

Korisničko uputstvo

Stem Tool - Program za ručno označavanje stemova

Ključne reči

Stem - osnovni oblik reči, nastao (uglavnom) odsecanjem sufiksa.

Stemer - programski alat koji automatski označava stem za svaku reč ulaznog dokumenta.

Stem Tool - program na koji se odnosi ovaj dokument, služi za ručno označavanje stemova.

Namena

Cilj programa je da se ručno označi stem za svaku reč u ulaznom dokumentu. To je značajan deo istraživanja jer se dobija ljudski označen rezultat koji će u kasnijim fazama služiti za određivanje performansi stemera.

Način rada

Ulaz

Ulaz programa je fajl koji u svakoj liniji ima po jednu rečenicu. Tokeni u rečenici su odvojeni razmacima. To je dokument koji treba označiti. Ako već ima označenih reči koje pripadaju tom dokumentu, program će automatski učitati i mapirajući fajl (videti poglavlje Način rada - Izlaz).

Izlaz

Izlaz programa su dva fajla:

1) Mapirajući fajl - Sadrži mapiranje reč - stem, da bi se za svaku reč sačuvao stem. Takođe za svaku reč se čuva i kontekste u kojima se ta reč pojavljuje, kao niz stringova razdvojen uspravnim crtom. Fajl je u TSV (Tab Separated Value) formatu. Kolone su reč, stem i konteksti.

Format linije izlaznog mapirajućeg fajla: reč \t stem \t kontekst₁|kontekst₂|...|kontekst_n

2) Zamenski fajl - Sadrži isti sadržaj kao i originalni, ulazni fajl, samo je umesto svake reči čiji je stem označen upisan upravo taj označeni stem.

GUI komponente

Menu

File - Open

Učitava se ulazni fajl sa korpusom za anotaciju. Pored tog fajla, učitava se i mapirajući fajl sa imenom <ulazni_fajl_bez_ekstenzije>_word_stem_map.txt. Ako mapirajući fajl ne postoji, učitava se samo korpus.

File - Save

Čuva rezultat u izlaznim fajlovima (videti poglavlje Način rada - Izlaz). Nazivi izlaznih fajlova su:

1) Mapirajući fajl: <ime ulaznog fajla>_word_stem_map.txt

2) Zamenski fajli: <ime ulaznog fajla>_replace.txt

View - Language

Bira se jezik na kojem će biti ispisane poruke.

Kontekst

U ListView komponenti su ispisani konteksti trenutne reči. U svakoj liniji se nalazi po jedan kontekst. Kontekst se formira od 5 reči sa leve i 5 reči sa desne strane trenutne reči. Ovo ne važi u situacijama kada se pojavi kraj rečenice pre nego što se izbroji 5 reči, sa bilo koje strane. U tom slučaju kontekst ide do kraja rečenice.

Dugme Izdvoji iz grupe

Služi za situacije kada se pojave reči koje imaju isti oblik u ulaznom dokumentu, a različit stem. Primer:

1. Selo ima hiljadu stanovnika. Stem za imenicu "selo" je "sel" (razmotriti: od sela, ka selu, seoski)
2. Dete je selo na stolicu. Stem za glagol "sesti" je "se" (razmotriti: sednem, sedneš, seo, sela, selo, sestu)

Šta je ispravan stem za reč "selo" zavisi od toga o kojoj vrsti reči se zapravo radi.

U ovakvim situacijama moramo da razdvojimo reči sa različitim kontekstima. Dugme Izdvoji iz grupe radi upravo to. Da bi dobio rezultat potrebno je izabrati bar jedan kontekst iz grupe konteksta trenutne reči (pogledati poglavlje Način rada - GUI komponente - Kontekst). Klikom na dugme kreiraju se dve nove reči, sa sufiksima 1 i 2. Ako se deli već podeljena reč, sufiks se postavlja na prvi naredni broj.

Da bi dugme proizvelo željenu akciju, potrebno je:

- 1) Izabrati bar jedan red iz ListView komponente sa kontekstima trenutne reči.
- 2) Ostaviti bar jedan element neizabran, da bi se pravilno formirale barem dve nove reči.

Dugme će biti aktivirano samo ako su uslovi ispunjeni.

Trenutna reč

Trenutno izabрана reč. Svako slovo trenutne reči ima svoju "kućicu". Klikom na kućicu reč se deli na stem i ostatak. Stem je obojen plavom bojom, ostatak crvenom. Promena stema se vidi u TextView polju (videti poglavlje Način rada - GUI komponente - TextView). Inicijalno, za svaku reč stem je jednak samoj reči.

Strelice

Strelice služe za "pomeranje" stema na levo ili desno. Mogu se koristiti alternativno kliku na kućice.

Text View

U ovom polju stoji tekstualni prikaz stema trenutne reči. **Tekst koji je ovde ispisan, predstavlja stem koji će biti označen kada se klikne na dugme Označi stem.** Razlog je to što ponekad nije moguće označiti stem prostim odsecanjem, već je potrebno napisati šta je ispravan stem.

Primer: Reč uzročni, stem uzrok.

Dugme Ukloni stem

Ako je trenutna reč u listi označenih reči, njen stem se uklanja i reč se prbacuje u listu neoznačenih reči.

Dugme Označi stem

Trenutna reč, sa svojim stemom zapisanim u TextView komponenti prelazi u listu označenih reči.

Dugme Spoji grupe

Radi inverznu operaciju od dugmeta Izdvoji iz grupe. Reči koje se javljaju u istom obliku a javljaju se u različitim kontekstima mogu se ponovo spojiti. **Prilikom spajanja stem se resetuje, postaje jednak originalnom obliku reči. Takođe, resetuje se redosled reči, automatski, tako da lista uvek ide od 1 do n.** Npr. ako imamo listu reči

a_1, a_2, a_3, a_4, spajanje a_2 i a_4 rezultovaće novom listom: a_1, a_2, a_3. **Prilikom spajanja proveriti novu numeraciju reči.** Da bi dugme proizvelo akciju, potrebno je:

- 1) Iz liste neoznačenih ili označenih reči izabrati najmanje dva elementa, da bi se postojale grupe za spajanje.
- 2) Izabrani elementi moraju imati istu osnovu. Npr. *selo_1*, *selo_2* je validan izbor, dok *a_1* i *ne_1* nije.

Dugme će biti aktivirano samo ako su uslovi ispunjeni.

Lista neoznačenih reči

Reči iz ulaznog dokumenta koje nisu označene.

Lista označenih reči

Reči čiji su stemovi označeni. Sa leve strane je reč, sa desne stem. **Potrebno je sačuvati rezultat u fajlovima, da ne bi nestao prilikom zatvaranja programa.**

Prečice na tastaturi

ENTER - Označavanje stema.

Ctrl + S - Čuvanje rezultata.

Left arrow, right arrow - Pomeranje stema levo i desno.

Pokretanje

Iz komandne linije: stemTool.jar

Korisnička podrška

Za uočene nepravilnosti, bagove i sugestije obratite se mejlom na dejan.golubovic.27@gmail.com