

Projekt JiMP 2

Automatyczna generacja tekstu

Szymon Maslowski

Semestr letni 2015

Specyfikacja funkcjonalna programu Markov

Nazwa

markov - program generujący tekst losowy na podstawie zadanego tekstu/tekstów
– interfejs linii poleceń

Przykładowe użycie :

\$ markov 2 tekst1.txt tekst2.txt 130 wynik.out

\$ markov test2 200 4 -S statystyki.out

Streszczenie

markov [-d](opcjonalne) nazwa_pliku_z_baza_ przejsc n {nazwy plików wejściowych} -długość_tekstu_wyjściowego -(opcjonalna)ilość_akapitów -plik_wynikowy -(opcjonalna)długość_n-gramu [-S]-S statystyki-(opcjonalna)nazwa_pliku [-c](opcjonalne) nazwa_pliku_do_zapisu_bazy_przejsć

Opis

Markov to program generujący, na podstawie analizy tekstów, losowe ciągi słów w języku naturalnym.

-n

Ilość plików wejściowych zawierających tekst, domyślnie 1.

-d nazwa_pliku_z_baza_przejsc

Możliwe jest wywołanie programu podając gotową już bazę przejść, stworzoną przy poprzednich wywołaniach programu.

-nazwy plików wejściowych

N plików wejściowych binarnego tekstu, zawierających tekst w języku naturalnym, na podstawie, których zostanie wygenerowany tekst wyjściowy

-długość tekstu wyjściowego

Docelowa długość wygenerowanego tekstu. Margines błędu 5 %

-ilość akapitów

Ilość akapitów, które ma mieć tekst, objętość akapitów będzie taka sama z niepewnością 5%. Domyślnie 1 akapit

-plik wynikowy

Nazwa pliku, do którego zostanie zapisany efekt działania programu. Jeżeli nie podano program wypisze wynik do konsoli

-długość n-gramu

Długość n-gramu na podstawie, którego budowane są zależności w słowach tekstu. Domyślnie 3

-S statystyki

Jeżeli wybrano program z tym parametrem, to oprócz normalnego wygenerowania tekstu losowego, program wypisze też do zadanego pliku (jeżeli nie podano to wypisze do konsoli) statystyki dotyczące najczęściej wykorzystywanych słów, najczęściej powtarzających się n-gramów, PMI (różnica między oczekiwaną częstotliwością występowania ciągów słów, a oczekiwaną)

-c nazwa_pliku_do_zapisu_bazy_przejsć

Możliwe jest zachowanie bazy możliwych przejść w pliku o podanej nazwie. Zawartość zostanie nadpisana! W celu tylko zwiększenia istniejącej bazy należy wywołać program podając teksty wejściowe i bazy danych, tekst wyjściowy długości 0 i wybrać ten parametr.

Błędy w użyciu

W razie wystąpienia błędów, komunikat zostanie wypisany do pliku "markov_errs" i na strumień błędów. Program stara się wygenerować tekst bez błędów, dlatego w razie wystąpienia nie właściwych parametrów wejściowych, stosowna informacja zostanie wydrukowana na standardowy strumień błędów a program spróbuje

działać dalej przyjmując wartości domyślne.

Np.:

```
$ markov 2 130 wynik.out
```

markov: blad! nie podano plikow wejsciowych!

```
$ markov 2 tekst1.in tekst2.in -130 wynik.out
```

markov: blad! ilosc tekstu nie moze byc ujemna!

Inne

Kontakt : szymmaslowski@gmail.com

Wykonane jako praca zaliczeniowa na projekt z JiMP2.
Politechnika Warszawska 2015.