SVEUČILIŠTE U ZAGREBU
GRAFIČKI FAKULTET

SEMIR RESIMOVIĆ

METODA PRILAGODBE ENCIKLOPEDIJSKOGA SADRŽAJA INTERNESKOM IZDANJU

DIPLOMSKI RAD

Zagreb, 2017.

SEMIR RESIMOVIĆ

METODA PRILAGODBE ENCIKLOPEDIJSKOGA SADRŽAJA INTERNESKOM IZDANJU

DIPLOMSKI RAD

Mentor:
Prof. dr. sc. Nikola Mrvac

Student:
Semir Resimović

Zagreb, 2017.

SAŽETAK

Postoji potreba da se enciklopedijska građa prenese u digitalni oblik i da na novi način prezentiranja sadržaja učini enciklopedijsku građu kako zanimljivijom tako i dostupnu većem broju korisnika. Cilj ovog diplomskoga rada je pojasniti jednu od metoda prilagođavanja enciklopedijske građe za internetsko izdanje te upoznati čitatelje s problemima u postupku i njihovo rješavanje. U empirijskom dijelu bit će pojašnjeno koji programski alati su korišteni i na koji način. Navedeni su rezultati analize podataka o posjetiteljima prikupljenih putem servisa Google Analytics za Proleksis enciklopediju.

KLJUČNE RIJEČI: enciklopedijski sadržaj, online enciklopedija, konverzija podataka

ABSTRACT

The need to convert content of encyclopedia into digital form and the need to make presenting of that content more interesting and more accessible to a bigger number of users is present. The goal of this research is to explain one of the many methods of adjusting encyclopaedic content for online edition and to show readers problems in the process and the way of solving them. In the empiric part it will be explained which program tools are used and how to use them. The results of the data analysis about visitors collected via Google Analytics servis Proleksis encyclopedia are listed.

KEYWORDS: content of encyclopedia, online encyclopedia, data conversion

SADRŽAJ

UVOD

Danas svi koristimo internet kao sredstvo informiranja, učenja i zabave. Sve tekstualne i slikovne informacije koristimo a da se ne zapitamo kako su došle tamo gdje jesu. Većina korisnika će na to pitanje odgovoriti da je to netko napisao ili metodom kopiraj i zalijepi (copy-paste) uzeo iz nekog digitalnog izvora i objavio na nekoj internetskoj stranici ili blogu. Da, to se zaista tako i radi, ali to je primjenjivo na kraćim tekstovima koji ne zahtijevaju dodatno dorađivanje. Znamo da su na internetu cijele enciklopedije sa obilnom ilustrativnom građom. Ako gledamo opseg sadržaja koje nudi enciklopedija onda to može biti dugotrajan posao.

A što su zapravo enciklopedije? Prema Hrvatskoj enciklopediji „…Enciklopedije su djela u kojim se, abecednim ili kakvim drugim metodičkim slijedom, okupljaju i sustavno obrađuju činjenice i spoznaje o svim ljudskim znanjima…“[1]. Konverzija takve građe u internetsko izdanje predstavlja veliki izazov kako zbog svog obujma tako i zbog svoje kompleksnosti.

Nakon ekspanzije internetskih tehnologija, prema Nikoli Mrvcu „… Enciklopedije su među prvima bile te koje je trebalo prilagoditi i ponuditi korisnicima putem nterneta…“ .[2] Internetske enciklopedije nisu više novost. Tako inozemni izdavači poput *Encyclopaedia Britannica* su svoju digitalnu enciklopediju predstavili na internetu 2012., a poznata njemačka enciklopedija *Brockhaus* je svoju internetsku inačicu predstavila još 2007. Ove enciklopedije su komercijalne što znači da puni sadržaj natuknica nije dostupan besplatno.[3]

Za razliku od ovih najpoznatija besplatna i otvorena enciklopedija je *Wikipedija* koja je startala sa svojim radom 15. siječnja 2001. Wikipedija je započeta kao projekt u čijem su uređivanju mogli svi sudjelovati, ali je to imalo implikacije na pouzdanost podataka. Pored tih problema Wikipedija je ostala jedna od najvećih enciklopedija do danas s više od deset milijuna natuknica na svim jezicima.[4]

Hrvatska ne zaostaje za drugim svjetskim izdavačima i također sudjeluje u projektu digitalizacije kulturne i znanstvene građe, tako su se pojavile i domaće enciklopedije kao *Proleksis enciklopedija* koja se na internetu pojavila u 2012[1]. kao otvorena besplatna online

[1] Pojavila se godinu ranije ali je bila limitirana na Carnetove korisnike, što se prelaskom 2012. na domenu Leksikografskoga zavoda Miroslav Krleža promijenilo te je postala javno dostupna.

enciklopedija. Nedugo zatim pojavljuje se i *Hrvatska enciklopedija,* koja je po opsegu sadržaja natuknica veća od Proleksis enciklopedije.

Do sada je grafička struka bila u potpunosti uključena u proces izvedbe tiskane inačice jedne enciklopedije. Pojavom internetskih izdanja grafička struka evolvira u drugom smjeru iako ostaje potreba za klasičnim tiskom. Danas su grafičari i dalje važan dio u procesu izvedbe internetskih enciklopedija, jer aktivno sudjeluju u dizajniranju stranica, određivanju tipografije, obradbi ilustrativne građe i pripremi audio vizualnih dodataka. Možemo reći da su grafičari ključna karika u lancu konverzije enciklopedija iz inačice pripremljene za tisak u internetsku inačicu. Novi način rada donosi i svoje probleme i metode rješavanja istih, teško je doći do metoda kojim su drugi izdavači svoje enciklopedije konvertirali za internetsko izdanje pa je u ovom radu predstavljena jedna metoda kao plod višemjesečnoga razvoja i istraživanja.

Hipoteze:

- Sa današnjom tehnologijom je moguće digitalizirati tiskane enciklopedije i prilagoditi enciklopedije pripremljene za tisak za potrebe prikaza na internetu.
- Metoda koja se koristi zahtijeva tehničko predznanje.
- Metoda koja se koristi je interdisciplinarna.
- Broj korisnika digitalne inačice enciklopedije ima tendenciju rasta.
- Enciklopedije na internetu će zamijeniti tiskane inačice.

1. TEORIJSKI DIO

1.1. Anatomija enciklopedija pripremljenih za tisak

Enciklopedije su jednosveščana a nerijetko i višesveščana izdanja. Najmanja jedinica enciklopedije je natuknica koja može biti neilustrirana ili ilustrirana s jednom ili više ilustracija. Natuknica može sadržavati od jednog do više stotina redaka. U natuknicama se mogu pojaviti uputnice. Na primjer „Å → angstrem“ je uputnica koja nam govori da je simbol „Å“ šire pojašnjen u natuknici „angstrem“, a simbol „→“ ima značenje „vidi pod …“.

Enciklopedija može brojati od nekoliko stotina do nekoliko desetaka tisuća natuknica, ilustrativna građa također može dosezati broj od više tisuća.

Kada se priprema enciklopedijski tekst za tiskano izdanje uzima se više parametara u obzir: Svezak ne smije biti s prevelikim brojem stranica iz praktičnih i financijskih razloga

Praktični:

a) Lakše je manipulirati s knjigom koja ima manji broj stranica.
b) Čvršći je uvez u hrptu kod sveska s manjim brojem stranica.
c) Može se uporabiti deblji papir kako bi se izbjegla preduboka penetracija boje u papir kada se otiskuju veće obojene površine pa time i pojava tamnih područja na poleđini papira.

Financijski:

a) Manje utrošenog papira.
b) Manje utrošene boje.
c) Kraće vrijeme izvedbe, manja ukupna cijena rada. Što u konačnici daje jeftiniji pojedinačni svezak za krajnjeg kupca.

Iz navedenih razloga dolazi do intervencija u tekstu već na razini kreiranja natuknica:

a) Koriste se kratice kako bi što više teksta stalo u zadani format.
b) Tipografija koja se koristi često je u graničnim veličinama čitljivosti (7-9 pt)
c) Ograničava se broj ilustracija kako se ne bi povećao broj stranica, ponekad se radi selekcija ilustracija a u nekim slučajevima se čak moraju izostaviti.

1.2.Prednost internetskih u odnosu na tiskana izdanja

1.2.1. Tiskana izdanja

a) Naklada tiskanih izdanja

Poznato je da su tiskana izdanja limitirana nakladom koja se kreće od 1000 do 5000. Ovime smo već suzili broj korisnika, jedan dio naklade ide u knjižnice, dio po ustanovama a dio ide krajnjem kupcu.

b) Mala dostupnost

Broj svezaka u opticaju je ograničavajući faktor pristupu informacijama, dvije osobe ne mogu koristiti isti svezak u isto vrijeme. Ako nemate enciklopediju onda morate posuđivati iz knjižnice.

c) Kad je otisnuta već je zastarjela

Podatci koji su u tiskanom obliku se ne mogu ažurirati, često se izvode zadnji urednički zahvati prije samoga tiska kako bi podatci bili što svježiji. Problem ažuriranja se rješava tiskanjem dopunskih svezaka no izdavanje istih nije frekventno.

d) Broj ilustracija

Ilustracije u tiskanom izdanju su limitirane na razumnu količinu i prilagođenih su dimenzija iz razloga da ne povećaju broj stranica.

1.2.2. Internetska izdanja

a) Pretraživanje

Višestruko brže pronalaženje željenog podatka.

b) Dostupnost

Građa dostupna u svako doba putem računala ili mobilnih uređaja.

c) Novi sadržaji

Oplemenjivanje građe novim sadržajima (audio-vizualnim elementima) što u tiskanim izdanjima nije moguće.

d) Ažurnost podataka

Gotovo u realnom vremenu je moguće napraviti izmjenu nekog podatka. Na primjeru biografskih natuknica: datum smrti, osvojena nagrada, novo djelo itd.

e) Povezanost

Interna povezanost natuknica i povezanost s drugim internetskim izdanjima i referencama.

f) Interaktivni elementi

Uvođenje interaktivnih elemenata u svrhu edukacije kao što su npr. kvizovi.

g) Filtriranje podataka

Prikaz filtriranih podataka, npr. izdvajanje struke iz ukupne građe – književnost.

1.3. Višejezičnost u tekstovima enciklopedija

Strana imena i nazivi zastupljeni u tekstovima enciklopedija se pišu u izvornom obliku ali se daje i pojašnjenje kako se neka strana riječ izgovara. Tamo gdje je potrebna transliteracija onda se i primjenjuje, na primjer izvornici pisani devanagrijem (क = ka) ili katakanom (ベッド =*beddo*).

1.3.1. Računalna tipografija

Tehnologija pripreme za tisak se mijenjala pa tako i način prikaza slovnih znakova drugih jezika. Od olovnog do računalnog sloga oblik znakova se nije mijenjao samo se pojednostavnio način na koji se slaže tekst. Danas je vrlo lako napraviti tekst veličine 12,2345 pt što je nekada bilo nemoguće jer su se matrice radile na propisane tipografske veličine 4 pt, 5 pt, 6 pt itd. Moguće je u trenu promijeniti font, rez i veličinu pisma.

Na drugu stranu novi način unosa teksta je otvorio jedan drugi problem prikaza slovnih znakova pojedinih jezika. Kada je priprema teksta predviđena za tisak onda je svejedno na koji način se dobije neki znak drugog jezika ako takav nije u izabranom fontu (kombiniranjem znakova i akcenata ili nekim drugim načinom). Taj znak je samo interpretacija, ali ako takav tekst ide na WEB stranicu onda je jako bitno kako se znakovi drugih jezika unose i na koji način. Prilikom pretraživanja nekog pojma pisat ćete ga onako kako se očekuje na primjer ako je traženi pojam „Müller“ tada ćete u tražilicu i upisati Müller no ako je u tekstu posebno upisan slovni znak „u“ i nakon njega umlaut „¨“ pa je smanjeno doslagivanje između u i umlauta, optički izgledaju kao jedan slovni znak. To će sustav i dalje tumačiti kao dva slovna znaka, što znači „Mu¨ller“ neće biti u rezultatima pretrage. Stoga je nužno imati gotove slovne znakove s akcentima unutar fonta kojim se unosi tekst.

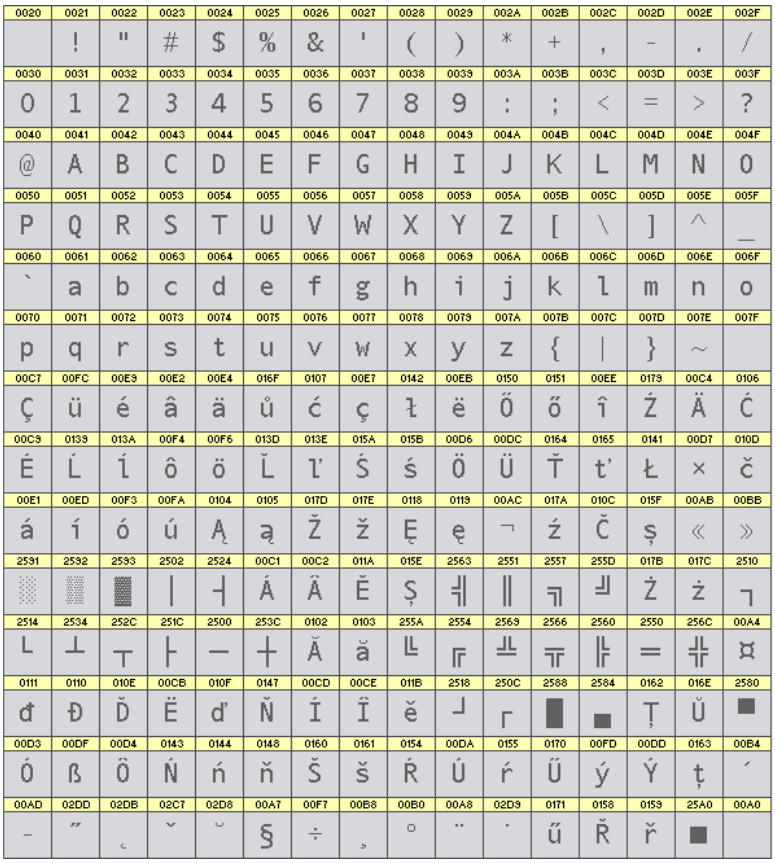
Font je skup znakova jednog pismovnog reza, broj znakova unutar fonta varira. To ovisi o dizajnu fonta, ponekad je u fontu samo skup znakova osnovne latinične abecede i to samo verzali. Nije nužno da u fontu budu slovni znakovi, u fontu se mogu naći simboli, interpunkcija, matematički znakovi itd. Bilo je više pokušaja kako bi se uveo standard po kojemu bi raspored slovnih znakova bio jednoznačno definiran. Za sada su to encoding setovi

po kojima se organizirano i sustavno razmještaju slovni znakovi jezika, skupine simbola i ostalih znakova.

1.3.2. Kodne stranice

U početku je bio standard za razmjenu informacija (ASCII) koji je bio 7-bitni i sadržavao slovne znakove engleske abecede a-z verzale i kurente, brojeve 0-9 te osnovnu interpunkciju.

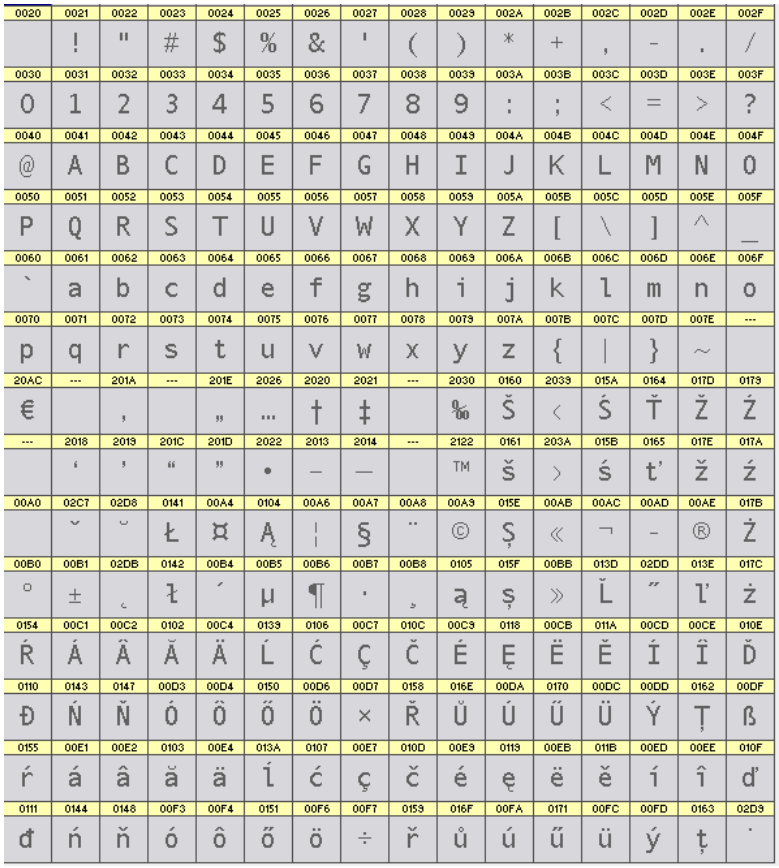
Kako bi se mogli koristiti znakovi drugih jezika onda se standard proširio na 8-bitni kako bi dao podršku i drugim jezicima osim engleskom. Prema operacijskim sustavima postojale su više inačica standarda tako je za MS-DOS i za starije inačice Windowsa[2] ako ste željeli imati hrvatske znakove morali koristiti kodnu stranicu „852 Latin 2“. Kako izgleda raspored možemo vidjeti na slici 1.



Slika 1. Raspored slovnih znakova u kodnoj stranici „852 Latin 2“ za MS-DOS

[2] Inačice Windows 3, 3.1, 3.11 su bile grafičko sučelje MS-DOS operativnog sustava pa je stoga i naslijeđen princip kodnih stranica. Windows 95 i 98 su se još naslanjale na MS-DOS kao glavni operativni sustav ali s nekim preinakama u korištenju kodnih stranica.

Na Windowsima novije generacije[3] korištena stranica Windows-1252. Ona ima malo drugačiji raspored što se vidi na slici 2.



Slika 2. Raspored slovnih znakova u kodnoj stranici „Windows 1252“ za Windows platforme.

1.3.3. UNICODE standard[4]

Stari način podrške raznim jezicima bio je organiziran kao takozvane kodne stranice koje su određivale kako će se koji znak tumačiti. Kodne stranice su bile podijeljene na datoteke pa je time djelomično riješen problem višejezičnosti. Fontovi su mogli sadržavati određeni broj slovnih znakova i još uvijek nedovoljan da se pokriju sve potrebe. Taj problem je uspješno riješen uspostavom UNICODE standarda.

Međunarodna neprofitna organizacija UNICODE konzorcij je uveo standardizaciju sustava kodiranja gdje se smještaju slovni znakovi pojedinih jezičnih skupina.[5] Taj standard se koristi od 1991. do sada se razvio do inačice 10.

[3] Windows XP, Vista, 7, 8, 10 su platforme koje su pravi operacijski sustavi koji ne trebaju MS-DOS.
[4] UNICODE standard rasporeda slovnih znakova unutar fonta, URL: http://www.unicode.org/standard/standard.html (pristupano 14. I. 2017.)

Zanimljivost ovog standarda je što ne samo da rješava problem prikaza slovnih znakova određenih jezičnih skupina nego je ujedno riješen problem sortiranja.

Standard se ne bavi samo živim jezicima, zastupljena su pisma jezika koji su van svakodnevne uporabe npr. Glagoljica je dobila svoje mjesto u UNICODE standardu u inačici 5.1

Da bi mogli pristupiti znakovima koji su spremljeni u tom standardu koristimo se sustavom kodiranja za određenu potrebu, jezik, skupinu znakova itd. Što znači da ako izrađujemo font sa hrvatskim znakovima za UNICODE standard onda slovne znakove hrvatskoga jezika stavljamo u UNICODE područje od 0100-017F prema slici 3.

Slika 3. Smještaj znakova hrvatske latinice unutar UNICODE područja 0100-017F

UNICODE standard je omogućio da jedan font može sadržavati 1.114.111 (10FFFF) znakova što je dostatno za slovne znakove svih jezika, simbole raznih vrsta, interpunkcije itd. To područje je podijeljeno na razine njih ukupno 17. Nulta razina koja sadrži 65.535 znakova (0000–FFFF) je osnovna višejezična razina BMP (Basic Multilingual Plane).[6]

1.4.Digitalizacija

Digitalizacija je prema redakciji Hrvatske enciklopedije „… u najširem smislu, prevođenje analognoga signala u digitalni oblik…“.[7]

Svi sadržaji koji se nalaze na internetskim stranicama, a koji potječu iz tiskanih medija, su na neki način prošli digitalizaciju. Digitalizacijom dobivamo mogućnost da sadržaje distribuiramo, pretražujemo, i sačuvamo od propadanja. Više nije upitno da li je potrebno provoditi digitalizaciju, nego kojim tehnikama i tehnologijom je kvalitetno provesti u što kraćem vremenu. Velika sredstva i napori se ulažu u proces digitalizacije kulturne baštine. Na kongresu bibliotekara, Dunja Seiter-Šverko i Lana Križaj u svom predstavljanju teme su navele par zanimljivih činjenica koje govore u prilog tom nastojanju da se građa sustavno digitalizira i da to postane nacionalna strategija.[8]

1.4.1. Metode digitalizacije

Digitalizaciju možemo provoditi na razne načine koji se mogu grupirati na dvije skupine manualna i automatska.

1.4.1.1. Manualna metoda

Osim što zahtjeva veće ljudske resurse od ostalih, ova metoda zahtjeva najviše vremena. Metoda se zasniva na prepisivanju teksta uz pomoć računala. Metoda je korištena u vrijeme kada su stolni skeneri bili rijetkost i njihova namjena je bila prvenstveno za obradbu ilustracija za tisak. Programska podrška u vidu aplikacija za optičko prepoznavanje teksta nije postojala.

1.4.1.2. Poluautomatska i automatska digitalizacija

Od ovog postupka očekuje se velika brzina obradbe podataka i mali broj operanada koji opslužuju sustav. Za ovakvu metodu mogu se koristiti: mobilni uređaj s kvalitetnom kamerom, plošni skener, fotografski aparat i specijalizirani skener odnosno uređaj posebno konstruiran za digitalizaciju.

Knjigu s velikim brojem stranica možemo skenirati na više načina:

a) Skeniranje bez razrezivanja knjige je sporije i donekle destruktivna metoda jer se knjiga mora otvoriti do kraja kako bi stranice „nalegle“ na staklo skenera, primjer na slici 4, što kod starih knjiga dovodi do pucanja hrpta. Metoda je sporija i zahtijeva stalni angažman operatera. Ovo je najjeftiniji način skeniranja što se tiče opreme ali je dugoročno neisplativ.

Slika 4. Uvezana knjiga se nedovoljno otvara za skeniranje plošnim skenerom

b) Razrezivanje knjige po hrptu će omogućiti skeniranje knjige u plošnom skeneru koji ima ADF (Automatic Document Fider) tj. jedinicu za automatsko ulaganje dokumenata. Izgled postupka skeniranja razrezane knjige je na slici 5.

Slika 5. Skeniranje razrezane knjige uz pomoć ADF dodatka na skeneru Epson.

Veća količina stranica se može obraditi u kraćem roku. Metoda je destruktivna i nije primjenjiva na vrijednim unikatnim izdanjima.

c) Najmanje destruktivna i najbrža metoda je korištenje specijaliziranih sustava za skeniranje koji se sastoje od stalka za knjigu digitalnih fotografskih aparata visoke razlučivosti prozirne planarne površine za izravnavanje stranica i mehaničkoga sustava za prelistavanje knjige. Na slici 6 je prikazan jedan eksperimentalni sustav velike brzine skeniranja i OCRiranja skoro u realnom vremenu razvijen u laboratoriju Ishikawa Watanabe Laboratory na Tokijskom univerzitetu.[9]

Slika 6. Eksperimentalni OCR sustav sa brzinom okretanja stranica od 250 u minuti

Iz navedenih prijedloga je moguće zaključiti da je najoptimalniji i najbrži specijalizirani sustav za digitalizaciju međutim ulaganja su znatna i potrebna je procjena isplativosti. Moguće je spojiti nekoliko priručnih komponenti i dobiti vlastiti sustav za digitalizaciju sa relativno malim ulaganjima. U eksperimentalnom dijelu će biti naveden jedan primjer.

Kada se digitalizacija provodi klasičnim skenerima onda se mogu koristiti dvije vrste skenera, jedna vrsta nešto jeftinijih koji imaju veću brzinu skeniranja. To ćemo izabrati ako je bitno iz knjige izdvojiti samo tekst bilo ona ilustrirana ili ne. Drugu vrsta skenera koja može dati kvalitetan izlaz je preporučljivo koristiti u slučaju kada nam je bitno da digitaliziramo ilustrativnu građu.

OCR programi znaju prepoznati i raščlaniti tekst od ilustracija. Optimalna granica razlučivosti za kvalitetan OCR oko 400 dpi ali ta razlučivost nije dostatna za ilustracije koje će biti korištene u tisku. Da bi bile podobne za tisak ilustracije moraju imati najmanju razlučivost 600 dpi da bi se mogle doraditi u nekom programu za obradbu ilustracija. Kvalitetniji skeneri to mogu izvesti međutim brzina skeniranja opada. Stoga je preporučljivo skenirati knjigu za potrebe OCR na nižoj rezoluciji i nakon toga skenirati ilustracije većom rezolucijom.

1.5. Prezentiranje podataka putem WEB stranica

1.5.1. Baze podataka

Praksa je pokazala da se s velikim količinama podataka najlakše manipulira ako su spremljeni u neku bazu. Baze omogućavaju brz pristup podatcima, pretraživanje, sortiranje, filtriranje po raznim kriterijima itd. Postoje razna izvedbena rješenja za spremanje podataka u bazu poznatija u računalnom žargonu kao engini. Neki od engina koriste SQL naredbeni jezik, neki koriste Java a neki Python script jezik. Među najraširenijim enginima je MySQL zahvaljujući tome što je besplatan i ima mogućnost manipulacije velikim brojem podataka.

Podatci unutar baze su organizirani putem tablica. Tablice se sastoje od zapisa (records). Zapis se sastoji od polja. Jedan zapis može sadržavati jedno ili više polja, npr.

Tablica: podatci

ID	Sadržaj
1	podatak prvi
2	drugi podatak

U ovom primjeru „ID“ i „Sadržaj“ su nazivi polja dok su „1 podatak prvi“ i „2 drugi podatak“ zapisi. Pojedina polja se indeksiraju kako bi engine mogao brzo pronaći željeni podatak.

1.5.2. HTML i CSS

Za prikaz podataka na nekoj internetskoj stranici nije dovoljno imati bazu. Podatke treba prikazati kao uređenu WEB stranicu. Uređena WEB stranica podrazumijeva da se podatci moraju „obući“ u HTML[5] i CSS[6] kod. HTML meta jezik služi tome da se elementima na WEB stranici odredi točna struktura i hijerarhija a CSS je skup definicija stilova za te elemente.

Kako to funkcionira? Želimo neki tekst prikazati kao naslov onda moramo ispred teksta i iza teksta postaviti oznake (tagove) <h1> za početak naslova i </h1> za kraj naslova.

```
<h1>Ovo je naslov</h1>
```

[5] Akronim od engleskoga naziva HyperText Markup Language, URL: https://www.w3.org/html/ (pristupljeno 14. I. 2017.)

[6] Akronim od engleskoga naziva Cascading Style Sheets URL: https://www.w3.org/Style/CSS/ (pristupljeno 14. I. 2017.)

Ovako pripremljen tekst svi internetski preglednici će protumačiti na isti način tj. kao naslov (heading 1). Ako se koristimo CSSom onda dodavanjem parametara za oznaku „h1“ precizno definiramo način prikaza naslova. Što to konkretno znači slijedi u primjeru. Želimo da sve što je označeno oznakom „h1“ bude prikazano pismom od 12 pt i crvene boje onda u moramo to napisati na slijedeći način:

h1 {font-size: 12pt; color: red}

Ova definicija stila može biti spremljena unutar html dokumenta unutar oznaka <style></style> ili kao poseban css dokument koji se poziva putem reference unutar osnovnog html dokumenta.

1.5.3. Serverska i korisnička strana internetske stranice

Dolazimo do novog dijela koji uvodi pojmove serverska strana i korisnička strana. Serverska strana je onaj dio koji se izvodi na samom serveru a to su izvedba pretraživanja vađenja podataka iz baze, dodavanje HTML i CSS koda podatcima koji se žele prikazati. Za ovakve intervencije na podatcima koristimo se nekim od script jezika kao što su Rubi, PHP, Java Script.

Korisnička strana je onaj dio koji se izvodi na mobilnim uređajima ili desktop računalima krajnjih korisnika. Korisnička strana obuhvaća interakciju s HTML elementima, pokretanje i interpretaciju video ili audio elemenata. U tu svrhu, na strani klijenata, obično se koriste Java Script, JQuery programski jezici.

1.5.4. CMS sustav

Vidjeli smo da za prikaz podataka na internetskoj stranici trebamo različitu programsku podršku. Ovako objedinjena programska podrška se zove sustav za upravljanje sadržajem ili Content Management System (CMS). U uporabi je više gotovih CMS sustava koji se razlikuju po kompleksnosti, mogućnostima proširivanja i po vrsti primjene.

Koji CMS odabrati (Drupal, Joomla!, WordPress, October...). Zaista nije bitno, bitno je da dobro poznajete izabrani CMS tj. njegove mogućnosti.

Ako razvijate svoj sustav onda ste pred velikim zadatkom da dostignete razinu funkcionalnosti nekog gotovog CMS sustava i pred još većim da napravite bolju funkcionalnost. Nerijetko je manji broj programera uključen u takav razvoj.

Gotovi CMS sustavi i još ako su otvorenoga koda, su daleko bolji izbor. Uvijek su dobro dokumentirani i podrška je dobra.

Ovo nikako ne znači da se ne smiju razvijati vlastita rješenja za neki CMS sustav, to ovisi s kojim vremenom, sredstvima i brojem angažiranih članova razvoja raspolažete. Za neka brza rješenja procjena ide u korist nekog gotovog sustava. Ako na tržištu ne postoji sustav po mjeri naručitelja onda se prilazi problemu sa vlastitim rješenjima.

2. EKSPERIMENTALNI DIO

U ovom dijelu će biti predstavljen postupak, od analize ulaznih podataka, izbora aplikacija, mogući problemi i kako ih riješiti i na kraju sâm postupak prebacivanja obrađenih natuknica u bazu i konačna analiza posjeta WEB stranicama.

2.1. Izvorna građa za konverziju

Ovo je bitan dio analize kako bi mogli odlučiti kojom opremom i programskim alatima treba izvršiti konverziju.

2.1.1. Tiskano i uvezano

Kao što je prije navedeno tiskana i uvezana građa mora proći proces digitalizacije.

Kao jednu od mogućnosti kojom se možemo poslužiti je jeftino priručno rješenje prikazano eksperimentom.

2.1.1.1. Digitalizacija – eksperiment

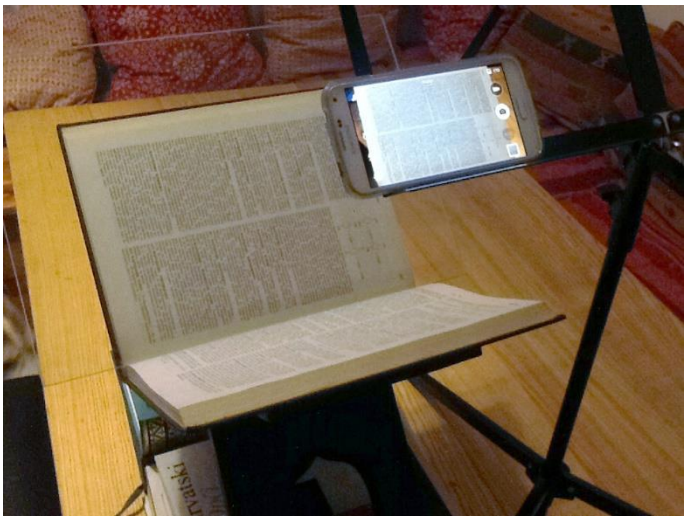
Današnji mobilni uređaji imaju ugrađene kamere velike razlučivosti, riječ je o redu veličine iznad 8 MP što je za digitalizaciju teksta dovoljno. Postoji razlika u tehničkim specifikacijama kamera kod mobilnih uređaja. U tablici1Tablica 1 navode se samo nekoliko pametnih telefona radi dobivanja uvida o kakvim je razlikama riječ.

Tablica 1. Tehničke specifikacije kamera pametnih telefona - uzorak

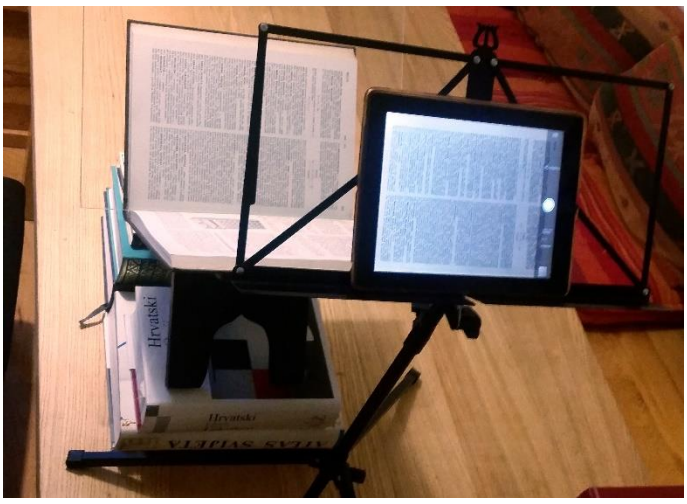
				
Naziv modela	HTC Desire 550	Huawei Y3 2017	Apple iPhone 7	Samsung Galaxy S8+ Exynos
Model senzora	-	-	Sony Exmor RS	Sony IMX333 Exmor RS
Tip senzora	CMOS BSI	CMOS	CMOS	CMOS
Otvor objektiva	f/2.4	f/2.0	f/1.8	f/1.7
Žarišna duljina	-	-	3.99 mm	4.2 mm
Tip bljeskalice	LED	LED	-	Dual LED
Razlučivost slike	3264 x 2448 pixels 7.99 MP	3264 x 2448 pixels 7.99 MP	4032 x 3024 pixels 12.19 MP	4032 x 3024 pixels 12.19 MP

Postoje i dodatne leće za mobilni uređaj koje se jednostavnom klipsom (kvačicom) pričvrste za mobilni uređaj na kameru. Ovaj sustav mobilni uređaj + dodatna leća nije

profesionalna oprema za digitalizaciju i može poslužiti kao priručno rješenje. Jednostavan sustav za digitalizaciju knjiga moguće sastaviti bez velikih financijskih ulaganja. Potrebni su stalak za knjigu stalak za mobilni uređaj, staklo koje služi za izravnavanje stranice knjige. Primjeri kako izgledaju sastavljeni priručni studiji za digitalizaciju knjiga na slici 7 i 8



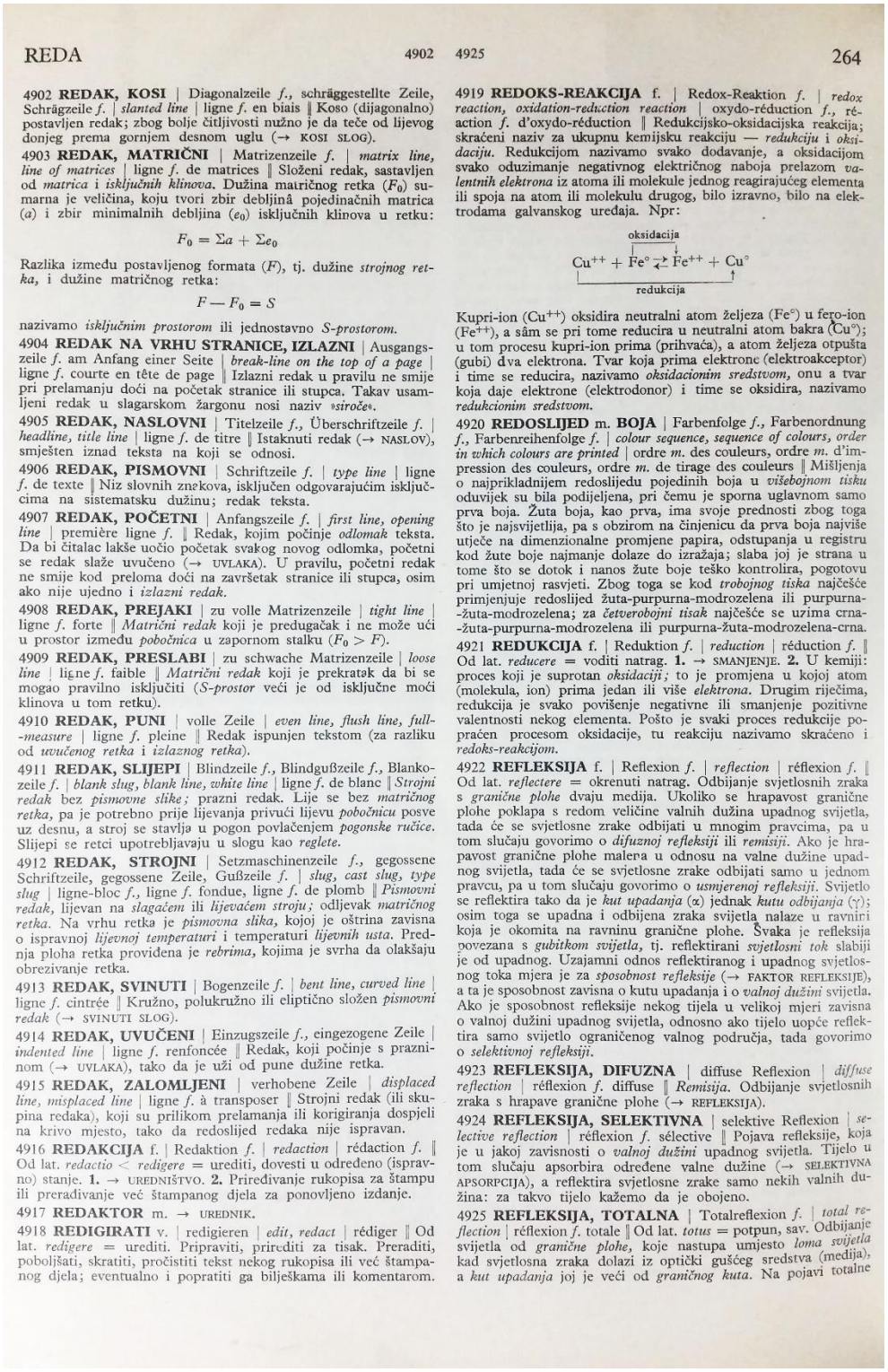
Slika 7. Priručno rješenje, digitalizacija mobilnim uređajem – smartphone Samsung Galaxy S6.



Slika 8. Priručno rješenje, digitalizacija mobilnim uređajem - tablet iPad 2

Uzeta je za primjer nasumična stranica iz Grafičke enciklopedije. Postupak zahtjeva slijedeće korake: postavljanje knjige u stalak, određivanje stranice koja se skenira, postavljanje stakla za poravnavanje stranice, snimanje. Trajanje cijele procedure je do 10 s dok bi za postupak uz pomoć skenera trajao do 30 s. Dobili smo trostruko brži postupak od standardnog skeniranja. Fotografija je zadovoljavajuće kvalitete za potrebe OCR programa. Ovim postupkom smo izbjegli problem iskrivljenog teksta prema hrptu. Inače taj problem uspješno rješava OCR aplikacija, taj korak smo ovdje izbjegli što u konačnici ubrzava

postupak OCRiranja. Kako izgleda nekorigirana fotografija dobivena ovim postupkom vidimo na slici 9.



REDA 4902

4902 **REDAK, KOSI** | Diagonalzeile *f.*, schräggestellte Zeile, Schrägzeile *f.* | *slanted line* | ligne *f.* en biais ‖ Koso (dijagonalno) postavljen redak; zbog bolje čitljivosti nužno je da teče od lijevog donjeg prema gornjem desnom uglu (→ KOSI SLOG).

4903 **REDAK, MATRIČNI** | Matrizenzeile *f.* | *matrix line, line of matrices* | ligne *f.* de matrices ‖ Složeni redak, sastavljen od *matrica* i *isključnih klinova*. Dužina matričnog retka (F_0) sumarna je veličina, koju tvori zbir debljinâ pojedinačnih matrica (a) i zbir minimalnih debljina (e_0) isključnih klinova u retku:

$$F_0 = \Sigma a + \Sigma e_0$$

Razlika između postavljenog formata (F), tj. dužine *strojnog retka*, i dužine matričnog retka:

$$F - F_0 = S$$

nazivamo *isključnim prostorom* ili jednostavno *S-prostorom*.

4904 **REDAK NA VRHU STRANICE, IZLAZNI** | Ausgangszeile *f.* am Anfang einer Seite | *break-line on the top of a page* | ligne *f.* courte en tête de page ‖ Izlazni redak u pravilu ne smije pri prelamanju doći na početak stranice ili stupca. Takav usamljeni redak u slagarskom žargonu nosi naziv »siroče«.

4905 **REDAK, NASLOVNI** | Titelzeile *f.*, Überschriftzeile *f.* | *headline, title line* | ligne *f.* de titre ‖ Istaknuti redak (→ NASLOV), smješten iznad teksta na koji se odnosi.

4906 **REDAK, PISMOVNI** | Schriftzeile *f.* | *type line* | ligne *f.* de texte ‖ Niz slovnih znakova, isključen odgovarajućim isključcima na sistematsku dužinu; redak teksta.

4907 **REDAK, POČETNI** | Anfangszeile *f.* | *first line, opening line* | première ligne *f.* ‖ Redak, kojim počinje *odlomak* teksta. Da bi čitalac lakše uočio početak svakog novog odlomka, početni se redak slaže uvučeno (→ UVLAKA). U pravilu, početni redak ne smije kod preloma doći na završetak stranice ili stupca, osim ako nije ujedno i *izlazni redak*.

4908 **REDAK, PREJAKI** | zu volle Matrizenzeile | *tight line* | ligne *f.* forte ‖ *Matrični redak* koji je predugačak i ne može ući u prostor između *pobočnica* u zapornom stalku ($F_0 > F$).

4909 **REDAK, PRESLABI** | zu schwache Matrizenzeile | *loose line* | ligne *f.* faible ‖ *Matrični redak* koji je prekratak da bi se mogao pravilno isključiti (*S-prostor* veći je od isključne moći klinova u tom retku).

4910 **REDAK, PUNI** | volle Zeile | *even line, flush line, full--measure* | ligne *f.* pleine ‖ Redak ispunjen tekstom (za razliku od *uvučenog retka* i *izlaznog retka*).

4911 **REDAK, SLIJEPI** | Blindzeile *f.*, Blindgußzeile *f.*, Blankozeile *f.* | *blank slug, blank line, white line* | ligne *f.* de blanc ‖ *Strojni redak* bez *pismovne slike;* prazni redak. Lije se bez *matričnog retka*, pa je potrebno prije lijevanja privući lijevu *pobočnicu* posve uz desnu, a stroj se stavlja u pogon povlačenjem *pogonske ručice*. Slijepi se retci upotrebljavaju u slogu kao *reglete*.

4912 **REDAK, STROJNI** | Setzmaschinenzeile *f.*, gegossene Schriftzeile, gegossene Zeile, Gußzeile *f.* | *slug, cast slug, type slug* | ligne-bloc *f.*, ligne *f.* fondue, ligne *f.* de plomb ‖ *Pismovni redak*, lijevan na *slagaćem* ili *lijevaćem stroju;* odljevak *matričnog retka*. Na vrhu retka je *pismovna slika*, kojoj je oštrina zavisna o ispravnoj *lijevnoj temperaturi* i temperaturi *lijevnih usta*. Prednja ploha retka predviđena je *rebrima*, kojima je svrha da olakšaju obrezivanje retka.

4913 **REDAK, SVINUTI** | Bogenzeile *f.* | *bent line, curved line* | ligne *f.* cintrée ‖ Kružno, polukružno ili eliptično složen *pismovni redak* (→ SVINUTI SLOG).

4914 **REDAK, UVUČENI** | Einzugszeile *f.*, eingezogene Zeile | *indented line* | ligne *f.* renfoncée ‖ Redak, koji počinje s prazninom (→ UVLAKA), tako da je uži od pune dužine retka.

4915 **REDAK, ZALOMLJENI** | verhobene Zeile | *displaced line, misplaced line* | ligne *f.* à transposer ‖ Strojni redak (ili skupina redaka), koji su prilikom prelamanja ili korigiranja dospjeli na krivo mjesto, tako da redoslijed redaka nije ispravan.

4916 **REDAKCIJA** f. | Redaktion *f.* | *redaction* | rédaction *f.* ‖ Od lat. *redactio* < *redigere* = urediti, dovesti u određeno (ispravno) stanje. **1.** → UREDNIŠTVO. **2.** Priređivanje rukopisa za štampu ili prerađivanje već štampanog djela za ponovljeno izdanje.

4917 **REDAKTOR** m. → UREDNIK.

4918 **REDIGIRATI** v. | redigieren | *edit, redact* | rédiger ‖ Od lat. *redigere* = urediti. Pripraviti, prirediti za tisak. Preraditi, poboljšati, skratiti, pročistiti tekst nekog rukopisa ili već štampanog djela; eventualno i popratiti ga bilješkama ili komentarom.

4925 264

4919 **REDOKS-REAKCIJA** f. | Redox-Reaktion *f.* | *redox reaction, oxidation-reduction reaction* | oxydo-réduction *f.*, réaction *f.* d'oxydo-réduction ‖ Redukcijsko-oksidacijska reakcija; skraćeni naziv za ukupnu kemijsku reakciju — *redukciju* i *oksidaciju*. Redukcijom nazivamo svako dodavanje, a oksidacijom svako oduzimanje negativnog električnog naboja prelazom *valentnih elektrona* iz atoma ili molekule jednog reagirajućeg elementa ili spoja na atom ili molekulu drugog, bilo izravno, bilo na elektrodama galvanskog uređaja. Npr:

Kupri-ion (Cu^{++}) oksidira neutralni atom željeza (Fe°) u fero-ion (Fe^{++}), a sâm se pri tome reducira u neutralni atom bakra (Cu°); u tom procesu kupri-ion prima (prihvaća), a atom željeza otpušta (gubi) dva elektrona. Tvar koja prima elektrone (elektroakceptor) i time se reducira, nazivamo *oksidacionim sredstvom*, onu a tvar koja daje elektrone (elektrodonor) i time se oksidira, nazivamo *redukcionim sredstvom*.

4920 **REDOSLIJED** m. **BOJA** | Farbenfolge *f.*, Farbenordnung *f.*, Farbenreihenfolge *f.* | *colour sequence, sequence of colours, order in which colours are printed* | ordre *m.* des couleurs, ordre *m.* d'impression des couleurs, ordre *m.* de tirage des couleurs ‖ Mišljenja o najprikladnijem redoslijedu pojedinih boja u *višebojnom tisku* oduvijek su bila podijeljena, pri čemu je sporna uglavnom samo prva boja. Žuta boja, kao prva, ima svoje prednosti zbog toga što je najsvijetlija, pa s obzirom na činjenicu da prva boja najviše utječe na dimenzionalne promjene papira, odstupanja u registru kod žute boje najmanje dolaze do izražaja; slaba joj je strana u tome što se dotok i nanos žute boje teško kontrolira, pogotovu pri umjetnoj rasvjeti. Zbog toga se kod *trobojnog tiska* najčešće primjenjuje redoslijed žuta-purpurna-modrozelena ili purpurna--žuta-modrozelena; za *četverobojni tisak* najčešće se uzima crna--žuta-purpurna-modrozelena ili purpurna-žuta-modrozelena-crna.

4921 **REDUKCIJA** f. | Reduktion *f.* | *reduction* | réduction *f.* ‖ Od lat. *reducere* = voditi natrag. **1.** → SMANJENJE. **2.** U kemiji: proces koji je suprotan *oksidaciji;* to je promjena u kojoj atom (molekula, ion) prima jedan ili više *elektrona*. Drugim riječima, redukcija je svako povišenje negativne ili smanjenje pozitivne valentnosti nekog elementa. Pošto je svaki proces redukcije popraćen procesom oksidacije, tu reakciju nazivamo skraćeno i *redoks-reakcijom*.

4922 **REFLEKSIJA** f. | Reflexion *f.* | *reflection* | réflexion *f.* ‖ Od lat. *reflectere* = okrenuti natrag. Odbijanje svjetlosnih zraka s *granične plohe* dvaju medija. Ukoliko se hrapavost granične plohe poklapa s redom veličine valnih dužina upadnog svijetla, tada će se svjetlosne zrake odbijati u mnogim pravcima, pa u tom slučaju govorimo o *difuznoj refleksiji* ili *remisiji*. Ako je hrapavost granične plohe manja u odnosu na valne dužine upadnog svijetla, tada će se svjetlosne zrake odbijati samo u jednom pravcu, pa u tom slučaju govorimo o *usmjerenoj refleksiji*. Svijetlo se reflektira tako da je *kut upadanja* (α) jednak *kutu odbijanja* (γ); osim toga se upadna i odbijena zraka svijetla nalaze u ravnini koja je okomita na ravninu granične plohe. Svaka je refleksija povezana s *gubitkom svijetla*, tj. reflektirani *svjetlosni tok* slabiji je od upadnog. Uzajamni odnos reflektiranog i upadnog svjetlosnog toka mjera je za *sposobnost refleksije* (→ FAKTOR REFLEKSIJE), a ta je sposobnost zavisna o kutu upadanja i o *valnoj dužini* svijetla. Ako je sposobnost refleksije nekog tijela u velikoj mjeri zavisna o valnoj dužini upadnog svijetla, odnosno ako tijelo uopće reflektira samo svijetlo ograničenog valnog područja, tada govorimo o *selektivnoj refleksiji*.

4923 **REFLEKSIJA, DIFUZNA** | diffuse Reflexion | *diffuse reflection* | réflexion *f.* diffuse ‖ *Remisija*. Odbijanje svjetlosnih zraka s hrapave granične plohe (→ REFLEKSIJA).

4924 **REFLEKSIJA, SELEKTIVNA** | selektive Reflexion | *selective reflection* | réflexion *f.* sélective ‖ Pojava refleksije, koja je u jakoj zavisnosti o *valnoj dužini* upadnog svijetla. Tijelo u tom slučaju apsorbira određene valne dužine (→ SELEKTIVNA APSORPCIJA), a reflektira svjetlosne zrake samo nekih valnih dužina: za takvo tijelo kažemo da je obojeno.

4925 **REFLEKSIJA, TOTALNA** | Totalreflexion *f.* | *total reflection* | réflexion *f.* totale ‖ Od lat. *totus* = potpun, sav. Odbijanje svijetla od *granične plohe*, koje nastupa umjesto *loma svijetla* kad svjetlosna zraka dolazi iz optički gušćeg sredstva (medija), a *kut upadanja* joj je veći od *graničnog kuta*. Na pojavi totalne

Slika 9. Nekorigirana fotografija dobivena snimanjem pomoću smartphonea Samsung Galaxy 6.

2.1.1.2. OCR programska podrška

Pretvorba skeniranog predloška u tekst koji se može uređivati na računalu se provodi s OCR aplikacijama. U tablici 2 dati su usporedni podatci o mogućnostima OCR aplikacija njihov cjenovni razred i na kojim platformama rade.

Tablica 2. Usporedba OCR aplikacija

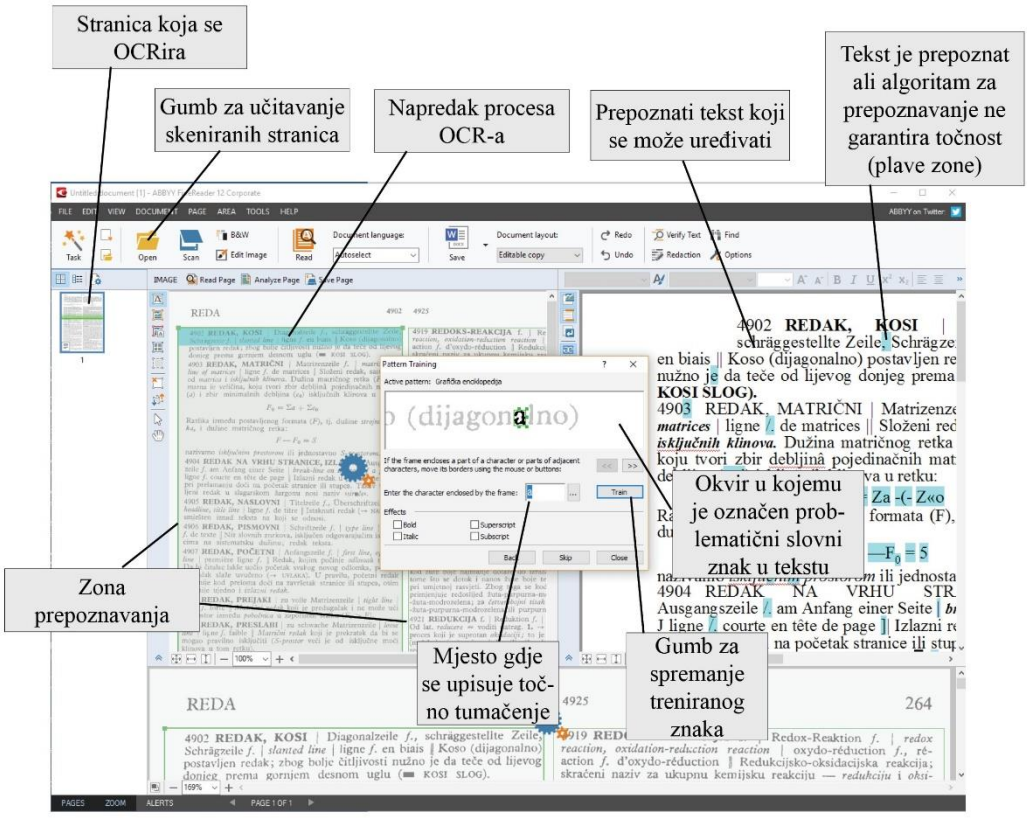
	OmniPage Standard	**Presto! OCR**	**Microsoft OneNote**	**PDF Transformer**	**Adobe OCR**	**ABBYY Fine Reader**
Cijena	$149.99	$89.95	Free	$79.99	$299.00	$143.99
Zadržava izgled	✓	✓	✓		✓	✓
Zadržava font	✓	✓		✓	✓	✓
Formatiziranje u pretraživi PDF	✓	✓		✓	✓	✓
Provjera teksta	✓	✓				
Konvertiranje tablica	✓	✓				✓
Prepoznatih jezika	120	40	52	184	60	198
Windows	✓	✓	✓	✓	✓	✓
Mac OS X		✓			✓	✓

Pravilan izbor aplikacije skratit će vrijeme obradbe, za potrebe enciklopedije prednost ima Abbyy Fine Reader zbog broja jezika koje može prepoznati.

Aplikacija je koncipirana tako da se u nju učitavaju fotografije dobivene skeniranjem ili fotografiranjem. Aplikacija ima poseban algoritam preko kojeg procesira stranicu ispitujući razne anomalije i pokušava ih eliminirati kao micanje šuma, ispravljanje iskrivljenog retka pri hrptu koji se događa pri skeniranju plošnim skenerom, rotiranje stranice koje se može dogoditi pri ulaganju stranice u skener. Očišćena stranica se nakon toga analizira kako bi se prepoznala koja područja sadrže tekst a koja ilustracije te zone aplikacija označi okvirima i numerira ih tako da nama bude jasno kojim slijedom će aplikacija obrađivati označene zone. Zone se mogu prenumerirati ako je to potrebno ili čak ukloniti pojedinačne ili ako nismo zadovoljni kako je to algoritam izračunao onda ih možemo proizvoljno označiti.

Ako pustimo da aplikacija radi OCR onda će uzeti generički algoritam za prepoznavanje slovnih znakova što u nekim slučajevima daje zadovoljavajuće rezultate pogreška od svega

nekoliko slovnih znakova po stranici. Kada algoritam OCR prepozna slova onda se dodatno analizira tekst preko rječnika kako bi se mogao odrediti jezik i uz njegovu pomoć dodatno smanjile pogreške u čitanju. Ako je i pored ovih provjera puno pogrešaka u prepoznavanju onda je potrebno načiniti „trening“. Trening je postupak kojim se OCR aplikacija uči kako da čita određene znakove. Trening se radi za svaku knjigu posebno i postupak se treba provesti na najmanje 1 stranici, po potrebi se može naknadno nastaviti trening. Primjer treninga je prikazan na slici 10.



Slika 10. Trening OCR-a u aplikaciji ABBYY Fine Reader

Nakon završenog treninga se može pustiti da aplikacija izvrši OCR na učitanim skeniranim stranicama. Ako je rezultat dobar prepoznati tekst se može spremiti u neki od popularnih formata: običan tekst bez zadržavanja oblika (bold, italic), Microsoft Word document format sa zadržavanjem formata i oblika (tekst u dva stupca sa svim njegovim karakteristikama veličinom pisma, bold, italic, eksponent, indeks itd.).

OCRirani tekst se dalje obrađuje tako da se učita u aplikaciju za prijelom. Današnje aplikacije za prijelom imaju mogućnost da se gotovi prijelom može jednostavno prilagoditi za tisak, elektroničku publikaciju ili za WEB. Dvije kompanije koje su za sada najbolje u tom

području su Adobe (InDesign) i Quark inc. (QuarkXPress). Izbor je prema afinitetima korisnika u ovom radu je korišten Adobeov InDesign.

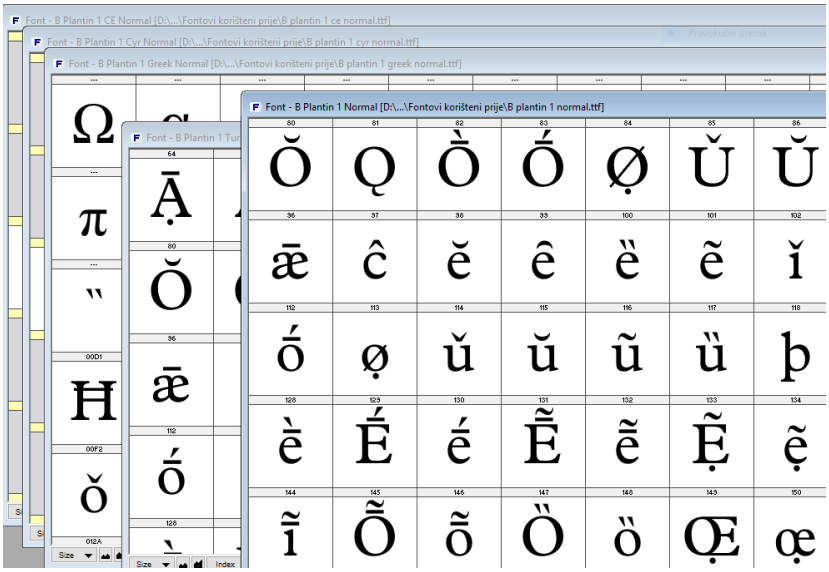
Razlog zbog kojeg se preporuča učitavanje teksta u program za prijelom a ne u program za obradbu teksta jer programi za prijelom imaju jednostavniji izlazni HTML format.

2.1.2. Prijelom na računalu

Ako je izvornik za konvertiranje u internetsko izdanje gotovi prijelom onda je postupak puno kraći od prethodnog jer nema digitalizacije.

2.1.2.1. Prijelom na računalu koji nije u UNICODE standardu

Starije inačice aplikacija kao što su QuarkXpress 3, 4, 5 Zatim InDesign 1, 1.1 nisu imale podršku za UNICODE standard. Fontovi koji su imali potpuni set znakova su bili rijetki, a postojeći se nisu mogli elegantno doraditi postojećim font aplikacijama. Izrađivani su zato pseudo multilingualni fontovi koji su imali do 255 znakova i u tom rasponu su bili postavljeni znakovi koji bi inače pokrivali neku kodnu stranicu UNICODE-a. Tako da ako ste nekome trebali predati prijelom morali ste obavezno predati i posebno prilagođene fontove kako bi se prijelom ispravno prikazao. Na primjer pseudo multilingvalni sustav je bio sastavljen od šest zasebnih datoteka koje su činile jednu cjelinu i to za potrebe europskih jezika, zatim dvije datoteke za proširenu latinični skup znakova, zatim jedna datoteka za IPA standard (izgovorni standard). Za svako pismo koje je korišteno u prijelomu bilo je nužno napraviti ovakav set datoteka, primjer izgleda rasporeda prikazan je na slici 11.



Slika 11. Font čini više datoteka, na primjeru „B Plantin 1“ otvorene datoteke u aplikaciji FontLab

Takvi prijelomi nisu se mogli koristiti za direktnu konverziju. Najprije je potrebno remapirati stari raspored znakova u UNICODE standard.

Za potrebe ove prilagodbe nužno je načiniti tablicu konverzije. To se jednostavno može izvesti u VBA[7] skript jeziku koji se izvršava unutar aplikacije Microsoft Word.

Osnovna ideja skripta je da kroz zadani tekst ide znak po znak, analizira kojemu fontu pripada i koja mu je kodna vrijednost zatim da ga zamijeni znakom drugog fonta iz UNICODE sustava.

Drugim riječima potrebno je imati font sa svim znakovima koji su vam potrebni i sa ispravnim UNICODE rasporedom da biste mogli prekodirati tekst.

Zašto je ovo bitno? Jednostavno – ako bi ostali znakovi na starim pozicijama i kada se podatci učitaju u bazu onda će se dogoditi krivo sortiranje podataka. Pretraživanje će biti otežano a u nekim slučajevima neće biti nikakvog rezultata u pretraživanju

Postupak se sastoji od eksportiranja teksta iz InDesigna u datoteku RTF formata. Zatim otvorimo eksportirani sadržaj u aplikaciji Microsoft Word, pokrenemo skript za remapiranje. Nakon završenog remapiranja tekst ponovno učitamo u InDesign.

2.1.2.2. VBA script za remapiranje slovnih znakova

Izvorni kôd se sastoji od više procedura:

a) Glavna procedura: „main“
b) Pomoćne procedure: „A_Znakovi“, „A_IzmjenaKoda“, „Grcki_dopune“, „Mat_Symbol“, „Plantin“, „B_Plantin_1“, „B_Plantin_1_Baltic“, „B_Plantin_1_CE“, „B_Plantin_1_Cyr“, „B_Plantin_1_Greek“, „B_Plantin_1_Tur“, „B_Plantin_Baltic“, „B_Plantin_CE“, „B_Plantin_Cyr“, „B_Plantin_Greek“, „B_Plantin_Tur“, „B_Simbol“ i „B_Transkripcija“

[7] Akronim od engleskoga naziva Visual Basic for Aplications, URL: https://msdn.microsoft.com/en-us/library/office/gg264383.aspx (pristupano 14. I. 2017.)

Procedura „main“ je startna procedura koja poziva druge

```
Sub main()
    Dim atEnd As Boolean
    While Not atEnd
    DoEvents
    If Selection.Type = wdSelectionIP And Selection.End =
ActiveDocument.Content.End - 1 Then atEnd = True
    Selection.MoveRight Unit:=wdCharacter, Count:=1, Extend:=wdExtend
       A_Znakovi AscW(Selection()), Selection.Font.Name
    Selection.MoveRight Unit:=wdCharacter, Count:=1
    StatusBar = Selection.Information(wdFirstCharacterColumnNumber)
    Wend
End Sub
```

Sub i End Sub su početak i kraj svake procedure u VBA scriptu. Dim je ključna riječ kojom se deklariraju varijable iz koje slijedi naziv varijable i nakon toga se određuje tip varijable u ovom slučaju varijabla atEnd je tipa Boolean (logička varijabla koja može imati vrijednost True ili False, inicijalno logičke varijable imaju vrijednost False).

While i Wend su početak i kraj petlje koja se izvršava sve dok varijabla atEnd ima vrijednost False.

DoEvents je sistemska funkcija koja omogućava da se izvršavaju drugi procesi izvan petlje. To je potrebno kako bi se korisniku omogućila daljnja interakcija sa glavnom aplikacijom.

„If Selection.Type = wdSelectionIP And Selection.End = ActiveDocument.Content.End - 1 Then atEnd = True„ je uvjet koji ispituje da li je pokazivač na kraju dokumenta i ako jeste onda varijabla atEnd dobiva vrijednost True što znači da će While Wend petlja završiti s radom.

„Selection.MoveRight Unit:=wdCharacter, Count:=1, Extend:=wdExtend“ je komanda kojom se kaže kursoru da se pomakne u desnu stranu za jedan slovni selektirajući ga.

„A_Znakovi AscW(Selection()), Selection.Font.Name“ Ovdje se poziva procedura „A_Znakovi“ koja ima dva ulazna parametra kôd slovnoga znaka i naziv fonta. Proceduri se proslijeđuju kod znaka koji je trenutno selektiran i to u numeričkom formatu dobivenog pozivom AscW komande VBA scripta i naziv fonta selektiranog znaka dobiven pozivom funksije Selection.Font.Name.

„Selection.MoveRight Unit:=wdCharacter, Count:=1“ ova linija koda će izvršiti deselektiranje znaka pomicanjem kursora u desnu stranu.

„StatusBar = Selection.Information(wdFirstCharacterColumnNumber)“ Ova linija koda će u statusnoj crti MS Worda ispisivati broj slovnog znaka na kojem se kursor trenutno nalazi. Ova komanda je informativnoga karaktera.

Slijedeća procedura je „A_Znakovi“ čiji je zadatak da provjeri kojeg je fonta selektirani znak i koju kodnu vrijednost ima i prema vrsti fonta da pozove odgovarajuću proceduru.

Procedura „A_Znakovi“ se poziva iz procedure „main“ i ima ulazne parametre „Kod“ i „Fnt“

```
Private Sub A_Znakovi(Kod As Long, Fnt As String)
    Select Case Fnt
        Case "B Grcki dopune"
            Grcki_dopune Kod
        Case "B Mat Symbol"
            Mat_Symbol Kod
        Case "B Plantin":  Plantin Kod
        Case "B Plantin 1: B_Plantin_1 Kod
        Case "B Plantin 1 Baltic":  B_Plantin_1_Baltic Kod
        Case "B Plantin 1 CE":  B_Plantin_1_CE Kod
        Case "B Plantin 1 Cyr":  B_Plantin_1_Cyr Kod
        Case "B Plantin 1 Greek":  B_Plantin_1_Greek Kod
        Case "Plantin 1 Tur":  Plantin_1_Tur Kod
        Case "B Plantin Baltic":  B_Plantin_Baltic Kod
        Case "B Plantin CE":  B_Plantin_CE Kod
        Case "B Plantin Cyr":  B_Plantin_Cyr Kod
        Case "B Plantin Greek":  B_Plantin_Greek Kod
        Case "B Plantin Tur":  B_Plantin_Tur Kod
        Case "B Simbol":  B_Simbol Kod
        Case "B Transkripcija":  B_Transkripcija Kod
        Case Else
            Selection.Font.ColorIndex = wdRed
    End Select
End Sub
```

Select Case – Case – End Select su ključne riječi koje se koriste kada se želi ispitati da li neka varijabla ima određenu vrijednost i što da se izvrši kada je kriterij zadovoljen. U ovom slučaju ulazna varijabla Fnt prima vrijednost preko glavne procedure main i to naziv fonta selektiranog znaka, a varijabla Kod prima numeričku vrijednost koda selektiranog znaka. Ova procedura se svaki put poziva kada se selektira znak i onda se prolaskom kroz Select Case Fnt gleda da li Fnt varijabla odgovara nekom od slučajeva pa recimo da je selektirani znak iz fonta koji se zove „B_Simbol“ i da je kod selektiranog znaka „-4063“ tada će se pozvati pomoćna procedura „B_Simbol“ čija je ulazna varijabla Kod (-4063). U slučaju da selektirani znak ne spada niti u jedan od zadanih provjera (Case Else) onda se selektirani znak oboji u crvenu boju što nam naknadnim pregledom pomaže da eventualni propušteni znak uvedemo dodatno u listu za konverziju.

Procedure „Grcki_dopune“, „Mat_Symbol“, „Plantin“, „B_Plantin_1“, „B_Plantin_1_Baltic“, „B_Plantin_1_CE“, „B_Plantin_1_Cyr“, „B_Plantin_1_Greek“, „B_Plantin_1_Tur“, „B_Plantin_Baltic“, „B_Plantin_CE“, „B_Plantin_Cyr“, „B_Plantin_Greek“, „B_Plantin_Tur“, „B_Simbol“ i „B_Transkripcija“ su slične samo se razlikuju po kodnim vrijednostima pa će na primjeru procedure „B_Simbol“ biti opjašnjen princip.

```
Private Sub B_Simbol(Kod As Long)
    Select Case Kod
        Case -4064: A_IzmjenaKoda 32
        Case -4063: A_IzmjenaKoda -2784
        Case -4062: A_IzmjenaKoda 9697
        Case -4061: A_IzmjenaKoda 10017
        Case -4060: A_IzmjenaKoda -2297
        Case -4059: A_IzmjenaKoda -2296
        Case -4058: A_IzmjenaKoda -2295
        Case -4057: A_IzmjenaKoda 8646
        Case -4056: A_IzmjenaKoda 9675
        Case -4055: A_IzmjenaKoda 9671
        Case -4054: A_IzmjenaKoda -2294
        Case -4053: A_IzmjenaKoda 160
        Case -4052: A_IzmjenaKoda 160
        Case -4048: A_IzmjenaKoda 9450
        Case -4047: A_IzmjenaKoda 9312
        Case -4046: A_IzmjenaKoda 9313
        Case -4045: A_IzmjenaKoda 9314
        Case -4044: A_IzmjenaKoda 9315
        Case -4043: A_IzmjenaKoda 9316
        Case -4042: A_IzmjenaKoda 9317
        Case -4041: A_IzmjenaKoda 9318
        Case -4040: A_IzmjenaKoda 9319
        Case -4039: A_IzmjenaKoda 9320
        Case -4038: A_IzmjenaKoda -2291
        Case -4037: A_IzmjenaKoda -2290
        Case -4036: A_IzmjenaKoda 8219
        Case -4034: A_IzmjenaKoda -2288
        Case -4031: A_IzmjenaKoda -2287
        Case -4030: A_IzmjenaKoda -2286
        Case -4029: A_IzmjenaKoda -2285
        Case Else: Selection.Font.ColorIndex = wdRed
    End Select
End Sub
```

Ovdje je sličan princip ispitivanja kao i kod prethodne procedure s tom razlikom što sada kad znamo kojeg je fonta selektirani znak ispitujemo njegovu kodnu vrijednost i kada je određena kodna vrijednost pronađena poziva se procedura A_IzmjenaKoda sa ulaznom varijablom novog koda.

Slijedeća procedura koja se poziva u nizu je „A_IzmjenaKoda“ koja konačno postavlja pravi font i kod selektiranog znaka.

```
Private Sub A_IzmjenaKoda(Kod As Long)
    Selection.Font.Name = "A PlantinV2"
    Selection.InsertSymbol CharacterNumber:=Kod, Unicode:=True
    Selection.MoveLeft Unit:=wdCharacter, Count:=1
End Sub
```

„Selection.Font.Name ='A PlantinV2'“ je linija koda koja će selektiranom slovnom znaku promijeniti font u „A PlantinV2“

„Selection.InsertSymbol CharacterNumber:=Kod, Unicode:=True“ linija koda će umetnuti znak kodne vrijednosti „Kod“ na mjesto selektiranoga i to UNICODE formata. Ova akcija će kursor pomaknuti u desno što onda linijom „Selection.MoveLeft Unit:=wdCharacter, Count:=1“ korigiramo vraćajući kursor jedno mjesto ulijevo.

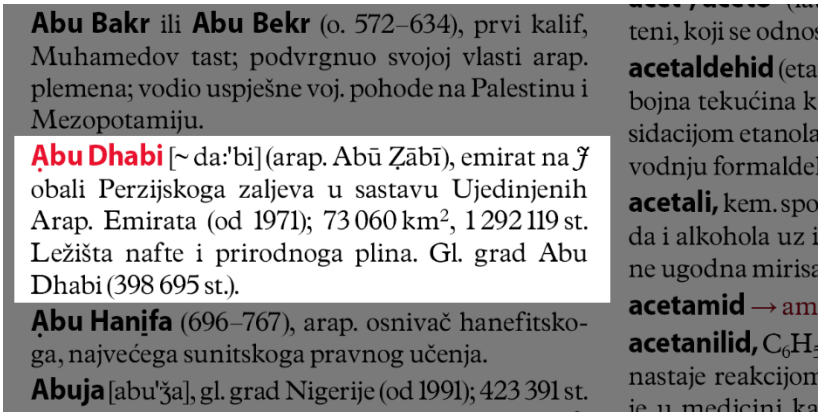
Postavlja se pitanje zašto ići znak po znak kada je moguće jednostavnim traži i zamjeni sustavom zamijeniti sve znakove? To bi značilo da bez obzira na duljinu teksta moramo napraviti 3291 prolaza kroz tekst jer toliko je u ovom slučaju kodova za konverziju. Na kratkim tekstovima to je izvedivo ali na tekstu jedne enciklopedije koja ima više stotina stranica teksta je problem. Dolazi do tzv. zagušenja nakon nekog vremena i aplikacija prestane funkcionirati što onda znači da postupak morate ponoviti iz početka jer ne znate na kojoj zamjeni je aplikacija prestala raditi. U načinu zamjene znak po znak ne dolazi do zagušenja, a ako se iz nekog razloga zaustavi script onda uvijek možete nastaviti od znaka na kojem se dogodio prekid.

2.2. Uređivanje stilova teksta prije eksportiranja u HTML

Stilovi odlomaka i znakova koje definiramo unutar aplikacije nam služe da zadržimo određene karakteristike isticanja kako bi tekst u HTML-u izgledao isto.

2.2.1. Stilovi u aplikaciji InDesign

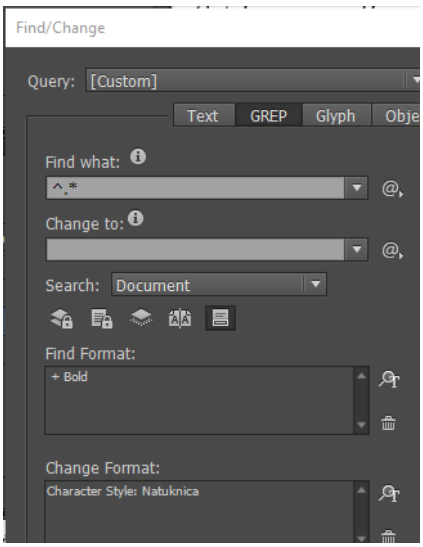
Pregledom sadržaja ustanovimo što sve možemo označiti kao specifično i prema tome gradimo listu stilova u InDesignu ako to nismo do sada učinili u fazi prijeloma. Poželjno je da izgradimo listu stilova za slova i za odlomke. Na slijedećim ilustracijama se vidi koji dio teksta u natuknici može dobiti poseban stil.



Slika 12. Primjer natuknice

Na slici 12 vidimo da naziv natuknice ima drugačiji font, veličinu pisma i boju. To znači da taj dio može dobiti svoj slovni stil, nazovimo ga „Natuknica“. Nadalje imamo i kurzivni tekst kojemu možemo dodijeliti stil „Italic“ i na kraju imamo i znakove u eksponentu kojima možemo dodijeliti stil „eksponent“. Nakon određivanja slovnih stilova možemo svakoj natuknici dodijeliti stil za odlomak, nazovimo ga „Normal“. Ne trebamo se zamarati oko detalja kao što su prored uvlake ili veličina pisma ovdje samo određenim dijelovima teksta dodjeljujemo imena.

Kada smo izgradili listu stilova potrebno je kroz cijeli prijelom primijeniti definirane stilove. To možemo izvesti uz pomoć jednostavnih funkcija ugrađenih u InDesign „traži i zamjeni“. Na primjer možemo tražiti tekst koji je kurzivan i zamjenjujemo ga s istom karakteristikom znači kurzivom + nazivom stila „Italic“. Kod složenijih zamjena možemo koristiti GREP opciju.



Slika 13. Primjer kako se koristi GREP opcija.

Na slici 13 je prikazano kako se traži tekst kojim počinje odlomak (^.*) ali takav da je masni (Bold) i dodjeljuje mu se stil „Natuknica“.

InDesign ima integriran GREP alat za traženje prema uzorku koji može poslužiti za kompleksnija pretraživanja. Nekoliko primjera uzoraka za traženje u tablici 3.

Tablica 3. Primjeri uzoraka za traženje pomoću GREP sustava

Uzorak	**značenje**
[[=u=]]	Pronađi sve slovne znakove u bez obzira ima li akcent ili ne (uúűü)
(?<=\().+?(?=\))	Pronađi sav tekst koji se nalazi u zagradama ali ne uključujući zagrade
^.*	Pronađi tekst od početka odlomka

2.3.Problem kratica u tekstu

Korištenje kratica za tiskano izdanje je opravdano, no kada se takva građa priređuje za internetsko izdanje onda kratice gube svoj smisao i naprotiv tekst mora biti raspisan do kraja.

Primjer teksta koji sadržava kratice (prema Proleksis enciklopediji):

> „željeznice, prom. sredstvo posebnih konstrukcijskih i organizacijskih osobina. Gl. elementi ž.: želj. kolosijek, lokomotiva, vagoni i vozni park. Ž. i iz nje nastao želj. promet jedna su od najznačajnijih prom. pa i ekon. grana većine država. Gl. prednosti ž. pred drugim prom. sredstvima: pouzdanost, razmjerno niski troškovi prijevoza nakon razvoja želj….“[8]

Kada se tekst „raspiše“:

> „željeznice, prometno sredstvo posebnih konstrukcijskih i organizacijskih osobina. Glavni elementi željeznice: željeznički kolosijek, lokomotiva, vagoni i vozni park. Željeznica i iz nje nastao željeznički promet jedna su od najznačajnijih prometala pa i ekonomska grana većine država. Glavne prednosti željeznice pred drugim prometnim sredstvima: pouzdanost, razmjerno niski troškovi prijevoza nakon razvoja željeznice…“

[8] Proleksis enciklopedija, natuknica: željeznice, URL: http://proleksis.lzmk.hr/51190/ (pristupljeno 14. I. 2017.)

Promjenu kratica u puni tekst je nemoguće provesti jednostavnim traži i zamjeni sustavom jer se kratice moraju raspisati u pravi padež. Tako da se ovaj dio provodi manualno.

Zašto raspisivati kratice kad ionako znamo značenje tih kratica. Mi znamo ali alati za pretraživanje to ne znaju. Pretraživači rade po principu da upisanu riječ smatraju kraćim oblikom nekog pojma tj. uzorkom i koriste ga kada se u pretraživanju ne pojavi rezultat iste duljine. Tako na primjer traži se pojam „glazba“ u rezultatima se pojavljuje sve što korijenski sadržava traženi pojam „glazba, glazbala, glazbaonica“ ali nikad nećete dobiti kraticu „glazb.“ unutar rezultata pretraživanja.

Natuknice koje obrađuju biografiju neke osobe samo na početku daju puno ime i prezime osobe na koju se odnosi natuknica, poslije se u tekstu kada je potrebno spomenuti ime osobe navodi se samo inicijal. Ovakva natuknica bi bila bolje pozicionirana unutar rezultata pretraživanja da ima ime osobe pisano u punom obliku jer se za relevantnost uzima u obzir duljina teksta plus broj pojavljivanja traženog termina. Tako da kad tražite Tesla onda će vam prije biti ponuđen članak koji u svom tekstu spominje Teslu u punom nazivu nego onaj koji je pisan inicijalom.

2.4.Prebacivanje velikog broja natuknica u bazu

Kada smo završili dodjeljivanje stilova u InDesignu potrebno je eksportirati tekstualni dio u HTML format. InDesign to načini tako da posebno spremi datoteku koja je u HTML formatu a koja sadržava tekst i zapiše još jednu datoteku s ekstenzijom CSS u kojoj se nalaze postavke naših stilova koje smo ranije definirali. Generirani HTML dokument otvorimo u Microsoft Wordu kao običan tekst i vidjet ćemo da su sve natuknice dobile očekivane HTML oznake, npr. natuknice počinju s HTML oznakom <p class=“Normal“> zatim natuknice imaju oznaku itd.

Baza podataka je mjesto u koje ćemo prebaciti natuknice obogaćene HTML kodom. Baza podataka koja je korištena za CMS Wordpress je MariaDB (jedna inačica MySQL baze). Baze podatke zapisuju u tablice. Tablice imaju zapise u obliku redaka, što znači da svaka naša natuknica je zapisana u svoj redak. To nas dovodi do problema kako razdvojiti HTML datoteku tako da dobijemo svaku natuknicu u posebnoj datoteci.

Za rješenje ovog problema koristimo VBA script jezik.

```
Sub Razlomi_na_natuknice_HTML()
Dim fso As New FileSystemObject
Dim doc As Document
Dim originalDoc As Document
Dim putanja$
Dim brojac As Long
    Set originalDoc = ActiveDocument
    putanja$ = originalDoc.Path
    Selection.Find.ClearFormatting
     With Selection.Find
        .Text = "<p class=""Normal""><span class=""Natuknica"">"
        .Replacement.Text = ""
        .Forward = True
        .Wrap = wdFindContinue
        .Format = False
        .MatchCase = False
        .MatchWholeWord = False
        .MatchWildcards = False
        .MatchSoundsLike = False
        .MatchAllWordForms = False
    End With
    Do Until originalDoc.Bookmarks("\Sel") =
originalDoc.Bookmarks("\EndOfDoc")
    Selection.Find.Execute
    Selection.MoveRight Unit:=wdCharacter, Count:=1
    Selection.Find.Execute
    Selection.MoveLeft Unit:=wdCharacter, Count:=1
    Selection.HomeKey Unit:=wdStory, Extend:=wdExtend
    Selection.Cut
    Set doc = Documents.Add
    doc.Activate
    Selection.Paste
    Selection.HomeKey Unit:=wdStory
    Selection.Find.ClearFormatting
     With Selection.Find
        .Text = "<p class=""Normal""><span class=""Natuknica"">"
        .Replacement.Text = ""
        .Forward = True
        .Wrap = wdFindContinue
        .Format = False
        .MatchCase = False
        .MatchWholeWord = False
        .MatchWildcards = False
        .MatchSoundsLike = False
        .MatchAllWordForms = False
    End With
    Selection.Find.Execute
    Selection.MoveRight Unit:=wdCharacter, Count:=1
    Do
    Selection.MoveRight Unit:=wdCharacter, Count:=1, Extend:=wdExtend
    A$ = Selection
    If Right(A$, 1) = "<" Then Exit Do
    Loop
    Selection.HomeKey Unit:=wdStory
    b$ = Mid(A$, 1, Len(A$) - 1)
    c$ = RTrim$(LTrim$(b$))
    Select Case Right(c$, 1)
        Case ",", ".", ":"
            D$ = Mid(c$, 1, Len(c$) - 1)
```

```
        Case Else
            D$ = c$
    End Select
    ime = D$ & "#"
    Selection.HomeKey Unit:=wdStory
    Selection.TypeText (ime)
brojac = brojac + 1
    D$ = brojac & ".txt"
doc.SaveAs2 FileName:=putanja$ & "\" & D$, FileFormat:=wdFormatText, _
        LockComments:=False, Password:="", AddToRecentFiles:=True,
WritePassword _
        :="", ReadOnlyRecommended:=False, EmbedTrueTypeFonts:=False, _
        SaveNativePictureFormat:=False, SaveFormsData:=False,
SaveAsAOCELetter:= _
        False, Encoding:=1200, InsertLineBreaks:=False,
AllowSubstitutions:=False _
        , LineEnding:=wdCRLF, CompatibilityMode:=0
    doc.Close
    originalDoc.Activate
    Loop
End Sub
```

Skripta radi na slijedećem principu: pronađi tekst „<p class=“Normal“> pomakni se za jedno mjesto udesno zatim ponovno pronađi isti tekst, to će nas dovesti na početak druge natuknice. Vrati se jedan znak ulijevo zatim selektiraj tekst do početka dokumenta što će nam obilježiti cijelu prvu natuknicu. Nakon toga naredbom „izreži“ uzimamo cijelu natuknicu u međuspremnik. Otvaramo novi dokument u koji uljepljujemo tekst natuknice iz međuspremnika. U novom dokumentu pronađemo kraj oznake za naziv natuknice i selektiramo naziv natuknice te ga ukopiramo na početak stavljajući mu na kraj oznaku #. Ta oznaka nam služi poslije pri učitavanju u bazu.

Spremamo taj novi dokument dodjeljujući mu za ime npr. redni broj 1 (ovaj brojač stalno povećavamo). Sada smo opet na glavnom dokumentu i taj postupak će se ponavljati dok ima teksta u glavnom dokumentu. Nakon toga je potrebno napraviti datoteku koja za sadržaj ima popis naziva datoteka natuknica.

Za učitavanje u bazu nam služi PHP skripta koja otvara datoteku s popisom naziva datoteka natuknica. Uzima preko for petlje naziv datoteke i otvara svaku pojedinačno i zapisuje u bazu njezin sadržaj izvorni kôd for petlje je dat na slici 14.

```php
foreach($popis_datoteka as $datoteka)
    {
        $tekst = file_get_contents($direktorij . $datoteka);        // dohvaćanje teksta datoteke
        $tekst = htmlspecialchars_decode(mb_convert_encoding($tekst, 'utf-8', 'utf-16'));   // konver
        $duzina = strlen($tekst);
        $$delimiter = strpos($tekst, '#');
        $post_title = trim(substr($tekst, 0, $$delimiter));
        $post_content = trim(substr($tekst, $$delimiter+1, $duzina-($$delimiter+1)));
        $query = $db->prepare("INSERT INTO wp_posts ( post_date , post_date_gmt , post_modified , pos
        '$datetime' , '$datetime' , '$datetime' , '$post_title' , '$post_content' , 1)");
        $query->execute();
    }

    $db = NULL;
```

Slika 14. For petlja PHP skripte kojom se učitavaju natuknice u bazu.

Ovako učitani sadržaj u tablicu se koristi u nekom CMS sustavu kao izvor podataka za prikaz i pretraživanje, izgled tablice na slici 15.

post_content	post_title	post_excerpt	post_status
<div class="PodrucjeSadrzaja1"...	Genova		publish
<div class="PodrucjeSadrzaja1...	Jiaozhou Wan		publish
<div class="PodrucjeSadrzaja1...	znanost		publish
<div class="PodrucjeSadrzaja1...	pamirski jezici		publish
<div class="PodrucjeSadrzaja1"...	Robbins, Harold		publish
<div class="DesniPanel" sty...	Marriner, Neville		publish
<div class="PodrucjeSadrzaja1"...	moralisti		publish
<div class= "DesniPanel" style=...	vjetrokaz		publish
<div class="PodrucjeSadrzaja1...	hidrostat		publish
<div class="PodrucjeSadrzaja1...	šipon		publish
<div class="DesniPanel" style="...	Tinbergen, Nikolaas		publish
<div class="PodrucjeSadrzaja1...	stupanj korisnosti		publish
<div class="DesniPanel" style="...	Škubonja, Fedor		publish
<div class="PodrucjeSadrzaja1...	leglica		publish

Slika 15. Izgled WordPressove tablice u bazi koja sadržava natuknice iz Proleksis enciklopedije.

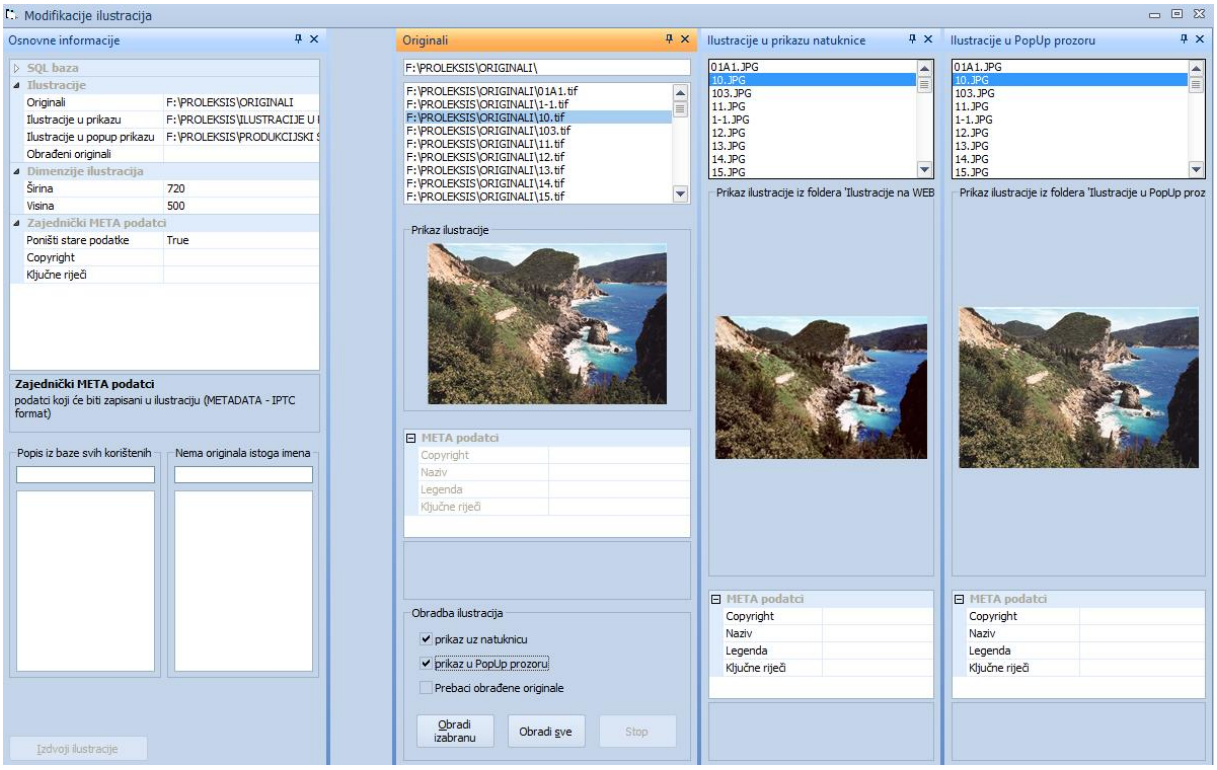
2.5. Obradba velike količine ilustracija

Ilustracije koje se koriste u prijelomu tj. za tisak su u TIF/CMYK/300 dpi formatu i kao takve se ne mogu koristiti za WEB jer su prevelike i taj format ne odgovara HTML standardu (TIF/CMYK). Slike bi trebale biti u JPG/RGB. [10]

Ovaj postupak se može automatizirati uz pomoć gotove aplikacije PhotoShop. Zbog specifičnih potreba da se ilustracijama mogu provjeriti i dodati EXIF određeni meta podatci napisana je aplikacija u Visual Basicu naziva „Modifikacije ilustracija“ koja konvertira veliku količinu ilustracija i manipulira meta podatcima. Još jedna funkcionalnost ove aplikacije daje joj prednost u odnosu na PhotoShop. U bazi podataka natuknice imaju reference na ilustracije. Može se dogoditi da je pogrešno unesena referenca na ilustraciju u tom slučaju u internetskom prikazu natuknica neće biti ilustrirana. Tu provjeru je teško provesti manualno jer se radi o više desetaka tisuća podataka. Postupak je da se iz baze eksportiraju sadržaji natuknica u

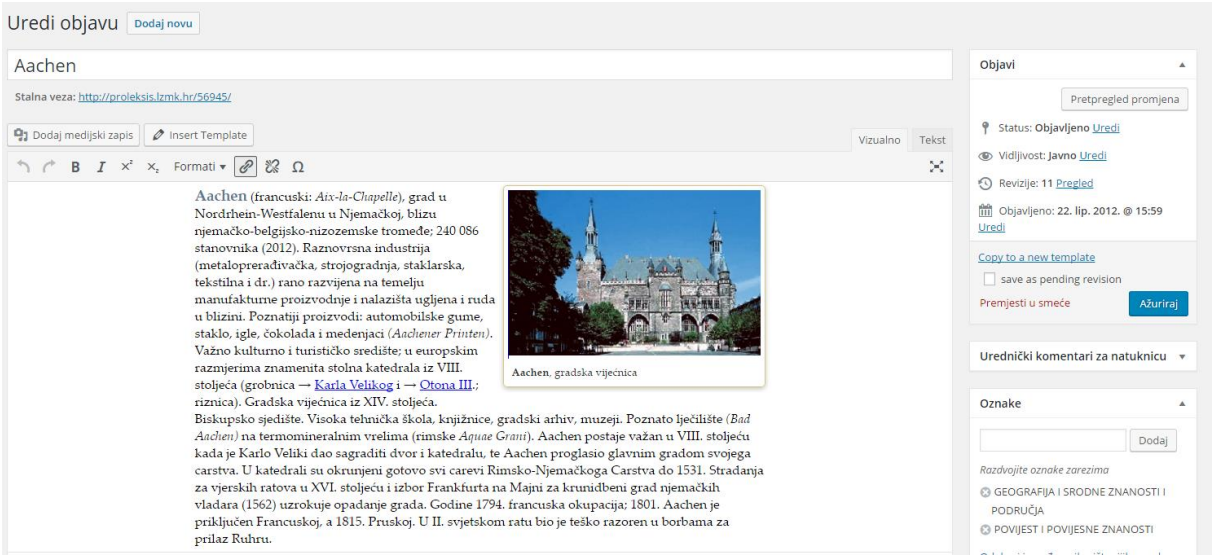
jednu SQL datoteku. Nakon toga tu datoteku učitamo u Modifikacije ilustracija. Zatim se navede putanja do mape sa ilustracijama i aplikacija će usporediti sve navedene reference u natuknicama sa stvarnim nazivima datoteka ilustracija. U posebnom okviru će biti izlistane razlike ako ih ima.

Formati ilustracija se mogu zadati kao konačna visina i konačna širina što znači da će ilustracije biti proporcionalno skalirane do vrijednosti koju prije dostignu. Na slici 16 je prikazano sučelje aplikacije.

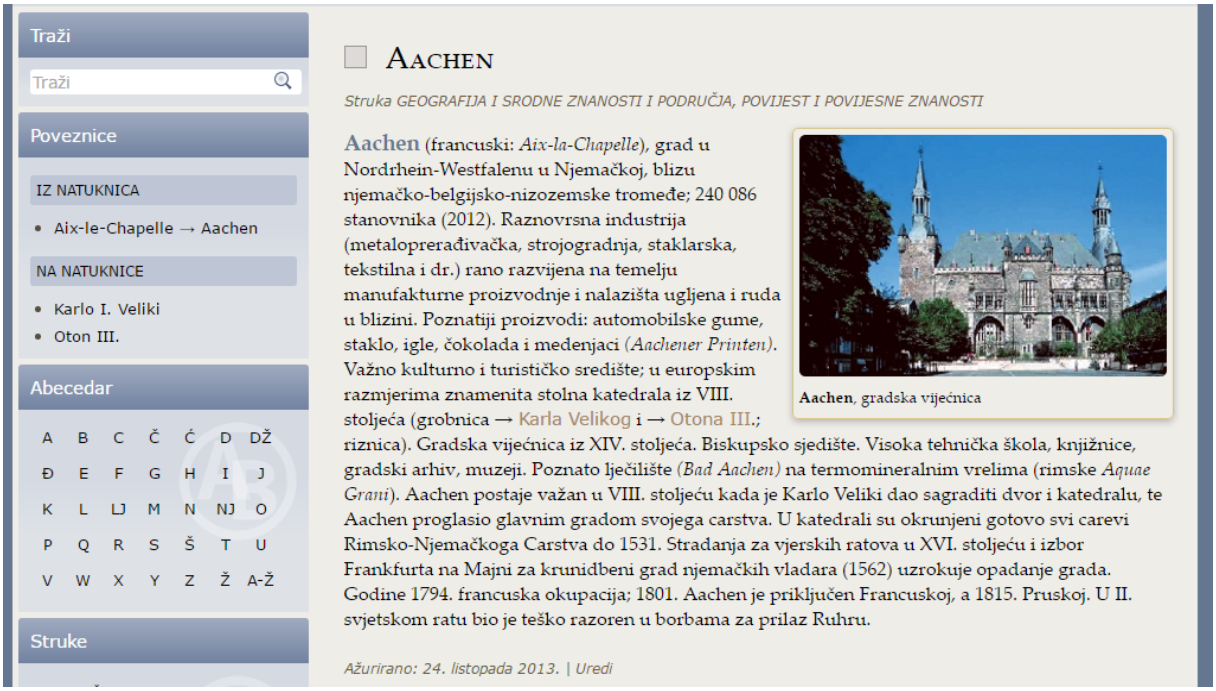


Slika 16. Aplikacija Modifikacije ilustracija

Konvertirane ilustracije se poslije koriste u CMS sustavu za ilustriranje natuknica na način da se u uredničkom sučelju ilustracije povezuju s natuknicama. Izgled uredničkog sučelja u CMS sustava ovom slučaju Wordpress je prikazan na slici 17 a kako izgleda prikaz iste natuknice na krajnjekorisničkom sučelju na slici 18.



Slika 17. Izgled uredničkog sučelja u Wordpressu, umetanje ilustracija u natuknicu



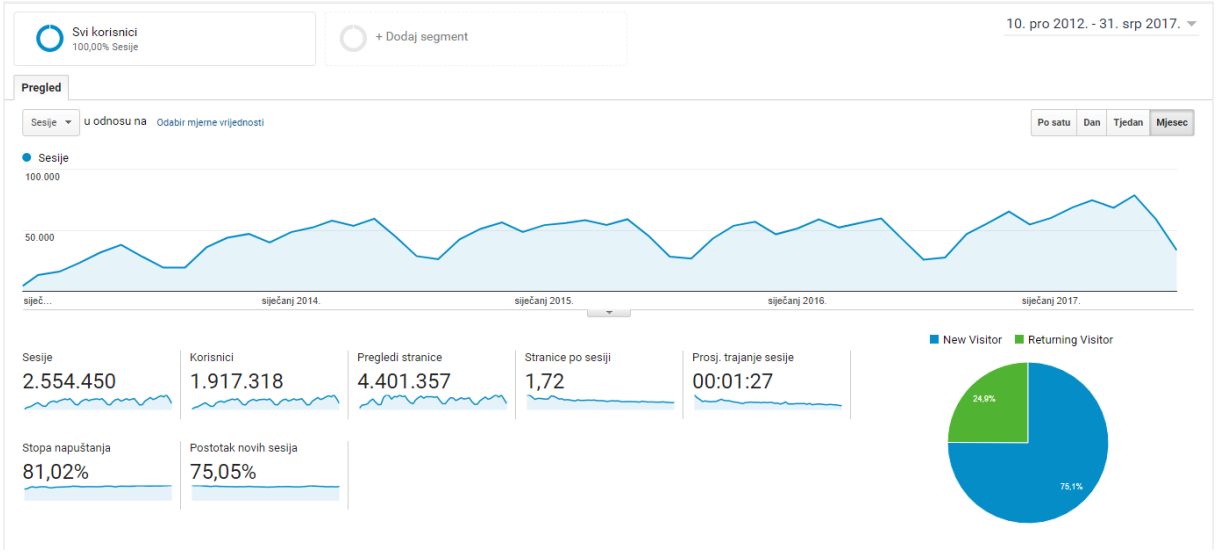
Slika 18. Izgled gotove natuknice u korisničkom sučelju

2.6. Posjećenost internetskih enciklopedija

Enciklopedije su inače po svom sadržaju izdanja koja ne čitate kao štivo nego ih koristite po potrebi. Kada je u pitanju analiza broja korisnika tiskanog enciklopedijskoga sadržaja onda podatke moramo prikupljati na razne načine kao što su ankete, podatci iz knjižnica o broju posuđivanja. Ta metoda je neprecizna jer nemamo uvid u stvarni broj korisnika.

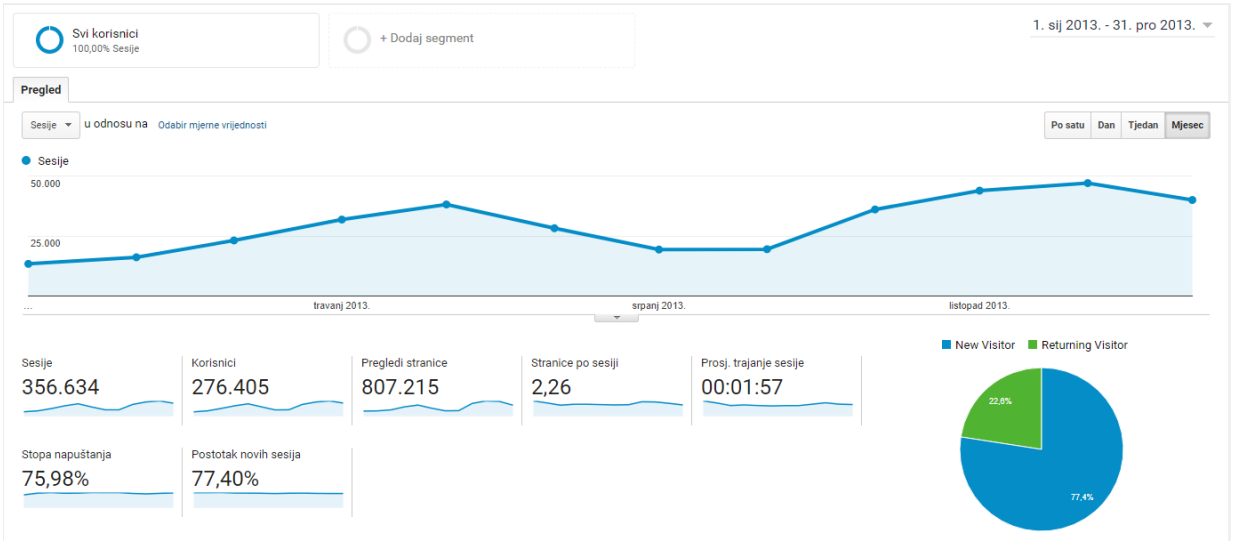
Internetska izdanja se mogu jednostavno pratiti putem Google Analytics servisa. Analiza u ovom diplomskom radu pokriva razdoblje od 2012. do 2017. za Proleksis enciklopediju.

Rezultati koje daje servis Google Analytics se mogu klasificirati po raznim parametrima tako na slici 19 imamo opći pregled korisnika u zadanom razdoblju od 10. prosinca 2012 kada je javno objavljena do 31. srpnja 2017.

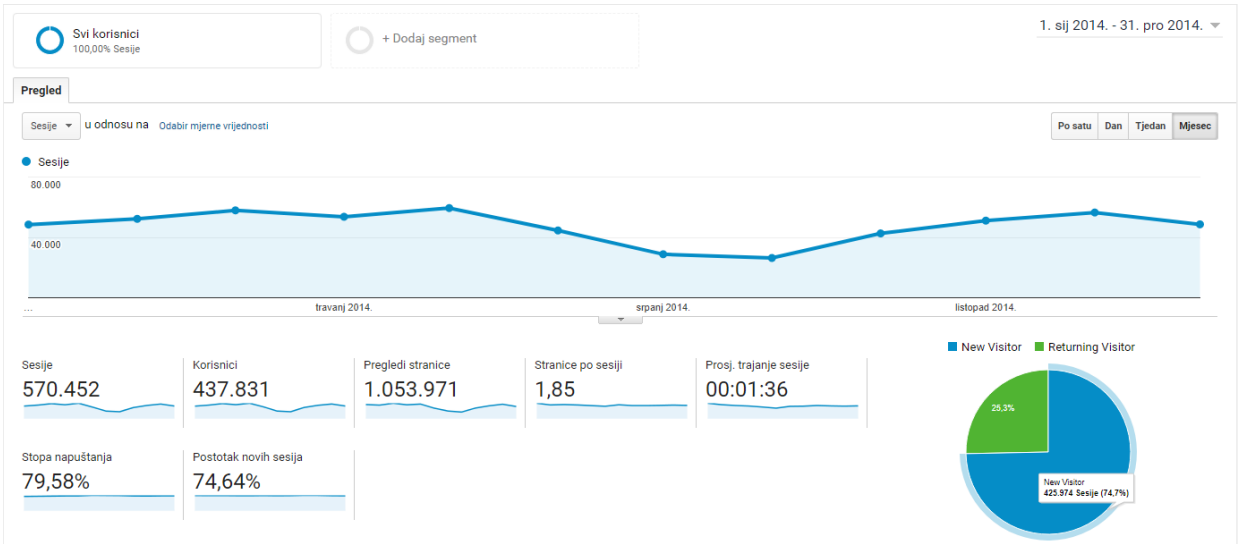


Slika 19. Opći pregled posjeta Proleksis enciklopedije u razdoblju 10. prosinac 2012 – 31. srpanj 2017.

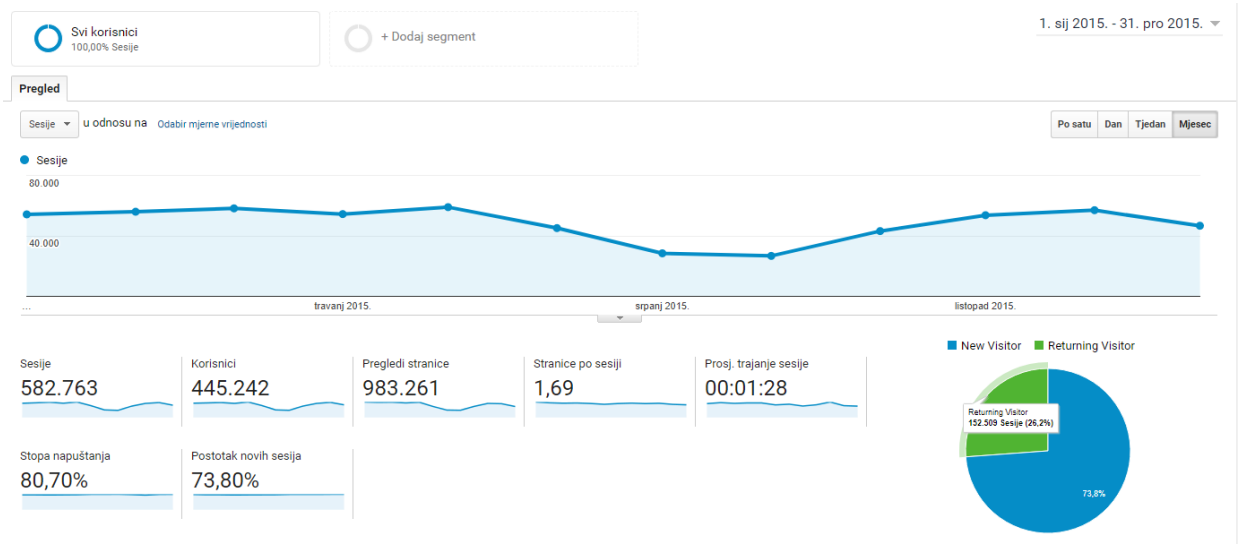
Iz priloženih informacija broj posjetitelja je u stalnom porastu, tijekom godine postoje oscilacije koje možemo vidjeti na slikama 20–23 Podatci za 2017. nisu cjeloviti pa su izuzeti.



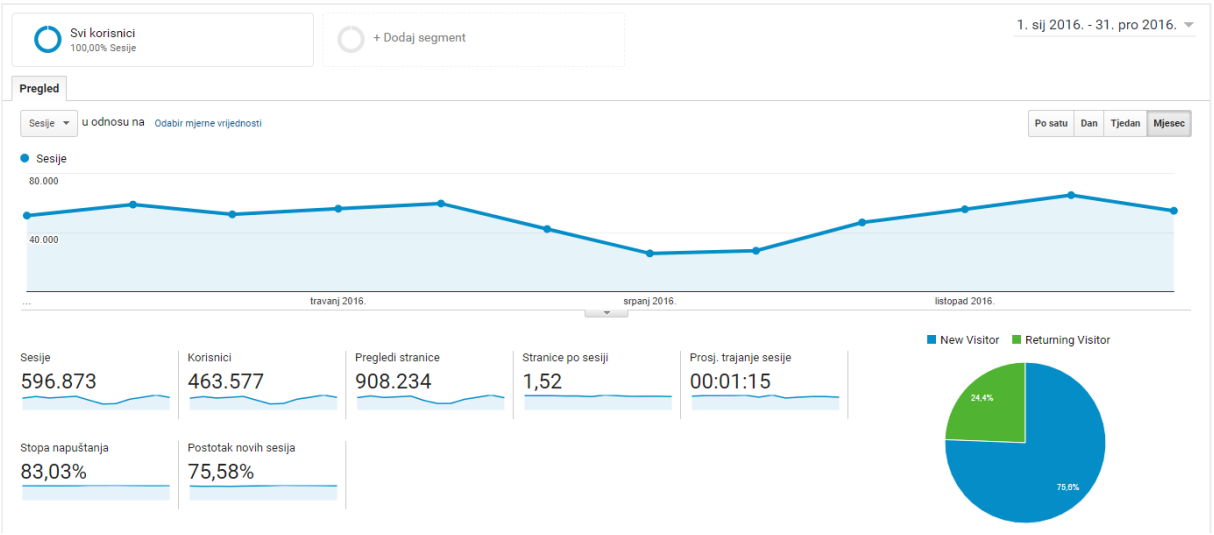
Slika 20. Opći pregled posjeta Proleksis enciklopedije u razdoblju 1. siječanj 2013 – 31. prosinac 2013.



Slika 21. Opći pregled posjeta Proleksis enciklopedije u razdoblju 1. siječanj 2014 – 31. prosinac 2014.



Slika 22. Opći pregled posjeta Proleksis enciklopedije u razdoblju 1. siječanj 2015 – 31. prosinac 2015.



Slika 23. Opći pregled posjeta Proleksis enciklopedije u razdoblju 1. siječanj 2016 – 31. prosinac 2016.

Nešto manji broj posjetitelja je u periodu srpanj kolovoz a najveći broj posjetitelja se događa u lipnju i studenom, ovi podatci su iznimno vrijedni jer nam govore da je posjet najveći u periodu pojačanih aktivnosti vezanih za školovanje. Kako bismo potvrdili pretpostavku potrebno je pogledati demografske podatke koje nam nudi Analytics. Uzet ćemo analizu podataka vezanu uz dobnu skupinu. Prije 2016. ti podatci nisu bili dostupni. Dobne skupine su od 18 godina pa naviše jer podatci o mlađim osobama također nisu dostupni.

☑	1. 18-24	**203.334** (26,84%)
☑	2. 25-34	**198.335** (26,18%)
☑	3. 35-44	**150.529** (19,87%)
☑	4. 45-54	**100.601** (13,28%)
☑	5. 55-64	**68.379** (9,03%)
☑	6. 65+	**36.301** (4,79%)

Slika 24. Dobna skupina posjetitelja Proleksis enciklopedije

Također možemo vidjeti na slici 25 podatke o spolu posjetitelja, vidi se da je broj posjetitelja ženskog spola skoro dvostruko veći.

☑	1. female	**517.396** (63,44%)
☑	2. male	**298.156** (36,56%)

Slika 25. Posjetitelji Proleksis enciklopedije prema spolu

Ovi podatci mogu poslužiti daljnjem razvoju enciklopedija u smislu privlačenja ciljnih skupina. Uvođenjem novih sadržaja i interakcija kao što su izdvojeni događaji na određeni datum, kvizovi, audio-vizualni prilozi itd.

Danas se korisnici služe mobilnim uređajima te je bitno da internetske enciklopedije budu u korak s tehnologijom što znači da WEB stranice budu prilagodljive (responsive) kako bi se lako mogle koristiti, usporedba podataka iz 2013. i 2017. godine govori u prilog tome. Korisnici mobilnih uređaja brojem su skoro izjednačeni sa korisnicima stolnih računala.

1.	desktop	**313.835** (86,93%)
2.	mobile	**38.872** (10,77%)
3.	tablet	**8.325** (2,31%)

Slika 26. Pregled tehnologije koju koriste posjetitelji Proleksis enciklopedije 2013.

1.	desktop	**250.243** (52,69%)
2.	mobile	**210.935** (44,42%)
3.	tablet	**13.721** (2,89%)

Slika 27. Pregled tehnologije koju koriste posjetitelji Proleksis enciklopedije 2017.

3. ZAKLJUČAK

Prednosti prilagodbe enciklopedijskoga sadržaja internetskome izdanju su višestruke i omogućuju krajnjim korisnicima brži pristup podatcima. Veći je broj korisnika koji mogu doći do vrijednih podataka, podatci su ažurniji. Korisnici mogu aktivno sudjelovati u poboljšanju sadržaja (ispravke netočnih informacija), koristeći se komentarima na natuknice. Povezanost putem hiperlinkova daje korisnicima dodatne izvore informacija. Pored svih automatiziranih postupaka, još uvijek se neke faze pretvorbe izvode manualno što ostavlja prostor za daljnji razvoj pomoćnih programa i skripti.

Ovim diplomskim radom potvrđeno je da se današnjom tehnologijom kojom raspolažemo može kvalitetno izvršiti digitalizacija tiskane enciklopedijske građe na primjeru već objavljenih enciklopedija Proleksis enciklopedije i Hrvatske enciklopedije.

Proces prilagodbe tiskanih enciklopedija i enciklopedija priređenih za tisak na internetsko izdanje zahtjeva određeno tehničko i stručno znanje iz oblasti grafičke struke i programiranja.

Interdisciplinarnost u postupku se ogleda u tome što su korištena znanja ne samo iz grafičke struke već i znanja iz oblasti informacijskih znanosti.

Prema statističkim podatcima prikupljenim servisom Google Analytics u razdoblju 2012–2017. jasno se vidi da je broj korisnika u konstantnom porastu i ima tendenciju daljnjega rasta.

Internetska izdanja enciklopedija neće u potpunosti zamijeniti tiskana jer za njih u manjoj mjeri postoji interes ciljanih korisnika kao što su razni arhivi, znanstvene institucije, istraživači, povjesničari, kolekcionari itd.

LITERATURA

[1] Hrvatska enciklopedija URL: http://enciklopedija.hr/Natuknica.aspx?ID=17879 (pristupljeno 29. VIII. 2017.)

[2] Mrvac, Nikola. Osobni komentar. I. 2017.

[3] Wikipedija: Internetske enciklopedije, URL: https://hr.wikipedia.org/wiki/Internetske_enciklopedije (pristupljeno 29. VIII. 2017.)

[4] Jecić, Zdenko, Damir Boras i Darija Domijan. 2008. Prilog definiranju pojma virtualna enciklopedija. Studia lexicographica br. 1 (2): 115–126

[5] Unicode inc., USA, Unicode 10.0 Character Code Charts, URL: http://www.unicode.org/charts/ (pristupljeno 29. VIII. 2017.)

[6] Unicode, URL: https://en.wikipedia.org/wiki/Unicode (pristupljeno 29. VIII. 2017.)

[7] Hrvatska enciklopedija, natuknica: digitalizacija, URL: http://enciklopedija.hr/Natuknica.aspx?ID=68025 (pristupljeno 14. I. 2017.)

[8] Seiter-Šverko, Dunja; Nacionalna i sveučilišna knjižnica, Zagreb, Hrvatska, Lana Križaj; Ministarstvo kulture Republike Hrvatske, Uprava za zaštitu kulturne baštine, Konzervatorski odjel u Krapini, Krapina, Hrvatska, Izlaganje sa skupa: Digitalizacija kulturne baštine u Republici Hrvatskoj: od trenutne situacije prema nacionalnoj strategiji, Vjesnik bibliotekara Hrvatske, vol. 55 No.2 2013, , URL: http://hrcak.srce.hr/106550 (pristupljeno 29. VIII. 2017.)

[9] BFS-Auto: High Speed & High Definition Book Scanner, Ishikawa Watanabe Laboratory, URL: http://www.k2.t.u-tokyo.ac.jp/vision/BFS-Auto/ (pristupljeno 29. VIII. 2017.)

[10] Information Technology Services, UNIVERSITY OF CALIFORNIA SANTA CRUZ, Preparing Images for Web or Print (2016), URL: http://its.ucsc.edu/fitc/tutorials/webimages.html (pristupljeno 29. VIII. 2017.)

POPIS ILUSTRACIJA

POPIS TABLICA