

Generator tekstu

Specyfikacja funkcjonalna

Cel:

- generacja tekstu na podstawie wczytanych do programu plików tekstowych;

Użytkownik docelowy:

- użytkownik posiadający co najmniej podstawową wiedzę na temat obsługi komputera i linii komend;

Opis programu:

- program generuje tekst z wykorzystaniem łańcuchów Markowa;
- parametrami tekstu wyjściowego jest ilość akapitów i ilość słów;
- do generacji tekstu wykorzystywane są n-gramy, gdzie n jest wybierane przez użytkownika;
- program generuje pliki pośrednie na podstawie których można wygenerować teksty i , pliki pośrednie można łączyć;

Opcje programu:

- wybór liczby słów (opcjonalnie, domyślna wartość);
 - wybór liczby akapitów (opcjonalnie);
 - wybór tekstów bazowych (opcjonalnie); *
 - wybór plików pośrednich (opcjonalnie); *
 - wybór parametru n (opcjonalnie, domyślna wartość);
 - wybór nazwy pliku wyjściowego (opcjonalnie, domyślna wartość);
 - wybór lokalizacji zapisu plików wyjściowych (opcjonalnie, domyślna lokalizacja);
- * - jedna z opcji musi zostać wybrana obowiązkowo;

Uruchomienia:

- uruchomienie i obsługa odbywa się z linii komend;

Sposób uruchomienia:

- w katalogu w który zlokalizowane są pliki programu wpisać komendę,
`./generator [-b nazwa_pliku] [-p nazwa_plik]
<-g nazwa_generowanego_tekstu> <-s liczba_całkowita>
<-a liczba_akapitów> <-gs>`

-b użytkownik podaje tekst bazowy (lub kilka);
-p użytkownik podaje plik pośredni (lub kilka);
-g będzie generowany tekst;
-s liczba słów;
-a liczba akapitów;
-gs generowany będzie plik pośredni;

[] - argumenty obowiązkowe
<> - argumenty opcjonalne

Opis działania programu

W zależności od podanych przy uruchomieniu programu parametrów program może obrać trzy główne ścieżki działania:

1. użytkownik podał tekst bazowy
 - a) generowany jest plik pośredni;
 - b) generowany jest tekst (opcjonalnie);
2. użytkownik podał tekst bazowy i plik pośredni
 - a) generowany jest plik pośredni i łączony jest drugim plikiem;
 - b) generowany jest tekst na podstawie połączonych plików pośrednich (opcjonalnie);
3. użytkownik podał plik pośredni
 - a) generowany jest tekst (opcjonalnie);

Format danych:

- teksty bazowe zapisane w formacie txt
- plik pośredni zapisany w postaci pliku o rozszerzeniu psr
Plik będzie zawierał wyrazy z tekstu bazowego, gdzie po każdym wyrazie są wypisane wyrazy, które bezpośrednio po nim występowały (wraz z liczbą ich wystąpień w tekście).
- tekst wyjściowy zapisany w formacie txt;

Obsługa błędów:

- nie podana wartość n - domyślna 2;
- nie podana nazwa pliku wyjściowego - domyślna "book";
- nie podana liczba słów - domyślna 100;
- nie podany żaden plik wejściowy - komunikat;
- plik o domyślnej nazwie "book" już istnieje - nastąpi dopisanie na końcu pliku po przerwie długości wiersza;

Testy:

- testy będą przeprowadzane na podstawie plików tekstowych zamieszczonych na stronach:
 1. <http://wolnelektury.pl> (język polski)
 2. <https://www.gutenberg.org> (język angielski)
- wyniki będą polegały na subiektywnej ocenie jakości wygenerowanych plików i ich zgodności z parametrami wprowadzonymi do programu;