

psi toolkit prezentacja

10 listopada 2017

zarys prezentacji

instalacja

tokenizacja

lematyzacja

struktura psi-toolkit

instalacja

1. zainstaluj dockera
2. docker pull skorzewski/psi-toolkit (run as root)
3. sprawdź czy działa:
echo "pies szczeka na kota sudo docker run -i skorzewski/psi-toolkit lamerlemma -lang pl

tokenizacja

tokenizacja: echo Ała ma kota"! sudo docker run -i
skorzewski/psi-toolkit tokenize --lang pl od teraz piszę tylko: echo
Ała ma kota"! tokenize --lang pl

tokenizacja

tokenizacja: echo Ała ma kota sudo docker run -i
skorzewski/psi-toolkit tokenize -lang pl od teraz piszę tylko: echo
Ała ma kota tokenize -lang pl

tokenizacja

lematyzacja: echo Ała ma kota sudo docker run -i
skorzewski/psi-toolkit lemmatize --lang pl

lematyzacja

kawałek przykładowego outputu:

##	beg.	len.	text	tags	annot.text
01	0000	03	Ala	!pl,token	Ala
02	0000	03	Ala	!pl,lamerlemma,morfologik-tagset	
03	0000	03	Ala	!pl,lamerlemma,lemma,morfologik-	
04	0000	03	Ala	!pl,lamerlemma,lexeme,morfologik-	
05	0000	03	Ala	!pl,lamerlemma,lemma,morfologik-	
06	0000	03	Ala	!pl,lamerlemma,lexeme,morfologik-	
07	0000	03	Ala	!pl,form,lamerlemma,morfologik-t	

używanie psi-toolkit

psi toolkit składa się z 3 podstawowych części:

- readery
- writery
- annotatory

przykład korzystania z komponentów przy użyciu pipe

poprzedni przykład można napisać z wygodnym formatowaniem dzięki writerowi:

```
echo Ała ma kota | head -text ! tokenize ! lemmatize ! write-simple  
-tags lemma
```

output:

```
A1 | A1a | A1o
```

```
mieć | mój : a
```

```
kot : s1 | kot : s2 | kota
```

przykład korzystania z komponentów przy użyciu pipe

jak widać w przykładzie linuxowe pipe — możemy zamienić na !
pipe pełni w psi-toolkit taką samą rolę jak w bashu
dzięki użyciu ! nie mamy problemów z pracą z dockerem echo Ała
ma kotahead-text ! tokenize ! lemmatize ! write-simple --tags
lemma
output:

```
A1|A1a|A1o
```

```
mieć|mój:a
```

```
kot:s1|kot:s2|kota
```

dokumentacja i tutoriale

psi-toolkit jest dobrze udokumentowany i ma proste w zrozumieniu tutoriale:

<http://psi-toolkit.amu.edu.pl>