

Kategorialny Parser Składniowo-Semantyczny „ENIAM”

instrukcja użytkownika

Wojciech Jaworski

21 stycznia 2017

1 Opis

Kategorialny Parser Składniowo-Semantyczny „ENIAM” jest narzędziem generującym formy logiczne dla zdań w języku polskim. Parser pracuje na niepreparowanych danych, realizuje kolejne etapy przetwarzania tekstu: tokenizację, lematyzację, rozpoznawanie związków składniowych, anotację sensami słów oraz rolami tematycznymi, częściową dezambiguację oraz tworzenie reprezentacji semantycznej.

Parser przyjmuje na wejściu niepreparowany tekst, a na wyjściu prezentuje reprezentacje semantyczne kolejnych zdań składających się na ten tekst. Reprezentacje semantyczne składają się z grafu semantycznego i formuły logicznej. Są one szczegółowo opisane w rozdziale 3.

Dla każdego zdania generowane jest do 10 alternatywnych propozycji reprezentacji. W przypadku, gdy podział tekstu na zdania jest niejednoznaczny (np. zdania zawierają skróty kończące się kropką) prezentowane są efekty przetwarzania wszystkich możliwych podziałów na zdania.

UWAGA: Wpisywany tekst powinien być poprawny. W szczególności zdania muszą kończyć się kropką lub innym znakiem interpunkcyjnym. Literówki są częściowo obsługiwane — są interpretowane jak nieznane słowa, a błędne końcówki fleksyjne prowadzą do uzyskania innych niż zamierzone powiązań semantycznych.

2 Użytkowanie

2.1 Wersja demonstracyjna

Oficjalne demo parsera znajduje się pod adresem <http://eniam.nlp.ipipan.waw.pl>. Niestety z przyczyn technicznych działa dość wolno (zwłaszcza przy wpisywaniu pierwszego zapytania). Dostępne jest też nieoficjalne demo (<http://students.mimuw.edu.pl/~wjaworski/eniam.html>), które działa istotnie szybciej.

2.2 Instalacja

2.3 Wersja wsadowa

Aby korzystać z parsera należy wpierw uruchomić serwer **pre** (znajduje się w podkatalogu **pre**). Serwer uruchamia się od 4 do 10 minut zależnie od sprzętu i wymaga przynajmniej 4GB RAM. Gdy serwer będzie gotowy do użycia wypisze „Ready!”.

Aby sparsować pojedyncze zdanie „Ala ma kota.” należy, będąc w katalogu **parser** wpisać komendę:

```
./eniam 'Ala ma kota.'
```

Spowoduje to wygenerowanie do katalogu wyjściowego formy logicznej dla zadanego zdania.

Aby sparsować korpus należy dostarczyć ENIAM-owi plik zawierający zdania rozdzielone znakiem nowej linii. Do parsowania korpusów przystosowana jest współbieżna wersja ENIAM-a, którą uruchamia się poleceniem

```
./eniam.distr file
```

gdzie **file** jest plikiem zawierającym korpus.

Plik **config** zawiera konfigurację ENIAM-a. Można w nim określić położenie katalogów z zasobami leksykalnymi i Walentym. Ustalić port i serwer na którym jest dostępne **pre**. Wskazać ścieżkę do katalogu wyjściowego. Określić liczbę generowanych rozwiązań, timeout parsera, ograniczenie pamięci dla parsera oraz liczbę współbieżnie uruchamianych procesów parsera przez **eniam.distr**.

W katalogu wyjściowym generowane są formy logiczne w postaci grafów semantycznych i formuł rozszerzonej logiki pierwszego rzędu. Grafy semantyczne zapisane są w XML oraz wizualizowane jako obrazki PNG. Z kolei formuły logiki pierwszego rzędu wyrażone są w MathML.

3 Definicja reprezentacji semantycznej