

Zbirka primerov rabe vejice Vejica 1.3

Peter Holozan
Amebis, d. o. o.

Napovednik

- Predstavitev zbirke primerov rabe vejice
- Izboljšave nove verzije
- Izboljšanje rezultatov zaradi izboljšav
- Popravljanje vejic v nestandardni slovenščini
- Rezultati Besane
- Kako naprej

Zbirka primerov rabe vejice

- Obstoječa Vejica 1.0
- Sestavljena iz štirih delov
 - KUST
 - Wikipedija
 - Šolar
 - Lektor

Format zbirke primerov

- Manjkajoče vejice označene z znakom α , odvečne nadomeščene z znakom \div
- Vsaka poved v svoji vrstici
- Spredaj s tabulatorjem ločena oznaka, sestavljena iz oznake dela, poddela in zaporedne številke v poddelu
- *KUST.de.4 Ko sva prišli do skupine α so vsi vprašali α kaj smo kupili in midve tudi.*
- *KUST.de.5 Danes α če gremo ven α bom oblekla mojo novo obleko, se že veselim.*

Izboljšave za novo verzijo

- Ročno preverjanje primerov
 - Preverjanje ni bilo sistematično, preverjeni primeri, na katere je opozorila Besana
 - *Določite§ kateri korporaciji pripadajo naslednje znamke:*
- Izločanje tujih povedi
- Popravki napak pri upoštevanju lektorskih popravkov
 - *Kaj vi mislite menite in zakaj?*
- Dodaten del iz korpusa Janes-Vejica 1.0
 - 500 tvitov v nestandardni slovenščini
 - Ročno označene vse odvečne in manjkajoče vejice

Izboljšanje popravljjanja vejic zaradi izboljšav

- Izboljšanje predvsem pri delu iz korpusa Lektor
 - Priklic: iz 40,5 % na 50,5 %
 - Natančnost: iz 18,8 % na 35,5 %
- Za odvečne vejice
 - Priklic: iz 26,8 % na 40,5 %
 - Natančnost: iz 29,7 % na 58,4 %

Sestava zbirke Vejica 1.3

Del	Število povedi	Število vejic	Delež manjkajočih	Delež odvečnih
KUST	388	221	66,16 %	31,37 %
Šolar	49438	49125	24,96 %	7,30 %
Lektor	52121	71204	1,57 %	1,00 %
Wikipedija	869	929	11,78 %	7,10 %
Janes	1368	646	37,46 %	3,00 %
SKUPAJ	104184	122125	13,10 %	3,77 %

Podpora za nestandardno slovenščino

- Podpora za tvite (začetni pozivi (@), #)
- Neuporaba strešic in velikih začetnic
- Nestandardno besedišče, tipične redukcije pri pisanju
- Koristno za Besano, da zna popraviti take elemente v besedilih
- Virtualni agenti lahko bolje odgovarjajo na vprašanja v nestandardni slovenščini

Trenutni rezultati Besane

del	priklic	natančnost
KUST	87,04 %	94,35 %
Šolar	77,62 %	90,76 %
Lektor	50,45 %	38,06 %
Wikipedija	79,84 %	86,73 %
Janes	49,48 %	79,26 %
SKUPAJ	75,59 %	85,78 %

Kako naprej

- Razširiti zbirko primerov (SSJ500k, KAS)?
- Nadaljnje dopolnjevanje obstoječih delov

- Izboljševanje Besane
 - Analizator povedi
 - Izpusti (elipse)
 - Oznake citatov