SmarTCHA: İnsan Hesaplama Kullanarak Oluşturulan Erişilebilir ve Kullanışlı İnsan Etkileşim İspat Sistemi

SmarTCHA: An Accessible and Usable Human Interaction Proof System Built Using Human Computation

Hakan Ezgi Kızılöz, Kemal Bıçakcı

Bilgisayar Mühendisliği
TOBB Ekonomi ve Teknoloji Üniversitesi
hakanezgi@etu.edu.tr, bicakci@etu.edu.tr

Özet

İnsanların kolaylıkla çözebileceği fakat bilgisayarların ya da programların çözemeyeceği testler olarak tasarlanan İnsan-Etkileşim İspatı (İEİ) günümüzde pek çok internet uygulamalarında yaygın olarak kullanılmaktadır. Ancak bu testlerin yaygın kullanımına karşın kullanışlılık ve erişilebilirlik problemleri vardır. Bu çalışmada, insan-hesaplama kavramından yararlanarak İEİ testlerinin bilgisayarlar tarafından çözülmesini kolaylaştırmadan kullanışlılığını ve erişilebilirliğini artırılmasını hedefleyen SmarTCHA sistemi tanıtılmaktadır.

Abstract

Tests for Human-Interaction Proof (HIP), which are assumed to be easily passed by humans but not by computers or automated programs, are widely used nowadays. Although being widely used, these tests have usabilty and accessibility problems. In this paper we introduce a new system, SmarTCHA, which aims to increase HIP tests' accessibility and usability by using human-computation, without easing automatic solubility.

1. Giriş

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, pratik olarak insanların kolay kullanımını merkeze alan bir yapıda değildir. En az kullanışlılık 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 metinsel olarak sunulan İEİler verilebilir. Fakat tamamen otomatik yöntemlerle üretilen bu tür İEİ testleri bilgisayarlar tarafından otomatik olarak çözülmeye daha elverişli olup, istenilen güvenliği sağlayamamaktadır.

İnsan hesaplama (human computation) kavramının filozofi ve psikoloji literatüründe kullanımı 1838 yıllarında görülmekteyken, bilgisayar bilimleri literatürüne girişi 1950 yılında Alan Turing ile olmuştur [1]. İnsan hesaplama kavramının modern anlamda kullanımına ise 2005 yılında von Ahn'ın "İnsan Hesaplama" başlıklı tezi ve sonrasındaki çalışmaları ile ilham olmuştur. İlgili tez çalışmasında insan hesaplama kavramı "insan işlem gücü kullanılarak bilgisayarların henüz çözemediği problemlerin çözülmesi" olarak tanımlanmıştır. Daha sonra yapılan çalışmalardaki tanımların ortak paydası bulunduğunda, insan hesaplama kavramı şu şekilde tanımlanabilir [1]:

- Genel olarak hesaplama yapısına uyan ve bir gün bilgisayarlar tarafından çözülebilir hale gelme ihtimali bulunan problemlerdir.
- İnsanların katkısı, hesaba dayalı sistemler veya işlemler tarafından yönlendirilmelidir.

Bu çalışmada, erişilebilirlik ve kullanışlılık problemlerini çözmekte umut vadeden metinsel İEİ'lerin güvenlik konusundaki zafiyeti kaldırmaya çalışılmaktadır. Bu amaçla metin İEİ testlerinin "tamamen otomatik" olarak oluşturulması yerine, "İEİ operatörleri" adını verdiğimiz kişilerce üretilmesi konusu araştırılmakta ve uygulanabilirlik, maliyet, hız, vb. temel parametreler incelenmektedir.

2. İlgili Çalışmalar

İnsan-Etkileşim İspatı (İEİ) kavramı ilk kez 1996 yılında Moni Naor tarafından ortaya atılmıştır [2]. İlk olarak bu kavramın pratik olarak Altavista şirketi tarafından 2001

yılında kullanıldığını görmekteyiz [3]. İlk defa 2003 yılında İEİ kavramını nitelemek için CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) terimi kullanılmaya başlanmıştır [4].

Günümüzde çoğu pratik uygulamada tercih edilen karakter veya sözcük tanıma tabanlı İEİ'ler gelişen optik karakter tanıma ve kesimleme (segmentation) teknolojisi karşısında, Şekil 1'deki gibi karmaşık olsalar bile, güvensiz hale gelmişlerdir [5][6][7]. Aralarında Msn, Yahoo ve Google gibi sitelerin de bulunduğu 7 farklı sistemin kullandığı İEİ testlerini, bilgisayarların en az normal görebilme yetisine sahip insanlar kadar iyi çözebildiği, hatta sorular karmaşıklaştıkça insanlardan daha iyi çözebildikleri tespit edilmiştir [7]. Alternatif olarak önerilen ve pek çok farklı türü olan imaj tanıma tabanlı sistemler [5] İEİ'lerin güvenlik problemine çözüm potansiyeline sahip olmakla birlikte kullanışlılık ve bilhassa erişilebilirlik problemlerine çare olamamaktadırlar. İmaj tabanlı İEİ'ler gözleri iyi görmeyen kullanıcılarca yine çözümü imkânsız testlerdir.

Şekil 1: Slashdot.org sitesinde kullanılan ve otomatik olarak çözülebilen bir CAPTCHA örneği

İEİ'lerin erişilebilirlik problemi uzun zamandır bilinmektedir. Bu problemin çözümü için WWW konsorsiyumu tarafından önerilen alternatiflerden [8] en yaygın olarak tercih edileni ses İEİ'leridir. Ses İEİ'leri gürültülü bir ortamda 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 ses İEİ'leri hem kullanışlılık problemleri içermekte [9] hem de başarılı güvenlik saldırılarına hedef olmaktadırlar [10] [11]. Yapılan yeni bir çalışmada [9] ses 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 algılanabildiği için ses İEİ'leri gibi görme engelli İnternet kullanıcılarının çözebildikleri testlerdir [12]. Bu tür testlerin tamamen otomatik yöntemlerle "güvenli" bir şekilde üretilebilmesi henüz çözümlenmemiş bir araştırma problemidir [13] [14]. Grafik unsurları içermeyen İEİ'lere verilecek enteresan bir örnek Hırvatistan'daki bir enstitü tarafından geliştirilen "Math CAPTCHA" yöntemidir. 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 [15].

Bizim çalışma konumuz ile en fazla benzerlik gösteren İEİ türü textCAPTCHA uygulamasıdır [16]. Uygulamanın web sayfasında 180 milyonun üzerinde soruya sahip bir veritabanından bahsedilmektedir. Bedava web servisi olarak WordPress ve bunun gibi pek çok web sitesine günlük ortalama 271.000 civarında soru sunulmaktadır. Soruların nasıl üretildiği konusunda bir bilgiye rastlanılmamakla birlikte, Çizelge 1'de örnekleri görülen soru türlerinden ve veri tabanının büyüklüğünden otomatik yöntemlerin kullanıldığı sonucunu çıkartmaktayız. Bu uygulamanın maalesef en büyük dezavantajı güvenlik konusundadır. İnternette açık kaynak kodlu geliştirilen bir proje [17] ile söz konusu İEİ sorularına otomatik olarak %99,5 başarı oranıyla doğru cevap vermek mümkün olmaktadır.

Çizelge 1: textcaptcha.com sitesinde yer alan İEİ sorularına örnekler

Soru:
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?

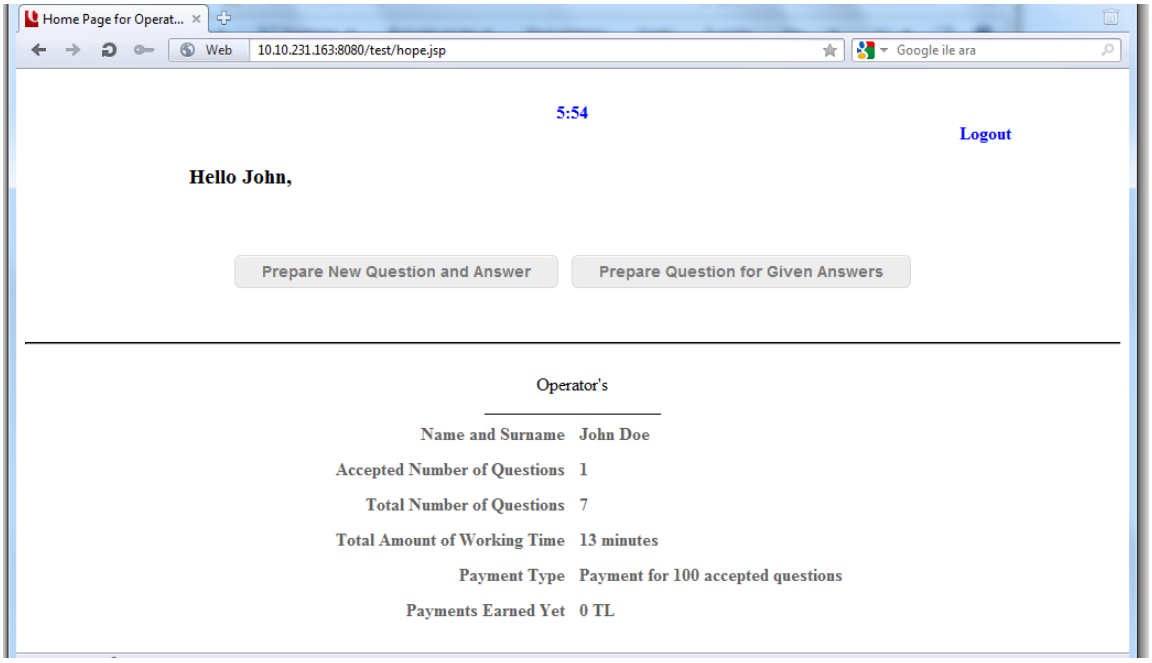
Çizelge 1'de sunulan örnek sorulara bakıldığında saldırıların bu kadar başarılı olmasının hiç de sürpriz olmadığı sonucunu çıkarabiliriz. Örnek sorulardan çıkarılabilecek diğer bir sonuç ise bazı soruların cevabının soru içerisindeki kelimelerden biri olduğu ve rastgele bu kelimelerden birinin cevap olarak seçilmesi durumunda dahi doğru cevap verme olasılığının hayli yüksek seviyelerde olduğudur. İEİ'lere otomatik olarak doğru cevap verilme olasılığının hangi seviyelerde olmasının kabul edilebilir olduğunu araştıran bir çalışmada %0,6'dan yüksek oranlarda başarılı sağlayan saldırıların "etkin" sayılması gerektiği sonucu çıkarılmıştır [5].

Bununla birlikte diğer bazı güvenlik araştırmacıları İEİ'lerin pratik ve teorideki sağladığı güvenliğin karıştırılmaması gerektiğini ve web sayfasında kullanıcıya metin olarak sunulan bir kelimenin (örneğin "orange" kelimesinin) bir metin kutusuna klavyeden girilmesini gerektiren ve düşünülebilen en basit İEİ'nin bile pratikte işe yarayabileceğini belirtmişlerdir [18]. Bizim bu çalışma öncesindeki pozisyonumuz teorik bir saldırının pratikte de gerçekleştirilmesinin başka faktörlere bağlı olduğunu kabul ile beraber bir İEİ'nin sahip olması gereken en temel özellik olan güvenliğinin belli bir seviyenin üzerine çıkarılması gerektiği yönündedir.

Konu ile ilgili literatürdeki çalışmalar genelde iki kampa ayrılmıştır. Bir tarafta yeni yöntemler öneren ve bu önerilerde zor yapay zeka problemlerinden faydalanan çalışmalar (örnek: [19] [20]), diğer tarafta ise bu yeni önerilere başarılı saldırılar yapılabileceğini gösteren çalışmalar vardır [21]. Bizim çalışmamız ile mevcut yöntemler arasında tamamen otomatik olma - olmama ve erişilebilirlik problemine çözüm sunma - sunmama gibi temel farkların bulunması sebebiyle bu söz konusu çalışmalardan daha detaylı bahsedilmesine gerek duyulmamıştır.

3. Çalışma Metodu

Bu çalışmada İEİ testlerinin insan hesaplama ile oluşturulmasının etkileri incelenmeye çalışılmıştır. Bunun için, "İEİ operatörü" adı verilen kullanıcıların internet üzerinden çalışabilecekleri bir sistem oluşturulmuştur ve operatörlerden bu sistemi kullanarak İEİ testleri hazırlamaları



Şekil 2: Operatör Giriş Sayfası

istenmiştir. Bu çalışmanın henüz bir pilot çalışma olması ve İngilizcenin evrensel bir dil olması sebebiyle, operatörlerden İEİ testlerini İngilizce hazırlamaları istenmiştir. Bununla birlikte, İEİ testlerinin Türkçe hazırlandığı çalışma da şu anda planlanmaktadır ve en kısa sürede o çalışmanın sonuçları da incelenecektir.

Bu çalışmada ücret karşılığı iş gücünden yararlanılan operatörler hakkında sadece yaşı, cinsiyeti, hangi bölümde okuduğu gibi kişisel olmayan temel bilgiler toplanmıştır. Operatörler, TOBB Ekonomi ve Teknoloji Üniversitesi lisans ve lisansüstü öğrencileri arasından seçilmiştir. Operatörlerin basit İngilizce sorular hazırlamaları gerektiğinden, öğrenciler seçilirken üniversitenin hazırlık programını geçmiş olmaları şartı aranmıştır.

Üniversite genelinde yapılmış olan duyuru 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 - 29 yaş aralığında değişen bu 29 operatörün 20 tanesi (yaklaşık %69'u) erkek öğrencilerken, 9 tanesi (yaklaşık %31'i) bayan 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 ederken geri kalan 29 operatör (yaklaşık %59 u) 100 geçerli soru yazma karşılığında 15 TL ücret almayı istemiştir.

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

Çalışmalar internet üzerinden yapılacağı ve üniversite bütün öğrencilere birer dizüstü bilgisayar verdiği için operatörlere çalışma günü, saati ve yeri kısıtı getirilmesine gerek görülmemiştir.

3.1. Sistem

Operatörler, web sitesi üzerinde kendilerine özel alanlara kendi kullanıcı adları ve parolaları ile giriş yaparak Şekil 2'de görülen operatör giriş sayfasını açarlar. Bu sayfa, operatörler için özelleşmiş bir sayfa olup operatörün adını soyadını, toplamda kaç soru yazabildiğini, bu sorulardan kaç tanesinin kabul edildiğini, operatörün saat karşılığı mı yoksa soru karşılığı mı ücreti tercih ettiğini, o anda ne kadarlık ücreti hak ettiğini görebileceği bir sayfadır. Operatörler ana sayfada "Prepare New Question & Answer" ya da "Prepare Question to Given Answers" düğmelerinden birisine tıklayarak soru girme sayfasına gidebilirler.

Operatör "Prepare Question to Given Answers" düğmesine tıklamışsa 1. tip soru girme sayfasına yönlendirilir ve kendisine gösterilecek cevap kümesinin belirlenmesi işlemini başlatır. Bu işlem, veritabanında bulunan cevapların kullanılma sayılarına göre azdan çoğa doğru sıralanmalarından sonra, kullanılma sayısı az olanlar öncelikli olmak üzere rastgele 20 tanesinin seçilerek belirlendiği bir kümenin oluşturulmasıdır. 1. tip soru girme sayfasında operatöre cevaplar kümesinin ilk elemanı gösterilir ve operatörden sadece bu cevabın doğru olduğu bir soru üretmesi istenir. Operatör bu cevap hakkında bir soru üretmek istemiyorsa "Next Answer" düğmesine tıklayarak kümedeki bir sonraki

cevaba geçebilir veya "Previous Answer" düğmesine tıklayarak kümedeki bir önceki cevap ile ilgili soru üretebilir. Operatör her zaman "Next Answer" veya "Previous Answer" düğmelerine tıklayarak cevap değiştirme hakkına sahip olduğu gibi "Give Me New Set of Answers" düğmesine tıklayarak 20 soruluk cevap kümesi seçim işlemini baştan başlatabilir. Operatör sadece bu cevabın doğru olduğu bir soruyu birinci yazı alanına yazdıktan sonra "Submit Question For The Given Answer" düğmesine tıklar ve kalite kontrol motorundan gelecek cevabı bekler.

Operatör "Prepare New Question & Answer" düğmesine tıklamışsa, kendisinin bir soru ve bu soruya uygun bir cevap yazabileceği 2. tip soru girme sayfasına yönlendirilir. Bu sayfada operatör, birinci yazı alanına yazmak istedikleri soruyu, ikinci yazı alanına ise bu sorunun cevabını yazar ve "Submit Question & Answer" düğmesine tıklayarak kalite kontrol motorundan gelecek cevabı bekler.

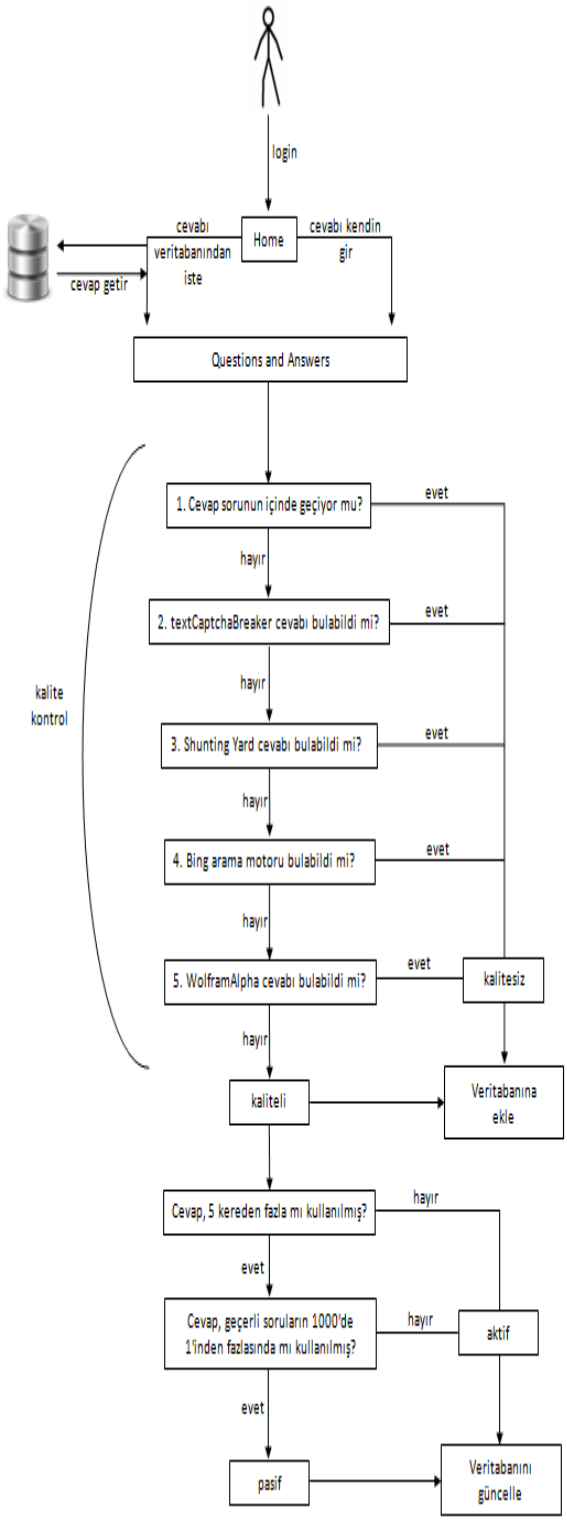
Herhangi bir tip soru girme sayfasında bulunan operatör, sayfada bulunan "Back" düğmesine tıklayarak operatör giriş sayfasına dönüş yapma ve diğer soru girme sayfasına geçme hakkına sahiptir.

Operatörlerin hangi tipte daha hızlı biçimde soru yazabildiklerini ölçmek amacıyla operatörlerin soru girme hızları da veritabanına saniye olarak kaydedilmiştir. Soru girme hızlarını ölçmek için; 1. tip soru girme sayfasında, operatör yazı alanını seçtikten sonra "Submit" düğmesine basana kadar geçen süre hesaplanırken 2. tip soru girme sayfasında, operatör yazı alanını veya cevap alanını seçtikten sonra "Submit" düğmesine basana kadar geçen süre hesaplanır ve veritabanına kaydedilir.

Bir sonraki kısımda ayrıntılı olarak nasıl çalıştığı anlatılacak olan kalite kontrol motorunun cevabı, her iki tip soru girme sayfasında da çok önemlidir. Eğer kalite kontrol motoru soruyu kabul etmişse, soru ve cevap veritabanına aktif soru olarak eklenerek operatöre soru eklemenin başarılı olduğuna dair bilgilendirme yazısı; kabul etmemişse de soru ve cevap veritabanına pasif soru olarak eklenerek operatöre soru eklemenin başarısız olduğuna dair bilgilendirme yazısı gösterilir. Eğer operatör 1. tip soru girme sayfasındaysa aynı cevap için başka bir soru ekleyebilmesi için yazı alanı boşaltılarak yeni soru girmesi için hazırlanır. Eğer operatör 2. tip soru girme sayfasındaysa hem birinci hem de ikinci yazı alanları boşaltılarak operatörün yeni soru girebilmesi için hazırlanır.
Operatörler istedikleri anda giriş sayfasına dönmekte ya da sistemden çıkış yapmakta serbesttirler. 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.

3.2. Kalite Kontrol Motoru

SmarTCHA ile bilgisayarların soruları kolayca çözememesinden ödün vermeksizin erişilebilir ve kullanışlı bir metin tabanlı İEİ sistemi oluşturulmak istediğimizden, halihazırda etkin biçimde kullanılan bir başka metin tabanlı



Şekil 3: Kalite Kontrol Motorunun Yapısı

servis olan textCAPTCHA'dan daha iyi/daha kaliteli bir sistem oluşturmamız hedeflenmektedir. textCAPTCHA'nın sunduğu soruların çoğunun bilgisayar programları ile otomatik olarak çözülebildiğini biliyoruz.

Bu sebeple SmarTCHA kalite kontrol motorunu birkaç aşamalı olarak tasarladık. Şekil 3'de görebileceğiniz bu aşamalardan ilkinde soru cevabının, büyük/küçük harf ayrımı olmaksızın, sorunun içerisinde geçip geçmediği kontrol edildi. Eğer sorunun cevabı, sorunun içerisinde geçmiyorsa bir sonraki aşamaya geçerek standart textCAPTCHA sorularını çözebilen algoritmaya sokuldu. Bu algoritma, bir önceki algoritmadan farklı olarak; yazıyla yazılmış sayıların rakamlara çevrilmesi, yazıyla veya rakamla yazılmış iki sayının toplam veya farkının bulunması, bir kelime veya sayı içerisindeki spesifik bir hanedeki sayının ya da harfin döndürülmesi, verilen bir kelimenin uzunluğunun döndürülmesi ile dün, bugün, yarın ve hafta sonu komutlarının anlaşılabildiği basit gün işlemlerini yapabilmektedir. Bu kısımda da doğru cevap algoritma tarafından elde edilememişse soru ve cevap, Shunting Yard algoritmasına yönlendirildi. Dijkstra'nın geliştirmiş olduğu Shunting Yard algoritması [22] kullanılarak ardışık olarak yapılan toplama, çarpma, çıkarma, bölme ve üs alma gibi basit aritmetik işlemlerin de kalite kontrol motoru tarafından reddedilmesi sağlanmıştır. Shunting Yard algoritması aslında sayılarla çalışabilen bir algoritma olmasına rağmen, kalite kontrol motoru için güncellenmiş olup "What is the result of (three hundred and three + five) / (four) + twelve?" sorunun cevabını 89 olarak verebilmektedir. Doğru cevabı Shunting Yard algoritması da bulamadıysa soru, Bing arama motorunda aratılarak ilk 50 sonuca bakıldı. Bing arama motorunun soru için döndürmüş olduğu ilk 50 cevap içerisinden; sorunun içerisinde bulunan kelimeler ve İngilizce bir anlamı olmadığı halde cümle bütünlüğünü sağlamaya yarayan ve adına Stop Words denen "a", "the", "there", "are" gibi kelimeler çıkarılarak cevaplar üzerinde kelimelerin yer alma sıklıkları hesaplandı (frekans analizi yapılmış) ve en çok geçen ilk 10 kelime bulundu. Daha sonra sorunun gerçek cevabı ile elde edilmiş olan en sık geçen ilk 10 kelime karşılaştırıldı ve sorunun Bing arama motoru tarafından otomatik olarak çözülebilip çözülemediğine karar verildi. Eğer Bing arama motoru da soruyu otomatik olarak çözememişse, son olarak soru WolframAlpha servisine soruldu ve oradan gelen cevaplar ile doğru cevap karşılaştırıldı. Ticari olmayan kullanımda ücretsiz olarak aylık 2000 sorguya kadar izin veren WolframAlpha web servisi, Wolfram Araştırma Şirketi tarafından geliştirilen bir hesaba dayalı bilgi motorudur (computational knowledge engine). WolframAlpha servisinden gelen cevaplar arasında da sorunun doğru cevabı bulunamamışsa bu sorunun kaliteli bir soru olduğuna karar verilmektedir. Ancak sorunun kaliteli olduğuna karar verilmesi, hemen veritabanına aktif soru olarak eklenmesini birlikte getirmemektedir. Proje sonunda oluşan sistemde bazı cevaplarda yığılma olmasını engelleyebilmek adına, 5'ten fazla farklı soru ile eşleştirilmiş cevaplar tekrar kullanılmadan önce bu cevabın toplam soruların %0,1 (binde bir)inden daha az kez kullanılıp kullanılmadığı kontrol edildi. Eğer cevap %0,1'den daha az kullanılmışsa soru-cevap ikilisi sisteme aktif soru olarak, fazla kullanılmışsa da kaliteli fakat pasif soru olarak kaydedildi. Bu önlemi almamış olsaydık ve bir cevap üzerinde yığılma olup bu cevap açığa çıksaydı, saldırganlar başka hiç bir algoritma uygulamadan, sadece kaba kuvvet atak (brute force attack) kullanarak bu yığılma olan cevabın doğruluğunu test edebilir ve sistemi kırabilirlerdi. Bu yöntem sayesinde, cevapların kullanılma sayıları düzensiz değil, birbirlerine yakın olacaktır, yani herhangi bir cevapta yığılma olmayacaktır.

Kalite kontrol motoru, dinamik yapısı sayesinde sonradan yapılabilecek güncellemelere açık olup soruların kalitesini ölçmede yeni kıstasların getirilmesine uygun yapıdadır.

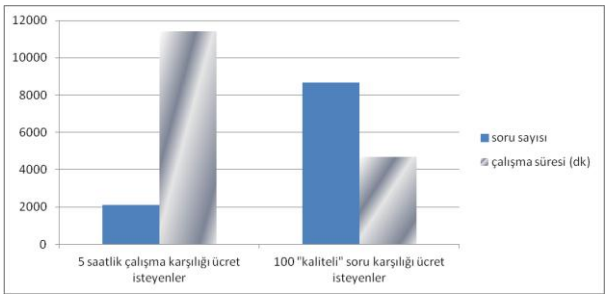
4. Sonuçlar

4.1. Operatör Çalışmaları

Operatörlerin soru girişi yapacakları sistem Mayıs 2012 sonuna doğru kullanıma açılmış ve 1 ay süresince operatörler soru girişleri yapmışlardır. Bu 1 ay süresince operatörler sisteme toplamda 10776 soru girmişlerdir. Sisteme girilmiş olan bu 10776 sorunun 8842 tanesi (yaklaşık %82'si) kalite kontrol motoru tarafından "kaliteli" olarak işaretlenmiş ve geriye kalan 1934 tanesi ise (yaklaşık %18'i) "kalitesiz" olarak işaretlenerek sisteme kaydedilmiştir.

Projede süre olarak en çok çalışan operatörün 7426 dakika çalışarak 200'ü kalitesiz ve 503'ü kaliteli olmak üzere toplamda 703 soru ürettiği gözlemlenmiştir. Yaklaşık olarak 10 dakikada 0,9 soru üreten bu operatör, her 5 saatlik çalışma karşılığı ücret almayı talep etmiştir. Diğer taraftan, 203'ü kalitesiz ve 5729'u kaliteli olmak üzere toplamda 5932 soru ile en çok soru üreten operatörün toplam 1990 dakika çalışmış olması dikkat çekmiştir. Yaklaşık olarak dakikada 2,98 soru üreten bu operatör ise, diğer operatörün aksine, yazmış olduğu her 100 kaliteli soru karşılığı ücret almayı talep etmiştir.

Diğer bir taraftan, 5 saatlik çalışma karşılığı ücret almayı seçen operatörlerin toplamda 11405 dakikada 2118 soru girdikleri (yaklaşık olarak dakikada 0,18 soru) gözlemlenmişken; 100 kaliteli soru karşılığı ücret almayı tercih eden operatörlerin ise toplamda 4697 dakikada 8658 soru ürettikleri (yaklaşık olarak dakikada 1,85 soru) gözlemlenmiştir (Şekil 4).



Şekil 4: Operatörlerin aldıkları ücrete tipine göre çalışma verimlilikleri

Bu gözlemler sonucu, operatörlerin hangi şekilde ücret almayı talep etmişlerse, kendilerini bu sistem içerisinde daha fazla ücret alacak şekilde çalışmaya adapte ettikleri görülmektedir.

4.2. Soruların Yapısı

Her ne kadar bir ay gibi kısa sürede %82'lik "kaliteli" oranıyla yaklaşık 11.000 soru üretilmiş olsa da, üretilmiş olan soruları incelediğimizde operatörlerin ürettikleri soruların hepsinin orijinal olmadığını; "kaliteli" olarak işaretlenmiş sorular üzerinde küçük değişiklikler yapılarak türetilmiş olduklarını fark ettik. Bu çalışmanın en büyük amaçlarından bir tanesi,

böyle bir sistemin ölçeklenebilir olup olmadığını tespit etmeye çalışmaktır. Yani başka bir deyişle, operatörlere yatırım yapmaya devam ettikçe yeni tür soruların bulunabilip bulunamadığını test etmektir. Bu sebeple operatörlerin, "kaliteli" olarak işaretlenmiş sorulardan yeni sorular türetmesi, sistemdeki soru sayısının artması açısından önemli olsa da ölçeklenebilirlik araştırmamız için bir katkı sağlamamaktaydı. Hem oryantasyon toplantısında hem de operatörlere dağıtılan kullanım kılavuzunda ısrarla belirtmemize rağmen, operatörlerin daha önce hazırlamış oldukları "kaliteli" sorular üzerinde küçük değişiklikler yaparak sisteme girmeye fazlasıyla eğilimli olduklarını gördük.

Bunun üzerine, operatörlerin yazmış oldukları her soruyu, kendi yazdığı diğer sorularla karşılaştırarak, soruların benzerlik hesabını yapan bir algoritma geliştirilmiştir. Bu algoritma gereğince her soru, aynı operatör tarafından yazılmış ve daha önce incelenmiş olan diğer bütün sorularla karşılaştırılarak içlerinde en benzer olduğu soru bulunmuştur. Eğer sorunun en benzer olduğu soru ile olan benzerliği %69'dan 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 3 tane hatalı tasnif işlemi yapıldığı ve bu oranın diğer oranlardan oldukça küçük olduğu gözlemlenmiştir.

Çizelge 2: Operatörlerimizin hazırlamış oldukları sorular ve cevaplara örnekler

Soru:	**Cevap:**
Which is the largest key on the computer keyboard?	space
Which animal is the king of the forest ?	lion
Complete the name of the famous social media website: Face.... ?	book
60 minutes refer to ___	hour

Bu algoritma sonucunda, sistemde bulunan 8842 "kaliteli" sorunun 1322 tanesinin (yaklaşık %15'i) orijinal, kalan 7520 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 tespit edilmiştir. Her ne kadar oluşturulmuş olan soruların büyük bir kısmı tekrar sorulardan oluşsa da, bu şekilde üretilen soruların hiç birisi kalite kontrol motorumuz tarafından otomatik olarak çözülemediğinden, aslında hepsi insan-etkileşim ispatı için kullanılabilecek sorulardır. Operatörlerin girmiş oldukları birkaç soru ve bu soruların cevapları Çizelge 2'de görülebilir.

Diğer bir taraftan, var olan 1322 orijinal soru kabaca incelendiğinde, soruların bir kısmının hatalı oldukları gözlemlenmiştir. Hatalı sorulara örnek olarak "What is your lucky number?" sorusuna verilmiş olan "786" cevabı verilebilir. Bu soru ve cevap tamamen öznel bir yapıda düzenlenmiş olup, operatörlerden istenmiş olan nesnellik kriteri sağlanmamıştır. Bu tarz soruların sisteme karışmasını otomatik olarak engelleyecek bir sistem, bu çalışma konusu olmaktan çıkarak doğal dil işleme alanına girmektedir. Fakat bunun gibi hatalı sorular veritabanında kısıtlı sayıda olduklarında, ki gözlemlerimiz bu yönde, sistem kullanıcıları bir sonraki soruya geçerek İEİ testini çözebilecektir. Buna ek olarak, geliştirilebilecek bir algoritma sayesinde, İEİ testi uygulanan kullanıcıların soruları doğru ya da yanlış çözmelerinin istatistikleri tutularak bu tarz hatalı sorular veritabanından elenebilir.

4.3. Çıkarılan Dersler

Bu bir aylık çalışma açıkça göstermiştir ki, koymuş olduğunuz kurallara veya yapmış olduğunuz uyarılara rağmen insanlar hata yapabilir. Bu sebeple sistemi tasarlarken, insan kaynaklı oluşabilecek hataları minimuma indirecek şekilde tasarlamak gerekmektedir. Bu çalışmada birkaç çeşit insan kaynaklı hata tespit edilmiştir. Örneğin, bir adet "kaliteli" olarak işaretlenmiş soru bulduktan sonra, bu soru üzerinde çok küçük değişiklikler yaparak sisteme girmeleri; soru veya cevabı yazarken yazım hataları yapmaları; "Prepare New Question & Answer" sayfasında cevap yazılacak kısma da soruyu ve soru yazılacak kısma cevabı yazmaları ile soru hazırlarken öznellikten kopamayarak nesnel olamamaları verilebilir.

Bu hataların çoğu, sistem tasarımı değiştirilerek otomatik giderilebilecek hatalardır. Bulunan bir kaliteli sorunun üzerinde küçük değişiklikler yapılarak yeni kaliteli sorular türetilmesini tespit eden sistemi kullanmak zorunda kalışımızdan bahsetmiştik. Bu sistem kolaylıkla kalite kontrol motoru içerisine dahil edilerek aynı hatanın tekrarı engellenebilir. Yazım hatalarını incelemek gerekirse, soruyu yazarken yapılan yazım hataları soruların çözülebilirliğini çok fazla etkilemeyebilir fakat cevabı yazarken yapılan hatalar sistemimizde kritik derecede önem arz etmektedir. Örneğin "Rio is in which country?" sorusunun cevabı İngilizce "Brazil" olması gerekirken "Brasil" olarak verilmiştir. Daha da kötüsü, bu cevap veritabanına yazım hatalı olarak kaydedildiğinden, başka operatörlerin soru girmesi için aktif hale gelmiştir. Bu durumu çözmek için cevap, Bing tercüme motorunda İngilizce'ye çevrilebilir ve bir farklılık varsa operatöre "Yazmak istediğiniz cevap aslında bu mu?" şeklinde sorulabilir. İkinci ve daha kolay bir seçenek ise, operatörlerin yeni cevap girmeleri engellenerek sadece veritabanında bulunan cevaplara soru girmeleri istenmesidir. Bu sayede operatörlerin cevap yazılacak alana soru ve soru yazılacak alana cevap yazmaları problemi de engellenmiş olacaktır.

Daha önce belirtildiği gibi, operatörlerin hazırladıkları soruları öznel değil de nesnel yazmalarını otomatik olarak sağlamak zordur. Bu problemin oluşma sıklığını minimuma indirmek için operatörlerin eğitilmesi, eğitimden sonra başlangıç testinden geçirilmesi ve testten geçebilen operatörlerin hazırladıkları soruların periyodik olarak incelenerek gereken uyarıların yapılması gerekmektedir.

Diğer önemli bir husus ise 100 "kaliteli" soru karşılığı ücret alacak olan operatörlerin daha verimli çalışmış olmalarıdır. Bu sebeple operatörlere yapılacak ödeme operatörün toplam çalışma saati baz alınarak değil, sisteme giriş yaptığı kaliteli soru sayısı baz alınarak yapılmalıdır.

5. Kapanış ve Planlanan Çalışmalar

Yaygın olarak kullanılan grafik tabanlı İnsan Etkileşim İspat (İEİ) sistemlerinin erişilebilirlik ve kullanışlılık problemleri

vardır. Örneğin görme engellilerin grafik tabanlı bir İEİ testini çözebilmesi mümkün değildir. Metin tabanlı İEİ sistemleri bu problemleri ortadan kaldırmakta umut vaat etmekle birlikte, bu sistemler henüz otomatik olarak güvenli biçimde oluşturulamamışlardır. Bu çalışmada, bahsedilen bu güvenlik açığı, testlerin insan hesaplama ile oluşturulması yöntemiyle giderilmeye çalışılmaktadır.

Bu amaçlar doğrultusunda oluşturulan SmarTCHA sistemi [23], bir pilot çalışma yapmak amacıyla, Mayıs 2012 sonuna doğru operatörlerin kullanımına sunulmuştur. Yapılmış olan bu pilot çalışmanın sonucu olarak, 49 operatörün yardımıyla 1 ay gibi kısa sayılabilecek bir sürede yaklaşık 11000 soru üretilmiş ve bu sayede sistemin uygulanabilirliği görülmüştür. Bu pilot çalışmada çıkarılan dersler doğrultusunda sistem güncellenecek ve bu sayede, devam edecek olan çalışmalarda hem İngilizce hem de Türkçe İEİ testlerinin daha verimli biçimde hazırlanmaları sağlanacaktır. Operatörlerin TOBB ETÜ öğrencileri olması ve dönemlerinin yoğun geçmesi sebebiyle bu bir aylık sürecin sonlarına doğru soru giriş miktarı epey azalmış olduğu gözlemlenmiştir. Bu iş üzerinde düzenli çalışabilecek operatörler bulunabildiği takdirde, kısa sürelerde yüksek miktarlarda sorular üretilebileceği düşünülmektedir.

Hazırlanmış olan bu veritabanı kullanılarak, soru ve cevap ikililerinin istemcinin tercihine göre xml veya json formatında gönderildiği bir web servisinin hazırlanması ve kullanıma sunulması planlanmaktadır.

Hazırlanan soruların kullanılabilirliğini ölçmeye yönelik yapılacak kullanıcı çalışmaları da en kısa zamanda planlanarak uygulanacaktır.

İleriki çalışmalarda, veritabanını sadece operatörler çalıştırıp yeni sorular girdirerek genişletmek yerine, yarı-otomatik yöntemler kullanılarak sistemin güvenilirliğini düşürmeden genişletebilmenin yolları da araştırılmaktadır. Veritabanını genişletmek için kullanılabilecek bir başka yol ise, sistem istemcilerine kullandıkları her 5 soru karşılığı 1 soru yazma zorunluluğu getirilmesi olabilir. Veritabanının sadece İEİ testi amacıyla kullanılmasına da gerek yoktur. Örneğin küçük çocukların eğitiminde oyun veya bir yarışma şeklinde sunulabilir.

6. Teşekkür

Bu çalışma 111E148 kodlu Tübitak 1002 projesi kapsamında gerçekleştirilmiş olup, katkılarından dolayı Tübitak'a ve operatör olarak çalışarak projeye katkıda bulunan TOBB ETÜ öğrencilerine teşekkür ederiz.

7. Kaynaklar

[1] A.J. Quinn and B.B. Bederson, "Human computation: a survey and taxonomy of a growing field", In Proceedings of the 2011 annual conference on Human factors in computing systems (CHI), 2011.

[2] Moni Naor. Verification of a human in the loop or identification via the turing test. Available electronically: http://www.wisdom.weizmann.ac.il/_naor/PAPERS/human.ps, 1996.

[3] C Pope and K Kaur. "Is It Human or Computer? Defending E-Commerce with CAPTCHA", IEEE IT Professional, March 2005, pp. 43-49

[4] L. von Ahn, M. Blum, N. J. Hopper, and J. Langford. Captcha: Using hard ai problems for security. In Springer, editor, Eurocrypt, 2003.

[5] B. B. Zhu, J. Yan, Q. Li, C. Yang, J. Liu, N. Xu, M. Yi, K. Cai, Attacks and design of image recognition CAPTCHAs, ACM CCS 2010.

[6] E. Bursztein, M. Martin, J. Mitchell, "Text-based CAPTCHA strengths and weaknesses", Proceedings of the 18th ACM conference on Computer and communications security (CCS), 2011.

[7] K. Chellapilla, K. Larson, P. Simard, M. Czerwinski, "Computers beat Humans at Single Character Recognition in Reading based Human Interaction Proofs (HIPs)", Microsoft Research, 2005. Available electronically: http://web.archive.org/web/20060613111749/http://www.ceas.cc/papers-2005/160.pdf, 11 Temmuz 2012'de erişildi.

[8] Inaccessibility of CAPTCHA, Alternatives to Visual Turing Tests on the Web W3C Working Group Note 23 November 2005. http://www.w3.org/TR/turingtest/ 10 Temmuz 2012'de erişildi.

[9] E. Bursztein, S. Bethard, J. C. Mitchell, D. Jurafsky, and C. Fabry. How good are humans at solving CAPTCHAs? a large scale evaluation. In IEEE S&P '10, 2010.

[10] J. Tam, J. Simsa, S. Hyde, and L. Von Ahn, Breaking Audio CAPTCHAs. Advances in Neural Information Processing Systems. 2008.

[11] E. Bursztein, S. Bethard, Decaptcha: breaking 75% of eBay audio CAPTCHAs, WOOT'09 Proceedings of the 3rd USENIX conference on Offensive Technologies, 2009.

[12] Fatoş Subaşıoğlu, Engellilerin Internet'e Erişimi Üzerine, Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi 40, 3-4 (2000), 203-216.

[13] Philip Brighten Godfrey. Text-based CAPTCHA algorithms. In First Workshop on Human Interactive Proofs, 2002. Unpublished Manuscript. Available electronically: http://www.aladdin.cs.cmu.edu/hips/events/abs/godfreyb_abstract.pdf.

[14] Bartosz Przydatek. On the (im)possibility of a text-only captcha. In First Workshop on Human Interactive Proofs, 2002. Unpublished Abstract. Available electronically: http://www.aladdin.cs.cmu.edu/hips/events/abs/bartosz_abstract.pdf.

[15] C.J. Hernandez-Castro, A. Ribagorda, Pitfalls in CAPTCHA design and implementation: The MATH CAPTCHA, a case study, Computers&Security (Elsevier) Journal, vol.29, p. 141-157, 2010.

[16] Accessible Text CAPTCHA Logic Questions, http://textcaptcha.com/, 10 Temmuz 2012'de erişildi.

[17] TextCaptchaBreaker: Python application that 'breaks' textCAPTCHA, https://github.com/kbhomes/TextCaptchaBreaker, 10 Temmuz 2012'de erişildi.

[18] Peter Gutmann, Engineering Security (Book Draft), http://www.cs.auckland.ac.nz/~pgut001/pubs/book.pdf, 10 Temmuz 2012'de erişildi.

[19] J Elson, JR Douceur, J Howell and J Saul. "Asirra: a CAPTCHA that exploits interest-aligned manual image categorization". Proceedings of the 14th ACM conference on Computer and communications security (CCS), 2007.

[20] Ritendra Datta, Jia Li and James Z. Wang, "IMAGINATION: A Robust Image-based CAPTCHA Generation System", Proceedings of the ACM Multimedia Conference, pp. 331-334, Singapore, ACM, November 2005.

[21] P. Golle, "Machine Learning Attacks Against the Asirra CAPTCHA", Proceedings of the 15th ACM conference on Computer and communications security (CCS), 2008.

[22] Shunting Yard Algorithm, http://en.wikipedia.org/wiki/Shunting_yard_algorithm, 10 Temmuz 2012'de erişildi.

[23] SmarTCHA web sitesi, http://smartcha.net, 10 Temmuz 2012'de erişildi.