

UNIVERZITET U KRAGUJEVCU
PRIRODNO-MATEMATIČKI FAKULTET
Institut za matematiku i informatiku

Predmet: Doktorska disertacija - teorijske osnove

Tehnologije Semantičkog Web-a u bioinformatici

Tokom prethodnih nekoliko decenija, rapidan razvoj hemijskih, bioloških i tehnologija molekularnog istraživanja, doprineli su nastanku velike količine informacija. Zajedno sa velikim brojem podataka o strukturi proteina, aktivnih supstanci i ekspresiji gena nastala je ogromna količina i velika raznovrsnost bioloških i hemijskih informacija. U cilju predstavljanja, obrade i analize ovih podataka računari su zauzeli primarno mesto. Otuda je i nastala potreba za bioinformatikom. Jednu od najsazetijih i najkonkretnijih definicija bioinformatike formulisao je NCBI (National Center for Biotechnology Information): “*Bioinformatika je polje nauke u kojoj su se biologija, kompjuterske nauke i informacione tehnologije spojile u jednu disciplinu*”.

Kako bi istraživači u domenu bioinformatike mogli lako da izvedu zaključke i donesu konkretne odluke neophodne za dalja istraživanja, oni moraju biti u mogućnosti da na jednostavan način pristupe svim relevantnim podacima. Međutim, pojavom heterogenih izvora podataka, pronalaženje i adekvatno korišćenje informacija od interesa je postalo pravi izazov za istraživače jer su podaci često predstavljeni korišćenjem različitih konvencija, formata i vokabulara. U takvom heterogenom okruženju, istraživači su prinuđeni da često menjaju web lokacije i prate odgovarajuće putanje međusobno povezanih podataka, kako bi došli do relevantnih informacija.

Sa ciljem da se što manje napora uloži u procese predstavljanja, integracije i pronalaženja relevantnih podataka, bioinformatička zajednica je primenila tehnologije Semantičkog Web-a u svojim projektima. Primena tehnologija Semantičkog Web-a omogućava agregaciju heterogenih podataka korišćenjem eksplicitne semantike, korišćenje bogatih i dobro definisanih modela za predstavljanje i integraciju podataka, kao i ponovnu upotrebu podataka i primenu logike u cilju izvođenja zaključaka.

Mnoga rešenja u domenu bioinformatike, kao što su Bio2RDF, Chem2BioRDF i EMBL-EBI, već su uspešno usvojila strategiju primena tehnologija Semantičkog Web-a za rešavanje postojećih bioinformatičkih problema. Želeći da budu u toku sa aktuelnim standardima bioinformatičke zajednice, istraživači CPCTAS (Centar za Prekliničko Ispitivanje Bioaktivnih Supstanci) laboratorije primenili su tehnologije Semantičkog Web-a za predstavljanje i pretragu svojih internih podataka. Za potrebe CPCTAS-a kreirana je PIBAS ontologija, koja opisuje strukturu eksperimenta, nastalu kao rezultat testiranja aktivnih supstanci nad ćelijskim linijama kancera. Na osnovu date ontologije kreirana je ontološka baza podataka, koja sadrži podatke pojedinačnih eksperimenata, takođe u formi ontologija. Vremenom, zahtevi CPCTAS-a su prošireni željom da se integrišu sa postojećim bioinformatičkim rešenjima (Bio2RDF, Chem2BioRDF i EMBL-EBI) u cilju izvođenja različitih Federated SPARQL upita, koji bi im pomogli u procesu pronalaženja novih i komplementarnih informacija neophodnih za dalja istraživanja.

Student: Marija Đokić-Petrović

Mentor: Prof. dr Vladimir Cvjetković

Kragujevac
Oktobar 2017.