

Univerza v Ljubljani
Fakulteta *za elektrotehniko*



Jezikovne in spletne tehnologije: zbirka nalog

Zbirka laboratorijskih nalog, prva izdaja

Vitomir Štruc



2014

UNIVERZA V LJUBLJANI

FAKULTETA ZA ELEKTROTEHNIKO

VITOMIR ŠTRUC

**JEZIKOVNE IN SPLETNE TEHNOLOGIJE:
ZBIRKA NALOG**

ZBIRKA LABORATORIJSKIH NALOG

PRVA IZDAJA

LJUBLJANA, 2014

Predgovor

Pričujoča zbirka nalog predstavlja dopolnilno študijsko gradivo pri predmetu *Jezikovne in spletne tehnologije*, ki se izvaja na Fakulteti za elektrotehniko in je namenjen predvsem študentom Filozofske fakultete Univerze v Ljubljani. Zbirka je nastala iz navodil za izvedbo laboratorijskih vaj pri tem predmetu v študijskem letu 2013/2014.

Namen gradiva je seznaniti študente z navodili laboratorijskih vaj in podati smernice za njihovo izvedbo. Zbirka nalog obsega šest poglavij, od katerih je prvih pet namenjenih opisom nalog, ki jih morajo študentje izvesti v okviru vaj, zadnje, šesto poglavje pa predstavlja vzorec poročila, ki ga morajo študentje v okviru vsake vaje izpolniti in oddati asistentu.

Avtor se zahvaljuje vsem sodelavcem Laboratorija za umetno zaznavanje, sisteme in kibernetiko na Fakulteti za elektrotehniko, ki so kakorkoli pripomogli k nastanku te zbirke.

Kazalo vsebine

LABORATORIJSKA VAJA A (do 10 točk).....	4
LABORATORIJSKA VAJA B (do 10 točk).....	9
LABORATORIJSKA VAJA C (do 10 točk).....	13
LABORATORIJSKA VAJA D (do 10 točk).....	18
LABORATORIJSKA VAJA E (do 10 točk).....	22
POROČILO O LABORATORIJSKI VAJI.....	26

LABORATORIJSKA VAJA A (do 10 točk)

Vaje potekajo v učilnici LUKS pod vodstvom asistenta. Študentje v prvem delu vaj skupaj z asistentom pregledajo in ponovijo teoretične vidike vaj, se spoznajo z osnovno terminologijo in rešijo nekaj ilustrativnih primerov. V drugem delu vaj, študentje na podlagi pridobljenega znanja sami rešijo nekaj nalog, ki so namenjene utrjevanju učne snovi.

V okviru laboratorijske vaje A se seznanimo z:

- osnovami zapisa znakov v računalniku in urejevalnikih besedil, in
- delom s spletnimi viri/orodji kot so SSKJ, Amebisovo orodje za generiranje pregibnih oblik in Evrokopus (Evroterm).

Znanje, ki ga študentje pridobijo v okviru Vaje A je predstavljeno v nadaljevanju.

KAJ BOM ZNAL/ZNALA PO IZVEDBI VAJE:

- Razumel/razumela bom pomen kodnih naborov (UTF - 8, WINDOWS -1250, etc.) pri zapisu znakov v računalniku.
- Znal/znala bom shranjevati in odpirati tekstovne datoteke z uporabo različnih kodnih naborov.
- Poznal/poznala bom vlogo kodnih naborov pri spletnih straneh.
- Znal/znala bom uporabljati:
 - Amebisovo spletno orodje za tvorjenje pregibnih oblik
 - Spletni korpus vzporednih prevodov EVROKORPUS
 - Spletno različico SSKJ

PRIKAZ ZNAKOV in KODNI NABORI

V učilnici je nameščena različica operacijskega sistema Linux, zato bomo kot osnovno orodje pri tem delu vaj uporabljali program »Gedit« in programski paket »OpenOffice«. Gedit predstavlja prosto-dostopen urejevalnik besedil, razmeroma podoben Microsoftovi Beležki (angl. Notepad), ki jo poznamo z operacijskega sistema Windows. Podobno predstavlja OpenOffice prosto-dostopno (brezplačno) alternativo Microsoftovemu programskemu paketu Office. Tako Gedit kot OpenOffice sta dosegljiva tudi za OS Windows in ju je mogoče prenesti z naslednjih povezav:

<http://gedit.en.softonic.com/>
<http://www.openoffice.org/download/>

Programa na učilniških računalnikih zaženemo tako, da bodisi kliknemo na ustrezno ikono med nameščenimi programi ali z vnosom naslednjih ukazov v terminal:

```
> gedit&
```

in

```
> soffice&
```

Naloga 1 – Pretvorbe med kodnimi nabori

V okviru prve naloge se bomo seznanili s kodnimi nabori, ki se uporabljajo za predstavitev znakov v računalniku. Za namene vaje zapišite tri do štiri stavke v slovenskem jeziku, pri čemer bodite pozorni, da izbrani stavki vsebujejo znake, ki so značilni za slovenščino (t.j., č, ž, š, Č, Š, Ž). Zapisano besedilo bomo vstavili v urejevalnik besedil, ga shranili in nato zopet odprli z uporabo različnih kodnih naborov.

Primer primerne besedila:

```
Ta pravilnik določa postopek reševanja pritožb posameznika, ki meni, da so
bile z ravnanjem pooblaščenih uradnih oseb carinske službe (carinika,
inšpektorja ali izterjevalca; v nadaljnjem besedilu: javni uslužbenec)
kratene njegove pravice ali svoboščine. Določa tudi evidentiranje pritožb,
varstvo podatkov in stroške postopka.
```

Besedilo iz: Pravilnik o reševanju pritožb v carinski službi, dostopen preko: www.pirs.si

Pripravljeno besedilo vnesite v urejevalnik besedil Gedit. Izvedite spodnje naloge in opazujte ter komentirajte rezultate:

- Besedilo shranite s privzetim kodnim naborom, datoteko shranite in jo ponovno odprite. Kaj opazite? Kako se vam prikaže vaše besedilo? So vidni vsi znaki, ki ste jih

vnesli? Utemeljite svoja opažanja - kateri kodni nabor je bil uporabljen pri shranjevanju?

- Z Geditom odprite novo, prazno datoteko in vanj znova vnesite svoje besedilo. Besedilo poskusite shraniti s kodnim naborom ISO-8859-15 (Western). Kaj opazite in zakaj?
- Z Geditom odprite novo datoteko in vanj vnesite svoje besedilo. Besedilo shranite s kodnim naborom CENTRAL EUROPEAN (Windows-1250) in ga ponovno odprite s privzetim kodnim naborom. Kako se vam prikaže besedilo? So vidni vsi znaki, ki ste jih vnesli? Komentirajte svoja opažanja in jih utemeljite!
- Ponovite vajo iz prejšnje točke, le da pri odpiranju uporabite kodni nabor, ki ste ga uporabili tudi za shranjevanje – CENTRAL EUROPEAN (Windows-1250). Kaj opazite tokrat? Kako se vam tokrat prikaže besedilo in zakaj?
- Besedilo, ki ste ga shranili v okviru prejšnje naloge poskusite pravilno odpreti še s programskim paketom OpenOffice. Kateri kodni nabor morate izbrati, da se vam besedilo prikaže pravilno?
- V urejevalniku besedil programskega paketa OpenOffice vnesite naslednji stavek (skupaj z naglasi):

»Nikóli in nikhár ne pústi pêti níz pôsla nekjé v vrísti!«

- Stavek skopirajte v urejevalnik besedil Gedit in ga shranite s primernim kodnim naborom, ki bo zagotavljal pravilni zapis in kasnejši prikaz vseh znakov vašega stavka. Kateri kodni nabor je primeren za zastavljeno nalogo in kateri ne? Zakaj?

Naloga 2 – Uporaba kodnih naborov - spletne strani

Za pravilni prikaz vseh znakov morajo primerne kodne nabore uporabljati tudi spletne strani. Uporabljeni kodni nabor spletne strani lahko preverite tako, da v brskalniku pogledate izvorno kodo strani (v Firefox-u se izvorna koda spletne strani lahko preveri z bližnjico Ctrl-U) in v glavi poiščete oznako »charset«. Pomen kodnih naborov pri spletnih straneh bomo spoznali skozi naslednje naloge:

- Preverite kodni nabor različnih spletnih strani (domačih in tujih), npr:
 - <http://luks.fe.uni-lj.si/>
 - <http://www.fe.uni-lj.si/>
 - <http://www.delo.si/>
 - <http://www.nbc.com/>
- Kateri kodni nabor se največkrat uporablja in zakaj?
- Obiščite spletno stran laboratorija za umetno zaznavanje, sisteme in kibernetiko (LUKS), ki je dostopna preko naslednje povezave: <http://luks.fe.uni-lj.si/>. Shranite spletno stran s pomočjo desnega miškega klika na primerno mesto v vašem računalniku. Spletno stran nato odprite s urejevalnikom besedil Gedit in poiščite oznako za uporabljeni kodni nabor. Zamenjajte navedeni kodni nabor z ISO-8859-15 in UTF-8, datoteko shranite in jo odprite z brskalnikom. Kaj opazite in zakaj?

SPLETNI VIRI

Na spletu obstaja vrsta orodij, ki se uvrščajo na področje jezikovnih tehnologij. V okviru te vaje si bomo ogledali nekaj primerov takih orodij in se seznanili z osnovami njihove uporabe.

Naloga 4 – Amebis

Amebis ponuja vrsto spletnih jezikovnih orodij, ki jih lahko uporabniki uporabljajo brezplačno. Med njimi je zanimiva storitev, ki omogoča generiranje pregibnih oblik za samostalnike, pridevnike, glagole, ipd. Storitev je dostopna preko naslednje povezave:

<http://besana.amebis.si/pregibanje/>

S pomočjo Amebisove spletne storitve:

- Sklanjajte samostalnika »breskev« ali »gospa«
- Poiščite pregibne oblike pridevnika »papeški« in jih nekaj zapišite.
- Poizkusite poiskati pregibne oblike osebnih imen, priimkov, slengovskih izrazov (npr, googlati). Vaše rezultate zabeležite.
- Kaj vrne storitev za netipične priimke kot je npr. Mithans? Zakaj?

Naloga 5 – Evrokopus

Za prevajalce in tudi druge jezikoslovce je zelo zanimiv korpus vzporednih prevodov EUROKORPUS, ki je dosegljiv preko naslednje povezave:

<http://www.evroterm.gov.si/evrokopus/index.php>

Korpus vsebuje vzporedne prevode besedil v različne evropske jezike in zato predstavlja pomembno orodje za prevajalce. Korpus omogoča iskanje izoliranih besed ali besednih zvez, pri tem pa se lahko uporablja standardne oblike zapisa iskalnih nizov, ki so znani tudi že z brskanja preko iskalnika Google. Izpostavimo nekaj pomembnejših:

- če iščemo posamično besedo, bo Evrokopus vrnil vse vnose z izbrano besedo,
- z znakom * lahko nadomestimo del besede (samo na koncu); Evrokopus bo vrnil vse besede z enakim korenem in različnimi končnicami,
- iskanje niza besed pomeni iskanje posameznih besed v nizu,
- za iskanje celotnega niza uporabimo narekovaje ali besede povežemo z znakom +.

Predstavljena pravila bomo uporabili za izvedbo preprostih nalog:

- Poiščite prevod besede *uradnik* v angleškem, nemškem in francoskem jeziku
- Poiščite prevode besedne zveze *družba z omejeno odgovornostjo* v angleškem, nemškem in francoskem jeziku
- Koliko zadetkov vrne beseda s korenem *financ*, če za iskanje uporabite vzporedni slovensko-angleški korpus. Zapišite po pet primerov za vsak jezik.

Naloga 6 – SSKJ

Eno od osnovnih orodij za vsakega, ki ga zanima področje jezikovnih tehnologij pri nas, predstavlja Slovar Slovenskega Knjižnega Jezika (SSKJ). Spletna oblika slovarja je dosegljiva preko spletnih strani Inštituta za slovenski jezik Frana Ramovša ZRC SAZU:

<http://bos.zrc-sazu.si/sskj.html>

Med drugimi spletnimi viri, ki jih najdemo na spletnih straneh ZRC SAZU, so še:

- Spletni seznam besed slovenskega jezika: <http://bos.zrc-sazu.si/besede.html>
- Odzadnji slovar slovenskega jezika: <http://bos.zrc-sazu.si/odzadnji.html>
- Besedišče slovenskega jezika: <http://bos.zrc-sazu.si/besedisc.html>
- Zbirka za določevanje osnovnih besednih oblik: http://bos.zrc-sazu.si/dol_lem1.html

Preko navedenih povezav lahko preverite funkcionalnost spletnih orodij, v okviru te vaje pa se bomo osredotočili na spletno različico SSKJ. SSKJ omogoča iskanje s preprostimi ukazi, med katerimi si velja zapomniti:

- Ukaze nizamo z uporabo logičnih operatorjev IN in ALI,
- Z znakom * lahko nadomestimo poljuben del besede,
- Poizvedbe v SSKJ izvajamo s pomočjo polj *ge* (geslo) in *op* (opis) in logičnih operatorjev₂
- Če v iskalno polje vnesemo besedo, slovar vrne seznam gesel in opisov, v katerih se beseda nahaja

Na podlagi zgoraj navedenih pravil zvedite naslednje naloge:

- V SSKJ poiščite vsa gesla, ki se začnejo na *rastlin* in jih zapišite. Zapišite tudi iskalni niz, ki ste ga uporabili.
- Poiščite vsa gesla, ki se začnejo na *rastlin* in so daljša od 15 znakov. Gesla in uporabljeni iskalni zapišite.
- Poiščite vsa gesla, ki se končajo s končnico *pedija* in vsebujejo opis, ki je daljši od 150 znakov. Zapišite rezultat iskanja in uporabljeni iskalni niz.
- Koliko pomenov ima po SSKJ beseda *student*?

LABORATORIJSKA VAJA B (do 10 točk)

Vaje potekajo v učilnici LUKS pod vodstvom asistenta. Študentje v prvem delu vaj skupaj z asistentom pregledajo in ponovijo teoretične vidike vaj, se spoznajo z osnovno terminologijo in rešijo nekaj ilustrativnih primerov. V drugem delu vaj, študentje na podlagi pridobljenega znanja sami rešijo nekaj vaj, ki so namenjene utrjevanju učne snovi in neposredni interakciji s potrebno tehnologijo.

V okviru laboratorijske vaje B se študentje seznanijo z:

- Slovenskim korpusom FIDAplus,
- Delom s spletnim vmesnikom korpusa FIDAplus,
- Delom s programom WordSmith.

Znanje, ki ga študentje pridobijo v okviru Vaje B je predstavljeno v nadaljevanju.

KAJ BOM ZNAL/ZNALA PO IZVEDBI VAJE:

- Seznanjen/seznanjena bom z namenom korpusov in poznal/poznala bom značilnosti slovenskega korpusa FidaPlus
- Znal/znala bom uporabljati spletni vmesnik slovenskega besedilnega korpusa FidaPlus za:
 - Napredno iskanje po korpusu,
 - Urejanje in čiščenje konkordančnih nizov,
 - Analizo konkordanc,
 - Itd.
- Znal bom uporabljati progam WordSmith za:
 - Analizo besedil,
 - Primerjalne analize,
 - Itd.

FIDApplus

FIDApplus je slovenski besedilni korpus, ki vsebuje preko 100 milijonov besed. Korpus predstavlja reprezentativni vzorec slovenskega jezika in omogoča raziskovanje slovenskega jezika. Korpus je nastal v okviru projekta, v katerem so sodelovali: Filozofska fakulteta UL, Inštitut Jožef Štefan, DZS d.o.o. in Amebis d.o.o. Dostopen je preko naslednje povezave:

<http://www.fidaplus.net/>

Uporaba korpusa je v povsem brezplačna, a zahteva predhodno registracijo. Uporaba korpusa in pripadajočega konkordančnika je dokumentirana v priročniku, ki je dostopen preko naslednje povezave:

http://www.fidaplus.net/Files/Kaj_poceti_s_korpusom_FidaPLUS_knjizica.pdf

Naloga 1 – uporaba spletnega vmesnika za iskanje

Konkordančnik (namenska programska oprema za delo s korpusi) korpusa FIDApplus ponuja različne možnosti za iskanje po korpusu, od enostavnih poizvedb na podlagi izoliranih gesel, do zahtevnejših iskanj z uporabo fraz, morfoloških deskriptorjev in logičnih operatorjev. Namen prve naloge je seznaniti študente z osnovami iskanja po besedilnem korpusu FIDApplus z uporabo namenskega spletnega vmesnika. Podrobnejša navodila o uporabi vmesnika za namene iskanja bodo podana na samih vajah.

Pridobljeno znanje poizkusite aplicirati na naslednjih nalogah:

- Poiščite besedne zveze, ki vsebujejo poljubne pregibne oblike besed »vsak« in »sreča«. Zapišite iskalni niz in nekaj reprezentativnih zadetkov.
- Kaj vrne iskalni niz »igra&~#2s*«? Zapišite nekaj reprezentativnih primerov zadetkov v korpusu in razložite pomen posameznih delov iskalnega niza.
- Kolikokrat se v besedilnem korpusu pojavi besedna oblika »podreti« in kolikokrat lema »podreti«? Podajte iskalni niz, ki ste ga uporabili.
- Poiščite vsaj 10 pridevnikov, ki se pojavljajo v bližini lem »muca« in »mačka«. So med njimi opazne razlike? Primerjajte rezultate s pridevniki, ki se pojavljajo v bližini lem »pes« in »psiček«. Podajte iskalni niz, ki ste ga uporabili.
- V kakšnih besedilih se najpogosteje uporablja beseda »prevarant«? Navedite nekaj primerov.
- Poiščite vsaj 3 pomene besede »krokar« in ugotovite ali obstaja kakšna povezava med pomenom besede in tipom besedila, v katerem je dani pomen besede uporabljen.

Naloga 2 – uporaba spletnega vmesnika za analizo

Poleg osnovnega iskanja in ogleda oz. brskanja po rezultatih iskanja omogoča spletni vmesnik besedilnega korpusa FIDAplus tudi čiščenje oz. urejanje konkordančnih nizov in obdelavo urejenih nizov. V tem kontekstu se posvetite naslednjim nalogam:

- V besedilnem korpusu poiščite:
 - 3 besede, ki se najpogosteje pojavijo pred besedo »krokar«; navedite vse tri besede in navedite kolikokrat se vsaka od njih pojavi pred konkordančnim jedrom; katera od teh treh besed se najpogosteje pojavi v celotnem besedilnem korpusu?
 - 3 pridevnike, ki se najpogosteje pojavijo pred besedo »krokar«; navedite vse tri pridevnike in navedite kolikokrat se vsaka od njih pojavi pred konkordančnim jedrom?
 - 3 besede, ki se najpogosteje pojavljajo za besedo »krokar«; navedite vse tri besede in navedite kolikokrat se vsaka od njih pojavi pred konkordančnim jedrom; za kakšne besedne vrste gre?
 - Preveriti pojavnost besede »tajkun« od leta 2003 do leta 2010. Kako se je pogostost uporabe besede spreminjala skozi čas? V kakšnih vrstah besedil se beseda najpogosteje uporablja? Se vrsta besedil, v katerih se beseda pojavlja, spreminja med 2003 in 2010?
 - Preverite ali se v besedilnem korpusu pogosteje pojavlja besedna zveza »rit in srajca« ali »rit in hlače«. V kakšnih besedilih se uporablja besedna zveza?

Naloga 3 – leksikalna analiza

V okviru prvih dveh nalog smo spoznali možne načine uporabe slovenskega besedilnega korpusa FidaPlus is se seznanili z orodji za podrobnejšo analizo rezultatov naših poizvedb. V okviru tretjega dela bomo na podlagi korpusa izvedli semantično analizo izbranih besed slovenskega jezika. S pomočjo korpusa FIDAplus identificirajte različne pomene besed in rezultate primerjajte z opisom gesel v SSKJ (Vaja A). Posebej bodite pozorni na slovnične lastnosti besed kot tudi njihove kolokacije. Za analizo si izberite par navidezno sinonimnih besed ter analizirajte razlike in podobnosti v njihovem pomenu in rabah. Predstavite potek dela, analizo in podajte svoje ugotovitve.¹

¹ Vaje povzete po vajah pri predmetu *Jezikovne tehnologije*, Politehnika Nova Gorica, 2005/2006; doc. dr. Tomaž Erjavec, doc. dr. Špela Vintar, Irena Srdanović.

WordSmith

Program WordSmith predstavlja prosto dostopno programsko opremo, ki jo lahko uporabljamo za analizo besedil. Program je žal mogoče namestiti zgolj na operacijske sisteme Windows in OSX, zato ga na učilniških računalnikih ni. Tretji del vaje študentje tako opravijo doma.

Prenesite programsko opremo WordSmith:

<http://www.lexically.net/wordsmith/index.html>

Domača stran programa WordSmith vsebuje obsežno dokumentacijo s praktičnimi primeri njegove uporabe. V okviru tretje vaje:

- Se seznanite s programom WordSmith in proučite glavne funkcije programa; predstavite osnove funkcionalnosti programa in navedite kako bi ga uporabili za primerjalno analizo besedil,
- Na spletu poiščite »Cvetje v jeseni« v obliki datoteke »txt« (Vaja C – napredno iskanje) in besedilo shranite,
- Odprite program WordSmith in izdelajte besedilni seznam za besedilo Cvetja v jeseni. Preverite seznam prvih najpogostejših besed. Katere besedne vrste prednjačijo na seznamu? Navedite nekaj najpogostejših besed.
- Kako različne oblike besed v slovenščini vplivajo na pogostost besednih oblik?
- Oglejte si konkordance za besedo »cvetje«.

LABORATORIJSKA VAJA C (do 10 točk)

Vaje potekajo v učilnici LUKS pod vodstvom asistenta. Študentje v prvem delu vaj skupaj z asistentom pregledajo in ponovijo teoretične vidike vaj, se spoznajo z osnovno terminologijo in rešijo nekaj ilustrativnih primerov. V drugem delu vaj, študentje na podlagi pridobljenega znanja sami rešijo nekaj vaj, ki so namenjene utrjevanju učne snovi in neposredni interakciji s potrebno tehnologijo.

V okviru laboratorijske vaje C se študentje seznanijo z:

- Naprednim iskanjem z spletnimi iskalniki;
- Postopki za snemanje in kodiranje govornih posnetkov;
- Orodji za izrezovanje in označevanje govornih posnetkov;
- Postopki za analizo lastnosti govora.

Znanje, ki ga študentje pridobijo v okviru Vaje C je predstavljeno v nadaljevanju.

KAJ BOM ZNAL/ZNALA PO IZVEDBI VAJE:

- Znal/znala bom bolje uporabljati spletne iskalnike in učinkoviteje usmerjati moje iskanje
- Znal/znala bom posneti zvočni posnetek in ga pretvarjati med različnimi formati.
- Razumel/a bom proces označevanja govornih zbirk in govornih signalov in poznal/a ustrezna orodja, ki jih za to potrebujem.
- Znal/a bom izvesti osnovno analizo govorne zbirke z orodji kot je OpenOffice SpreadSheets ali Excel

GOVORNE TEHNOLOGIJE

V učilnici je nameščena različica operacijskega sistema Linux, zato bomo kot osnovno orodje pri tem delu vaj uporabljali programe »Audacity«, »WaveSurfer« in programski paket »OpenOffice«. Audacity predstavlja prosto-dostopno orodje za editiranje zvočnih posnetkov, ki je uporabno širše, tudi izven področja govornih tehnologij. Program WaveSurfer prav tako predstavlja prosto-dostopno orodje in spada med najpogosteje uporabljena orodja za analizo govornih signalov. Program je nastal v okviru Inštituta za govor, glasbo in sluh s Štokholma na Švedskem. Vsi programi so dostopni tudi za Os Windows in sicer preko naslednjih povezav:

<http://audacity.sourceforge.net/?lang=sl>

<http://sourceforge.net/projects/wavesurfer/>

<http://www.openoffice.org/download/>

Naloga 1 – iskanje s spletnimi iskalniki

Spletne iskalnike uporabljamo tako rekoč vsak dan tako pri našem delu kot tudi v prostem času. Osnovna uporaba spletnih iskalnikov kot so Google, Yahoo ali Bing je izredno preprosta in za večino uporabnikov pomeni zgolj vnos iskalnega niza v iskalno polje in pritisk na tipko Enter. Takšen način iskanja s spletnimi brskalniki pogosto privede do množice rezultatov, ki zahtevajo zamuden pregled vseh (oz. prvih nekaj) prikazanih zadetkov iskanja.

Da si olajšamo delo s spletnimi iskalniki, se bomo v okviru prvega dela vaje seznanili z napredno uporabo spletnih iskalnikov in pridobljeno znanje uporabili za bolj usmerjeno iskanje po spletu. Navodila in predstavitev napredne uporabe spletnih iskalnikov bodo podana na samih vajah.

Pridobljeno znanje poskusite aplicirati na naslednjih nalogah:

- Na spletu poiščite datoteko tipa »mp3«, ki vsebuje besedo »adele«. Pri tem zahtevajte še, da med prikazi rezultatov iskanja ni prikazov, ki bi vsebovali besedo »rolling«. Zapišite vaš iskalni niz in spletni naslov prvega zadetka.
- Kako se rezultat iskanja spremeni, če iskano vrsto datoteke spremenite na »wav«?
- Na spletu poiščite datoteko tipa PDF s natančnim naslovom »Introduction to Digital Speech Processing«. Zapišite iskalni niz in navedite kdo je avtor knjige, ki jo iskalnik vrne kot prvi zadelek.

Naloga 2 – snemanje in kodiranje zvočnih posnetkov

Zvočne datoteke lahko zapišemo v različni formatih, ki se med seboj razlikujejo v uporabljenem postopku kodiranja signala, zapisu glave datoteke, ipd. Različni programi, ki jih po navadi uporabljamo za predvajanje in urejanje zvočnih datotek pogosto podpirajo le določene, pogosteje uporabljene formate zapisa. V primeru, da smo pri našem delu vezani na uporaba določenega orodja, ki ne podpira naših zvočnih datotek, je pomembno, da poznamo orodja in načine kako zvočno datoteko zapisano v enem formatu pretvorimo v drugega.

V okviru drugega dela vaje se bomo seznanili z načini za zapis zvočnih posnetkov različnih formatih. Pri tem bomo uporabljali prosto-dostopni avdio program »Audacity«. Program poiščite me nameščenimi programi na vašem računalniku in ga zaženite. Če kot mikrofona uporabljate mikrofona slušalk, ki so priključene na monitor, je potrebno pred uporabo Audacity-ja ustrezno urediti nastavitve zvoka v namizju operacijskem sistemu. Podrobnejša navodila v zvezi z nastavljanje vhodnih in izhodnih zvočnih naprav bodo podana na vajah.

Odprite zvočni posnetek »govor.wav«, ki se nahaja na 'storage/vaje/databases/gst/govor.wav'.

- Posnetek shranite (ukaz izvozi/export pod menijsko izbiro datoteka/file) v formatih mp3 in ogg. Preverite velikost izvirne datoteke in velikost shranjenih datotek in podajte opažanja.
- S programom Audacity posnemite lasten testni zvočni posnetek in ga prav tako shranite v različnih formatih (mp3 in ogg). Znova preverite velikost shranjenih datotek in podajte svoja opažanja.

Pretvarjanje med formati je mogoče izvesti tudi s spletnimi orodji. Eno takšnih orodij je dosegljivo na spletnem naslovu:

<http://media.io/>

- Ponovite pretvorbo zvočnega posnetka »govor.wav« v mp3 in ogg format še s spletnim orodjem. Primerjajte velikost datotek po pretvorbi s spletnim orodjem in po pretvorbi s programom Audacity. Podajte svoja opažanja in jih argumentirajte!

Naloga 3 – izrezovanje in označevanje govornih posnetkov ter shranjevanje rezultatov

Za potrebe analize govornih signalov, študija lastnosti govornega lastnika, fonetičnih karakteristik, ipd. je po navadi potrebno proučevano govorno zbirko ustrezno označiti, rezultate označevanja pa ustrezno shraniti za kasnejšo analizo. V okviru tretjega dela vaje si bomo zato ogledali postopke in orodja za označevanje govornih posnetkov. Pri našem delu bomo uporabljali program WaveSurfer, ki je že nameščen na učilniških računalnikih.

Za namene označevanja potrebujemo govorni material oz. govorne posnetke, ki jih bomo označili in nato v primerni obliki tudi shranili. S programom Audacity zato posnemite zvočni posnetek (v mono načinu), v katerem glasno in razločno izgovorite stavek:

»Nikóli in nkdár ne pústi pèti níz pôsla nekjé v vrísti!«

Posneti stavek »Nikóli in nikdár ne pústi pêti níz pôsla nekjé v vřsti!« vsebuje vseh osem dolgih naglašanih samoglasnikov slovenskega govornega jezika, ki jih v računalniški različici mednarodne fonetične abecede IPA zapišemo (po vrsti pojavljanja v stavku na podčrtanih mestih) s simboli: o:, a:, u:, E:, i:, O:, e:, in @:.

- S program Audacity ustvarite datoteke, ki vsebujejo le samoglasnike in jih shranite v formatu WAV. V vaši ciljni mapi naj se po opravljeni nalogi nahaja 8 krajših zvočnih datotek, ki vsebujejo posnetke samoglasnikov, ter dodatna datoteka, ki vsebuje posnetek celotnega izgovornega stavka. Pri izrezovanju bodite pozorni, da posnetki vsebujejo tudi predvsem stacionarne dele signala.

Po izvedbi do zdaj opisanega dela vaj imate na voljo zvočne datoteke, ki jih lahko obdelujete s programi za obdelavo in analizo govornih posnetkov. Zaženite program WaveSurfer (uredite nastavitve za vhodne in izhodne avdio naprave) in odprite posnetek celotnega stavka, ki ste ga pripravili v prvem delu Naloga 3.

- Izvedite transkripcijo stavka, shranite rezultate transkribiranja in v poročilu prikažite rezultate – vsebino datoteke s transkripcijami. Povzemite svoje izkušnje s tem delom vaje (zahtevnost, časovna zahtevnost, ...)
- Naložite še vseh osem krajših posnetkov samoglasnikov in shranite podatke o formantnih frekvencah za vsak glas posebej. Na koncu te vaje morate imeti pri sebi osem datotek s podatki o formantnih frekvencah.

Naloga 4 – analiza in prikaz rezultatov

Tekstovne datoteke s podatki o formantnih frekvencah bomo uporabili pri naši nadaljnji analizi, kjer bomo poskusili ponazoriti kako izgleda graf, ki prikazuje prvih dveh formantnih frekvenc naših osmih samoglasnikov.

- Zaženite program OpenOffice SpreadSheet in vanj eno za drugo uvozite tekstovne datoteke s podatki o formantnih frekvencah, ki ste jih ustvarili v okviru Naloga 3. To storite tako, da v menijski izbiri programa izberite »Insert -> Sheet From File«. Pri uvozu podatkov bodite pozorni na nastavitve uvoza (pravilno nastavljena ločila, simbolične oznake za decimalne vejice, ipd.). Po uvozu ohranite le prva dva stolpca podatkov za vsako uvoženo datoteko. V tretji stolpec vpišite oznako samoglasnika, ki ste ga pravkar uvozili.
- Po uvozu vseh osmih datotek podatke skopirajte na skupni list (spreadsheet) ter za vsak samoglasnik izračunajte povprečno vrednost prve in druge formantne frekvence, F1 in F2. V poročilu zapišite izračunana povprečja za F1 in F2 za vseh osem samoglasnikov.

Povprečje naših meritev lahko izračunamo s pomočjo funkcij, ki jih OpenOffice podpira (postopek je identičen v primeru Excela). Najprej izberemo prazno polje, v katerem želimo prikazati rezultat. Kliknemo na menijsko izbiro »Insert->Function« in poiščemo želeno

funkcijo »AVERAGE«, jo izberemo in nastavimo obseg podatkov, ki jih želimo vključiti v izračun.

- Ko izračunate povprečja F1 in F2 za vse samostalnike jih prikažite v obliki grafa (scatter plot). Pri čemer orientirajte osi tako, da bo na y osi F1 in na x osi F2. Os F1 naj bo orientirana kot prikazuje spodnja slika:



Dodajte oznake glasov in komentirajte rezultat. Kaj smo narisali?

LABORATORIJSKA VAJA D (do 10 točk)

Vaje potekajo v učilnici LUKS pod vodstvom asistenta. Študentje v prvem delu vaj skupaj z asistentom pregledajo in ponovijo teoretične vidike vaj, se spoznajo z osnovno terminologijo in rešijo nekaj ilustrativnih primerov. V drugem delu vaj, študentje na podlagi pridobljenega znanja sami rešijo nekaj nalog, ki so namenjene utrjevanju učne snovi in neposredni interakciji z ustrežno tehnologijo.

V okviru laboratorijske vaje D se študentje seznanijo z:

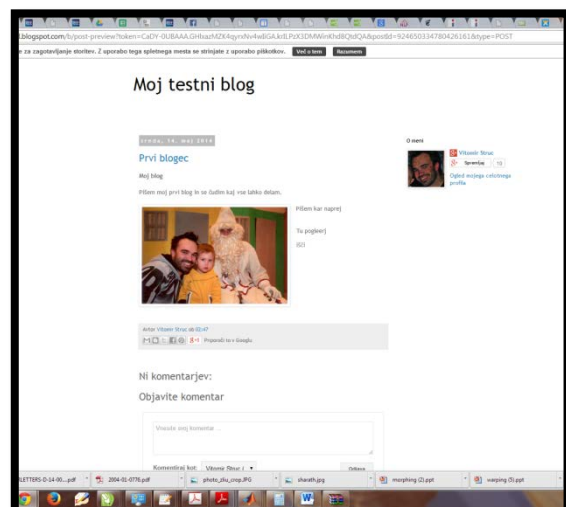
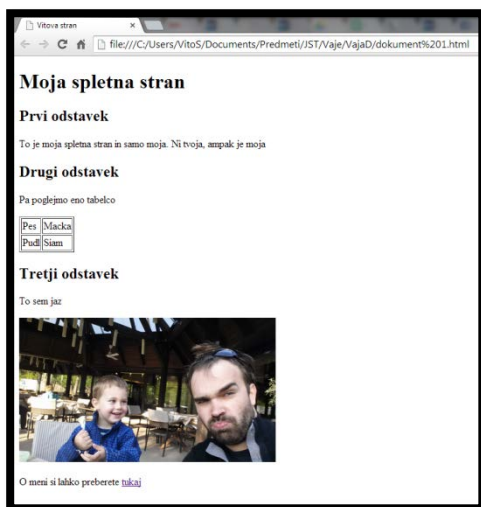
- Osnovami označevalnega jezika HTML kot so:
 - Osnovna struktura HTML dokumentov,
 - Naslovi, odstavki, tabele in slike,
 - Povezave, ipd.
- Osnovnimi orodji za delo z blogi

Znanje, ki ga študentje pridobijo v okviru Vaje D je predstavljeno v nadaljevanju.

KAJ BOM ZNAL/ZNALA PO IZVEDBI VAJE:

- Razumel/razumela bom osnovno zgradbo spletnih strani
- Poznal/poznala bom značilnosti in temeljne elemente spletnega označevalnega jezika HTML, s katerimi se gradi spletne strani,
- Znal/zнала bom sestaviti osnovno spletno stran v tekstovni obliki brez uporabe naprednejših CMS orodij
- Znal/zнала bom postaviti svoj blog, ga urejati in mu dodajati večpredstavnostne vsebine

Študentje v okviru vaje izdelajo preprosto spletno stran in blog:



OSNOVE HTML

Spletni označevalni jezik HTML (Hyper Text Markup Language) je označevalni jezik za izdelavo spletnih strani. Čeprav se dan danes pri izgradnji spletnih strani le redko pojavi potreba po sestavljanju spletnih strani v tekstovni obliki - obstajajo namreč naprednejša CMS (Content Management System) orodja, ki ta postopek bistveno poenostavijo - pa je poznavanje osnov jezika HTML vseeno izrednega pomena.

V okviru Vaje D se bomo spoznali z osnovnimi značilnostmi spletnega označevalnega jezika HTML in na podlagi pridobljenega znanja poskusili zgraditi preprosto spletno stran brez uporabe naprednejših CMS orodij.

Pri našem delu v tem sklopu vaje bomo uporabljali prosto dostopni urejevalnik besedil Gedit, s katerim bomo tudi zgradili spletno stran (html datoteko).

Naloga 1 – osnovne značilnosti in elementi jezika HTML

Pri spoznavanju osnovnih značilnosti in elementov spletnega označevalnega jezika HTML bomo uporabljali spletne strani W3schools. Spletne strani so dosegljive na naslednji povezavi:

<http://www.w3schools.com/html/DEFAULT.asp>

Spletne strani W3schools so namenjene spoznavanju in učenju različnih spletnih tehnologij kot so HTML, CSS, Javascript, PHP, JQuery in drugi. V okviru tega dela vaj bomo pozornost namenili zgolj spoznavanju jezika HTML.

V okviru vaje si oglejte in sami preizkusite (povezava »Try it yourself«):

- HTML Introduction – ponazarja zgradbo html strani
- HTML Basic – ponazarja osnovne gradnike
- HTML Elements – razlaga osnovne elemente
- HTML Headings – razlaga naslove
- HTML Paragraphs – razlaga odstavke
- HTML formatting – razlaga formatiranje besedila
- HTML Comments – razlaga komentiranje
- HTML Links – razlaga povezave
- HTML Head – razlaga glavo strani
- HTML Images – razlaga slikce
- HTML Tables – razlaga tabele
- Na kratko preglejte še ostale strani tutoriala.

Ker vaja zahteva pregled in izvedbo večjega števila krajših vaj v poročilu navedite zgolj svoje izkušnje in vtise pri vašem delu z jezikom HTML. Ste se z njim srečali prvič, ste dobili dodaten vpogled v zgradbo spletnih strani in vlogi brskalnikov, ... ?

Poiščite še preprosto spletno stran (brez CSS), npr. <http://nlp.stanford.edu/fsnlp/promo/>, in s pomočjo bližnjice »Ctrl+U« v brskalniku (Chrome ali Firefox) prikažite kodo spletne strani. V

primeru, da uporabljate katerega od drugih spletnih brskalnikov (npr. IE, Opera, Safar, ipd.) se lahko pojavi potreba po uporabi druge bližnjice za prikaz kode spletne strani. V kodi nato:

- identificirajte nekaj značilnih gradnikov spletnih strani, ki ste jih spoznali v okviru vaje. V poročilu izpostavite nekaj identificiranih gradnikov (del kode) in prikažite grafični ekvivalent kode.

Naloga 2 – pisanje spletnih strani

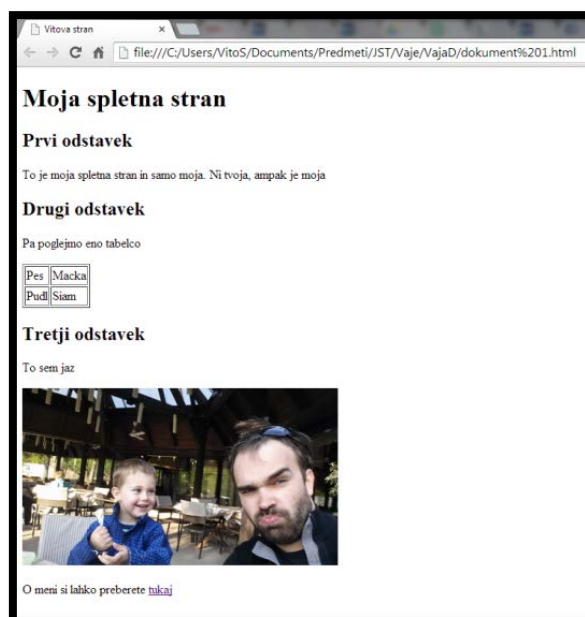
V drugem delu vaje bomo s pomočjo spletnega označevalnega jezika HTML zgradili preprosto spletno stran. Gradnike spletnih strani, ki smo jih spoznali v okviru prvega dela vaje bomo združili v celotno spletno stran, ki jo bomo v nato prikazali v spletnem brskalniku. Kot osnovno orodje pri našem delu bomo uporabljali prosto dostopni urejevalnik besedil Gedit, ki ga lahko prenesete preko naslednje povezave:

<https://wiki.gnome.org/Apps/Gedit>

S pomočjo Gedita sestavite preprosto spletno stran, ki mora:

- v telesu strani (body) vsebovati:
 - naslov strani (h1)
 - vsaj tri podnaslove (h2)
 - pod vsakim podnaslovom vsaj en odstavek besedila
 - eno tabelo z vsaj dvema vrsticama in dvema stolpcema in okvirjem
 - eno sliko, in
 - povezavo na kakšno drugo stran
- v glavi strani (head) vsebovati:
 - naslov strani (title)
 - nastavljen kodni nabor (UTF-8)

Primer izgleda:



BLOGI

Pri študijskem delu kot tudi kasneje, v okviru zaposlitve (še posebej na področju humanistike in družboslovja) se pogosto pojavi potreba po objavljanju in deljenju določenih vsebin. Kot priročno orodje se tukaj ponujajo spletni blogi, ki uporabnikom omogočajo preprosto ustvarjanje spletnih vsebin in deljenje s širšo množico bralcev.

Na spletu obstaja več platform, ki uporabnikom omogočajo pisanje blogov, precej teh platform pa je celo brezplačnih. Med najbolj priljubljene platforme spada tudi Googlov blogger, ki je dostopen preko naslednjega spletnega naslova:

<https://www.blogger.com/>

Med slovenskimi blogi pa so popularne strani RTV-jevega MMC-ja:

<http://www.rtv slo.si/blog/>

ali Siolov blogos:

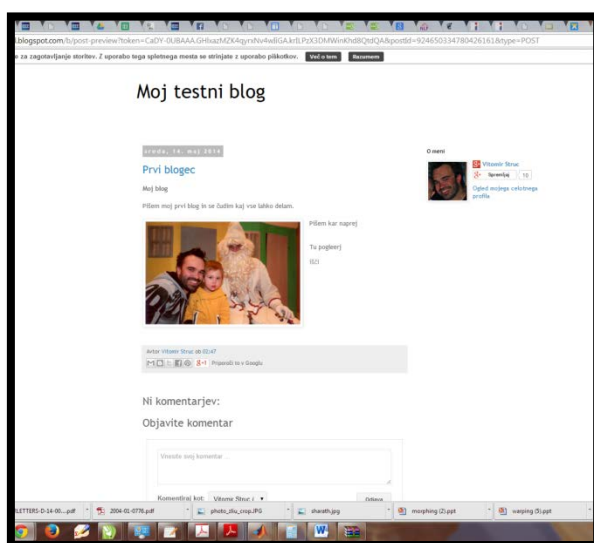
<http://blog.siol.net/>

Naloga 3 – pisanje blogov

V okviru tega dela vaje bomo sestavili svoj lasten blog na straneh Googlovega bloggerja. Na spletnem naslovu boggerja se z vašim Googlovim računom vpišite v storitev in sestavite lasten blog, ki naj vsebuje:

- naslov, sliko, povezavo, nekaj besedila, video posnetek in druge elemente

Ilustrativni primer bloga je prikazan spodaj:



V poročilu podajte vaše izkušnje in spletno povezavo do vašega bloga.

LABORATORIJSKA VAJA E (do 10 točk)

Vaje potekajo v učilnici LUKS pod vodstvom asistenta. Študentje v prvem delu vaj skupaj z asistentom pregledajo in ponovijo teoretične vidike vaj, se spoznajo z osnovno terminologijo in rešijo nekaj ilustrativnih primerov. V drugem delu vaj, študentje na podlagi pridobljenega znanja sami rešijo nekaj nalog, ki so namenjene utrjevanju učne snovi in neposredni interakciji z ustrežno tehnologijo.

V okviru laboratorijske vaje E se študentje seznanijo z osnovnimi principi postavljanja spletnih strani. Študentje izdelajo lastno spletno stran na podlagi storitev in orodij, ki jih nudi podjetje Google, v okviru Google sites. Drugače kot pri vaji D, kjer študentje sestavijo preprosto spletno stran neposredno z uporabo označevalnega spletnega jezika HTML, si pri tej vaji ogledamo naprednejša orodja, s katerimi je moč zgraditi grafično dodelane spletne strani, do katerih je dejansko mogoče dostopati preko spleta.

Znanje, ki ga študentje pridobijo v okviru Vaje D je predstavljeno v nadaljevanju.

KAJ BOM ZNAL/ZNALA PO IZVEDBI VAJE:

- Znal/znala bom sestaviti lastno spletno stran
- Znal/znala bom uporabljati spletna orodja storitve Google Sites

Študentje v okviru vaje izdelajo preprosto spletno stran:



GOOGLE SITES

Podjetje Google v okviru svoje storitve Google sites uporabnikom omogoča postavitev lastnih spletnih strani. Za izgradnjo spletnih strani hkrati ponuja v naprej pripravljene grafične predloge spletnih strani in namenska orodja za njihovo urejanje. Storitve je namenjena postavitvi spletnih strani za:

- Različne klube,
- Skupinske projekte,
- Intraneta podjetij,
- Razrede,
- Igralne skupine,
- Družine, ipd.

Do storitve Google sites lahko dostopamo preko:

<https://sites.google.com/?pli=1>


Ustvarjanje spletne strani


Prvi korak pri ustvarjanju spletnih strani je klik na gumbek USTVARI, ki se nahaja v zgornjem levem kotu storitve Google Sites. V naslednjem oknu se nastavijo parametri spletne strani, kot je:

- Ime spletne strani (title),
- URL spletne strani znotraj storitve Google sites,
- Ustrezna grafična podoba strani

Google Sites **USTVARI** Prekliči

Izberite predlogo, ki jo želite uporabiti:

 Prazna predloga

 Classroom site 🔍

[Prebrskajte galerijo, kjer je na voljo več](#)



Poimenujte spletno mesto:

Lokacija spletnega mesta – V URL-jih lahko uporabite le te znake: A–Z, a–z in 0–9
 https://sites.google.com/site/

▸ Izberi temo

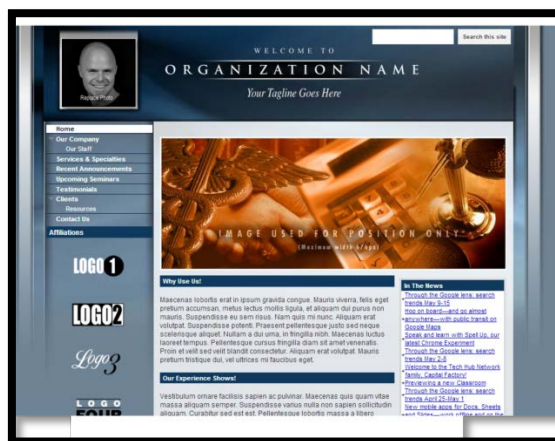
▸ Več možnosti

Vnesite prikazano kodo:

Ko so parametri strani izbrani in je storitev preverila, da je izbrani URL na voljo, s klikom na gumb USTVARI MESTO ustvarimo svojo spletno stran.

Če smo izbrali že obstoječo grafično podobo spletne strani, se nam po nekaj trenutkih prikaže stran, ki jo lahko nato urejamo:



Urejanje spletne strani

V ustvarjeni predlogi spletne strani je mogoče zamenjati prav vse elemente in stran tako preurediti za lastne potrebe.

Stran začnemo urejati s klikom na gumb v obliki v svinčnika ali kolesčka, ki se nahajata v zgornjem desnem kotu strani. Po kliku na gumb se nam prikaže meni, s katerim je mogoče urediti elemente strani. Primer zgornje strani po preurejanju je prikazan spodaj²:



² <https://sites.google.com/site/testjstx/>

Naloga – sestavi svojo spletno stran

Poglavitni cilj vaje E je sestaviti lastno spletno stran s pomočjo storitve Google sites. Spletna stran naj bo izdelana na podlagi ene od obstoječih grafičnih predlog in naj vsebuje:

- Več podstrani,
- Več področij z besedilom,
- Ustrezno glavo in nogo,
- Nekaj slik,
- Povezave na pod-strani in zunanje strani,
- En zemljevid, ki naj kaže na vašo fakulteto,
- Koledar,
- En video posnetek v obliki video povezave na zunanjo stran,
- Gumb +1
- En foto album,
- ...

V poročilu podajte svoje izkušnje pri pisanju strani in povezavo na spletno stran, ki ste jo ustvarili.

POROČILO O LABORATORIJSKI VAJI

Podatki o študentih:

Vpisna številka	Ime in priimek	Datum

Namen vaje je bil seznaniti študente s/z

Najprej smo se seznanili z ... in potem z ...

Spodaj so podani komentarji, opažanja in rezultati, ki se nanašajo na posamezne naloge te vaje.

Naloga 1

Nalogo smo izvedli tako, da smo najprej ...

Naloga 2

Nalogo smo izvedli tako, da smo najprej ...

Naloga 3

Nalogo smo izvedli tako, da smo najprej ...

Naloga 4

Nalogo smo izvedli tako, da smo najprej ...