Sveučilište u Zadru

Odjel za informacijske znanosti

1. godina diplomskog studija

Matija Baković

**Ocjenski zadatak**

Izabrana poglavlja iz informacijskih znanosti

Profesor: dr. sc.Franjo Pehar

Zadar, rujan 2019.

# U zadatku će se izdvojiti kompetencije i vještine od kojih bi informacijski stručnjaci imali najviše koristi u svakodnevnom radu te u istraživačkim procesima kroz koje prolaze, sve te sposobnosti i vještine se nalaze tj. nudi ih kroz svoj program organizacija The Carpentries koja se sastoji od sljedećih ogranaka: Library Carpentry, Data Carpentry te Software Carpentry. Cilj The Carpentries je obučiti informacijske i druge stručnjake kvalitetnom i efikasnom radu s podacima, metapodatkovnim formatima i dokumentima - i to kroz programiranje i različite alate koji bi im olakšali svakodnevne poslove i istraživački proces.

# Library Carpentry

Library Carpentry je rezultat rada volontera diljem svijeta kojima je cilj stvarati „lekcije“ i radionice koje će pomoći u radu knjižničarima i drugim informacijskim stručnjacima, te lekcije su vezane uz znanost o podacima - upravljanje podacima, bazama podataka te uz rad na određenim sustavima. Cilj je naučiti tu ciljanu publiku informacijskih stručnjaka novim vještinama, osnažiti ih za korištenje tih vještina i kompetencija u svakodnevnom radu te ih opremiti s dovoljno znanja da sami mogu druge kolege naučiti raditi s određenim sustavima, podacima, formatima itd. Nadaljenjihov temeljni kurikulum se sastoji od četiri lekcije koje su iznimno važne za informacijsku djelatnost, prva lekcija je uvod u znanost o podacima; uvod u rad s podacima koja zapravo približava korisnicima sam leksikon koji se koristi u ovom području, uče se termini i fraze koje se učestalo koriste u radu s podacima i softverima pri npr. pretraživanju itd. Također u ovoj lekciji se daje smjer korisnicima te se raspravlja o temeljnim konceptima i najboljim praksama vezano uz strukture podataka, upravljanje podataka, metapodatkovne formate, različite sustave itd. Sve je to iznimno važno za sobe koje rade u gore spomenutom sektoru jer im se na određen način daje putokaz, smjernice za samostalan rad u budućnosti – kroz tu lekciju mogu shvatiti svrhu rada te ciljeve u radu. Drugu lekciju je važno istaknuti jer govori o UNIX ljusci (ljuska je zapravo program koji čita naredbe i pokreće druge programe) koju je važno poznavati jer ona omogućava automatizaciju zadataka koji se često ponavljaju i inače oduzimaju dosta vremena, pomoću poznavanja UNIX ljuske olakšava se cijeli proces, radi se učinkovitije i dobiva na vremenu za druge zadatke koji zahtijevaju više misaone energije. Nadalje opet se ističe automatizacija, ali i lakše manipuliranje podacima uz pomoć naredbi kao što su sort ili grep za pretraživanje i dr. Također još jedna prednost UNIX ljuske je to što se koraci lako mogu zabilježiti u skripti te ih je jednostavno ponovo koristiti. Važno je poznavanje ljuske i u smislu poznavanja programiranja tj. kada se radi u ljusci mogu se primijetiti i kako su to posložene funkcije kada se programira, nadalje vidi se i to rascjepkavanje na manje povezane dijelove – što je ideja i u ljusci i u programiranju. Korištenje ljuske omogućava specifične alate i resurse kojima informacijski stručnjaci inače ne mogu pristupiti, a izuzetno su korisni za različite preformanse. Za kraj korisnici ove lekcije kroz sve navedeno mogu vidjeti važnost svog unosa, instaliranja i uređivanja, izdavanja naredbi, povezivanja naredbi uz pomoć „loops“ i „pipes“, upravljanja podacima, kontroliranja s udaljenog mjesta itd. Treća lekcija obuhvaća rad s Open Refineom, a OpenRefine služi za čišćenje i transformaciju podataka što je iznimno važno danas za informacijske stručnjake kada su podaci dosta neuredni, pomoću OpenRefinea je moguće izvršiti podjelu i granuliranje seta podataka te je moguće na određen način poboljšati podatke uz pomoć iz drugih izvora (vanjski izvori), također moguće je eksportirati podatke u različitim formatima - što su sve važne vještine za rad s podacima u knjižnici ili drugim informacijskim ustanovama koje korisnici mogu naučiti uz pomoć ove lekcije. Četvrta lekcija obuhvaća Uvod u Git, Git/GitHub je važan radi suradnje s kolegama, ističe se po svom suradničkom radu koji je danas potreban i olakšava posao, lakše je npr. objaviti svoj rad ili kod negoli slati mailom; takav način rada i prijenosa podataka ojačava znanstvenu zajednicu. Nadovezujući se na prethodno, informacijski stručnjaci i sami naglašavaju važnost otvorenog koda i dijeljenja radi unaprjeđenja poslovanja i smanjivanja resursa. Git je privlačan za korištenje radi svoje jednostavnosti u smislu da ga je lako postaviti za korištenje te nije potreban server, također u Gitu-u se može pratiti vlastiti prosec i napredak te jednostavno ispravljati greške nadalje nije potrebno naučiti previše naredbi, već se i sa znanjem određenih temeljnih naredbi može dosta ostvariti. Ukratko - važno je korištenje repozitorija te povijesti (version control, commit) za efikasan rad. Sve četiri temeljne lekcije su ovdje navedene radi važnosti za informacijsku djelatnost te se vidi kako se navedene lekcije/kompetencije za informacijske stručnjake nadovezuju - OpenRefine za čišćenje i transformacijju podataka, UNIX Shell je dobar za povezivanje (linking), dok se Git ističe po dijeljenju podataka i radova. Library Carpentry osim temeljnog kurikuluma ima i prošireni kurikulum, u tom proširenom kurikulumu se izdvaja SQL i Webscraping od važnosti za informaciju djelatnost. Naime SQL je bitan za informacijsku djelatnost jer radi s relacijskim bazama podataka s kojima se informacijski stručnjaci susreću u praksi, nadalje uz pomoć SQL-a se može raditi s velikim i kompleksnim skupovima podataka, bazama podataka te se u ovom programskom jeziku ističu alati prepoznati po učinkovitom radu na analizi podataka, a rad s podacima i skupovima podataka te rad na analizi istih je jedan od važnijih zadataka informacijskih stručnjaka te postoji potreba za jezikom koji radi na velikim skupovima podataka koji nisu lagani za analizirati. SQL se ističe i po alatima za pretraživanje te po tome da može služiti i iza izvješćivanje o određenim pojavama i obrascima. Web scraping je jedna od iznimno korisnih vještina za informacijske stručnjake u ovom proširenom kurikulumu, kao što i samo ime kaže Web scraping obuhvaća ekstrahiranje potrebnih podataka s određenih web stranica, najjednostavniji oblik web scrapinga je copy paste, a složeniji i sofisticiraniji oblici obuhvaćaju specifične alate i tehnike koji rade na velikim skupovima podataka – automatizirani proces u kojem je određeno gdje tražiti podatke, koje podatke tražiti, kada stati s pretragom, u kojim intervalima ponavljati radnje kako bi se tražile nove informacije, ali i kako bi se uvidjele promjene u podacima. Naravno za Webscraping je potrebno poznavanje DOM-a i HTML-a, a za posebne situacije (npr. specijalne knjižnice) i poznavanje drugih programskih jezika.

# Data Carpentry

Data Carpentry izrađuje lekcije te provodi radionice koje su potrebne različitim stručnjacima/znanstvenicima kako bi proveli istraživanje, na radionicama korisnici imaju priliku usvojiti nove vještine i kompetencije vezane uz upravljanje podacima te analizu podataka koji su potrebni za istraživanje, zapravo korisnici mogu vidjeti životni ciklus informacija i te spoznaje iskoristiti za svoja istraživanja. Data Carpentry je projekt koji kao i i ostali opisivani projekti u ovom zadatku spadaju pod The Carpentries organizaciju. Data Carpentry lekcije se odvajaju po područjima, tako postoji kurikulum za ekologiju, društvene znanosti, genomsku medicinu, geospacijalne podatke, ekonomiju, astronomiju, digitalnu humanistiku itd. Jedna od lekcija koja se nalazi na ovoj stranici je vezana uz analizu i vizualizaciju podataka što je potrebno informacijskim stručnjacima kako bi iz većih skupova podataka izvukli zaključke te vizualizacija kako bi se rezultati bolje razumjeli, ne samo kako bi informacijski stručnjaci sebi stavili sve u perspektivu pomoću vizualizacije podataka, već da bi i osobe koje su van područja informacijskih znanosti bolje razumjeli određene koncepte. Ta analiza i vizualizacija podataka se može provoditi u Pythonu i R-u, Python je programski jezik koji sadrži alate koji omogućavaju interdisciplinarnost, prilagodljivost (integracija s drugim aplikacijama), efikasnu obnovljivost, te je lagan za naučiti i pomoću njega se stvara snažna zajednica koja može pomoći početnicima u učenju – sve su to razlozi za korištenje ovog programskog jezika u svakodnevnom radu s podacima. Nadalje kada je riječ o Pythonu, na ovoj stranici se nalazi i lekcija o analizi slika u Pythonu što govori o potrebi za raznovrsnosti; pogotovo kada je riječ o informacijskim znanostima jer se građa mijenja te dolazi u različitim oblicima koji zahtijevaju nove pristupe. Lekcija sadrži upute o načinu upravljanja s digitalnim slikama te govori o tome kako slike mogu poslužiti u istraživačkim projektima i na drugačiji način odgovoriti na pitanja. R je također programski jezik otvorenog koda (moduliranje+automazizacija+enkapsulacija) i jednostavan za korištenje s kojim se olakšava proces istraživačkog rada te je izrazito popularan radi zanimljivih grafikona, R se ističe po dobroj vizualizaciji podataka i po svojoj velikoj bazi gdje se nalaze paketi iz vanjskih izvora koji su nadasve praktični pri obavljanju različitih zadataka. Jedna od lekcija se odnosi na organizaciju podataka, a dobra organizacija je ključna u informacijskim djelatnostima, detaljnije ovdje se radi o organizaciji podataka u proračunskim tablicama (spreadsheets) što je jedan od prvih koraka u radu s različitim programskim jezicima, nadovezujući se na prethodno knjižničari i informacijski stručnjaci za početak unose podatke te je tu važno formatiranje tablica kako bi se kasnije moglo upravljati podacima te ih na kraju i eksportirati. Naravno informacijski stručnjaci to već rade, ali je bitno ne izgubiti fokus i prenositi to znanje dalje. Data Wrangling and Processing je vještina koja bi bila korisna informacijskim stručnjacima u radu s velikim skupovima podataka kada bi trebalo identificirati varijance u određenom uzorku koji ima svoj slijed i temu, cilj navedenog je vizualizacija te efikasno korištenje command-line alata s čime se poboljšava produktivnost i učinkovitost. Nadalje lekcija se ističe i po vještinama vrednovanja podataka s čime se informacijski stručnjaci susreću svakodnevno3 i savjeti vezani uz vrednovanje su im i dalje važni jer informacije dolaze u različitim i novim oblicima koji zahtijevaju nove pristupe. Posebno bitne lekcije koje bi trebale biti istaknute zbog svoje važnosti za informacijske stručnjake, a koje su u sklopu Data Carpentry su programiranje u Pythonu te programiranje u R-u što najčešće može poslužiti za analizu i vizualizaciju podataka.

# Software Carpentry

# „Teaching basic lab skills for research computing!“ – cilj organizacije Software Carpentry je naučiti istraživače/znanstvenike računalnim vještinama i kompetencijama kako bi si olakšali zadatke, obavljali ih u manje vremena - zapravo kako bi njihov posao bio učinkovitiji te kako bi se mogli usredotočiti na ono što je bitno, a ne toliko na repetitivne, lagane poslove koji oduzimaju dosta vremena i drugih resursa. Version Control with Git spada u jedne od temeljnih lekcija koje se nude putem ove organizacije, ta lekcija se izdvaja po svojoj praktičnosti za osobe koje rade u informacijskoj djelatnosti jer može pomoći radnicima u ovom sektoru u praćenju svog rada i napretka, sve promjene se automatski zabilježavaju pomoću Automated Version Control alata te se je i jednostavno vratiti na prethodne faze ako je to potrebno. Ovaj alat se zapravo fokusira na verzioniranje, stvaranja različitih verzija koji su nam dostupni jednim klikom, moguće je npr. napraviti više različitih promjena na jednom dokumentu koje će ovaj alat znati organizirati i omogućiti njihovu vidljivost, iz navedenog se vidi kako je ovo alat za promjene, alat koji bilježi i upravlja promjenama što je potrebno u radu jer se tijekom rada mnogo korigira te u dosta slučajeva bez ovog alata dolazi do gubljenja dijelova rada, nemogućnosti pristupa i sl. Još jedna prednost ovog alata za informacijske stručnjake je mogućnost suradničkog rada te otvoreni kod, naime u isto vrijeme paralelno može raditi više ljudi ako je to potrebno. Suradnički način rada je bitan informacijskim stručnjacima ne samo kako bi možda brže došli do novih spoznaja, već i zbog svakodnevne praktičnosti, smanjenja resursa i pametnog iskorištavanja radnog vremena. Također otvoreni kod nije samo za dijeljenje svog rada, već i za bolje mogućnosti u budućem radu, naime radovi koji su dostupni, open access su često vidljiviji i samim time više citiraniji što znanstvenicima otvara nove mogućnosti. Make alat (radi se o jednom malom programskom jeziku koji je poznat po svojim built-in funkcijama kojih ima mnogo) se također izdvaja kao zanimljiv alat za informacijske stručnjake jer je pomoću Makea i naredbi moguće procesirati dokumente te ih takve eksporitrati za vlastite potrebe, također Make služi i za analizu i stvaranje izvješća, vizualizaciju, sažimanje te manipulaciju podataka/dokumenata, u biti Make olakšava proces tako što sam reproducira rezultate te informacijski stručnjaci ne moraju razmišljati o tome, već je automatizirani. Još jedan od razloga za upotrebu ovog alata je to što može raditi u ljusci te može raditi s tri najveća/najrelevantnija operacijska sustava.

|  |
| --- |
|  |

Za kraj možda se u svim gore navedenim lekcijama posebno izdvaja programiranje; važnost programiranja u svakodnevnom radu za informacijske stručnjake i to programiranja u Pythonu koji je jedan od programskih jezika kojeg je jednostavno koristiti, prilagodljiv je i interdisciplinaran te je poznat i po snažnoj zajednici ljudi koji programiraju u Pythonu te programiranja u R-u koji se ističe po sposobnosti rada s velikim skupovima podataka, relacijskim bazama podataka te vizualizaciji podataka. Nadalje ovdje je važno spomenuti i UNIX shell koja omogućava pristup specifičnim alatima određenim kojima informacijski stručnjaci inače ne mogu pristupiti. Make i Automated Version Control bi mogli poslužiti kao najpraktičniji alati pri radu informacijskim stručnjacima.

# Literatura:

1. Data Carpentry . URL: <https://datacarpentry.org> (09-09-2019)
2. Library Carpentry. URL: <https://librarycarpentry.org> (09-09-2019)
3. Software Carpentry. URL: <https://software-carpentry.org> (09-09-2019)