

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 2324

**Orthobalancer: aplikacija za
kreiranje skupova bioloških vrsta
usporedive taksonomijske širine**

Ivan Slijepčević

Zagreb, svibanj 2012.

*Umjesto ove stranice umetnite izvornik Vašeg rada.
Da biste uklonili ovu stranicu obrišite naredbu \izvornik.*

zahvale

SADRŽAJ

Popis slika	v
Popis tablica	vi
1. Uvod	1
2. Teoretski uvod	2
2.1. Homologija proteina	2
3. Podaci	3
4. Metode	4
5. Implementacija	5
5.1. Cjevovod	5
5.2. tax	6
5.3. server	6
6. Rezultati	7
7. Zaključak	8
Literatura	9

POPIS SLIKA

POPIS TABLICA

1. Uvod

2. Teoretski uvod

ključna uloga proteina

srodnost / evolucija / otkrivanje

2.1. Homologija proteina

Homologija u biološkom smislu predstavlja slične osobine među vrstama na različitim razinama organizacije života, poput organa, tkiva, stnice ili molekule. Homologne osobine uočene među jedinkama različitih vrsta obično upućuju na zajedničke pretke tih vrsta u evoluciji. Međutim, u molekularnoj biologiji termin homolog se često koristi i za naznačavanje sličnosti, bez obzira na genetsko srodstvo [1]

Za homologne sekvence proteina kažemo da su ortologne kad su direktni potomci neke sekvence u zajedničkom pretku, bez da su prošle duplikaciju gena. Drugim riječima, ortologne sekvence se mogu naći u jedinkama različitih vrsta, a obavljaju istu funkciju u svim tim vrstama. Paralogne sekvence su homologne sekvence koje su nastale od dvije različite kopije nekog gena koji je prošao kroz proces duplikacije gena u nekom zajedničkom evolucijskom pretku. Paralozi se mogu naći u jedinkama jedne ili više vrsta te obavljaju slične funkcije.

chart ortho-para

ideja...

3. Podaci

s FASTA formatu

blast nr baza

taxonomy baza

ulaz

izlaz

4. Metode

pipeline neki dijagram za pipeline (sequence / activity / state)

tax dio pseudokod

neki sequence dijagram za sve

izlaz

(prebaciti u neki teex file za server) server slike (ulaz, zamjenjivi, izvršavanje, kraj, error)

5. Implementacija

Aplikacija je pisana u programskom jeziku Python verzije 2.7. Aplikacija se dijeli u nekoliko zasebnih cjelina. U središtu aplikacije nalazi se cjevovod koji poziva alate poput BLAST-a i fastacmd-a za komuniciranje sa NCBI-jevom neredundantnom bazom, zatim dio aplikacije za odabir i balansiranje vrsta na taksonomskom stablu te alat mafit za poravnanje sekvenci. Pored cjevovoda implementirana je web aplikacija kao korisničko sučelje za cijeli program. Web aplikacija je implementirana koristeći Flask microframework, dok su operacije na klijentskoj strani implementirane u javascriptu uz korištenje biblioteke jQuery.

5.1. Cjevovod

Cjevovod je arhitektonski programski obrazac u kojem prolaze kroz filtre koji su postavljeni jedan za drugim. Time se simulira jedan tok koji ulazne podatke transformacijom kroz filtre generira izlazne podatke. U ovome projektu cjevovodna arhitektura je samo logički kostur koji se enkapsulira unutar razreda *Pipeline*. Iako je u začetku razvoja aplikacije svaki filter bio zaseban proces, vrlo ubrzo je ustanovljeno kako većina filtera generira podatke koji su potrebni na raznim mjestima u cijeloj aplikaciji te se činilo lakše imati sve podatke u memoriji pojedinog cjevovoda. To je omogućilo da razred *Pipeline* naslijedi razred *Thread* iz modula *threading* te se može pozivati kao zasebna dretva.

Tok cjevovoda se može vidjeti na slici 5.1. Ulaz u cjevovod predstavljaju paralogni proteini u FASTA formatu koje zadaje korisnik. Ti se podaci zadaju pri stvaranju objekta *Pipeline* kako bi se mogli zapisati na disk u direktorij vezan za instancu *Pipeline-a*. Stvarni objekt kojeg prima konstruktor *Pipeline-a* je riječnik prilagođen uporabi servera, što je detaljnije opisano u sekciji 5.3.

Pri pokretanju cjevovoda za svaku se od unešenih sekvenci stvara objekt razreda *ProteinHolder* prilikom čega se obavljaju pozivi filtera nezavisnih za svaku pojedinu sekvencu. Prvi filter koji se koristi je alat *BLAST* te je izveden kao poziv zasebnog

programa *blastall* na sljedeći način: `language=Pascal blastall -p blastp`

i tako dalje...

SLIKA CJEVOVODA (prikazuje podatke a ne filtre radi boljeg shvaćanja)

tekst o objektima

SLIKA objekata (class dijagram)

5.2. tax

5.3. server

dio po dio

jQuery

komunikacija pipelinea i klijenta, log, dekoratori

automatizacija (posla kroz cjevovod->treba ići u neko uvodno poglavlje o namjeri)

robusnost

6. Rezultati

7. Zaključak

LITERATURA

- [1] Andreas D. Baxevanis. *Bioinformatics and The Internet*. John Wiley & Sons, Inc., 2002. ISBN 9780471223924.

**Orthobalancer: aplikacija za kreiranje skupova bioloških vrsta usporedive
taksonomijske širine**

Sažetak

Ključne riječi:

**Orthobalancer: web application for creation of taxonomically balanced sets of
orthologous protein sequences**

Abstract

Keywords: