**Obrada prirodnih jezika**

Projekat – dokumentacija

Vladimir Milijić 3366/20

Dušan Janić 3317/19

Lena Jokanović 3299/20

Projekat je podeljen na tri glavne faze:

Faza 1: prikupljanje podataka

Faza 2: anotacija podataka

Faza 3: modeli

Svaka od ovih faza je detaljnije opisana u nastavku ovog dokumenta.

FAZA 1 – prikupljanje podataka

U prvoj fazi je bilo potrebno naći parove funkcija-komentar. Zbog toga što je u našem timu bilo tri osobe, bilo je potrebno prikupiti 750 parova funkcija-komentar.

Ovi parovi funkcija-komentar su se uglavnom dohvatali sa projekata sa predmeta “Principi softverskog inženjerstva“ sa osnovnih studija na smeru za softversko inženjerstvo.

Skupili smo 4 projekata iz ovog predmeta na kome je rađen isključivo jezik php i koji je morao biti izkomentarisan odgovarajućim doc blokovima. Pored toga iskorišćen je i privatan repozitorijum u kome se nalazio programski kod isključivo za vežbanje jezika PHP.

Zbog toga što kolege nisu želele da njihovi projekti budu okačeni javno, smatramo ih kao da su privatni repozitorijumi.

Svaki od članova iz tima je imao zadatak da skupi 225 parova funkcija-komentar i da prilagodi već postojeće komentare na nekin način sa datim upitima.

Svaki par funkcija-komentar je smešten u jedinstveni fajl sa unikatnim imenom fajla po obrazcu: redni broj upita \_ ime funkcije.

Naravno, postoji tabela sa nazivom faza\_1 u kojoj su izlistani svi ovi parovi.

Pored gore navedenih parova, postoji jos njih 75 koje smo skupili zajedno da bi se zajednički anotirali. To je ukupno 10% svih parova funkcija-komentar.

Svi prikupljeni podaci su u skladu sa specifikacijama izloženim u postavci projekta.

FAZA 2 – anotacija podataka

Nakon faze prikupljanja podataka, došli smo do trenutka kada je bilo potrebno anotirati prikupljene parove funkcija-kometar.

Podaci su bili anotirani sa četri klase:

0 - Nema sličnosti

1- Slaba sličnost

2- Jaka sličnost

3- Identičnost

Svaki član grupe je samostalno anotirao po 225 podataka onako kako je on mislio da je ispravno.

Ovi podaci su smešteni u tabeli sa nazivom faza\_2.

Preostalih 10% parova funkcija-komentar su anotirani zajednički.

Svrha anotiranja 75 podataka zajednički je bila da se mi kao osobe koje anotiraju podatake na neki način usaglasimo. Svako od članova tima je anotirao podatake kako je on mislio da je ispravno i zatim se vodila diskusija koja od anotacija je bila najviše ispravna. Uglavnom su se mišljenja poklapala izuzev na malom skupu parova funkcija-komentar.

U takvim slučajevima smo birali za anotaciju klasu sa najvećim brojem glasova članova tima.

Ovih 10% od ukupno prikupljenih podataka su izdvojeni u posebnu tabelu sa nazivom…

Time je bila završena druga faza i u tom trenutku smo bili spremi na pređemo na fazu 3.

FAZA 3 - modeli