

Algoritmi korišćeni pri izradi projekta - **student 1: RA99/2016 Igor Gligorević**

Algoritam dodavanja nove reči u trie stablo:

- 1) Instancira se globalna promenljiva koja predstavlja trie stablo.
- 2) Od korisnika se traži unos absolutne putanje do foldera koji će biti posmatran kao "root" direktorijum.
- 3) Iterira se kroz fajls sistem počevši od root direktorijuma koji je unešen u prethodnom koraku. Proverava se da li je u pitanju html fajl. Ukoliko jeste prelazi se na korak 3, a ako nije nastavlja se dalje sa proverom ostalih fajlova u hijerarhiji.
- 4) Sledi parsiranje html dokumenta koje vraća sve reči koje se nalaze unutar html tagova parsiranog dokumenta.
- 5) Za svaku reč koju je parser pronašao poziva se metoda trie stabla, namenjena za dodavanje reči u strukturu podataka trie.
 - a. Ako se ne nalazi dodaje se u rečnik sa slovom kao ključem a za vrednost se instancira novo stablo. I onda se prelazi na korak 7.
- 6) Algoritam dodavanja započinje proverom da li se prvo slovo prosleđene reči nalazi u ključevima rečnika koje predstavlja polje trie stabla. Ako se nalazi prelazimo na korak broj 7.
 - a. Ako se ne nalazi dodaje se u rečnik sa slovom kao ključem a za vrednost se instancira novo stablo. I onda se prelazi na korak 7.
 - b. Ukoliko jeste duža od jednog karaktera onda se rekurzivno poziva, nad vrednošću rečnika(rečnik je polje klase) za ključ koji predstavlja prvo slovo reči koja je parametar funkcije, ista funkcija samo što se ovaj put prosleđuje reč koja predstavlja podstring gde je prvi karakter izopšten. I onda se opet ponavlja korak 6 i 7 i tako sve dok se ne dođe do koraka 7a.
- 7) Vrš se proveru da li je reč koju dodajemo duža od 1 karaktera.
 - a. Ukoliko nije duža tj. sadrži 1 karakter, unutar rečnika koji predstavlja polje stabla(tj. trenutnog čvora koji je takodje predstavljen istom klasom) se inkrementira vrednost koja se odnosi na broj pojavljivanja reči, ta vrednost vezana je za ključ koji predstavlja dokument u kojem se ta reč nalazi. Ukoliko ne postoji u rečniku dokument za koji se reč dodaje u trie stablo, on se dodaje kao ključ sa početnom vrednosti 1. Algoritam dodavanja se ovde zaustavlja.
 - b. Ukoliko jeste duža od jednog karaktera onda se rekurzivno poziva, nad vrednošću rečnika(rečnik je polje klase) za ključ koji predstavlja prvo slovo reči koja je parametar funkcije, ista funkcija samo što se ovaj put prosleđuje reč koja predstavlja podstring gde je prvi karakter izopšten. I onda se opet ponavlja korak 6 i 7 i tako sve dok se ne dođe do koraka 7a.

Algoritam pretrage trie stabla:

- 1) Od korisnika se traži unos reči za pretragu (Pretraga se vrši nad trie stablom koje je prethodno kreirano), gde format unesene reči može biti:
 - a. Samo jedna reč – proverava da li se zadata reč nalazi u nekom od html dokumenata koji se nalaze u okviru fajl sistema čiji je root odabran u početku rada s aplikacijom.
 - b. Više reči odvojenih razmakom - proverava da li se bar jedna od reči nalazi u nekom od html dokumenata
 - c. Dve reči odvojene oznakom OR - proverava da li se bar jedna od dve reči nalazi u nekom od html dokumenata
 - d. Dve reči odvojene oznakom AND - proverava da li se bar jedna se obe reči nalaze u nekom od html dokumenata
 - e. Dve reči odvojene oznakom NOT- proverava da li se prva reč nalazi a druga ne nalazi u nekom od html dokumenata
- 2) Po unosu korisnika poziva se funkcija za pretraku koja obrađuje korisnicki unos tako da se ispuni traženo ponašanje navedeno u nabranju koraka 1. U slučajevima 1.c, 1.d i 1.e nakon obrade teksta poziva se funkcija koja vraća setove u kojima se nalaze putanje do fajlova u kojima se pojavljuje tražena reč i broj ponavljanja za svaki od fajlova. Funkcija je opisana u koraku 3. U slučajevima pretrage reči odvojenih razmakom ili samo jedne reči korak 7.
- 3) Unutar funkcije koja kreira setove, pozivom funkcije za pretragu nad trie stablom vrši se pretraga trie stabla. Funkciji se prosleđuje reč za koju se pretražuje trie stablo.
- 4) Algoritam pretrage započinje proverom da li se prvo slovo prosleđene reči nalazi u ključevima rečnika koje predstavlja polje trie stabla nad kojim se funkcija poziva. '
 - a. Ukoliko se ne nalazi, algoritam se završava i vraća se prazan rečnik. Zaključuje se da reč nije sadržana ni u jednom html fajlu.
 - b. Ukoliko se nalazi prelazi se na korak 5
- 5) Vrš se prover da li je reč koja je parametar funkcije duža od jednog karaktera.

- a. Ako nije vraća se polje koje je rečnik unutar kog ključevi predstavljaju nazive dokumenata u kojima se reč pojavljuje, a vrednost predstavlja broj pojavljivanja te reči. Algoritam se završava.
 - b. Ako je duža od jednog karaktera onda se rekurzivno poziva, nad vrednošću rečnika(rečnik je polje klase) za ključ koji predstavlja prvo slovo reči koja je parametar funkcije, ista funkcija samo što se ovaj put prosleđuje reč koja predstavlja podstring gde je prvi karakter izopšten. Proces se ponavlja, ide se u korak 4 i 5, sve dok se ne dođe do koraka 5.a
- 6) Povratne vrednosti funkcije za pretragu trie stabla se smestaju u setove i vraćaju se kao povratna vrednost funkcije. U zavisnosti od formata korisnickog unosa razlikuje se tretman vraćenih setova. Tu su sledeći slučajevi:
- a. 1.c OR – Radi se unija dva vraćena seta
 - b. 1.d AND – Radi se presek dva vraćena seta
 - c. 1.e NOT – Radi se komplement dva vraćena seta
- 7) U slučaju pretrage jedne reči ili više reči odvojenih razmakom, poziva se funkcija za pretragu trie stabla za svaku reč koju je korisnik uneo. Funkcija pretrage trie stabla opisana je u koracima 4 i 5. Rezultat pretrage se smešta u set.

