

# Pregled i poređenje zaštita od automatskog slanja podataka na veb sajtovima

Željko Gavrić, Vanja Mišković, Dragoljub Pilipović

Fakultet za informacione tehnologije

Slobomir P Univerzitet

Doboj, BiH

zeljko.gavric@spu.ba, vanja.elcic@spu.ba, dragoljub.pilipovic@spu.ba

**Sadržaj**—Ovaj rad opisuje zaštitu od automatskog slanja podataka ka veb sajtovima. Ta zaštita zasniva se na rješavanju *captcha* testa od strane korisnika koji želi da pošalje podatke ka veb sajtu. U radu je objašnjen pojam *captcha* testova kao i područja primjene ovih testova. Napravljena je podjela na tekstualne, grafičke i audio *captcha* testove. Opisane su neke od podvrsta ovih testova, te navedene njihove prednosti i mane. Poređenje *captcha* testova u radu opisano je kroz parametre kvaliteta *captcha* testova.

**Ključne riječi**—*captcha*; automatsko slanje podataka; sigurnost veb sajta;

## I. UVOD

Ekspanzijom Interneta došlo je do povećanja broja veb sajtova, ali i do povećanja broja zlonamjernih korisnika i raznih zlonamjernih programa. Veliki problem za veb sajtove predstavlja automatsko slanje podataka od strane nestvarnih osoba. Nestvarne osobe ustvari predstavljaju računare i računarske programe koji su napravljeni da šalju lažne podatke ka veb sajtovima, a najčešće u cilju njihovog preopterećenja. Da bi veb sajt bio osiguran od automatskog slanja podataka potrebno je da koristi neku vrstu zaštite od automatskog slanja. *Captcha* predstavlja sigurnosni mehanizam koji na osnovu testa treba da utvrdi da li je korisnik koji želi da pošalje podatke ka veb sajtu čovjek ili računar [1]. Riječ *captcha* nastala je od engleskih riječi *Completely Automated Public Turing test to tell Computers and Humans Apart*, što u prevodu znači potpuno automatizovani javni Turingov test za razlikovanje računara i ljudi. Postupak utvrđivanja da li je neko ili nešto čovjek ili računar svodi se na jednostavan test od strane servera, koji bi svaki čovjek trebao da razumije i da riješi, a koji ni jedan računarski program ne može riješiti. Najjednostavniji *captcha* test sastoji se od slike na kojoj je određeni tekst, te koja eventualno sadrži neku jednostavnu pozadinu. Od korisnika se zahtijeva da prepozna tekst sa slike i unese ga u određeno tekstualno polje. Osnovni problem ovakvih *captcha* testova je ekspanzija OCR programa. To su programi koji se koriste za prepoznavanje teksta na slici. Zbog toga je došlo do pojave raznih vrsta *captcha* testova, koji zahtijevaju od korisnika da koristi razne vrste čula, ali i da ima određen nivo opšteg obrazovanja i inteligencije da bi uspješno riješio test.

## II. PRIMJENA CAPTCHA TESTOVA

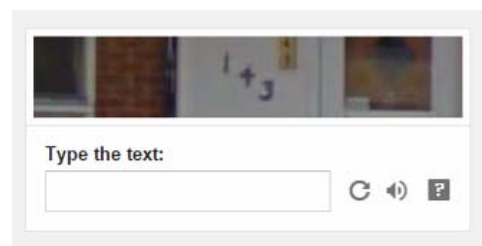
Obzirom da postoje razne vrste napada na veb sajtove *captcha* testovi su našli nekoliko praktičnih primjena. U nastavku su opisane primjene *captcha* testova.

### A. Sprječavanje spam komentara u blogovima

Većina blogera je upoznata sa programima koji ostavljaju lažne poruke na blogovima. Svrha ovih programa je podizanje rejtinga kod pretraživača, ostavljanjem raznih linkova. Korištenjem *captcha* testova kod ostavljanja komentara na blogovima postiže se zaštita od spam poruka. Na ovaj način se ostavlja mogućnost pisanja i objave komentara samo stvarnim korisnicima, bez potrebe za registrovanjem ili prijavljivanjem na blog.

### B. Sprječavanje registracije nestvarnih naloga

Mnogi veb sajtovi imaju probleme sa automatskim pravljenjem naloga od strane računarskih programa. Takav problem najčešće se javljao kod registracije besplatnih *e-mail* naloga, gdje se dešavalo da se za jednu minutu pojavljivalo po nekoliko hiljada naloga. Ovaj problem je riješen korištenjem *captcha* testova koji omogućavaju da samo ljudi koji uspješno riješe test mogu izvršiti registraciju naloga. Na Sl. 1 prikazan je *captcha* test koji koristi Google prilikom kreiranja besplatnog *e-mail* naloga.



Slika 1. *Captcha* test - Gmail

### C. Zaštita e-mail adrese od spamera

Spameri pretražuju internet u potrazi za *e-mail* adresama koje su napisane kao tekst. *Captcha* rješenje za zaštitu *e-mail* adresa sakriva *e-mail* adresu u tekstu, te od korisnika zahtijeva da riješi *captcha* test da bi vidio adresu.

#### D. Zaštita online ankete

Svrha *online* ankete je da na jednostavan način dođe do velikog broja realnih odgovora na postavljena anketna pitanja. Problem nastaje kada se u rješavanje ankete upuste razni programi koji u potpunosti mogu da naruše rezultate ankete. Ovakav problem rješava se korištenjem *captcha* testa, koji korisnik mora da riješi prije nego što odgovori na anketna pitanja.

#### E. Automatsko unošenje lozinki

*Captcha* se takođe koristi i za prevenciju napada na sisteme lozinki. Potrebno je napraviti zaštitu koja onemogućava višestruke iteracije različitih lozinki. Problem se rješava korištenjem *captcha* testa. Da bi se utvrdilo da li čovjek pokušava višestruku iteraciju, nakon određenog broja pogrešnih pokušaja od korisnika se traži da riješi *captcha* test. Ovo rješenje je dosta bolje nego zaključavanje naloga nakon višestrukog unošenja pogrešnih lozinki, jer je kod zaključavanja naloga napadaču omogućeno da namjerno zaključa nalog.

#### F. Zaštita od indeksiranja stranice

Ponekad je potrebno da veb stranica ostane neindeksirana kod pretraživača. To se rješava postavljanjem HTML taga koji će spriječiti softver pretraživača da indeksira sadržaj sa veb stranice. Ovaj tag ne garantuje da softver pretraživača neće pročitati sadržaj veb stranice, jer tag samo daje jednu vrstu instrukcije pretraživačima da ne indeksiraju veb stranice. Obično pretraživači velikih kompanija poštuju želje pojedinih stranica da se njihov sadržaj ne indeksira, ali svakako postoje i oni pretraživači koji se ne obaziru na preporuke. Jedina garancija da pretraživači ne indeksiraju veb stranice je upotreba *captcha* testova, prilikom pokušaja indeksiranja.

#### G. Zaštita od neželjene pošte

*Captcha* testovi nude prihvatljivo rješenje protiv *e-mail* crva i neželjene pošte. Ideja je da se prihvataju samo one *e-mail* poruke za koje je sigurno da su poslone od strane ljudi. Nekoliko velikih *e-mail* servisa rade na uvođenju *captcha* testova u ovu svrhu.

### III. VRSTE CAPTCHA TESTOVA

Tokom vremena korištenja *captcha* testova došlo je do pojavljivanja raznih vrsta testova. Osnovni razlog pojave novih testova je razvoj kompjuterskih algoritama koji uspijevaju da riješe test bez posredstva čovjeka. Važan razlog pojave brojnih testova je i kompleksnost pojedinih testova koja dovodi do neuspjeha pri rješavanju kod nekolicine ljudi. Postoje razne podjele *captcha* testova od strane raznih autora. U [1] *captcha* testovi su podijeljeni u tekstualne, grafičke i audio testove.

#### A. Tekstualni captcha testovi

Tekstualni *captcha* testovi su najjednostavniji za implementaciju [3]. Jedan od često korištenih tekstualnih *captcha* testova je tzv. novinarski pristup. Kod ovog pristupa

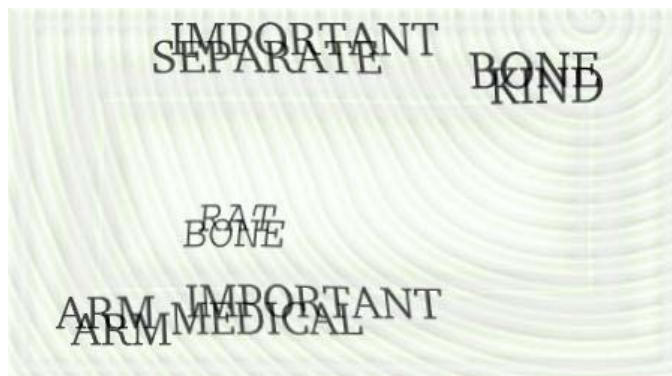
korisniku se postavlja jednostavno tekstualno pitanje na koje treba da odgovori. Primjeri pitanja koja se postavljaju su:

- Koje je treće slovo u riječi Univerzitet?
- Ako je danas ponedjeljak, koji je sutra dan?
- Kolika je razlika između dvadeset i petnaest?

Drugi pristupi tekstualnih *captcha* testova podrazumijevaju da se korisniku prikaže izobličen tekst, koji korisnik treba da prepozna i tekst unese u odgovarajuće tekstualno polje.

##### 1) Gimpy captcha

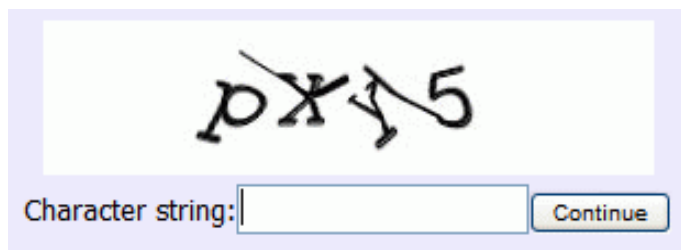
Predstavlja veoma pouzdan *captcha* test. Razvijen je za potrebe Yahoo servisa, u saradnji sa CMU (Carnegie Mellon University). Zasniva se na ljudskoj sposobnosti čitanja izuzetno izobličenog teksta, i nemogućnosti računara da prepozna takav test. Gimpy test zasnovan je na nasumičnom biranju deset riječi iz riječnika, prikazujući ih korisniku na iskrivljen i preklopljen način. Korisnik je sposoban da identifikuje prikazane riječi, dok računarski program nije u stanju da ih prepozna. Potrebno je da korisnik prepozna bar 3 riječi da bi uspješno riješio test. Na Sl. 2 prikazan je izgled Gimpy *captcha* testa.



Slika 2. Primjer Gimpy *captcha* testa [3]

##### 2) Ez-Gimpy captcha

Predstavlja pojestavljenu verziju Gimpy *captcha* testa. Koristi se prilikom prijavljivanja na Yahoo servise. Ez-Gimpy *captcha* test zasnovan je na nasumičnom izboru jedne riječi iz riječnika, te njenom izobličavanju, tako da ne bude prepoznatljiva za računarski program. Sl. 3 prikazuje Ez-Gimpy *captcha* test koji se koristi Yahoo.

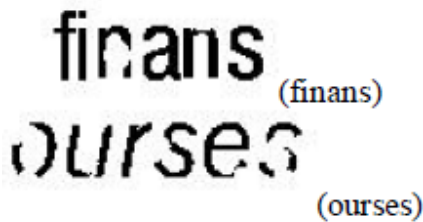


Slika 3. Primjer Ez-Gimpy *captcha* testa - Yahoo

##### 3) Baffle text captcha

Baffle *captcha* test predstavlja varijaciju Gimpy testa, ali za razliku od Gimpy testa, ne sadrži riječnik iz koga se biraju riječi, nego se slučajnim izborom slova kreira nesmislena riječ.

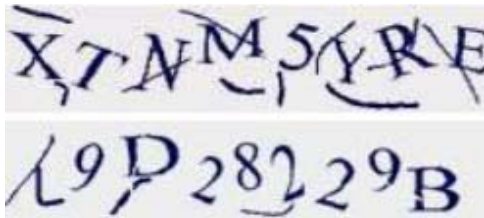
Pažljivim izborom lokacije samoglasnika i suglasnika u riječi postiže se čitljivost prikazane riječi. Na ovaj način korisnik jednostavnije pročita ponuđenu riječ. Ova vrsta testa prevazilazi probleme Gimpy testova sa riječnicima, koje ugrožavaju pametni računarski programi koji skupljaju riječi i kreiraju svoje riječnike. Sl. 4 prikazuje primjer Baffle *captcha* testa.



Slika 4. Primjer Baffle *captcha* testa [3]

#### 4) MSN *captcha* test

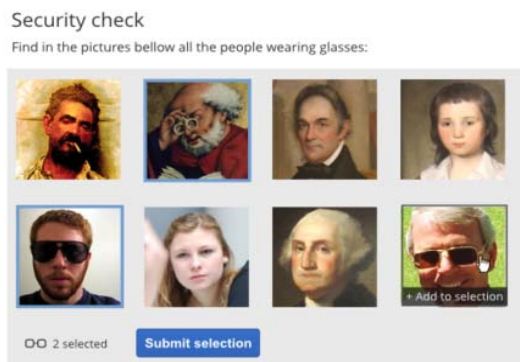
MSN *captcha* test napravljen je za potrebe Microsoft-ovih servisa. Korisniku se prikazuje osam znakova, koji mogu biti velika pisana slova ili brojevi. Znakovi se prije prikazivanja korisniku izobličavaju tako da ne budu lako prepoznatljivi za računarski softver. Pozadina iza znakova je sive boje a slova su tamno plava. Na Sl. 5 prikazan je primjer MSN *captcha* testa.



Slika 5. Primjer MSN *captcha* testa [3]

### B. Grafički *captcha* testovi

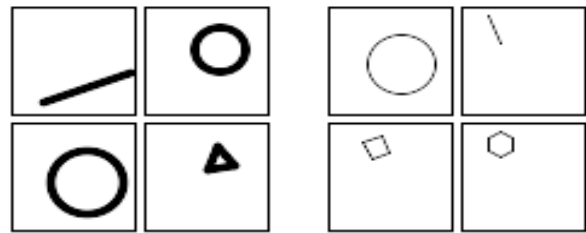
Nastankom i razvojem OCR alata koji se uspješno koriste za rješavanje pojedinih tekstualnih *captcha* testova, pojavili su se grafički testovi. Kao i kod tekstualnih *captcha* testova i kod grafičkih postoje razne vrste testova. Na Sl. 6 prikazan je jedan primjer grafičkog *captcha* testa u kome se od korisnika traži da izabere slike na kojima su osobe koje nose naočale. Korisnik treba da klikne na tražene slike i da potvrdi unos.



Slika 6. Primjer Grafičkog *captcha* testa

#### 1) Bongo *captcha* test

Bongo *captcha* predstavlja test koji je zasnovan na prepoznavanju grafičkih šablona. Korisniku se prikazuju dva bloka, lijevi i desni. Blokovi se sastoje od sličica na kojima se nalaze određeni oblici. Korisnik treba da prepozna razliku između desnog i lijevog bloka. Kada korisnik prepozna razliku, prikazuje mu se nekoliko sličica sa različitim oblicima. Od korisnika se tada traži da prepozna kojoj strani pripada svaka od ponuđenih sličica, lijevoj ili desnoj. Na Sl. 7 prikazan je primjer Bongo blokova, koji se sastoje od po četiri sličice sa određenim oblicima. Od korisnika se zahtijeva da uvidi razliku između ova dva bloka, a razlika je u debljini linija. Nakon ovog korisniku se nude oblici sa podebljanim i nepodebljanim linijama, te za svaki od oblika korisnik treba da odabere lijevu ili desnu stranu. Dakle, podebljane oblike na lijevu stranu, a nepodebljane na desnu.



Slika 7. Primjer Bongo *captcha* testa

#### 2) PIX *captcha* test

PIX *captcha* test koristi veliku bazu fotografija i animiranih slika koje se sreću u svakodnevnom životu [5]. Korisniku se nudi grupa slika koje su u osnovi povezane sa istim objektom ili konceptom. Korisnik treba da identifikuje šta je to zajedničko za sve slike i da iz padajućeg menija izabere pojam koji opisuje sve slike. Na Sl. 8 prikazan je primjer slika kod PIX *captcha* testa. Korisnik na osnovu malih slika u testu treba da identifikuje loptu kao zajednički koncept svih slika.



Slika 8. Primjer slika istog koncepta kod PIX *captcha* testa

### C. Audio *captcha* testovi

Da bi slijepe i slabovide osobe bile u prilici da uspješno riješe *captcha* test, razvijeni su audio *captcha* testovi. Audio *captcha* slučajnim odabirom bira riječ ili broj iz riječnika. Riječ se rastavlja na slova, te se slovo po slovo čita korisniku. Broj se

rastavlja na cifre koje se čitaju korisniku. Tokom čitanju slova ili cifara u izgovor se ubacuje distorzija kako zvuk ne bi mogao biti prepoznat od strane računara. Kod ove vrste testova koristi se sposobnost čula sluha da iz distorziranog glasa raspozna slova ili brojeve. Osnovna mana ove vrste *captcha* testova je to što korisnik mora da razumije jezik i akcenat izgovora u kome se reprodukuje audio snimak.

#### D. ReCaptcha test

ReCaptcha predstavlja vrstu testa koja najviše podsjeća na tekstualne testove. Svojim brojnim doprinosima i svrsishodnošću mnogi autori, sa pravom, predstavljaju ovu vrstu *captcha* testa odvojeno od ostalih. ReCaptcha predstavlja koncept koji osim što razlikuje osobe od računara, ima i dodatne mogućnosti. Te mogućnosti se svode na pomoć pri digitalizovanju knjiga i tekstova. Naime, rješavajući ovaj tip *captcha* testa, korisnik pomaže u digitalizaciji starih knjiga ili članaka. Korisniku se nudi riječ iz članka ili knjige, koju alat za prepoznavanje teksta nije uspješno prepoznao. Rješavanjem ovog pojma riječ se uspješno digitalizuje. Da bi program bio siguran da je čovjek uspješno prepoznao riječ, koju ni on sam ne razumije, od njega se traži da unese dvije riječi. Kako je jedna od riječi poznata, tj. ranije je digitalizovana, na osnovu nje se određuje da li je korisnik čovjek, te ako korisnik uspješno unese poznatu riječ, nepoznata riječ se tada pretvara u poznatu. Drugim riječima pretpostavlja se da će korisnik koji pročita i unese jednu riječ ispravno, pročitati i unijeti i drugu riječ ispravno. Zahvaljujući velikom broju *captcha* testova koji se svakodnevno rješavaju, dolazi se do informacije da se svaki dan na ovaj način uspješno digitalizuje nekoliko miliona riječi [6]. Na Sl. 9 prikazan je primjer ReCaptcha testa.



Slika 9. Primjer ReCaptcha testa

ReCaptcha posjeduje i mogućnost izgovora riječi koje korisnik treba da unese. Ova činjenica je takođe dovela do toga da se ReCaptcha testovi ne mogu svrstati u tekstualne testove.

#### IV. POREĐENJE CAPTCHA TESTOVA

U prethodnom dijelu rada navedene su razne vrste *captcha* testova. Svaki od tih testova ima svoje prednosti i nedostatke. Da bi se testovi mogli porediti prvo je potrebno definisati kriterijume po kojima se mogu porediti. Neki od kriterijuma za poređenje su:

- Razumljivost
- Brzina rješavanja
- Težina rješavanja za računar
- Uspješnost rješavanja testova od strane korisnika
- Postojanje tajne baze podataka - riječnika

TABELA I. POREĐENJE CAPTCHA TESTOVA

	Tekstualni test				Grafčki test		Audio test	ReCaptcha test
	Gimpy	Ez-Gimpy	Baffle	MSN	Bongo	PIX		
<b>Razumljivost</b>	Dobra – mala distorzija	Prihvatljiva – povećana distorzija	Prihvatljiva – dodatni elementi preko teksta	Prihvatljiva – dodatni elementi preko teksta	Prihvatljiva Jednostavan [8]	Prihvatljiva Subjektivno [8]	Dobra	Prihvatljiva
<b>Brzina rješavanja</b>	Potrebno unijeti 3 riječi	Potrebno unijeti 1 riječ	Potrebno unijeti 1 riječ	Potrebno unijeti 8 znakova	Potrebno identifikovati razliku između blokova i odrediti kom bloku pripada svaka od 4 slike	Potrebno je identifikovati sa kojim objektom je povezana grupa slika	Potrebno je slušati izgovorenu riječ ili cifre te ih unijeti	Potrebno je unijeti 2 riječi
<b>Teško rješavanje od strane računara</b>	Ne	Ne	Da	Da	Ne [8]	Da [8]	Ne	Ne
<b>Prezentovanje testa korisniku</b>	Preklapanje slova	Distorzija slova	Distorzija uz preklapanje sa šablonima	Zakrivljenje, dodavanje dodatnih elemenata i dodavanje pozadine	Prikaz jednostavnih šablona	Prikaz jasnih slika	Čitanje znakova uz malu distorziju	Distorzija i precrtavanje riječi
<b>Tajna baza podataka (teksta, slika, snimaka)</b>	Da	Da	Ne	Ne	Da	Da	Da	Ne

U tabeli I prikazano je poređenje *captcha* testova koji su opisani u ovom radu, po definisanim kriterijumima. Neke od ovih kriterijuma nije moguće jasno obrazložiti, obzirom da su subjektivnog karaktera, tj. razlikuju se od čovjeka do čovjeka.

Tekstualni testovi se po mnogim referencama navode kao testovi koje računarski programi najlakše rješavaju. Ova konstatacija je postala činjenica pojavom OCR alata koji u velikom broju slučajeva, baš kao i čovjek rješavaju zadani test. Povećanjem broja testova koji su riješeni od strane računarskih alata, razvijali su se i tekstualni *captcha* testovi. Njihov razvoj uglavnom se ogledao u preklapanju slova i povećanju distorzije, tj. iskrivljenosti slova koja se prikazuju korisniku. Iskrivljenost je išla do te mjere da su pojedini testovi postajali neprepoznatljivi ljudima. Drugi način za poboljšanje tekstualnih *captcha* testova napravljen je kod Baffle i MSN testova, tako što se umjesto riječi iz riječnika, kao kod Gimpy testova, znakovi postavljaju nasumično te predstavljaju nesmislene riječi. Naime kada se koristi riječnik za prikaz riječi koje je potrebno identifikovati, postoji velika mogućnost prikupljanja riječi iz riječnika od strane zlonamjernih računarskih programa, koje će kasnije biti iskorištene za automatsko rješavanje testova.

Za razliku od tekstualnih *captcha* testova, grafički i audio testovi su teži za prepoznavanje od strane računarskih programa. Problem kod ove dvije vrste testova predstavlja teža implementacija nego kod tekstualnih. Grafički testovi su vizuelno dosta prihvatljiviji od tekstualnih, međutim neki od njih zahtijevaju poznavanje engleskog jezika, da bi se moglo razumjeti šta je potrebno uraditi da bi se test uspješno riješio. U posljednje vrijeme dosta sajtova prelazi sa tekstualnih na grafičke testove.

Audio *captcha* nudi nezaobilaznu pomoć slijepim i slabovidim osobama kod korištenja servisa koji zahtijevaju razlikovanje čovjeka i računara. Ipak, audio *captcha* predstavlja problem za ljude kojima engleski jezik, na kome je najčešće realizovana audio *captcha*, nije poznat. Postoje na sreću i audio *captcha* testovi koji podržavaju više jezika. Jedan od najpoznatijih audio *captcha* testova koji podržava izgovor na više jezika je e-Bay *captcha*. E-Bay *captcha* test korisnici rješavaju sa uspjehom od 95% [7]. Kao i kod tekstualnih testova i kod audio testova problem predstavljaju razne vrste softvera koji su napravljeni za prepoznavanje zvuka.

## V. ZAKLJUČAK

Zaštita od automatskog slanja podataka je svakako potrebna na svim sajtovima gdje je omogućeno slanje podataka od strane

korisnika. *Captcha* testovi djelimično rješavaju problem lažnog predstavljanja računara u svojstvu čovjeka. Kao i svaki sistem, tako su i *captcha* testovi pretrpjeli mnogobrojne izmjene i proširenja od svog nastanka do danas. Međutim, pred sistemima za zaštitu od automatskog slanja podataka postavljeni su teški zadaci, koji svakim danom postaju sve teži. Težina koju će *captcha* testovi morati da podnesu u budućnosti svodi se na svakodnevna poboljšanja računarskih programa i algoritama koji su u svojstvu prepoznavanja teksta, slika i zvuka.

## LITERATURA

- [1] Roshanbin, N., Miller, J., "ADAMAS: Interweaving unicode and color to enhance CAPTCHA security", Future Generation Computer Systems, 2014.
- [2] Von Ahn, L., Blum, M., Langford, J., "Telling humans and computers apart automatically", Communications of the ACM, 47(2) 2004, pp. 56-60.
- [3] Choudhary, S., Saroha, R., Dahiya, Y., Choudhary, S., "Understanding Captcha: Text and Audio Based Captcha with its Applications", International Journal of Advanced Research in Computer Science and Software Engineering, 3(6), 2013.
- [4] Chew, M., Baird, H. S., "Baffletext: A human interactive proof", International Society for Optics and Photonics, 2003, pp. 305-316.
- [5] Pope, C., Kaur, K. "Is it human or computer? Defending e-commerce with captchas", IT professional, 7(2), 2005, pp. 43-49.
- [6] Zvanični ReCaptcha sajt, [www.recaptcha.net](http://www.recaptcha.net), pristupano 15.01.2015.
- [7] Soupionis, Y., Dimitris G., "Audio CAPTCHA: Existing solutions assessment and a new implementation for VoIP telephony", Computers & Security, 29(5), 2010, pp. 603-618.
- [8] Zhu, B. B., et al., "Attacks and design of image recognition CAPTCHAs.", Proceedings of the 17th ACM conference on Computer and communications security, ACM, 2010, pp. 187-200.

## ABSTRACT

The paper discusses the protection of the automatic data sending to websites. This protection is based on solving the captcha test by the user who wants to send the data. The paper explains the concept of captcha tests and the areas in which they can be used. They are divided into textual, graphic and audio captcha tests. We have described several subtypes of these tests, and mentioned their advantages and disadvantages. The comparison of captcha tests is seen through the captcha tests quality parameters.

## REVIEW AND COMPARE THE PROTECTION OF THE AUTOMATIC SENDING DATA ON WEB SITES

Željko Gavrić, Vanja Mišković, Dragoljub Pilipović