırlayamamıştır. Bu sebeplerle operatörlerin hem soruyu hem de cevabı kendilerinin hazırladıkları sayfa yayından kaldırılarak ikinci ve üçüncü çalışmalarda kullanılmamıştır.

Birinci ve ikinci çalışmalarda hazırlanan sorular incelendiğinde soruların bir kısmının güvenli fakat kullanışlı olmayan sorular oldukları gözlemlenmiştir. Bu tarz sorulara örnek olarak "What is your lucky number?" sorusuna verilmiş olan "786" cevabı verilebilir. Bu soru-cevap ikilisi tamamen öznel bir yapıda düzenlenmiş olup, operatörlerden istenmiş olan nesnellik kriteri sağlanmamıştır. Bir başka ör-



nek olarak ise “226” cevabı için yazılmış olan “in which year Lu Kang, son of Lu Xun born?” sorusu verilebilir. Bu sorunun dünya genelinde hemen çözülemeyeceği, ancak Wikipedia gibi ek kaynaklar kullanılarak çözüme ulaşılabileceği aşıkardır. Bu da sorunun kullanışlılığını olumsuz yönde etkilemektedir.

Her ne kadar bu tarz kullanışlı olmayan sorular girmeleri durumunda sorular güvenli bile olsa kendilerine ücret ödenmeyeceği belirtilmiş olsa da; hatta daha da ileriye giderek, ikinci çalışmada uygulandığı gibi kullanışsız sorular hazırlayan operatörler elenerek sadece kullanışlı sorular hazırlayan operatörlerin çalışmasına izin verilmiş olsa da; çalışma motivasyonu hazırladığı güvenli soru karşılığı ücret almak olan operatörlerin bir noktadan sonra soruların kullanışlılığını göz önünde bulundurmaktan ziyade güvenlik motorunu geçebilecek sorular üretmeye çalıştığı gözlemlenmiştir.

Bu tarz soruların sisteme girilmesini otomatik olarak engelleyecek bir sistemin kurgulanması imkansız olmasa da, farklı uzmanlık alanları ve çok uzun çalışmalar gerektiren bir konudur. Fakat bu gibi kullanışlı olmayan soruların veritabanında kısıtlı sayıda bulunmaları durumunda sistem kullanıcıları bir sonraki İEİ testine geçerek kullanışlı bir soru çözebilecektir. Buna ek olarak, daha önce de bahsedilmiş olduğu gibi, İEİ testi uygulanan kullanıcıların soruları doğru ya da yanlış çözmelerinin istatistikleri tutularak bu tarz kullanışlı olmayan sorular veritabanından elenebilir. Bir başka yöntem ise filtreleme tekniği uygulamaktır ki bu yöntem bu çalışmada uygulanmıştır.



## 5. KULLANICI ÇALIŞMALARI

Bu bölümde yapılmış olan üç farklı kullanıcı çalışmasının yöntemi, sonuçları ve analizleri verilmektedir.

### 5.1 textCAPTCHA Kullanıcı Çalışması

Teorik güvenlik ile pratik güvenlik her zaman örtüşmeyebilir. Sayfada bulunan belirli bir metnin bir metin kutusuna yazılmasını isteyen akla gelebilecek en basit İEİ testi bile pratikte işe yarayabilir [15]. Yaygın bir şekilde kullanılmış birçok İEİ'ler de başarılı saldırılara maruz kalmıştır [1, 2, 3]. textCAPTCHA servisi, bir İEİ sisteminin kullanıma girebilmesi için güvenli olma şartının olmadığı bir başka örneğidir.

Yaygın kullanılan İEİ çeşitlerine pratik bir alternatif olmasına rağmen textCAPTCHA servisi tarafından sağlanan metin tabanlı İEİ'lerin kullanılabilirliğine dair daha önce yapılmış bir araştırmaya rastlanılmamıştır. Metin tabanlı İEİ'lerin aslında kullanılabilir ve erişilebilir olduğu fakat güvenli olmadıklarından günümüzde yaygın olarak kullanılmadıkları önceki bölümlerde bahsedilmiştir. Yapılan bu çalışma ile bir metin tabanlı İEİ sistemi olan textCAPTCHA servisinin kullanılabilirliği işitsel ve görsel reCAPTCHA servisiyle [23] karşılaştırılarak bu yorumun geçerliliği somut şekilde gösterilmek istenmiştir. Çalışmanın sonucunda elde edilecek olan somut sonuçlar, önerilen metin tabanlı İEİ sistemi olan SMARTCHA'nın kullanılabilirliği yönünde ipucu sağlayacaktır.

Yapılan kullanıcı çalışmasının sırasıyla yöntemi, çalışma öncesi oluşturulan hipotezleri, sonuçları ile bu sonuçların istatistiksel analizi, kısıtları ve son olarak sonuçların tartışması aşağıda yer alan alt bölümlerde verilmektedir.

#### 5.1.1 Yöntem

Bu kullanıcı çalışmasının Amazon Mechanical Turk [40] kullanıcılarıyla yapılması planlanmıştır. Çalışmanın yapılabilmesi için Amazon Mechanical Turk sistemiyle

uyumlu çalışan bir internet sitesi hazırlanmıştır. Çalışmaya katılım Amazon Mechanical Turk sistemine bir iş olarak yüklenmiş ve internet sitesi Mechanical Turk işçilerinin ziyaretine açılmıştır. Bu sitede kullanıcılara öncelikle çalışmanın amacından ve yaklaşık olarak ne kadar süreceğinden bahsedilmiştir. İş kabul ederek çalışmaya katılan kullanıcıların öncelikle Çizelge 5.1’de seçenekleri bulunan demografik sorulara cevap vermeleri istenmiştir.

Çalışmada kullanılan görsel ve işitsel İEİ’ler reCAPTCHA servisinden gerçek zamanlı olarak alınmıştır. Öte yandan, Bölüm 2.2.1’de anlatıldığı gibi textCAPTCHA servisinin İEİ testleri yedi başlıkta toplanabilen farklı soru tiplerinden oluşmaktadır. Çalışmada bazı soru tiplerinin yoğun kullanılmasının sonucu etkileyebileceği düşünülerek her soru tipinden eşit sayıda soru bulunması adına textCAPTCHA İEİ’leri yerel bir veritabanında saklanmış ve çalışma süresince kullanıcılara buradan servis edilmiştir.

Kullanıcıların incelenen İEİ sistemlere olan aşinalık seviyelerinin farklı olabileceği düşünülmüştür. Hem bu dezavantajı ortadan kaldırmak, hem de hazırlanan internet sitesine alışmalarını sağlamak için kullanıcıların her bir İEİ sisteminden çözdükleri ilk test alıştırmaya oturumu olarak planlanmıştır. Buna göre kullanıcılar gerçek oturumdan önce alıştırmaya oturumuna yönlendirilerek her sistemden (textCAPTCHA, görsel reCAPTCHA ve işitsel reCAPTCHA) birer tane olmak üzere üç tane İEİ testi çözmüşlerdir. Alıştırma oturumu bittikten hemen sonra gerçek oturum başlamaktadır. Gerçek oturumda kullanıcılardan yine her bir sistem için birer tane İEİ testi çözmeleri istenmiştir. Bu çalışmada her iki oturumun verileri de tutulmasına rağmen sadece gerçek oturumun verileri rapor edilmektedir.

Test arayüzü minimalist bir tasarıma sahiptir. Sayfalar sadece sistemin ismini, İEİ testini ve cevap alanını, soruyu değiştirmek için bir linki, yardım linkini, cevabı kaydetme düğmesini ve son olarak testi çözmeden atlamak için bir linki içermektedir. Ayrıca, anket sorularını yanıtlarken rahatça ayırt edilebilmeleri için

Çizelge 5.1: textCAPTCHA üzerine yapılan kullanıcı çalışmasında toplanan demografik bilgiler.

<b>Soru</b>	<b>Seçenekler</b>
Yaş	Bir tam sayı değeri
Cinsiyet	“Erkek” ya da “Kadın”
İngilizce Seviyesi	“Ana dil” ya da “Ana dil değil”
Görme Seviyesi	“Normal”, “Düzeltilmiş (Gözlük veya lensler ile)”, “Kısmi görme engelli” ya da “Görme engelli”

her sistem farklı renklerde gösterilmiştir. Alıştırma oturumu ile gerçek oturumun arayüzleri arasındaki tek fark, alıştırma oturumlarında bunun en üstte yer alan bir yazı ile belirtilmesidir. Görsel sistem için tasarlanan arayüz Şekil 5.1’de verilmiştir. Metin tabanlı ve işitsel sistemler için tasarlanan arayüzler de benzer olduklarından Şekil 5.2’de daha küçük olarak gösterilmiştir.

Kullanıcılardan bir testi yalnızca gerçekten çözemediklerine inandıkları zaman testi çözmeden atlama linkini kullanmaları rica edilmiştir. Herhangi bir kötüye kullanımı önlemek adına kullanıcılara bütün çalışma boyunca toplamda üç kere testi çözmeden geçme hakkı tanınmıştır (üçüncü haktan sonra testi çözmeden atlamak için verilen link sayfadan kalkar). Oluşabilecek herhangi bir sıralama etkisini ortadan kaldırmak için hem alıştırma oturumu hem de gerçek oturumda sistemler kullanıcılara farklı sıralarda gösterilmiştir.

Kullanıcılar iki oturum boyunca toplam altı adet İEİ testi çözdükten sonra kullanılan sistemlerin kullanılabilirliği ile ilgili bir anket sayfasına yönlendirilmişlerdir. Şekil 5.3’te görülen anket sayfasında kullanıcılardan sistem testlerinin çözülebilirliğini kolay ve eğlenceli olma bakımlarından 5’li Likert ölçeğinde oylamaları istenmiştir. Bunun yanı sıra, kullanıcıların hangi sistemi daha çok sevdiklerini anlayabilmek adına sistemleri “Birinci”, “İkinci” veya “Üçüncü” olarak sıralamaları istenmiştir.

System: Visual

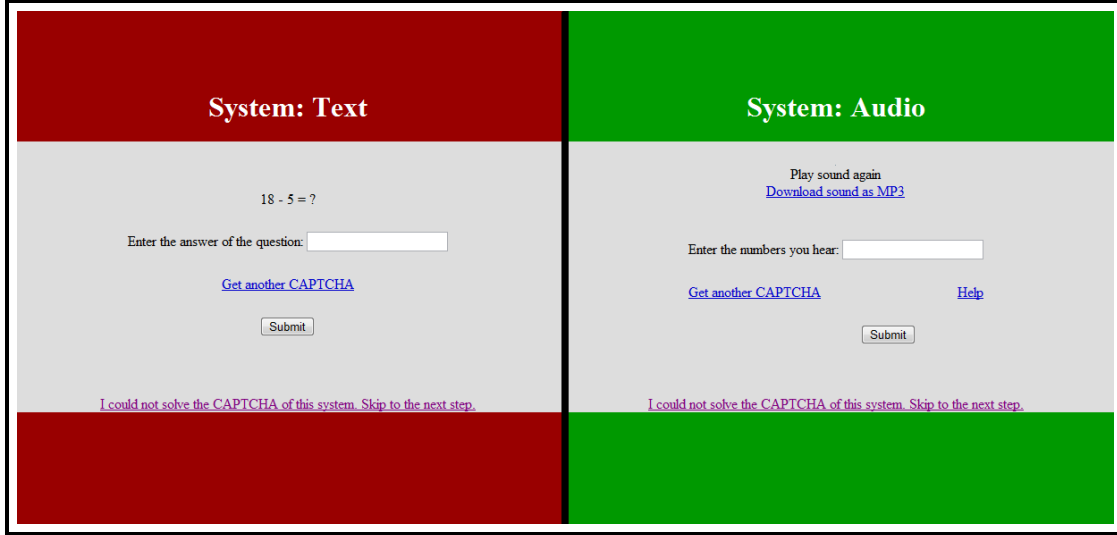
ory preferving

Enter the words above:

[Get another CAPTCHA](#) [Help](#)

[I could not solve the CAPTCHA of this system. Skip to the next step.](#)

Şekil 5.1: Kullanıcı çalışmasında kullanılan görsel İEİ arayüzü.



Şekil 5.2: Kullanıcı çalışmasında kullanılan metin tabanlı (solda) ve işitsel (sağda) İEİ arayüzleri.

### 5.1.2 Hipotezler

Çalışma öncesinde textCAPTCHA İEİ'ler üzerinde yapılan birkaç deneme sonrası bu testlerin kullanılabilirliği hakkındaki beklentiler oldukça yüksekti. Ayrıca otomatik saldırıların önlenmesi adına görsel İEİ'lerin bozulmaya uğratılıyor olmasının kullanıcıların hata yapmasına sebep olabileceği tahmin edilmekteydi. İşitsel İEİ'ler için, gürültülü ortamda kaydedilmiş olan sesi dinleme, karakterleri tanıma ve tanınan karakterleri metin kutusuna yazma sürelerinin yüksek olacağı düşünülerek bu İEİ tipi için beklentiler oldukça düşük tutulmuştu.

Bu çalışmada İEİ sistemi değiştirmenin kullanıcıların çözme süresi ile kullanıcılardaki kolaylık ve eğlenme algısı üzerine etkisi araştırılmıştır. Kullanıcı çalışması öncesinde sıfır hipotezleri aşağıdaki gibi belirlenmiştir:

- H1 textCAPTCHA İEİ'lerini çözme süresi ile görsel reCAPTCHA İEİ'lerini çözme süresi arasındaki fark anlamlı değildir.
- H2 textCAPTCHA İEİ'lerini çözme süresi ile işitsel reCAPTCHA İEİ'lerini çözme süresi arasındaki fark anlamlı değildir.
- H3 textCAPTCHA İEİ'lerini veya görsel reCAPTCHA İEİ'leri çözenin kullanıcıların kolaylık veya eğlenme algısına anlamlı bir etkisi yoktur.
- H4 textCAPTCHA İEİ'lerini veya işitsel reCAPTCHA İEİ'leri çözenin kullanıcıların kolaylık veya eğlenme algısına anlamlı bir etkisi yoktur.

You have encountered 3 systems. Please rate each system in a five-scale:

Audio :					
	Totally Agree	Agree	Undecided	Disagree	Totally Disagree
It was easy to solve questions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Solving these CAPTCHAs are enjoyable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Text :					
	Totally Agree	Agree	Undecided	Disagree	Totally Disagree
It was easy to solve questions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Solving these CAPTCHAs are enjoyable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Visual :					
	Totally Agree	Agree	Undecided	Disagree	Totally Disagree
It was easy to solve questions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Solving these CAPTCHAs are enjoyable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

---

Please rate three systems in terms of how much you like it:

	Best	Second	Third
Audio	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Text	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Visual	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Şekil 5.3: Kullanıcı çalışmasında kullanılan anket sayfası arayüzü.

### 5.1.3 Sonuçlar

Bu çalışmaya yaşları 18 ila 63 arasında değişen ve 80 tanesi (%58) 20 ila 30 yaş arasında olan 140 adet Mechanical Turk işçisi katılmıştır. Bu çalışmanın bedeli olarak her işçiye küçük miktarlarda (\$0,05 - \$0,10) para verilmiştir.

Katılımcıların 53'ü (%38) kadındır. 84 katılımcı (%60) görme seviyesi sorusuna “Normal” olarak cevap vermiştir. Geriye kalan 56 katılımcının tamamı (%40) bu soruyu “Düzeltilmiş (Gözlük veya lensler ile)” olarak cevaplarlarken hiçbiri “Kısmi görme engelli” veya “Görme engelli” seçeneklerini işaretlememiştir. Katılımcıların 59 tanesi (%42) İngilizce’yi ana dil olarak konuşmaktayken, kalan 81’i (%58) İngilizce’yi yabancı dil olarak kullandıklarını belirtmişlerdir.

Bu çalışmada soruları atlamadan çözen kullanıcılar o test için başarılı olarak değerlendirilmiştir. Bir kişi dışında bütün kullanıcılar textCAPTCHA İEİ'lerini başarılı bir şekilde çözmüşlerdir. Kullanıcılar her bir textCAPTCHA testi için en az 3 ve en çok 92 saniye olmak üzere ortalama 15,59 saniye harcamışlardır. Kullanıcılar toplam yedi kere (kullanıcı başına 0,05 kere) textCAPTCHA İEİ'si testi değiştirmiş ve toplam 13 kere (kullanıcı başına 0,09 kere) bu İEİ testlerine yanlış cevap vermişlerdir.

Görsel İEİ'lerin tamamı kullanıcılar tarafından başarıyla çözülmüştür. Kullanıcılar görsel İEİ'ler için en az 4 ve en çok 101 saniye olmak üzere ortalama 16,33 saniye harcamışlardır. Kullanıcılar toplam 16 kere (kullanıcı başına 0,11 kere) görsel İEİ testi değiştirmiş ve toplam sekiz kere (kullanıcı başına 0,06 kere) bu İEİ testlerine yanlış cevap vermişlerdir.

Birçok kullanıcı işitsel İEİ'leri çözememiş ve bu testleri atlamayı seçmiştir. Sadece 21 kullanıcı (%15), en az 12 ve en çok 174 saniye olmak üzere ortalama 52,47 saniyede işitsel İEİ'leri doğru bir şekilde çözebilmiştir. Bir kullanıcının işitsel İEİ testini 378 saniye boyunca çözmeye çalıştıktan sonra atlamaya karar vermesi ilginç bir anekdot olmuştur. Kullanıcılar işitsel İEİ'leri toplam 32 kere (kullanıcı başına 0,23 kere) değiştirmiş ve bu testlere 134 kere (kullanıcı başına 0,96 kere) yanlış cevap vermişlerdir. Başka bir deyişle, neredeyse her dört kullanıcıdan biri işitsel İEİ testlerini bir kere değiştirmiş ve neredeyse her kullanıcı testi bir kere yanlış çözmüştür.

Anket sonuçları şu şekilde sonuçlanmıştır: 5'li Likert ölçeğinde sistemler kolaylık bakımından oylandığında görsel İEİ'ler ortalama 4,4 ve işitsel İEİ'ler ortalama 1,79 alırken textCAPTCHA İEİ'ler ortalama 4,62 almıştır. Eğlenme bakımından oylama sonuçlarında ise görsel İEİ'ler ortalama 4,03 ve işitsel İEİ'ler ortalama 1,98 alırken textCAPTCHA İEİ'lerin ortalaması 4,35 çıkmıştır. Sistemleri kendi beğenilerine göre sıralamaları sorusuna verilen cevaplar incelendiğinde; textCAPTCHA İEİ'ler 102 kullanıcı (%73) tarafından "Birinci", 35 kullanıcı (%25) tarafından "İkinci" ve üç kullanıcı (%2) tarafından "Üçüncü" olarak seçilmiştir. Görsel İEİ'ler 36 kullanıcı (%26) tarafından "Birinci", 99 kullanıcı (%70) tarafından "İkinci" ve beş kullanıcı (%4) tarafından "Üçüncü" olarak seçilmiştir. İşitsel İEİ'ler iki kullanıcı (%1) tarafından "Birinci", altı kullanıcı (%4) tarafından "İkinci" ve 132 kullanıcı (94%) tarafından "Üçüncü" olarak sıralanmıştır.

Çizelge 5.2'de sonuçların ortalama ve standart sapma değerleri (uygulanabilir olduğu durumlarda, parantez içinde) verilmiştir.



Çizelge 5.2: İlk kullanışlılık çalışmasında toplanan verilerin ortalama ve standart sapma değerleri (parantez içinde).

	metin	işitsel	görsel
Sistemi geçmek için harcanan süre (sn)	15,59 (15,43)	52,47 (57,82)	16,33 (12,92)
Kullanıcı başına soru değişim oranı	0,05 (0,37)	0,23 (1,63)	0,11 (0,74)
Kullanıcı başına hata oranı	0,09 (0,38)	0,96 (1,65)	0,06 (0,23)
Başarılı çözüm oranı	0,99 (0,08)	0,15 (0,38)	1,00 (0,00)
Çözmenin ne kadar kolay olduğu	4,62 (0,63)	1,79 (1,16)	4,44 (0,75)
Çözmenin ne kadar eğlenceli olduğu	4,35 (0,93)	1,98 (1,29)	4,03 (1,12)
Birinci seçilme oranı	0,73	0,01	0,26
İkinci seçilme oranı	0,25	0,04	0,71
Üçüncü seçilme oranı	0,02	0,94	0,04

Elde edilen sonuçlar bir uyum iyiliği testi olan Kolmogorov-Smirnov testi ile analiz edildiğinde verilerin normal dağılıma uymadığı saptanmıştır. Bunun sonucu olarak analizlerde parametrik olmayan testler uygulanmıştır. Bütün analizler grup içi karşılaştırmalardan oluşmaktadır. Süre karşılaştırmalarında (H1 ve H2) Wilcoxon işaretli sıra testi, Likert ölçeği içeren anket verileri karşılaştırmalarında ise Friedman testi uygulanmıştır. Friedman testine göre anlamlı bir fark bulunduğu takdirde bu testi post hoc Wilcoxon işaretli sıra testi takip etmiştir. Anketin sonunda yer alan sıralama sorusu için Kendall uyum katsayısı (W) incelenmiştir. Analizlerde 0,05 değerinden küçük bir p-değeri sıfır hipotezini reddetmek için yeterli görülmüştür.

textCAPTCHA testlerinin ortalama çözülme süresi (15,59 saniye) görsel reCAPTCHA testlerinin çözülme süresinden (16,33 saniye) anlamlı derecede daha kısadır ( $W = 3460, p = 0,047$ ). Benzer şekilde, textCAPTCHA testlerinin ortalama çözülme süresi, işitsel reCAPTCHA testlerinin ortalama çözülme süresinden (52,47 saniye) anlamlı derecede daha kısadır ( $W = 1173, p < 0,001$ ). Bu sonuçlara göre H1 ve H2 hipotezleri reddedilmektedir.

Friedman testine göre textCAPTCHA ile görsel ve işitsel reCAPTCHA testlerinin çözülme kolaylığı arasında anlamlı bir fark bulunmaktadır ( $\chi^2 = 212,005, p <$

0,001). Bunun üzerine yapılan post hoc Bonferroni düzeltmeli Wilcoxon işaretli sıra testine göre kullanıcılar textCAPTCHA testlerini görsel ( $W = 637, p = 0,003$ ) ve işitsel ( $W = 7.996,5, p < 0,001$ ) reCAPTCHA testlerini çözmeye göre daha kolay bulmuşlardır. Benzer şekilde, Friedman testine göre farklı İEİ testlerini çözmenin ne kadar eğlenceli olduğu da anlamlı şekilde farklıdır ( $\chi^2 = 177,311, p < 0,001$ ). Bonferroni düzeltmeli Wilcoxon işaretli sıra testi textCAPTCHA testlerini çözmenin görsel ( $W = 1.018, p < 0,001$ ) ve işitsel ( $W = 7.236, p < 0001$ ) reCAPTCHA testlerini çözmekten daha eğlenceli olduğunu belirtmektedir. Bunun sonucu olarak H3 ve H4 hipotezleri de reddedilmektedir. Ek olarak, ankette yer alan sıralama sorusunda textCAPTCHA sistemi 102 kullanıcı tarafından birinci sıraya koyulmuşken, görsel ve işitsel sistemler sırasıyla 36 ve iki kez birinci sırada yer bulmuştur. Kendall uyum testine ( $W$ ) göre bu fark ileri derecede anlamlıdır ( $W = 0,706, \chi^2 = 197,586, df = 2, p < 0,001$ ). Bu sonuç da H3 ve H4 hipotezlerinin reddedilmesi yönünde ek bir kanıt olarak gösterilebilir.

Sonuç olarak bu çalışma için tanımlanan bütün hipotezler textCAPTCHA lehine reddedilmiştir.

#### 5.1.4 Kısıtlar

Çalışma öncesinde “Kısmi görme engelli” veya “Görme engelli” katılımcıların da çalışmada yer alması beklenmekteydi. Ancak katılımcılardan hiçbiri bu seçeneklerden birini işaretlememiştir. Bu sebeple textCAPTCHA servisinin erişilebilirliği hakkında herhangi bir sonuca ulaşamamıştır.

Sistemin arayüzünde kullanıcıların bir testi çözmeden geçebilmeleri için bir link bulunmaktadır. Bu link, çalışma boyunca en fazla üç kere kullanılacak şekilde düzenlenmiştir ve kullanıcıların çözemedikleri bir testte çalışmayı tamamen bırakmaları yerine diğer kısımları çözerek çalışmayı tamamlayabilmeleri amacıyla eklenmiştir. Ancak bu durum bazı kullanıcıların işitsel testleri çözmeyi gerçekten denemeden onları atlamalarına sebep olmuş olabilir.

#### 5.1.5 Tartışma

Bursztein *ve diğerleri* [8] tarafından 2010 yılında yapılmış bir kullanıcı çalışmasında reCAPTCHA başarı oranı %47 olarak bulunmuştur. Bu oran, bu kullanıcı çalışmasında gözlenen başarı oranından oldukça fazladır. reCAPTCHA servisinin

işitsel İEİ sistemi güvenlik açıklarına [9, 26] karşı sıklıkla değiştirilmektedir. Bu çalışmada kullanılan işitsel İEİ testleri Bölüm 2.2.2’de anlatılan eski sürüm olup, Bursztein *ve diğerlerinin* [8] kullanıcı çalışmalarında kullandıkları sürümden farklıdır. İşitsel reCAPTCHA İEİ’lerinin başarı oranındaki düşüş, bu testler üzerinde yapılan ve kullanışlılığı düşüren değişimlere [43] bağlanmaktadır.

Bölüm 2.2.1.1’de de bahsedildiği gibi, textCAPTCHA İEİ’leri bilgisayar programları tarafından otomatik olarak çözülebildikleri için güvenli değildirler. Ancak bu kullanışlılık çalışmasıyla textCAPTCHA, özellikle işitsel reCAPTCHA İEİ’leri ile karşılaştırıldığında, kullanışlı ve tercih edilir bir seçenek olarak kendini kanıtlamıştır (işitsel İEİ’leri çözmek için harcanan süre textCAPTCHA İEİ’leri çözmek için harcanan süreden üç kat daha fazladır). Bu sonuçlara göre, eğer güvenli olsaydı, textCAPTCHA sisteminin kullanılan İEİ sistemlere geçerli bir alternatif olabileceği görülmektedir.

## 5.2 SMARTCHA Kullanıcı Çalışması

SMARTCHA sisteminin tasarımı ve metin tabanlı İEİ testlerinin hazırlanması süreçleri sırasıyla Bölüm 3 ve 4’te ayrıntılı şekilde anlatılmıştır. Yapılan geniş bir kullanıcı çalışmasıyla SMARTCHA sisteminin kullanışlılığı görsel reCAPTCHA İEİ’leri ile karşılaştırılarak değerlendirilmiştir. Sonuçlar bu bölümde rapor edilmiştir.

SMARTCHA İEİ’lerinin kullanışlılığı incelenirken farklı durum senaryoları da test edilmiştir. Bu durum senaryoları sayesinde soru sayısını artırmak için otomatik tekniklerin kullanılmasının (*yarı-otomasyon*) kullanışlılığa etkisi ile insan hesaplama kullanarak daha kullanışlı soru kümelerinin seçilmesinin (*soru ön filtreleme*) kullanışlılığa etkisi incelenebilmiştir.

Yarı-otomasyonun etkisini inceleyebilmek için Bölüm 3.1.3’te anlatılan örüntü modülünde tanımlanan yarı-otomasyon teknikleri seçilerek kullanılmıştır. Soru ön filtreleme işlemi için Bölüm 4’te anlatılan insan hesaplama çalışmalarına katılan iki İEİ operatörüne yardımcı olmaları rica edilmiştir. Bu iki operatör toplamda bir saat çalışarak güvenli sorular arasından kullanışlı olduğunu düşündükleri 100 tanesini seçmişlerdir. Seçilen bu 100 sorunun yarı-otomasyon teknikleri uygulanmış ve uygulanmamış halleri ile yapılan testler sırasıyla Durum1 ve Durum2 senaryolarında incelenmiştir. Durum3 ve Durum4 senaryolarında ise soru ön filtreleme kullanılmamıştır. Bunun yerine, SMARTCHA veritabanında güvenli ve aktif

olarak işaretlenmiş bütün testlerin arasında rastgele yapılan seçimle elde edilen sorulara yarı-otomasyon tekniklerinin uygulandığı ve uygulanmadığı senaryolar sırasıyla Durum3 ve Durum4'te incelenmiştir.

Sonuç olarak, SMARTCHA sisteminin kullanılabilirliği dört farklı durum senaryosu ile incelenmiştir. Kolaylık açısından Çizelge 5.3'te gösterilen bu durum senaryolarında SMARTCHA testleri yukarıda anlatıldığı şekilde elde edilmiş ve her bir senaryo görsel İEİ'ler ile ayrı ayrı karşılaştırılmıştır. Yarı-otomasyon tekniklerinin uygulanması İEİ testlerinin sayısını artırırken büyük olasılıkla kullanılabilirlikleri düşürmektedir. Diğer yandan, soru ön filtrelemenin kullanılabilirliği artırabileceği fakat baz soruların sayısını azaltabileceği düşünülmektedir. Bunun yanı sıra soru ön filtreleme işlemi ekstra insan gücü gerektirmektedir. Bahsedilen durum senaryoları detaylı incelenmesi, kullanılabilirlik ve ölçeklenebilirlik arasında denge kurulabilmesi açısından önem arz etmektedir.

### 5.2.1 Yöntem

Bu çalışma üniversitenin e-posta listesine gönderilen bir ilan ile duyurulmuş ve gönüllülük esasına göre yürütülmüştür. Çalışmaya katılımı teşvik amacıyla TÜBİTAK tarafından desteklenen bir proje için yardıma ihtiyaç duyulduğu belirtilmiştir. Çalışmanın internetten ulaşılabilecek linki de ilan metninde verilmiştir. Katılımcılara sistemde yer alan Türkçe ve İngilizce testlerinden istediklerini seçerek çözme hakkı tanınmıştır fakat neredeyse bütün kullanıcılar Türkçe İEİ testlerini tercih etmişlerdir. Bu sebeple sadece Türkçe testlerin sonuçları incelenmiştir.

Bu çalışmanın yöntemi bir önceki bölümde bahsedilen çalışmanın yöntemine çok benzemektedir. Orada olduğu gibi kullanıcılar ilk önce demografik sorulara cevap verecek, ardından alıştırma oturumu ve gerçek oturum ile devam edecek, son olarak anket sorularına verecekleri cevaplarla çalışmayı tamamlayacaklardır. Bir önceki çalışmada işitsel reCAPTCHA İEİ'lerin sonuçları kullanılabilirlik yönünden

Çizelge 5.3: Kullanıcı çalışmasında test edilmiş SMARTCHA sistemindeki durumlar.

	Yarı - Otomasyon Teknikleri	Soru Ön Filtreleme
Durum1	✓	✓
Durum2	✗	✓
Durum3	✓	✗
Durum4	✗	✗

oldukça kötü çıktığından bu çalışmada sadece iki sistem kullanılmıştır: metin tabanlı SMARTCHA İEİ'ler ve görsel reCAPTCHA İEİ'ler. Alıştırma oturumunda kullanıcılardan yine her sistem için birer test çözmeleri istenmiştir. Bir önceki çalışmadan farklı olarak, gerçek oturumda kullanıcıların ikişer test çözmeleri gerekmektedir. Sonuç olarak kullanıcılar toplamda altı test çözmüşlerdir.

Çalışmada kullanılan arayüz bir önceki çalışmanın arayüzüyle aynıdır. Yalnız bu çalışmada metin tabanlı İEİ'lerin içeriği textCAPTCHA testlerinden değil SMARTCHA testlerinden oluşmaktadır.

SMARTCHA sistemi için dört farklı durum senaryosu oluşturulduğundan katılımcılar kabaca dört eşit sayıda gruba bölünmüş ve sadece kendi grubunda yer alan testleri çözmüşlerdir. Diğer bir deyişle, her kullanıcı SMARTCHA'nın tek bir durumuna ait soruları çözmüştür. Diğer yandan, görsel İEİ'leri bütün kullanıcılar çözmüştür.

### 5.2.2 Hipotezler

Çalışma öncesinde özellikle Durum3 ve Durum4 senaryolarında yer alacak olan kullanıcıların SMARTCHA sistemindeki bazı sorulara ilk denemelerinde doğru cevap veremeyecekleri tahmin edilmiştir. Yine de, bütün senaryolar için ilk denemede başarılı çözüm oranları incelendiğinde SMARTCHA İEİ'lerin ilk denemede başarı oranlarının görsel İEİ'ler ile karşılaştırılabilir olacağı öngörülmüştür. Ayrıca yarı-otomasyon teknikleri uygulandığında soruyu anlamak için daha fazla süre gerekebileceği tahmin edilmiştir. Bu sebeple yarı-otomasyon teknikleri kullanmanın soru çözme süresini artıracığı düşünülmüştür. Diğer bir hipotez de soru ön filtreleme ile daha kullanışlı sorular elde edileceği ve sonuç olarak soru çözme sürelerinin azalacağı yönündedir. Son olarak oluşturulan senaryolardan Durum2'nin en kullanışlı testleri üreten senaryo olacağı tahmin edilmiştir. Bu sebeple Durum2 testlerinin ortalama çözme süresi ile kullanıcıların kolaylık ve eğlenme üzerine algısı bakımından görsel İEİ'lere göre daha iyi performans sergileyeceği düşünülmüştür.

Bu çalışma öncesinde oluşturulan sıfır hipotezleri aşağıdaki gibidir:

H5 SMARTCHA İEİ'lerinin bütün dört durum senaryosunda da kullanıcıların ilk denemelerindeki başarı oranları ile görsel İEİ'lerinin ilk denemede başarı oranları arasında fark yoktur.

H6 (Yarı-otomasyon Etkisi) Durum2 İEİ'leri çözülme süresi ile Durum1 İEİ'leri çözülme süresi arasında fark yoktur. Benzer şekilde, Durum4 İEİ'lerinin çözülme süresi ile Durum3 İEİ'lerinin çözülme süresi arasında fark yoktur.

H7 (Soru Ön Filtreleme Etkisi) Durum2 İEİ'lerinin çözülme süresi ile Durum4 İEİ'lerinin çözülme süresi arasında fark yoktur. Benzer şekilde, Durum1 İEİ'lerinin çözülme süresi ile Durum3 İEİ'lerinin çözülme süresi arasında fark yoktur.

H8 Durum2 İEİ'lerinin çözülme süresi ile görsel İEİ'lerin çözülme süresi arasında fark yoktur.

H9 Durum2 İEİ'lerini çözmek ile görsel İEİ'leri çözmek arasında kolaylık ve eğlenme yönlerinden kullanıcı algısı farkı yoktur.

### 5.2.3 Sonuçlar

Bu çalışmaya yaşları 18 ila 63 arasında değişen 372 kullanıcı katılmıştır. Bunlardan 290'ı (%78) 19 ila 24 yaş aralığındadır. Katılımcıların 160'ı (%43) kadındır. Görme seviyesi sorusuna 204 katılımcı (%55) "Normal" şeklinde cevap vermişken, 163 katılımcı (%44) "Düzeltilmiş (Gözlük veya lensler ile)" cevabını vermiştir. Kalan beş katılımcıdan üçü "Kısmi görme engelli" şikkını ve ikisi "Görme engelli" şikkını seçmişlerdir. Katılımcılar arasında sadece dokuz kişi (%2) Türkçe'yi yabancı dil olarak konuşmaktadır, geriye kalan 363 kişinin (%98) ana dili Türkçe'dir.

SMARTCHA'nın bütün senaryolarında elde ettiği sonuçlar aşağıda toplu halde verilmiştir. Tanımlanmış olan dört senaryonun her biri için elde edilen sonuçların dağılımı Çizelge 5.4'te ayrıca incelenebilir.

Katılımcıların çözmeleri gereken 744 SMARTCHA testinin 737 tanesi başarılı şekilde çözülmüş, sadece yedi test çözülmeden geçilmiştir. Katılımcılar SMARTCHA İEİ'ler için en az 2 ve en çok 307 saniye olmak üzere ortalama 19,27 saniye harcamışlardır. Kullanıcılar toplam 64 kere (kullanıcı başına 0,09 kere) İEİ testi değiştirmiş ve toplam 208 kere (kullanıcı başına 0,28 kere) İEİ testlerine yanlış cevap vermişlerdir.

Görsel İEİ'lerin 740 tanesi katılımcılar tarafından başarıyla çözülmüş, sadece dört tanesi çözülmeden geçilmiştir. Katılımcılar görsel İEİ'ler için en az 4 ve en çok 186 saniye olmak üzere ortalama 16,09 saniye harcamışlardır. Kullanıcılar toplam

65 kere (kullanıcı başına 0,09 kere) görsel İEİ testi değiştirmiş ve toplam 51 kere (kullanıcı başına 0,07 kere) bu testlere yanlış cevap vermişlerdir.

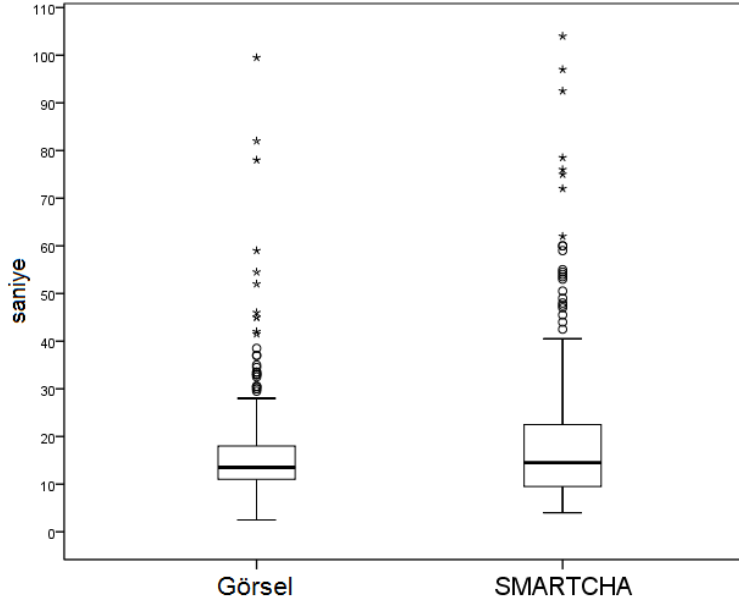
Katılımcıların anket sorularına verdikleri cevaplar da toplu halde şu şekilde özetlenebilir: 5'li Likert ölçeğinde İEİ'ler kolaylık yönünden oylandığında SMARTCHA ortalama 4,11 ve görsel İEİ'ler ortalama 3,94 almıştır. Eğlence yönünden yapılan oylama sonuçlarında ise SMARTCHA ortalama 3,29 alırken görsel İEİ'ler ortalama 2,75'te kalmıştır. Katılımcıların sıralama sorusuna verdikleri cevaplara göre SMARTCHA 234 katılımcı (%63) tarafından "Birinci" seçilmişken, kalan 138 katılımcı (%37) görsel İEİ'leri "Birinci" ve SMARTCHA'yı "İkinci" olarak işaretlemişlerdir.

Bu sonuçlar Şekil 5.4, 5.5, 5.6 ve 5.7'de de görülebilir. Şekil 5.4'te sistemlerin çözülme süreleri kutu grafiği ile verilmiştir. Şekil 5.5 iki sistemi anketlere verilen cevaplara göre karşılaştırmaktadır. Şekil 5.6'da katılımcıların ankette yer alan sıralama sorusuna verdikleri cevaplar oransal olarak gösterilmektedir. Şekil 5.7 ise kullanıcıların her bir sistemdeki başarı oranını vermektedir.

Çalışmada elde edilen sonuçların analizini yapabilmek için ilk önce iki farklı metrik tanımlanmıştır: "ilk denemede başarılı çözüm" ve "başarılı çözüm". Bir İEİ testi ile karşılaştığında o testi değiştirmeden ya da yanlış cevap vermeden soruya

Çizelge 5.4: İkinci kullanışlılık çalışmasında toplanan verilerin ortalama ve standart sapma değerleri (parantez içinde).

	(97 kullanıcı)		(96 kullanıcı)		(76 kullanıcı)		(103 kullanıcı)	
	Durum1	Görsel	Durum2	Görsel	Durum3	Görsel	Durum4	Görsel
Sistemi geçmek için harcanan süre (saniye)	19,48 (14,28)	14,21 (5,09)	12,58 (7,74)	17,29 (13,36)	24,38 (15,97)	15,74 (7,94)	21,57 (24,21)	17,08 (11,31)
Kullanıcı başına soru değişim oranı	0,09 (0,41)	0,14 (0,56)	0,03 (0,17)	0,20 (0,76)	0,30 (0,83)	0,22 (0,64)	0,28 (0,77)	0,15 (0,51)
Kullanıcı başına hata oranı	0,63 (1,88)	0,14 (0,38)	0,25 (0,54)	0,08 (0,28)	0,45 (0,68)	0,18 (0,45)	0,86 (2,33)	0,15 (0,41)
İlk denemede başarılı çözüm oranı	0,83 (0,31)	0,89 (0,22)	0,88 (0,23)	0,88 (0,27)	0,75 (0,30)	0,86 (0,23)	0,69 (0,34)	0,87 (0,25)
Başarılı çözüm oranı	0,99 (0,07)	0,99 (0,05)	0,99 (0,05)	0,99 (0,10)	0,99 (0,06)	1,00 (0,00)	0,98 (0,10)	1,00 (0,05)
Çözmenin ne kadar kolay olduğu	4,07 (1,10)	3,87 (1,11)	4,48 (0,78)	3,97 (1,11)	3,79 (1,26)	3,87 (1,17)	4,03 (1,11)	4,04 (1,08)
Çözmenin ne kadar eğlenceli olduğu	3,23 (1,33)	2,82 (1,34)	3,59 (1,41)	2,93 (1,41)	3,20 (1,32)	2,62 (1,36)	3,15 (1,39)	2,60 (1,26)
Birinci seçilme oranı	0,62	0,38	0,74	0,26	0,53	0,47	0,61	0,39



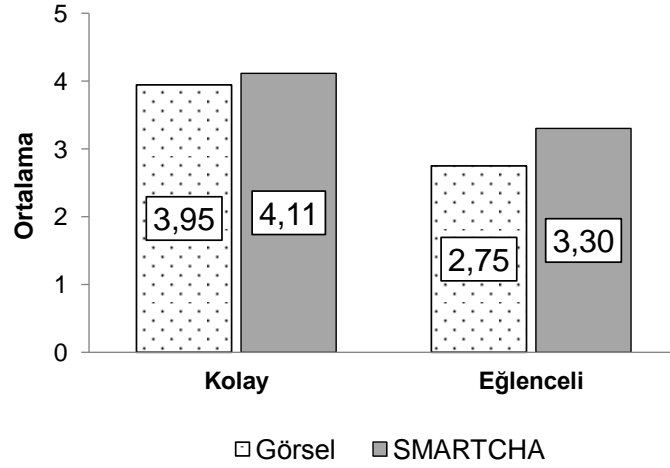
Şekil 5.4: SMARTCHA ve görsel İEİ'lerin çözülme süreleri.

doğru cevap verebilen katılımcılar “ilk denemede başarılı çözüm” metriğine göre başarılı sayılmışlardır. Diğer taraftan, hata ya da değişiklik yapsa da yapmasa da, eğer bir katılımcı bir testi atlamayıp çözdüyse o kullanıcı “başarılı çözüm” metriğine göre başarılıdır.

Kullanıcı çalışmasında katılımcılardan her bir sistemden iki test çözmeleri istenmiştir. Aşağıdaki veriler ortalamaların ortalaması hesaplanarak paylaşılmıştır. Örneğin, katılımcıların ilk test için ortalama başarı oranı 0,75 ve ikinci test için ortalama başarı oranı 0,8 ise başarı ortalaması 0,78 olarak paylaşılmıştır.

Bu çalışmada elde edilen veriler Kolmogorov-Smirnov testine göre normal dağılıma uymamaktadır. Bu sebeple analizlerde parametrik olmayan testler kullanılmıştır. Grup içi zaman ve başarı oranı karşılaştırmalarında (H5 ve H8) Wilcoxon işaretli sıra testi, Grup içi Likert ölçeği içeren anket verileri karşılaştırmalarında (H9) Friedman testi uygulanmıştır. Friedman testine göre anlamlı bir fark bulunmuşsa bu testi post hoc Wilcoxon işaretli sıra testi takip etmiştir. Gruplar arası zaman karşılaştırmalarında ise Mann-Whitney U testi kullanılmıştır. 0,05 değerinden küçük bir p-değeri sıfır hipotezini reddetmek için yeterli görülmüştür. Yapılan analizlerde alıştırma oturumu verileri kullanılmamıştır. Gerçek oturum verileri ise verinin bağımsızlığı ilkesini koruyabilmek adına birleştirilmiştir. Birleştirme işlemi, kullanıcıların başarı oranlarını toplayarak, testleri çözme sürelerinin ise ortalaması alınarak yapılmıştır.



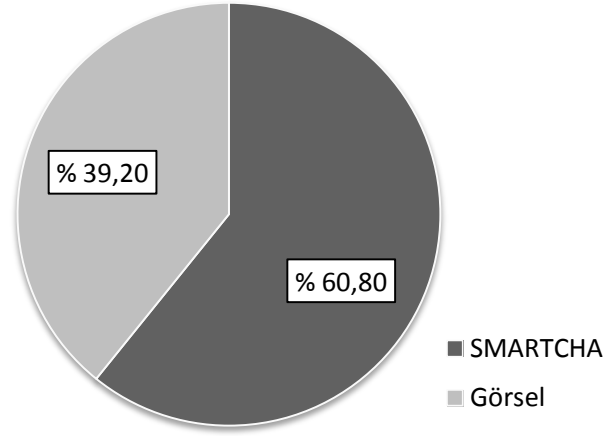


Şekil 5.5: SMARTCHA ve görsel İEİ'lerin anket sonuçları (Ankette verilen '5' cevabı tamamen katılıyorum, '1' cevabı ise tamamen katılmıyorum anlamına gelmektedir).

Wilcoxon işaretli sıra testine göre Durum1 ve Durum2'nin ilk denemede başarılı çözüm oranları görsel İEİ'lerin ilk denemede başarılı çözüm oranları ile karşılaştırıldığında arada anlamlı bir fark bulunmamaktadır (Durum1 ve Durum2 için sırasıyla  $W = 275, p = 0,166$  ve  $W = 361,5, p = 0,88$ ). Öte yandan Durum3 ve Durum4 İEİ'lerinin ilk denemede başarılı çözüm oranları görsel İEİ'ler ile karşılaştırıldığında arada anlamlı bir fark bulunmuştur (Durum3 ve Durum4 için sırasıyla  $W = 240,5, p = 0,037$  ve  $W = 331, p < 0,001$ ). Ancak Durum3 ve Durum4 İEİ'lerinin başarılı çözüm oranları görsel İEİ'lerin başarılı çözüm oranıyla karşılaştırıldığında arada anlamlı bir fark bulunmamaktadır (Durum3 ve Durum4 için sırasıyla  $W = 0, p = 0,317$  ve  $W = 0, p = 0,107$ ). Bu sonuçların ışığında H5 hipotezini reddetmek için yeterli kanıtın olmadığı söylenebilir.

Durum2 (12,58 saniye) ile Durum1'in (19,48 saniye) ortalama çözme süreleri karşılaştırıldığında aradaki fark ileri derecede anlamlıdır ( $U = 2.774,5, p < 0,001$ ). Benzer şekilde, Durum4 (21,58 saniye) ile Durum3'ün (24,38 saniye) ortalama çözme süreleri arasındaki fark da ileri derecede anlamlıdır ( $U = 2.817, p = 0,001$ ). Her iki sonuca göre H6 hipotezini reddetmek için yeterli kanıt vardır, yani soru sayısını artırmak için otomatik yöntemlerin kullanılması soruların çözülebilme süresini artırmaktadır.

Durum2 (12,58 saniye) ile Durum4'ün (21,58 saniye) ortalama çözme süreleri karşılaştırıldığında aradaki fark ileri derecede anlamlıdır ( $U = 3.412, p < 0,001$ ). Benzer şekilde, Durum1 (19,48 saniye) ile Durum3'ün (24,38 saniye) ortalama çözme süreleri arasında fark da ileri derecede anlamlıdır ( $U = 2.682,5, p = 0,002$ ).



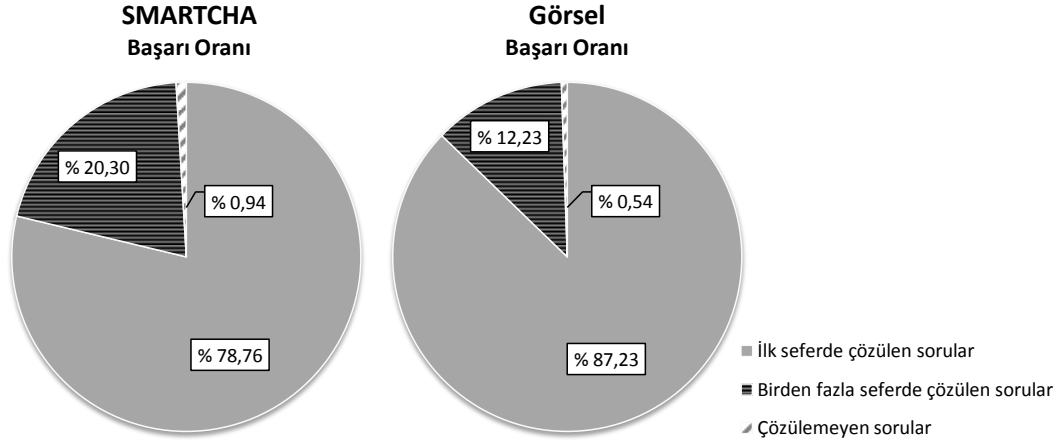
Şekil 5.6: Kullanıcıların SMARTCHA ve görsel İEİ'leri "Birinci" olarak seçme oranları.

Her iki sonuca göre H7 hipotezini reddetmek için yeterli kanıt vardır, yani soru ön filtreleme uygulamak soruların çözülebilme sürelerini azaltmaktadır.

Wilcoxon işaretli sıra testine göre Durum2 (12,58 saniye) ile görsel İEİ'lerin (17,29 saniye) ortalama çözme süreleri arasındaki fark ileri derecede anlamlıdır ( $W = 3.204,5, p < 0,001$ ). Bu sonuca göre H8 hipotezini reddetmek için yeterli kanıt vardır, Durum2 testlerinin ortalama çözme süresi görsel İEİ'lerin ortalama çözme süresinden kısadır.

Friedman testine göre Durum2 ile görsel İEİ testlerinin çözülme kolaylığı arasında ileri derecede anlamlı bir fark bulunmaktadır ( $\chi^2 = 13,298, p < 0,001$ ). Bunun üzerine yapılan post hoc Bonferroni düzeltmeli Wilcoxon işaretli sıra testine göre kullanıcılar Durum2 testlerini çözmeyi görsel testleri çözmeye göre daha kolay bulmuşlardır ( $W = 929,5, p < 0,001$ ). Benzer şekilde, Friedman testine göre farklı sistemlerin İEİ testlerini çözenin ne kadar eğlenceli olduğu da ileri derecede anlamlı şekilde farklıdır ( $\chi^2 = 14,878, p < 0,001$ ). Bonferroni düzeltmeli Wilcoxon işaretli sıra testi, Durum2 testlerini çözenin görsel testleri çözmekten daha eğlenceli olduğunu belirtmektedir ( $W = 982,5, p < 0,001$ ). Bu iki testin sonucu olarak H9 hipotezi de reddedilmektedir, kullanıcılar Durum2 testlerini çözmeyi daha kolay ve daha eğlenceli bulmuşlardır.

Sonuç olarak H5 hipotezi hariç bütün hipotezler reddedilmiştir. H5 hipotezini reddetmek için yeterli kanıt bulunamamıştır.



Şekil 5.7: SMARTCHA ve görsel İEİ'lerin başarılı çözüm oranları.

#### 5.2.4 Kısıtlar

Kullanıcı çalışmasında yer alan katılımcıların çoğu üniversite öğrencileri veya personeli olduklarından genel internet kullanıcısı nüfusunu temsil edemeyebilirler.

İstatistiksel olarak anlamlı olmayan birkaç kullanıcı dışında bütün kullanıcılar Türkçe İEİ testlerini çözmeyi seçmişlerdir. Bu yüzden, yapılan analizlerde bahsedilen küçük kısım ile ilgili veriler çıkarılmıştır. Bu çalışmanın kısıtlarından birisi de İngilizce İEİ testlerinin kullanılabilirlik analizinin olmamasıdır.

Kullanıcıların görme düzeyi ile ilgili olan demografik soruya cevap olarak üç kullanıcı "Kısmi görme engelli" ve iki kullanıcı "Görme engelli" seçeneklerini işaretlemiştir. Fakat bu verinin gerçeği yansıtıp yansıtmadığı bilinmemekle birlikte, beş örnek istatistiksel olarak çok küçük olduğundan bu katılımcılara ait veriler ayrıca analiz edilmemiştir.

#### 5.2.5 Tartışma

Bu çalışmada dört senaryonun tamamında da SMARTCHA İEİ'lerinin kabul edilebilir bir başarılı çözme oranına sahip olduğu gözlemlenmiştir. Testi çözmeden geçen en fazla kullanıcı yalnızca %2'lik oranla Durum4 senaryosunda yer almıştır.

Oldukça ilginç bir gözlem ise, her ne kadar SMARTCHA İEİ'leri görsel İEİ'lerden daha fazla değiştirilmiş ya da yanlış çözülmüş olsalar da, ankette yer alan sıralama

sorusuna göre dört senaryonun dördünde de kullanıcıların yarısından fazlasının SMARTCHA'yı "Birinci" olarak seçmesidir. Bu sonuç, daha zor çözümler bile metin tabanlı İEİ testlerinin kullanıcılar üzerinde daha çok toleransa sahip olduğunun bir göstergesi olarak düşünülebilir.

Ek olarak, görsel İEİ'ler ile dört senaryodaki SMARTCHA İEİ'leri çözenin eğlencesi karşılaştırıldığında, Friedman testi aradaki farkın ileri derecede anlamlı olduğunu göstermektedir; sırasıyla Durum1 için ( $\chi^2 = 8,018, p = 0,005$ ), Durum2 için ( $\chi^2 = 14,878, p < 0,001$ ), Durum3 için ( $\chi^2 = 11,308, p = 0,001$ ) ve Durum4 için ( $\chi^2 = 17,308, p < 0,001$ ). Yapılan post hoc Bonferroni düzeltmeli Wilcoxon işaretli sıra testine göre bütün senaryolardaki kullanıcılar SMARTCHA İEİ'leri çözerken görsel İEİ'leri çözmeye oranla daha çok eğlenmişlerdir; sonuçlar sırasıyla Durum1 için ( $W = 1.190,5, p = 0,004$ ), Durum2 için ( $W = 982,5, p < 0,001$ ), Durum3 için ( $W = 630,5, p = 0,001$ ) ve Durum4 için ( $W = 1.080,5, p < 0,001$ ) şeklindedir. Bu sonuçlar SMARTCHA'nın daha kullanışlı bir İEİ sistemi olmaya aday olduğunu göstermektedir.

### 5.3 Görme Engelli Kullanıcılarla Yapılan Kullanıcı Çalışması

SMARTCHA'nın görme engelli kullanıcılar tarafından kullanılabilirliğini test edebilmek için bir kullanıcı çalışması daha yapılmıştır. Yapılan bu kullanıcı çalışmasının yöntemi, hipotezleri, sonuçları, analizleri, kısıtlamaları ve tartışmaları takip eden alt bölümlerde verilmiştir.

#### 5.3.1 Yöntem

Bölüm 5.1'de anlatılan ilk kullanıcı çalışmasında Amazon Mechanical Turk [40] servisini kullanan görme engelli kullanıcı sayısının yetersizliğinden dolayı text-CAPTCHA servisinin erişilebilirliği test edilememiştir. Bu sebeple bu çalışmada Mechanical Turk servisi tercih edilmemiştir. Bunun yerine bir adet çalışmaya davet mesajı hazırlanmış ve bu mesaj Yahoo! Groups'da [44] başlığında veya konusunda "blindness" (körlük) ve "visual impairment" (görme engellilik) olan gruplara gönderilmiştir. Davet mesajında çalışmanın amacından bahsedilmiş ve çalışmaya katılmak isteyen kullanıcılar için çalışmanın bulunduğu internet adresi paylaşılmıştır. [www.smartcha.net](http://www.smartcha.net) adresinde yer alan çalışmanın erişilebilirliği JAWS ekran okuyucu [45] ve WebAim WAVE erişilebilirlik değerlendirme aracı [46] ile çalışma öncesinde test edilmiştir.

Diğer kullanıcı çalışmalarında olduğu gibi bu çalışmada da kullanıcılar ilk önce çalışmanın yöntemini anlatan bir karşılama mesajı ile karşılanmış ve sonrasında yaş, cinsiyet, İngilizce yeterlilik bilgileri ve görme durumlarını sorgulayan bir kullanıcı bilgisi anketine katılmışlardır. Bu çalışmada da kullanıcılardan altı adet İEİ testi çözmeleri istenmiştir. Görme engelli kullanıcılar görsel testleri çözemeyeceklerinden bu çalışmada görsel reCAPTCHA İEİ'leri kullanılmamıştır. Bunun yerine kullanıcılardan üç adet SMARTCHA ve üç adet işitsel reCAPTCHA testi çözmeleri istenmiştir. Burada vurgulanması gereken önemli bir bilgi, bu çalışmada kullanılan işitsel reCAPTCHA testlerinin birinci çalışmada kullanılan işitsel reCAPTCHA testlerinden farklı olmasıdır. Bu çalışmada, Bölüm 2.2.2'de bahsedilen yeni sürüm işitsel reCAPTCHA testleri kullanılmıştır. Önceki çalışmalarda olduğu gibi kullanıcılar bu çalışmada da her bir sistem ile ilk karşılaşmalarının alıştırmaya oturumu olduğu konusunda bilgilendirilmiştir. Dolayısıyla bu veriler analizlerde kullanılmamıştır. Son olarak kullanıcılardan, önceki çalışmalardakine uygun şekilde tasarlanmış bir ankete cevap vermeleri istenmiştir. Bu ankette kullanıcılardan her iki sistemin İEİ testlerini çözmenin 5'li Likert ölçeğine göre kolaylık ve eğlenme yönlerinden değerlendirmeleri ile bu sistemleri "Birinci" ve "İkinci" şeklinde sıralamaları istenmiştir. Önceki çalışmalardan farklı olarak, anket sonrasında isteyen kullanıcılara çalışma hakkındaki yorumlarını yazabilecekleri bir alan verilmiştir.

### 5.3.2 Hipotezler

Yeni sürüm işitsel reCAPTCHA testleri çalışma öncesinde test edildiğinde çözümenin çok kolay olduğu görülmüştür. Dolayısıyla bu testlerin çözme süresinin çok kısa olacağı tahmin edilmektedir. Buna rağmen kullanıcıların otomatik bir ses yerine düz metin ile çalışırken daha rahat edecekleri ve bu sebeple SMARTCHA'nın çözme süresinin daha kısa olacağı öngörülmektedir<sup>6</sup>. Benzer şekilde, gerçekten anlamlı bir soruya cevap vermenin vereceği hazzın, otomatik bir şekilde arka arkaya sıralanmış numaraları bir kutuya yazmanın vereceği hazdan daha fazla olacağı düşünülmektedir. Dolayısıyla kullanıcıların SMARTCHA testlerini çözmeyi daha eğlenceli bulacağı tahmin edilmektedir. Ancak her iki sistemin sorularını çözmek de çok kolay olduğundan çözme kolaylığı yönünden arada anlamlı bir fark beklenmemektedir. Aynı sebepten ötürü kullanıcıların soruları başarılı şekilde çözebilme oranları arasındaki farkın da anlamlı olmayacağı öngörülmektedir.

<sup>6</sup>Görme engelli kullanıcılar internet sitelerine erişebilmek için ekran okuyucu araçlar kullanırlar. Ekran okuyucular internet sitesinin ilgilenilen kısmının HTML kodunu ayrıştırır ve içeriği kullanıcıya sesli biçimde aktarır.

Bu durumda, çalışma öncesinde sıfır hipotezleri aşağıdaki şekilde belirlenmiştir:

H10 SMARTCHA İEİ'lerinin çözme süresi ile işitsel reCAPTCHA İEİ'lerinin çözme süresi arasındaki fark anlamlı değildir.

H11 Hangi İEİ sisteminin testlerinin çözüldüğünün kullanıcıların soruları çözme kolaylığı algısına etkisi yoktur.

H12 Hangi İEİ sisteminin testlerinin çözüldüğünün kullanıcıların soruları çözmenin verdiği eğlence algısına etkisi yoktur.

H13 Hangi İEİ sisteminin testlerinin çözüldüğünün soruların başarılı çözümüne etkisi yoktur.

### 5.3.3 Sonuçlar

Bu kullanıcı çalışmasına toplamda 31 kullanıcı katılmıştır. Kullanıcıların yedisi (%23) kadındır. Katılımcıların yaşı 16 ila 67 arasında değişmektedir ve yaş ortalaması 40,7'dir. Hiç bir katılımcı normal ya da düzeltilmiş görüye sahip değildir; yalnızca üç kullanıcı kısmi görüye sahip olup, geri kalan 28 katılımcı görme engellidir. Yirmi üç katılımcı (%74) İngilizce'yi ana dili olarak konuşmaktadır, kalan sekizi ise İngilizce'yi sonradan öğrenmişlerdir.

Yöntem bölümünde anlatıldığı üzere, kullanıcılardan her bir İEİ sistemi için üçer test çözmeleri istenmiştir. Bunlardan ilki alıştırmaya oturumu, kalan ikisi ise gerçek oturum olarak kaydedilmiştir. Yapılan analizlerde alıştırmaya oturumu verileri kullanılmamıştır. Gerçek oturum verileri ise verinin bağımsızlığı ilkesini koruyabilmek adına birleştirilmiştir. Birleştirme işlemi kullanıcıların başarı oranlarını toplayarak ve çözme sürelerinin ortalamasını alarak yapılmıştır.

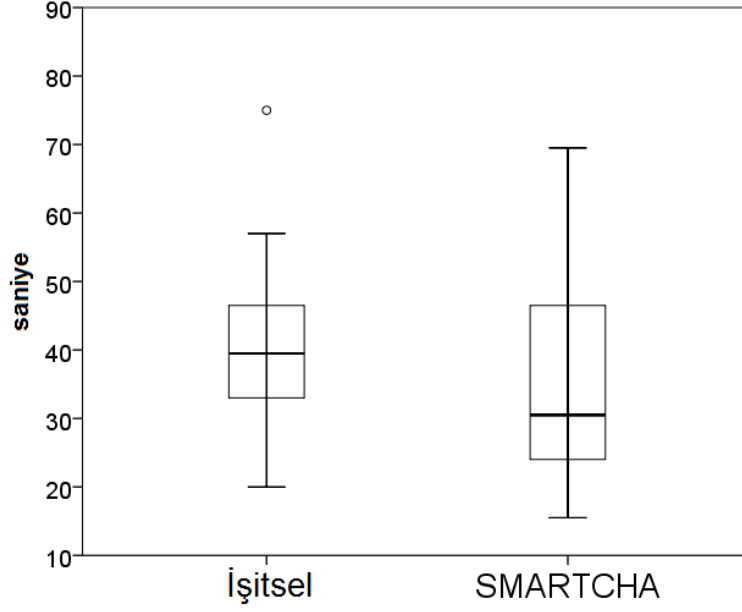
Bu çalışmada toplanan veriler Kolmogorov-Smirnov testine göre normal dağılıma uymamaktadır. Bu sebeple analizlerde parametrik olmayan testler kullanılmıştır. Yapılan bütün karşılaştırmalar grup içi karşılaştırmalar olup, zaman ve başarı oranı karşılaştırmalarında (H10 ve H13) Wilcoxon işaretli sıra testi, Likert ölçeği içeren anket verileri karşılaştırmalarında (H11 ve H12) ise Friedman testi uygulanmıştır. Friedman testine göre anlamlı bir fark bulunuyorsa, bu testi post hoc Wilcoxon işaretli sıra testi takip etmiştir. 0,05 değerinden küçük bir p-değeri sıfır hipotezini reddetmek için yeterli görülmüştür.

Şekil 5.8’de SMARTCHA ve işitsel reCAPTCHA sistemlerinin çözülebilmek süreleri kutu grafiği ile gösterilmiştir. SMARTCHA için ortalama çözme süresi 39,18 saniye iken bu süre işitsel reCAPTCHA için 43,24 saniyedir. Wilcoxon işaretli sıra testine göre arada anlamlı bir fark gözlemlenmemektedir ( $W = 194, p > 0,2$ ), dolayısıyla H10 hipotezini reddedebilmek için yeterli kanıt yoktur. Ancak veriler dikkatle incelendiğinde ilginç bir durum göze çarpmaktadır: kullanıcılar SMARTCHA testlerini çözmeye alıştıkça daha hızlı şekilde çözmeye başlamalarına rağmen işitsel reCAPTCHA testlerini çözme hızları sabit kalmıştır (Çizelge 5.5). Şekil 5.9’da kullanıcıların her bir sistemle üçüncü karşılaşmalarındaki soru çözme süreleri kutu grafiği ile verilmiştir. Kullanıcıların her bir sistemle ikinci ve üçüncü karşılaşmalarında o testi çözmek için harcadıkları süre Wilcoxon işaretli sıra testi ile karşılaştırıldığında aradaki fark SMARTCHA için anlamlı çıkmış ( $W = 349,5, p = 0,0134$ ), fakat işitsel reCAPTCHA için anlamlı bir fark çıkmamıştır ( $W = 259,5, p = 0,197$ ). Kendall Tau-b korelasyon testine göre kullanıcıların SMARTCHA sistemi ile ikinci ve üçüncü karşılaşmaları arasında zayıf, pozitif yönde anlamlı bir ilişki vardır ( $\tau_b = 0,226, p = 0,039$ ). Bu bilgiler ışığında kullanıcıların her bir sistemle üçüncü karşılaşmalarındaki testleri çözme süreleri karşılaştırılmıştır. Buna göre kullanıcıların SMARTCHA testlerini çözme süresi ile işitsel reCAPTCHA testlerini çözme süresi arasında ileri derecede anlamlı bir fark bulunmuştur ( $W = 107, p = 0,006$ ). Toparlamak gerekirse, kullanıcıların SMARTCHA sistemine alışma etkisi göz önünde bulundurulduğunda H10 hipotezini reddetmek için yeterli kanıt vardır. Yine de, SMARTCHA sistemine alışma etkisinin kalıcı olup olmadığını daha net tespit edebilmek için daha fazla kullanıcı ile yeni bir çalışma yapılması faydalı olabilir.

Kullanıcıların, her iki sistemin çözülme kolaylığı ve eğlenme algısı üzerine ankette yer alan sorulara verdikleri cevapların ortalamaları Şekil 5.10’da; iki sistemi sıralama sorusuna verdikleri yanıtlara göre sistemlerin birinci sıraya koyulma oranları da Şekil 5.11’de verilmiştir. Friedman testine göre sistem testlerinin çözülme kolaylığı arasında anlamlı bir fark bulunmamıştır ( $\chi^2 = 0,048, p = 0,827$ ). Bu sebeple

Çizelge 5.5: SMARTCHA ve işitsel İEİ’lerin saniye cinsinden ortalama çözme süreleri.

	SMARTCHA	İşitsel reCAPTCHA
İlk Karşılaşma (Alıştırma oturumu)	93,94	54,19
İkinci Karşılaşma	43,9	44,35
Üçüncü Karşılaşma	34,45	42,13
Ortalama	39,18	43,24



Şekil 5.8: SMARTCHA ve işitsel İEİ'lerin çözüme süreleri.

H11 hipotezini reddetmek için yeterli kanıt yoktur. Öte yandan testleri çözümlenme yönünden karşılaştırmasında Friedman testine göre arasında anlamlı bir fark bulunmaktadır ( $\chi^2 = 5, p = 0,025$ ). Post hoc Wilcoxon işaretli sıra testine göre kullanıcılar SMARTCHA testlerini çözmeyi daha eğlenceli bulmaktadır ( $W = 171, p = 0,012$ ). Bu durumda H12 hipotezini reddetmek için yeterli kanıt mevcuttur.

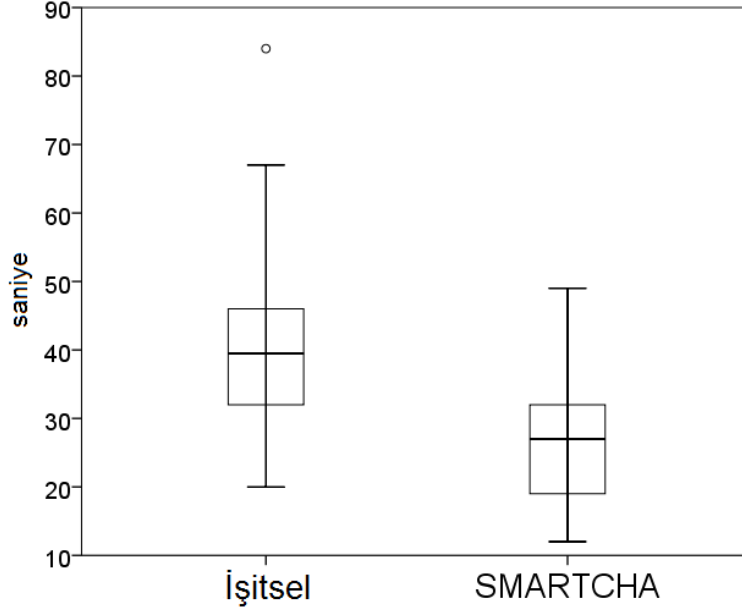
Her iki sistemin de başarılı çözüm oranları epey yüksek çıkmıştır; SMARTCHA İEİ'lerinden yalnızca bir tanesi çözülmeden geçilmesine karşın işitsel İEİ'lerin tamamı çözülmüştür. Wilcoxon işaretli sıra testine göre aradaki fark anlamlı değildir ( $W = 0, p > 0,2$ ), dolayısıyla H13 hipotezini reddetmek için yeterli kanıt yoktur.

Sonuç olarak H10 ve H12 hipotezleri reddedilmiştir. H11 ve H13 hipotezlerini reddetmek için yeterli kanıt bulunamamıştır.

### 5.3.4 Tartışma

Başarılı çözüm oranlarının her iki sistemin testleri için de oldukça yüksek çıkması sebebiyle başarılı çözüm oranı karşılaştırması tavan etkisine maruz kalmaktadır. Bu sebeple iki sistem, bir önceki çalışmada tanımlanmış olan “ilk denemede başarılı çözüm” metriği ile karşılaştırılmıştır. Bu karşılaştırmaya göre SMARTCHA ile işitsel testlerin ilk denemede başarılı çözümü arasında ileri derecede anlamlı fark bulunmaktadır ( $W = 5, p = 0,008$ ) ve kullanıcılar işitsel testleri ilk denemede



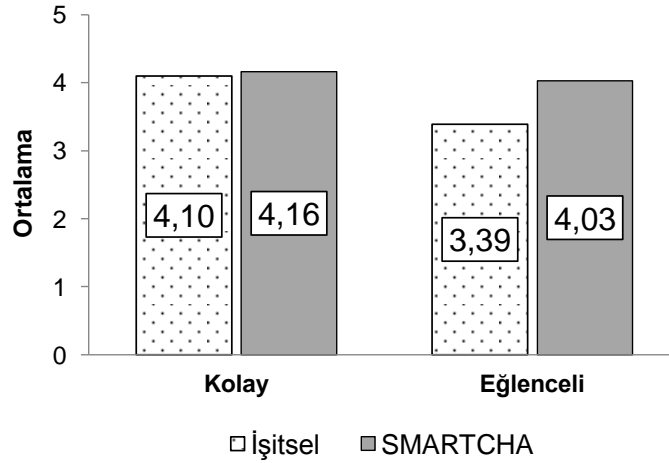


Şekil 5.9: Kullanıcıların SMARTCHA ve işitsel İEİ sistemleriyle üçüncü karşılaşmalarındaki çözme süreleri.

daha başarılı şekilde çözebilmektedirler. Buna rağmen her iki sistemin çözüm süreleri arasında anlamlı bir fark bulunmadığı, hatta alışma etkisi göz önünde bulundurulduğunda kullanıcıların SMARTCHA testlerini daha hızlı çözebildikleri unutulmamalıdır. Kullanıcıların yarısının (12’de altı) ilk denemede başarılı çözüme ulaşamamasının soru değişiminden kaynaklandığı tespit edilmiştir. Her ne kadar yanlış cevaplanan soruların nedenine dair bir veri bulunmasa da, hesaplama ya da yazım hatası gibi basit hatalar akla ilk gelen seçeneklerdir.

Yaş, cinsiyet ya da İngilizce seviyesi gibi demografik bilgilerin sistem testlerini çözmekteki süreye veya başarıya etkisini görebilmek adına gruplar arası Mann-Whitney U karşılaştırma testleri yapılmıştır. Yapılan bu karşılaştırmalara göre sadece dil yetkinliğinin işitsel testlerin çözüm süresine etkisinde anlamlı bir fark çıkmıştır. Buna göre işitsel testleri, ana dili İngilizce olan kullanıcıların (14,04 saniye) İngilizce’yi yabancı dil olarak bilen kullanıcılardan (21,63 saniye) daha hızlı çözdüğü ortaya çıkmıştır ( $U = 47, p = 0,042$ ). Ancak sadece ana dili İngilizce olan kullanıcılar üzerinde uygulanan grup içi Wilcoxon işaretli sıra testine göre SMARTCHA testleri (37,56 saniye) ile işitsel testlerin (41,13 saniye) çözüm süreleri arasında anlamlı bir fark bulunmamaktadır.

Çalışmanın sonunda kullanıcılara yorumlarını yazabilmeleri için bir alan verilmiştir. Bu yorumlara göre: beş kullanıcı bu alanda çalışma yapıldığı için teşekkür etmiştir. Bu kullanıcılardan bir tanesi işitsel İEİ’lerin hem işitme hem de görme engelli kullanıcıları tamamen dışarıda bıraktığını yazmıştır. Bir başka kullanıcı

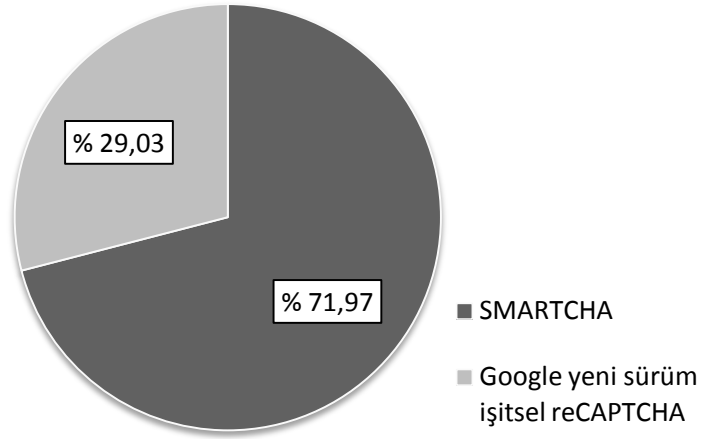


Şekil 5.10: SMARTCHA ve işitsel İEİ'lerin anket sonuçları (Ankette verilen '5' cevabı tamamen katılıyorum, '1' cevabı ise tamamen katılmıyorum anlamına gelmektedir).

genel kültür sorularının otomatik olarak çözülebileceği yönünde uyarıda bulunmuştur. Bir katılımcı SMARTCHA'yı "güzel ve çözmesi ilginç" olarak tanımlamıştır. Benzer şekilde bir başka kullanıcı arka planda gürültü olmadığı için işitsel testleri sevdiğini, fakat soruyu cevaplamak ve cevabı yazmak daha kolay olduğu için metin tabanlı İEİ'leri kullanmayı daha çok tercih edeceğini belirtmiştir. İki kullanıcı SMARTCHA testlerinin belli kültürlere özel sorular içerdiği konusunda uyarıda bulunmuş ve bu kullanıcılardan bir tanesi örnek olarak her kültürde yılbaşı gecesini hindi eti yenmediğini belirtmiştir.

Bölüm 5.2'de bahsedilen kullanıcı çalışmasında katılımcıların çoğunlukla Türkçe testleri tercih etmesi, İngilizce SMARTCHA testlerinin kullanılabilirliği hakkındaki soruları yanıtsız bırakmaktaydı. Her ne kadar bu çalışmada İngilizce testler kullanılmış olsa da, bu etkinin kalıcılığının araştırılması gerekmektedir. Bu sebeple, Bölüm 5.2'de bahsedilen çalışmayı birinci çalışma ve bu çalışmayı ikinci çalışma olarak adlandırmak üzere, bir karışık ölçümler için iki faktörlü ANOVA testi uygulanmıştır. Bu testte kullanıcıların hangi çalışmada yer aldığı (birinci çalışma ya da ikinci çalışma) gruplar arası değişken ve kullanıcıların her bir İEİ sisteminin (SMARTCHA ya da reCAPTCHA) sorusunu ne kadar sürede çözdüğü bilgisi ise grup içi değişken olarak belirlenmiştir.

Karışık ölçümler için iki faktörlü ANOVA testine göre kullanıcılar SMARTCHA İEİ'lerini reCAPTCHA İEİ'lerinden daha hızlı çözebilmektedir ve aradaki fark anlamlıdır ( $F(1, 125) = 12,391, p = 0,001$ , kısmi eta-kare = 0,09). Bunun yanı sıra, birinci çalışmadaki kullanıcılar SMARTCHA İEİ'lerini ikinci çalışmadaki kulla-



Şekil 5.11: Kullanıcıların SMARTCHA ve işitsel İEİ'leri “Birinci” olarak seçme oranları.

nıcılardan daha hızlı çözebilmektedir ve aradaki fark ileri derecede anlamlıdır ( $F(1, 125) = 6.062, 285, p < 0, 001$ , kısmi eta-kare= 0,980). Ancak çalışma grubu ve çözülen İEİ testi arasında anlamlı bir etkileşim yoktur ( $F(1, 125) = 0, 840, p = 0, 361$ , kısmi eta-kare= 0,007). Bunun anlamı, kullanıcının hangi çalışmada yer aldığı, çözdüğü İEİ'lerin çözüme süresine etkisinin bulunmamasıdır.

Her ne kadar iki çalışma arasındaki tek farklılık kullanılan SMARTCHA İEİ'lerin dili olmasa da, sonuçların tutarlı olduğu gözükmemekte ve bu da SMARTCHA İEİ'lerinin kullanılabilirliğinin dilden bağımsız olduğu hakkında ipucu vermektedir.

Son olarak, bu çalışmanın tasarımı aşamasında karşılaşılan zorluklar aşağıda özetlenmiştir:

- Görme engellilerin rahatça kullanabilmesi için yardımcı teknolojilerle uyumlu bir web sitesi hazırlamanın kolay olmadığı görülmüştür. W3C erişilebilir bir internet sitesinin nasıl hazırlanacağını anlatan bir kılavuz sağlamaktadır. Fakat bu kılavuz çok kapsamlıdır. Bu sebeple kılavuzda yer alan maddelerin tamamını anlamak ve uygulamak oldukça zordur. Çalışmalarda kullanılan sayfanın erişilebilirliğini sağlamak adına WebAIM'in hazırladığı WCAG 2.0 listesindeki maddelere uyulmuştur. Ayrıca WebAIM'in web erişilebilirlik değerlendirme uygulaması [46] ile hazırlanan internet sitesi test edilmiştir. Buna rağmen bir kullanıcı yorum olarak sistemin VoiceOver teknolojisiyle uyumlu olmadığı yönünde bir bildirimde bulunmuştur.

- Görme engelliler için hazırlanan yardımcı teknolojilerin standart bir kullanım yönteminin olmadığı görülmüştür. İnternet sitesinin kullanılabilirliğini test edebilmek için JAWS, NVDA, Windows Narrator ve Orca gibi değişik teknolojiler kullanılmaya çalışılmış, fakat bahsedilen her bir teknolojiyi doğru kullanmayı öğrenmek önemli bir miktar zaman gerektirdiğinden çoğu teknoloji verimli şekilde kullanılamamıştır.
- Bu çalışmaya gönüllü olarak katılabilecek görme engelli katılımcı bulmak zor olmuştur. Türkiye’de İngilizce’nin yabancı dil olması sebebiyle yerel kullanıcılardan hiç bir aşamada yardım alınamamıştır. Sistemi test edecek ya da kullanacak olan görme engelli kullanıcıların bilgisayarı rahatça kullanabilmeleri kadar İngilizce dilini rahatça anlayıp gerekli işlemleri uygulayabilmeleri de gerekmektedir. Benzer çalışmanın Türkçe testler kullanılarak yapılması da düşünülmüş, fakat bazı dernekler ve görme engellilerin kullanımı için özelleşmiş bir internet kafe ile kurulan iletişimlerde katılımcı bulma çalışmalarının bir sonuca bağlanamaması sebebiyle bu çalışma iptal edilmiştir.

#### 5.4 Sonuçların Tartışması

Kullanıcı çalışmalarında yazım, imla veya dil bilgisi hatalarına karşı herhangi bir önlem alınmamasına rağmen sonuçlar oldukça ümit vericidir. Bir otomatik yazım denetleme algoritması hem operatörlerin soru üretimi, hem de kullanıcıların sorulara cevap vermeleri aşamalarında kullanılarak her iki durumda da yardımcı olunabilir. Akıllı bir telefonun harfleri yazdıkça eşleşen kelimeleri önermesi gibi çalışarak operatör veya kullanıcı yazı yazdıkça arka planda çalışarak önerilerde bulunacak bir yazım denetim sistemi kullanılabilirliği artırarak sistemi iyileştirebilir.

Bir diğer iyileştirme, imla hatalarına sunucu tarafında müsamaha göstermek olabilir. Detaylandırmak gerekirse, öncelikle İEİ servislerinin testlerini hizmete sunmakta kullandıkları iki yaklaşımdan bahsetmek faydalı olacaktır: 1) İEİ sağlayıcısı, istekte bulunan web sitesi ile İEİ testini ve cevabın özetini paylaşır, başka haberleşme yapılmaz. Bu yaklaşım textCAPTCHA servisinde uygulanmaktadır. Bu yaklaşımda istemci web sitesiyle sadece cevabın özet bilgisi paylaşıldığından kullanıcının vermiş olduğu cevabın özet değeri hesaplanmalı, ardından hesaplanan bu değer cevabın özet değeri ile tam olarak uyuşup uyuşmadığı kontrol edilmelidir. Gerçek cevap istemci web sitesi tarafından bilinmediğinden imla hatalarına müsamaha gösterilemez. 2) İEİ sağlayıcısı, istemci web sitesiyle İEİ testini paylaşır ancak cevaba dair herhangi bir paylaşımda bulunmaz. Kullanıcının

İEİ testine vermiş olduğu cevap istemci web sitesinden İEİ sağlayıcısına gönderilir ve sağlayıcı cevabın doğruluğunu analiz ettikten sonra istemciye evet/hayır şeklinde bildirim yapar. Bu yaklaşım reCAPTCHA servisinde uygulanmaktadır. İmla hatalarına bu durumda müsamaha gösterilebilir. Bu tez çalışmasında en kötü durum senaryolarını analiz edebilmek adına birinci tipteki yaklaşım uygulanmıştır. İkinci tipteki yaklaşım kullanışlılığı artırabileceğinden bir iyileştirme olarak SMARTCHA sistemine eklenebilir.

Bu çalışmada soru ön filtrelemenin etkilerini test edebilmek için 100 tane kullanışlı olduğu düşünülen İEİ testi insan hesaplama kullanılarak seçilmiştir. Operatörlerin yeni bir İEİ testi hazırlayabilmek için önceden hazırlanmış testlerden çözmeleri sağlanarak bu işlem otomatikleştirilebilir. Bu yöntemde operatörlerin verdikleri cevaplar ile her sorunun başarı oranları analiz edilerek kullanışsız soru-cevap ikilileri daha genel kullanıma açılmadan tespit edilebilir ve sistemden ayıklanabilir.

Makine öğrenme teknikleri uygulamak da SMARTCHA'nın kullanışlılığını artırabilir. Nazir *ve diğerleri* [47] görsel İEİ'lerin özellik vektörünü çıkarmış ve kullanıcı girdileri ile birleştirerek bir sınıflandırıcıyı eğitmişlerdir. Bu sınıflandırıcı ile yeni hazırlanan İEİ testlerinin kullanışlı olup olmadığı tahmin edilebilmektedir. Benzer şekilde, SMARTCHA'nın metin tabanlı testleri için de böyle bir özellik vektörünün çıkarılması faydalı olabilir. Bu özellik vektörü ile eğitilmiş bir sınıflandırıcı kullanılarak operatörlerin hazırladıkları yeni testlerin kullanışlılığı proaktif şekilde tahmin edilebilir. Böyle bir özellik vektörü muhtemelen şu bilgileri içermelidir: sorunun uzunluğu, sorunun aritmetik işlem içerip içermediği bilgisi, soruda yazım hataları olup olmadığı bilgisi vb.

Makine öğrenme teknikleri SMARTCHA'nın ölçeklenebilirlik sorununu çözmeye de yardımcı olabilir. Google'ın reCAPTCHA servisi, kullanıcının davranışlarını inceleyen ve yapay zeka teknikleriyle kullanarak bir kullanıcının meşru olup olmadığını tahmin eden No CAPTCHA reCAPTCHA [48] servisine dönüşmüştür. Bu serviste İEİ testleri ancak yapay zeka tekniklerinin karar veremediği ve ikincil doğrulama gerektiren işlemlerde kullanılmaktadır. Google yapay zeka tekniklerini nasıl kullandığı konusunda herhangi bir açıklama yapmış olmasa da, SMARTCHA bu tarz yapay zeka tekniklerinden yararlanabilir. Bu amaçla, meşru ve meşru olmayan kullanıcıların özellikleri ile bir sinir ağı eğitilebilir. Eğitimde kullanılacak olan özelliklerde örneğin şu bilgiler kullanılabilir: kullanıcının IP adresi, istemci web sitesinin ismi, kullanıcının istemci web sitesindeki kullanıcı adı, kullanıcının klavyedeki tuşları kullanma verileri vb. Klavyedeki tuşları kullanma verileri olarak şu bilgiler örnek verilebilir: kullanıcının yazma hızı, yazı yazarken hata yapma

oranı, büyük harf yazmak için Caps Lock tuşunu mu yoksa Shift tuşunu mu kullandığı, hatta Shift tuşunu kullanırken sağ taraftaki tuşu mu yoksa sol taraftaki tuşu mu seçtiği bilgisi vb. Bu tarz veriler ile bahsedilen diğer özellikler birleştirilerek bir kullanıcının meşru bir insan mı yoksa meşru olmayan bir program mı olduğunu tahmin etmek mümkündür. Ölçeklenebilirlik problemi için örnek bir çözüm şu şekilde tanımlanabilir: Kullanıcılardan önce bir textCAPTCHA testi çözmeleri istenir. Kullanıcıların textCAPTCHA testine verdikleri cevap sinir ağı modeli tarafından analiz edilir. Eğer kullanıcı İEİ testine doğru cevap verir ve aynı zamanda sinir ağı modeli tarafından meşru bir insan olarak tahmin edilirse, o kullanıcı İEİ testini başarıyla geçmiş sayılır. Eğer kullanıcı sinir ağı modeli tarafından meşru bir insan olarak tahmin edilmesine rağmen İEİ testini başarılı şekilde çözememişse bu kullanıcıya yeni bir textCAPTCHA testi sorularak çözmesi beklenebilir. Diğer taraftan, eğer sinir ağı modeli kullanıcıyı meşru olmayan bir program olarak tahmin ederse, bu durumda kullanıcının ilk textCAPTCHA testini başarılı olarak çözüp çözmemesine bakılmaksızın bir SMARTCHA testi çözmesi istenebilir. Bunun sebebi SMARTCHA testlerini otomatik olarak çözmeyenin çok daha zor olmasıdır.

SMARTCHA'nın hem avantajları hem de dezavantajları vardır. Öncelikle, metin tabanlı İEİ sistemlerinin sahip olduğu avantajların tamamı SMARTCHA için de geçerlidir. İlk olarak, yukarıda anlatılmış olan ilk kullanıcı çalışması sonucunda işitsel reCAPTCHA'nın eski sürümünün (çalışmanın yapıldığı dönemde kullanılan sürüm) kullanışsız olduğu saptanmıştır. Bu çalışmada işitsel reCAPTCHA'nın başarı oranı %15'te kalmıştır. Ancak bu kullanıcı çalışması görme engeli olmayan kullanıcılar ile yapılmıştır. Olalere *ve diğerlerinin* [49] yapmış olduğu bir çalışmada, işitsel İEİ'lerde görme engelli kullanıcıların başarı oranının görme engeli bulunmayan kullanıcılara oranla yaklaşık %15 daha fazla olduğu saptanmıştır. İlk kullanıcı çalışmasında kullanılan metin tabanlı sistemin (textCAPTCHA) başarı oranı olan %99 düşünüldüğünde işitsel İEİ'lerin olası artmış başarı oranı bile hala düşük sayılabilir. İkinci olarak, metin tabanlı sistemler görsel, işitsel veya dokunsal olmak üzere üç farklı formda sunulabildiğinden diğer İEİ sistemlerine göre daha erişilebilir sistemlerdir. Örneğin, görme ve/veya işitme engeli bulunan kullanıcılar görsel ve/veya işitsel herhangi bir İEİ testini çözemezler. Fakat her iki engele de sahip olan kullanıcılar görme engelliler tarafından kullanılan Braille alfabesi yardımıyla bilgisayarla etkileşime geçebilir ve metin tabanlı bir İEİ'yi çözebilirler. Bunların dışında, metin tabanlı İEİ'lerde içerik anlayışı vardır ve bu sayede farklı tiplerde sorular kullanılabilir. Örneğin, aritmetik soruları çözemeyen diskalkuli bir kullanıcı için bu tarz bir test çok zorlayıcı olabilir. Böyle durumlarda kullanıcı kolayca yeni bir soru talep ederek genel kültür gibi tamamen farklı

bir içerikten bir test çözebilir. Oysa görme engelli bir kullanıcının görsel bir testi değiştirip çözebileceği başka bir görsel test talep etmesi gibi bir durum söz konusu değildir. Metin tabanlı İEİ'lerin bir başka avantajı da bir sistem üzerinde görsel veya işitsel bir arayüz gerektirmeden çalışabilmeleridir. SSH bağlantısı veya Lynx gibi metin tabanlı tarayıcılar bu tarz sistemlere örnek olabilir. Bir diğer avantaj ise transfer edilecek veri miktarıdır. Genellikle metin tabanlı içerikler görsel veya işitsel içeriklerden daha küçüktür.

SMARTCHA sisteminin dezavantajlarından birisi soru üretme aşamasındaki insan hesaplama gerekliliğidir. SMARTCHA sistemi tamamen otomatik yöntemlerle çalışabiliyor olsaydı, çok daha tercih edilebilecek bir sistem olabilirdi. Diğer bir önemli dezavantajı ise ortaya çıkan saldırılara karşı güvenlik motorunu güncel tutma ihtiyacı olabilir. Fakat saldırılara karşı önlem olarak yeni bir katman geliştirildiğinde bu katmanı güvenlik motoruna eklemek basit bir işlem olduğundan bu durumun üstesinden gelmek zor gözükmemektedir. Ayrıca yeni saldırıların sadece SMARTCHA için değil, bütün İEİ'ler için tehlike oluşturduğunu unutmamak gerekir. Son olarak, günümüzde hem kullanıcıların, hem de uygulama geliştiricilerinin görsel İEİ testlerini kullanmaya aşına oldukları düşünülürse SMARTCHA'nın piyasaya girebilmek için aşması gereken bir engel daha olduğu söylenebilir.

SMARTCHA'nın veritabanında şu anda 22.000 İEİ testi bulunmaktadır. Bu sayı, aktif kullanılacak bir İEİ testi için yeterli değildir. Ancak işitsel reCAPTCHA ve textCAPTCHA gibi güvenlik problemleri olan İEİ'ler dışındaki çoğu erişilebilir İEİ sistemleri için ölçeklenebilirlik temel bir kısıtlamadır. Birkaç erişilebilir İEİ sisteminin veritabanında yer alan soru sayısı Çizelge 5.6'da görülebilir. Çizelgede yer alan sistemlerden birçoğu Bölüm 2.1'de detaylı olarak anlatılmışlardır. Bahsedilmemiş sistemlerden HIPUU'da [50], kullanıcılara aynı nesnenin (örneğin bir piyano) görseli ve sesi aynı anda sunulmakta ve kullanıcılardan bu nesnenin ne olduğunu bilmeleri istenmektedir. SoundsRight [51] ise tamamen işitsel bir İEİ sistemidir. Bu sistemde kullanıcılara arka arkaya farklı nesnelere ait sesler dinletilmekte ve özel bir nesneye ait bir ses her çıktığında kullanıcılardan klavyedeki boşluk tuşuna basmaları istenmektedir.

Büyük çaplı saldırılara karşı İEİ testleri yerine güvenlik duvarı kullanılmalıdır. Örneğin bir sızma engelleme sistemi olan Fail2Ban [52], kaba kuvvet ya da sözlük saldırılarına karşı etkilidir. Fail2Ban sistemi kayıt dosyalarını inceleyerek gerekli durumlarda güvenlik duvarını güncellemekte ve arka arkaya belli bir eşik değerinden fazla kez başarısız işlem yapan IP adreslerinin isteklerini geçici olarak engellemektedir. Kötü niyetli eylemlerin kolayca tespit edilemediği durumlarda istatis-

Çizelge 5.6: Erişilebilir İEİ sistemlerindeki soru sayıları hakkında bilgi.

Sistem	Soru sayısı hakkında bilgi
HIPUU [50]	Veritabanında 35 adet işitsel/görsel ikili vardır. Yazarlar ikili sayısının toplamda 100'e kadar çıkartılabileceğini belirtmişlerdir.
SoundsRight [51]	Yazarlar veritabanının boyutu hakkında kesin bir bilgi vermemişlerdir. Ancak, işitsel kliplerin sisteme bir yönetici tarafından elle eklendiği belirtilmektedir.
Strangeness in Machine Translation [18]	200 adet geçerli cümle kullanılmıştır. Yazarlar güvenlik sorunlarına sebep olabileceğinden geçerli cümlelerin internet ortamından alınmamasını önermişlerdir. Ancak veritabanına güvenli şekilde geçerli cümle eklemek için herhangi bir yöntem önermemişlerdir.
Collaborative Filter [19]	Veritabanında 24.953 adet şaka bulunmaktadır. Çalışmalarda kullanılmak üzere 100 şaka yazarlar tarafından seçilmiştir.
SMARTCHA	Veritabanında insan hesaplama kullanılarak üretilmiş yaklaşık 22.000 soru bulunmaktadır. Soru ön filtrelemeyi test edebilmek için bu soruların içinden 100 tanesi insan hesaplama kullanılarak seçilmiştir.

tiksel analizler kullanışlı olabilir [53]. Meşru yollarla e-posta oluşturup göndermek, çevrimiçi anketlerde oy kullanmak veya bir blogda yorum yapmak gibi işlemler belli bir miktar zaman alan işlemlerdir. Stockade [53] istenmeyen e-postaları azaltmak için kullanılan bir ağ katmanı aracıdır. Bir bağlantı isteği geldiğinde istekte bulunan kişinin daha önceki bağlanma isteklerini kontrol ederek bu istemcinin istenmeyen e-posta gönderiyor olma ihtimaline karar verir. Bu kararı verdikten sonra otomatik olarak bağlantı isteğini kabul veya ret ettikten sonra kendi veritabanında bu istemciye ait olasılık değerini günceller. Bir istemcinin istenmeyen e-posta gönderiyor olma olasılığı, o istemcinin bağlantı istekleri arasında geçen süre arttıkça düşer. Tabii bu durumun tam tersi de geçerlidir.

Bu çalışmada kullanılan İEİ testlerinin tür, erişilebilirlik, güvenlik, kullanışlılık ve ölçeklenebilirlik yönlerinden karşılaştırması Çizelge 5.7'de verilmiştir.



Çizelge 5.7: Kullanıcı çalışmalarında kullanılan İEİ servislerinin karşılaştırması.

İEİ servisi	Tip	Erişilebilirlik	Güvenlik	Kullanışlılık	Ölçeklenebilirlik
SMARTCHA	Metin tabanlı	Evet	<ul style="list-style-type: none"> <li>•Güvenli.</li> <li>•Otomatik çözülebilirliğe karşı proaktif güvenlik motoru kontrolü vardır.</li> </ul>	<ul style="list-style-type: none"> <li>•Düşük çözmeye süresi.</li> <li>•Yüksek başarılı çözüm oranı.</li> <li>•Tercih edilebilir.</li> </ul>	Veritabanında yaklaşık 22.000 test bulunmaktadır.
textCAPTCHA	Metin tabanlı	Evet	<ul style="list-style-type: none"> <li>•Güvenli değil.</li> <li>•Testlerin %99,5'i otomatik olarak çözülebilmektedir.</li> </ul>	<ul style="list-style-type: none"> <li>•Düşük çözmeye süresi.</li> <li>•Yüksek başarılı çözüm oranı.</li> <li>•Tercih edilebilir.</li> </ul>	Veritabanında 180 milyon test bulunmaktadır.
İşitsel reCAPTCHA (eski sürüm)	İşitsel	Evet	<ul style="list-style-type: none"> <li>•Güvenli.</li> <li>•Rapor edilmiş başarılı bir saldırı yoktur.</li> </ul>	<ul style="list-style-type: none"> <li>•Yüksek çözmeye süresi.</li> <li>•Düşük başarılı çözüm oranı.</li> <li>•Daha az tercih edilebilir.</li> </ul>	Ölçeklenebilir.
İşitsel reCAPTCHA (yeni sürüm)	İşitsel	Evet	<ul style="list-style-type: none"> <li>•Güvenli değil.</li> <li>•Testlerin en az %62,5'i otomatik olarak çözülebilmektedir.</li> </ul>	<ul style="list-style-type: none"> <li>•Düşük çözmeye süresi.</li> <li>•Yüksek başarılı çözüm oranı.</li> <li>•SMARTCHA'ya karşı daha az tercih edilebilir.</li> </ul>	Ölçeklenebilir.
Görsel reCAPTCHA	Görsel	Hayır	<ul style="list-style-type: none"> <li>•Güvenli değil.</li> <li>•Testlerin %99,8'i otomatik olarak çözülebilmektedir.</li> </ul>	<ul style="list-style-type: none"> <li>•Düşük çözmeye süresi.</li> <li>•Yüksek başarılı çözüm oranı.</li> <li>•textCAPTCHA'ya karşı daha az tercih edilebilir.</li> </ul>	Ölçeklenebilir.



## 6. SONUÇ VE GELECEKTEKİ ÇALIŞMALAR

İnsanları bilgisayarlardan ayırt edebilen İnsan Etkileşim İspatı (İEİ) sistemleri günümüzde otomatik saldırılara karşı temel güvenlik bileşeni haline gelmişlerdir. Bilgisayarlar gün geçtikçe daha akıllı hale gelmekte ve eskiden sadece insanların yapabileceği düşünülen şeyleri yapabilmektedirler [3]. Bu sebeple İEİ'lerin güvenlik ve kullanılabilirliği arasındaki dengenin kurulabilmesi daha önemli hale gelmiştir.

İEİ'lerin tamamen metin tabanlı olması, diğer birçok iyi özelliğinin yanında, görme engelli kullanıcıların erişilebilirlik problemlerine de çözüm olma potansiyeline sahiptir. Fakat tamamen otomatik yöntemler kullanılarak güvenli bir metin tabanlı İEİ sistemi yaratma problemi henüz çözülememiştir. George Pólya'nın ünlü bir söylemi vardır: "Bir problemi çözemiyorsanız, çözebileceğiniz daha kolay bir problem vardır: onu bulun."<sup>7</sup> Bu çalışmada metin tabanlı güvenli İEİ'ler oluşturmak ile ilgili bulunan daha kolay problem, sistemin tamamen otomatik olma zorunluluğunu ortadan kaldırmaktır. Bu sebeple, insan hesaplama ve otomasyon tekniklerinin birleştirildiği bir süreç olan "yarı-otomasyon" kavramı tanıtılmıştır. SMARTCHA ismiyle önerilen sistemde, insan işçilerin oluşturduğu metin tabanlı İEİ sorularının otomatik olarak çözülüp çözülmediği güvenlik motoru tarafından kontrol edilmektedir. İnsan operatörler tarafından üretilen ve güvenlik motoru tarafından güvenli olarak işaretlenen sorular, otomatik yöntemler kullanılarak yeni sorular üretilebilmesi için bir veritabanında saklanmıştır.

Bu çalışmada öncelikle kullanılmakta olan textCAPTCHA ve reCAPTCHA İEİ sistemleri analiz edilmiştir. Bu sistemlerin kullanılabilirliklerini test edebilmek için bir kullanıcı çalışması yapılmıştır. Çalışma sonuçlarına göre metin tabanlı İEİ'ler, görsel ve eski sürüm işitsel reCAPTCHA İEİ'lerine göre daha iyi kullanılabilirlik özellikleri göstermektedir. Ancak veritabanındaki bütün testler az sayıda soru örneği kullanılarak üretildiğinden textCAPTCHA İEİ'leri güvenlik bakımından yetersizdir ve bu durum daha geniş çapta kullanılmasına engel teşkil etmektedir.

372 katılımcının yer aldığı bir kullanıcı çalışmasında kullanıcılar, SMARTCHA testleri görsel testlerden daha fazla değiştirilmelerine ve daha çok hatalı yanıtla-

---

<sup>7</sup>"If there is a problem you cannot solve, then there is an easier problem you can solve: find it."

malarına rağmen, istatistiksel olarak anlamlı bir şekilde SMARTCHA testlerini çözmenin daha eğlenceli olduğuna karar vermişlerdir. Görsel İEİ'ler kullanışlılık açısından kötü bir üne sahiptirler. Önerilen SMARTCHA sistemi, güvenlik ve kullanışlılık arasındaki gerginliği azaltacak bir çözüm olabilir.

SMARTCHA ile yeni sürüm işitsel reCAPTCHA'yı karşılaştırma amacıyla 31 görme engelli katılımcının yer aldığı bir başka kullanıcı çalışması daha yapılmıştır. Bu çalışmanın sonuçlarına göre kullanıcılar SMARTCHA İEİ'lerini işitsel reCAPTCHA İEİ'leriyle benzer başarı oranlarıyla çözebilmiş ve aynı kolaylıkta bulmuşlardır. Dahası, SMARTCHA İEİ'lerini daha kısa sürede çözebilmiş ve çözerken daha fazla eğlenmişlerdir.

SMARTCHA'nın bir dezavantajı ölçeklenebilir olmayışıdır. Her gün yaklaşık 200 milyon İEİ testi çözülmektedir [23] ve SMARTCHA bugünkü haliyle bu büyük yarışa çıkmaya hazır değildir. Fakat SMARTCHA, küçük veya orta ölçekli kuruluşlar için birincil İEİ testi, büyük kuruluşlar için ise ikincil İEİ testi olma potansiyeline sahiptir.

Uygulamaya konduğu zaman SMARTCHA'nın uzman bir firma tarafından yönetileceği öngörülmektedir. Günde 30 milyondan fazla İEİ testi sunan reCAPTCHA sistemi sürekli Google'ın gözetimi altında tutularak saldırılara karşı korunmaktadır. Benzer şekilde, SMARTCHA'nın güvenlik motoru yönetici firma tarafından sürekli güncel tutulmaya çalışılmalıdır. Bunun dışında yönetici firma, İEİ testi üretme aşamasını daha eğlenceli ve ilgi çekici hale getirebilmek için oyunlaştırma tekniklerinden yararlanabilir. Oyunlaştırma tekniğini uygulayabilmek için kullanılacak örnek sloganlar şu şekildedir: "Arkadaşımdan daha fazla güvenli soru hazırla", "500 adet güvenli soru hazırlayarak 500MB'lık bulut depolama alanı kazan" veya "100 adet güvenli soru hazırlayarak yeni çıkmış bu oyuna sahip olma şansını yakala". SMARTCHA, çok uluslu bir firma tarafından yönetildiği takdirde her bölgede yerel soruları içerecek şekilde güncellenebilir, böylelikle farklı bölgelerdeki kültür ve dil bariyerleri aşılabılır.

Bahsedildiği üzere, oyunlaştırma teknikleri gibi alternatif metot ve stratejiler kullanılarak insanlar tarafından üretilen İEİ sorularının kalitesi ve sayısı artırılabilir. Örneğin SMARTCHA sorularının kullanışlılığını test edebilmek için operatörlerden yeni bir İEİ sorusu hazırlamadan önce önceden hazırlanmış bir soruyu çözmeleri istenebilir. Sonuçlar incelenerek belli bir eşik değerinden daha fazla operatör tarafından çözülemeyen sorular kullanışsız olarak işaretlenebilir ve veritabanından otomatik olarak çıkarılabilir. Dahası, bu teknik sistemin kötüye kullanımını önleyebilir. Örneğin, bazı operatörler daha kısa sürede daha fazla güvenli soru üre-

tebilmek adına bilerek güvenli fakat kullanışsız sorular üretebilirler. Bahsedilen yöntemle hazırlanan soruların kullanışlı oldukları kanıtlandıktan sonra operatörlere ödeme yapmak bu problem için bir çözüm olabilir.

Sonuç olarak, SMARTCHA'nın iki adet önemli katkısı vardır. 1) Görsel ve işitsel reCAPTCHA İEİ'lerine göre daha kullanışlıdır: sırasıyla 372 katılımcı ve 31 katılımcı ile gerçekleştirilen iki kullanıcı çalışmasına göre kullanıcılar SMARTCHA İEİ'lerini çözmeyi görsel ve işitsel reCAPTCHA İEİ'lerini çözmeye göre daha eğlenceli bulmuşlardır. Ayrıca kullanıcılar, soru ön filtreleme tekniklerini kullanan SMARTCHA İEİ'lerini görsel ve işitsel reCAPTCHA İEİ'lerinden daha hızlı çözmüşlerdir. 2) Aynı zamanda SMARTCHA, textCAPTCHA'dan daha güvenlidir: SMARTCHA'nın güvenlik motoru textCAPTCHA benzeri soruların kullanımını engellemektedir. Bu güvenlik motoru, sorunun otomatik olarak çözülüp çözülmediğini tespit etmek ve çözülebilen soruları elemek için detayları bu tezde verilen birçok farklı tekniği beraber kullanmaktadır.

Son olarak İEİ'lerin geleceğinden bahsetmek gerekirse, öncelikle İEİ'lerin bir ihtiyaçtan ortaya çıktığını ve bu ihtiyacın henüz devam ettiğini hatırlatmak gerekmektedir. Öte yandan günümüzde kullanımda olan İEİ sistemlerin hiçbirinin güvenlik, kullanışlılık ve erişilebilirlik şartlarının tamamını aynı anda sağlayamadığı da bir gerçektir. Bu sebeple İEİ'lerin kısa ve orta vadede (belki bazı küçük iyileştirmelerle) aynen kullanılmaya devam edileceği ön görülmekte, uzun vadede ise daha farklı ve bütün gereksinimleri yerine getiren bir yöntemle geçileceği düşünülmektedir. Bu yeni yöntemin kullanılmasının maddi ya da teknik yetersizlikler nedeniyle şu anda imkanı olmayan fakat bilimin ve teknolojinin ilerlemesiyle birlikte ucuzlayacak ya da bulunacak yeni yöntemlerden oluşacağına inanılmaktadır.



## KAYNAKLAR

- [1] **Zhu, B. B.** ve diğeri. “Attacks and design of image recognition CAPTCHAAs”. In: *Proceedings of the 17th ACM conference on Computer and communications security*. ACM. 2010, pp. 187–200.
- [2] **Bursztein, E., Martin, M.,** ve **Mitchell, J.** “Text-based CAPTCHA strengths and weaknesses”. In: *Proceedings of the 18th ACM conference on Computer and communications security*. ACM. 2011, pp. 125–138.
- [3] **Chellapilla, K.** ve diğeri. “Computers beat humans at single character recognition in reading based human interaction proofs (HIPs)”. In: *Proceedings of the 2nd Conference on Email and Anti-Spam*. 2005, pp. 21–22.
- [4] **Naor, M.** “Verification of a human in the loop or Identification via the Turing Test”. In: *Unpublished draft from [http://www.wisdom.weizmann.ac.il/~naor/PAPERS/human\\_abs.html](http://www.wisdom.weizmann.ac.il/~naor/PAPERS/human_abs.html)* (1996).
- [5] **Pope, C.** ve **Kaur, K.** “Is it human or computer? Defending e-commerce with captchas”. In: *IT professional 7.2* (2005), pp. 43–49.
- [6] **Von Ahn, L.** ve diğeri. “CAPTCHA: Using hard AI problems for security”. In: *Advances in Cryptology—EUROCRYPT 2003*. Springer, 2003, pp. 294–311.
- [7] **May, M.** “Inaccessibility of CAPTCHA: Alternatives to visual Turing Tests on the Web”. In: *web page. <http://www.w3.org/TR/turingtest>* (2005).
- [8] **Bursztein, E.** ve diğeri. “How good are humans at solving CAPTCHAAs? a large scale evaluation”. In: *Security and Privacy (SP), 2010 IEEE Symposium on*. IEEE. 2010, pp. 399–413.
- [9] **Tam, J.** ve diğeri. “Breaking audio captchas”. In: *Advances in Neural Information Processing Systems 1.4* (2008).
- [10] **Bursztein, E.** ve **Bethard, S.** “Decaptcha: breaking 75% of eBay audio CAPTCHAAs”. In: *Proceedings of the 3rd USENIX conference on Offensive technologies*. USENIX Association. 2009, p. 8.
- [11] **Subaşıoğlu, F.** “Engellilerin İnternet’e Erişimi Üzerine”. In: *Türk Kütüphaneciliği* (2000).

- [12] **Godfrey, P. B.** “Text-based CAPTCHA algorithms”. In: *First Workshop on Human Interactive Proofs*. 2002.
- [13] **Przydatek, B.** “On the (im) possibility of a text-only CAPTCHA”. In: *First Workshop on Human Interactive Proofs*. 2002.
- [14] **Hernandez-Castro, C. J.** ve **Ribagorda, A.** “Pitfalls in CAPTCHA design and implementation: The Math CAPTCHA, a case study”. In: *Computers & Security* 29.1 (2010), pp. 141–157.
- [15] **Gutmann, P.** “Engineering Security”. In: *Book draft* (2013). <http://www.cs.auckland.ac.nz/~pgut001/pubs/book.pdf>, Son erişim: 10 Ağustos 2016.
- [16] **Bergmair, R.** ve **Katzenbeisser, S.** “Towards human interactive proofs in the text-domain”. In: *Information Security*. Springer, 2004, pp. 257–267.
- [17] **Ximenes, P.** ve diğerleri. “A CAPTCHA in the Text Domain”. In: *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*. Springer. 2006, pp. 605–615.
- [18] **Yamamoto, T., Tygar, J. D.,** ve **Nishigaki, M.** “Captcha using strangeness in machine translation”. In: *Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on*. IEEE. 2010, pp. 430–437.
- [19] **Chew, M.** ve **Tygar, J. D.** “Collaborative filtering captchas”. In: *Human Interactive Proofs*. Springer, 2005, pp. 66–81.
- [20] **Östlund, R.** ve **Johansson, P.** “Sentence First CAPTCHA: Proposal and study of a text based CAPTCHA scheme”. PhD thesis. University of Gävle, 2011.
- [21] **Tuley, R.** *Accessible Text CAPTCHA Logic Questions*. <http://textcaptcha.com/>, Son erişim: 10 Ağustos 2016.
- [22] **Anwar, S.** *TextCaptchaBreaker: Python application that ‘breaks’ textCAPTCHA*. <https://github.com/kbhomes/TextCaptchaBreaker>, Son erişim: 10 Ağustos 2016.
- [23] *Official web page*. <http://www.google.com/recaptcha>, Son erişim: 10 Ağustos 2016.
- [24] **Goodfellow, I. J.** ve diğerleri. “Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks”. In: *CoRR* abs/1312.6082 (2013). URL: <http://arxiv.org/abs/1312.6082>.
- [25] **Sano, S., Otsuka, T.,** ve **Okuno, H. G.** “HMM-based CAPTCHA Breaker for Overlapped Symbols”. In: *The 76th National Convention of IPSJ* 2 (2014), p. 3.



- [26] *Defcon Group 949*. <http://www.dc949.org/projects/stiltwalker/>, Son erişim: 4 Mart 2016.
- [27] *Audio recaptcha: completely impossible*. <https://groups.google.com/forum/#!topic/recaptcha/lkCyM34zbJo>, Son erişim: 10 Ağustos 2016.
- [28] *CAPTCHA: Inaccessible to Everyone*. <http://www.sitepoint.com/captcha-inaccessible-to-everyone>, Son erişim: 10 Ağustos 2016.
- [29] *CAPTCHAs that capture your heart*. <https://security.googleblog.com/2014/02/captchas-that-capture-your-heart.html>, Son erişim: 10 Ağustos 2016.
- [30] *Attacking Audio reCaptcha using Google's Web Speech API*. <http://www.debasish.in/2014/04/attacking-audio-recaptcha-using-googles.html>, Son erişim: 10 Ağustos 2016.
- [31] **Nielsen, J.** “How many test users in a usability study”. In: *Nielsen Norman Group 4* (2012).
- [32] **Wayland, F.** *The limitations of human responsibility*. D. Appleton, 1838.
- [33] **Quinn, A. J.** ve **Bederson, B. B.** “Human computation: a survey and taxonomy of a growing field”. In: *Proceedings of the 2011 annual conference on Human factors in computing systems*. ACM. 2011, pp. 1403–1412.
- [34] **Von Ahn, L.** “Human Computation”. AAI3205378. PhD thesis. Pittsburgh, PA, USA, 2005. ISBN: 0-542-54148-3.
- [35] **Dijkstra, E. W.** *ALGOL-60 translation*. Mathematisch Centrum, 1961.
- [36] *Web Search Engine*. <http://www.bing.com>, Son erişim: 10 Ağustos 2016.
- [37] *Computational Knowledge Engine*. <http://www.wolframalpha.com>, Son erişim: 10 Ağustos 2016.
- [38] **Motoyama, M.** ve diğerleri. “Re: CAPTCHAs-Understanding CAPTCHA-Solving Services in an Economic Context.” In: *USENIX Security Symposium*. Vol. 10. 2010, p. 3.
- [39] **Douceur, J. R.** “The sybil attack”. In: *Peer-to-peer Systems*. Springer, 2002, pp. 251–260.
- [40] *Marketplace for work*. <http://www.mturk.com>, Son erişim: 10 Ağustos 2016.
- [41] *MySQL: the world's most popular open source database*. <http://www.mysql.com>, Son erişim: 10 Ağustos 2016.
- [42] **Levenshtein, V. I.** “Binary codes capable of correcting deletions, insertions, and reversals”. In: *Soviet physics doklady*. Vol. 10. 8. 1966, pp. 707–710.

- [43] *Wikipedia article*. <http://en.wikipedia.org/wiki/ReCAPTCHA>, Son erişim: 10 Ağustos 2016.
- [44] *Yahoo Groups*. <http://groups.yahoo.com>, Son erişim: 10 Ağustos 2016.
- [45] *JAWS Screen Reader*. <http://www.freedomscientific.com>, Son erişim: 10 Ağustos 2016.
- [46] *WebAim, WAVE - Web Accessibility Evaluation Tool*. <http://wave.webaim.org>, Son erişim: 10 Ağustos 2016.
- [47] **Nazir, M.** ve diğerleri. “Poster: Captchæcker—Automating Usability-Security Evaluation of Textual CAPTCHAs”. In: *Proceedings of the Seventh Symposium on Usable Privacy and Security*. ACM. 2011.
- [48] *Are you a robot? Introducing “No CAPTCHA reCAPTCHA”*. <https://security.googleblog.com/2014/12/are-you-robot-introducing-no-captcha.html>, Son erişim: 10 Ağustos 2016.
- [49] **Olalere, A.** ve diğerleri. “Investigating the effects of sound masking on the use of audio CAPTCHAs”. In: *Behaviour & Information Technology* 33.9 (2014), pp. 919–928.
- [50] **Sauer, G.** ve diğerleri. “Towards a universally usable human interaction proof: evaluation of task completion strategies”. In: *ACM Transactions on Accessible Computing (TACCESS)* 2.4 (2010), p. 15.
- [51] **Lazar, J.** ve diğerleri. “The SoundsRight CAPTCHA: an improved approach to audio human interaction proofs for blind users”. In: *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*. 2012, pp. 2267–2276.
- [52] *Fail2ban*. <http://www.fail2ban.org>, Son erişim: 10 Ağustos 2016.
- [53] **Tran, M.** ve **Armitage, G.** “Mitigating Email Spam by Statistical Rejection of TCP Connections Using Recent Sender History”. In: *Australian Telecommunication Networks and Application Conference*. 2006.

## **EKLER**

**EK 1 :** SMARTCHA veritabanı tabloları.

**EK 2 :** SMARTCHA sistemindeki İngilizce soruların sayısını artırmak için kullanılacak ek örüntüler.

**EK 3 :** Birinci insan hesaplama çalışmasının duyurusu için kullanılan afiş.



## EK 1

### Çizelge Ek.1: SMARTCHA veritabanı tabloları.

<b>Answers</b>				
answer_id	int	auto inc	primary key	cevabın tanımlayıcısı
answer	varchar			cevap metni
timesUsed	int			cevabın kaç kere kullanıldığı

<b>Questions</b>				
question_id	int	auto inc	primary key	sorunun tanımlayıcısı
question	varchar			soru metni
answer_id	int		foreign key	sorunun cevabının tanımlayıcısı
isValid	tinyint			(0 - 1 - 2) 0: güvenli olmayan soru, 1: güvenli fakat pasif soru, 2: güvenli soru

<b>OperatorQuestions</b>				
question_id	int	auto inc	foreign key	sorunun tanımlayıcısı
operator_id	int		foreign key	soruyu soran operatörün tanımlayıcısı
typingTime	int			operatörün soruyu yazma süresi (saniye cinsinden)
typeOfQuestion	boolean			operatör hem soruyu hem de cevabı kendisi mi yazmış yoksa kendisine verilen cevaba mı soru hazırlamış bilgisi
howManyPossibleAnswers	int			güvenlik motorunun soru için kaç tane muhtemel cevap ürettiği bilgisi
howLongItTookToCheckSecurity	int			sorunun güvenli olup olmadığı belirlemek için güvenlik motorunun kaç milisaniye çalıştığı bilgisi
questionStateInSecurityEngine	int			sorunun güvenlik motorunda hangi aşamaya kadar gelebildiği bilgisi
bingTotalNumberOfWords	int			Bing arama motorunun bu soru için kaç kelime döndürdüğü bilgisi

<b>OperatorsPersonalInfo</b>				
operator_id	int	auto inc	primary key	operatörün tanımlayıcısı
opname	varchar(25)			operatörün adı
opsurname	varchar(25)			operatörün soyadı
age	int			operatörün yaşı
gender	boolean			operatörün cinsiyeti (true = erkek, false = kadın)
department	varchar(25)			operatörün bölümü
universityNo	varchar(12)			operatörün öğrenci numarası
nationalIdNo	varchar(12)			operatörün TC kimlik numarası
paymentType	boolean			operatörün saat başına mı yoksa hazırladığı güvenli soru sayısına göre mi ücret alacağı bilgisi (true: saat başına, false: güvenli soru sayısı karşılığı)

<b>OperatorsLoginInfo</b>				
operator_id	int	auto inc	foreign key	operatörün tanımlayıcısı
user name	varchar			operatörün kullanıcı adı
password	varchar			operatörün parolası
registration time	timestamp			operatörün sisteme kayıt olduğu zaman
totalAmountOfWorkingTime	int			operatörün toplam çalışma süresi (dakika)
lastLoginTime	timestamp			operatörün en son giriş yaptığı zaman

## EK 2

Çizelge Ek.2: SMARTCHA sistemindeki İngilizce soruların sayısını artırmak için kullanılabilecek ek örüntüler.


<b>Pattern</b>	<b>Type of answer</b>
Calculate the sum of digits in the answer to the question “Base Question”?	numeric
Calculate the product of digits in the answer to the question “Base Question”?	numeric
What would be the result if you would subtract the answer to the question “Base Question” from the number in which the digits of the same answer are written in reverse order?	numeric
Order the digits of the answer to the question “Base Question” in ascending/descending order?	numeric
Double type all digits in the answer to the question “Base Question”?	numeric
Double type all occurrences of digit '2' in the answer to the question “Base Question”?	numeric
Triple type the first occurrence of the digit '2' in the answer to the question “Base Question”?	numeric
What would be the result after removing all occurrences of digit '2' from the answer to the question “Base Question”?	numeric
What would be the answer to the question “Base Question” after removing all repeating occurrences of digits?	numeric
Round the answer of the question “Base Question” to the nearest tens digit?	numeric
Enter only the odd digits of the answer to the question “Base Question”?	numeric
Enter only the even digits of the answer to the question “Base Question”?	numeric
Double type all letters of the answer to the question “Base Question”?	character
Double type all occurrences of letter 'a' in the answer to the question “Base Question”?	character
Triple type the first occurrence of the letter 'a' in the answer to the question “Base Question”?	character
What would be the answer to the question “Base Question” after removing 2nd and 3rd letters?	character
What would be the answer to the question “Base Question” after removing all occurrences of letter 'a'?	character
What would be the answer to the question “Base Question” after removing all repeating occurrences of letters?	character
What would be the answer to the question “Base Question” after concatenating the first two letters of the answers in both back and front?	character
What would be result if you alphabetically order the characters of the answer to the question “Base Question”?	character
What would be result if you alphabetically reverse-order the characters of the answer to the question “Base Question”?	character
What are the vowels of the answer to the question “Base Question”?	character
Type only the odd enumerated characters in the answer to the question “Base Question” where first character is enumerated as 1?	character
Type only the even enumerated characters in the answer to the question “Base Question” where first character is enumerated as 1?	character
Starting from 4th character, type only the first 3 letters of the answer to the question “Base Question”?	character
Type the answer of the question “Base Question” in reverse order?	both


### EK 3


Bu afiş hem üniversite email gruplarına gönderilerek, hem de üniversitenin çeşitli yerlerine asılarak çalışmaya katılımcılar toplanmıştır.

Çoğumuzun CAPTCHA olarak bildiği İnsan-Etkileşim Sistemleri, günümüzde bir çok yerde kullanılmaktadır.


Duyurunun devamını okuyabilmek için lütfen aşağıdaki CAPTCHA'yı çözünüz.

No premium user. Please enter all letters having a  below.



Four letters with a  :  [Download via Level\(3\) #4](#)

Çok mu zor geldi? Peki bir de bunu deneyin:



Yine mi olmadı?

O zaman gelin, birlikte daha kullanışlı bir sistem oluşturalım.

TÜBİTAK desteğiyle yapılan projemizde daha kullanışlı bir sistemi oluşturmakta etkin rol oynamak ve aynı zamanda **para kazanmak** istiyorsanız **hakanezgi@etu.edu.tr** adresine mail atabilir ya da **B-11** numaralı ofiste benimle irtibata geçebilirsiniz.

Çalışmaların Mayıs ve Haziran aylarında olması planlanmaktadır. İnternet erişimine sahip olduğunuz herhangi bir yerden herhangi bir zamanda çalışabilirsiniz. Çalışma karşılığı ücret 15TL ve katları şeklinde olacak olup, kaç kata kadar çıkacağınızı **siz** belirleyeceksiniz. (tabi ki bütçenin elverdiği ölçüde ☺)

Doç. Dr. Kemal BIÇAKCI      TÜBİTAK 1002 Projesi - SmarTCHA      Hakan Ezgi KIZILÖZ

Şekil Ek.1: Birinci insan hesaplama çalışmasının duyurusu için kullanılan afiş.

## ÖZGEÇMİŞ

**Ad-Soyad** : Hakan Ezgi KIZILÖZ  
**Uyruđu** : TC  
**Dođum Tarihi ve Yeri** : 26 Mart 1986 Ankara  
**E-posta** : hakanezgi@etu.edu.tr

### ÖĐRENİM DURUMU:

- **Yüksek Lisans** : 2010, TOBB Ekonomi ve Teknoloji Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliđi
- **Lisans** : 2008, TOBB Ekonomi ve Teknoloji Üniversitesi, Fen Edebiyat Fakültesi, Matematik

### MESLEKİ DENEYİM VE ÖDÜLLER:

Yıl	Yer	Görev
2015-...	TED Üniversitesi	Araştırma Görevlisi
2011-2015	TOBB Ekonomi ve Teknoloji Üniversitesi	Tam Burslu Lisansüstü Öğrencisi
2008-2010	TOBB Ekonomi ve Teknoloji Üniversitesi	Tam Burslu Lisansüstü Öğrencisi

**YABANCI DİL:** İngilizce

**TEZDEN TÜRETİLEN YAYINLAR, SUNUMLAR VE  
PATENTLER:**

- Bicakci, K. and **Kiziloğ, H. E.**, “Leveraging Human Computation for Pure-Text Human Interaction Proofs,” in International Journal of Human-Computer Studies, Vol.92-93, August-September 2016, pp.44-54.
- **Kiziloğ, H. E.**, and Bicakci, K., “Towards Making Accessible Human-Interaction Proofs More Secure and Usable,” in 20th IEEE Symposium on Computers and Communication (ISCC2015), 6-9 July 2015, Larnaca, Cyprus.
- **Kızılöz, H. E.**, and Bıçakçı, K., “SmarTCHA: İnsan Hesaplama Kullanarak Oluşturulan Erişilebilir ve Kullanışlı İnsan Etkileim İspat Sistemi,” Elektrik-Elektronik ve Bilgisayar Mühendisliği Sempozyumu (ELECO) 2012, 29 Kasım-1 Aralık 2012, Bursa.