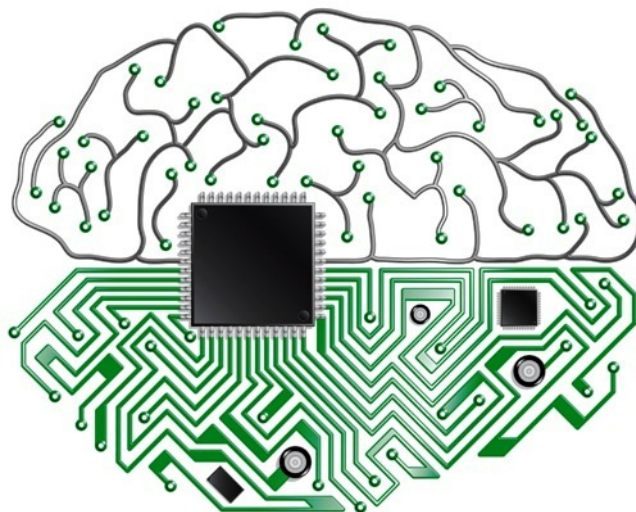


Specyfikacja funkcjonalna programu "Noobot".



Projekt na zajęcia: *"Języki i Metodyka Programowania"*
wydz. Elektryczny Politechniki Warszawskiej, rok akademicki 2014/2015
wykonał: Paweł Drapiewski
pod kierownictwem: mgr inż. Zuzanny Krawczyk

1. Opis ogólny programu

Program jest symulacją rozmowy ze sztuczną inteligencją za pomocą czatu, który będzie tworzył tekst na podstawie łańcuchów Markova. Dodatkowo program będzie pozwalać "zajrzeć" w algorytm programu generując odpowiednie dane w panelu "Umysł Noobot'a".

W tym przypadku użytkownikiem docelowym jest osoba nie zaawansowana w obsłudze systemów operacyjnych, gdyż komunikacja odbywa się za pomocą "okienek", dzięki czemu obsługa programu będzie się sprowadzać do wpisywania dowolnego tekstu, który zostanie przekazany do "Noobota".

2. Opis funkcjonalności

Program uruchamiany jest najprościej przy pomocy ikony. Jedynym wymogiem jaki musi spełnić użytkownik to posiadanie wirtualnej maszyny JAVA, którą można pobrać i zainstalować na wszystkich znaczących systemach operacyjnych. Komunikacja z nim będzie polegała na wpisywaniu w odpowiednie okienko tekstu, a następnie wysłanie tejże wiadomości przy pomocy guzika "wyślij" lub też klawisza ENTER. Program będzie analizował tekst, i na jego podstawie tworzył swój po czym w oknie rozmowy będzie ukazywała się odpowiedź. Dodatkowo w panelu "Umysł Noobot'a" po tej operacji będą generowane statystyki, m.in. drogę jaką szedł przy generacji tekstu, gdzie dla każdego n-gramu będą wypisywane możliwe przyrostki, oraz zaznaczanym będzie ten wybrany. Oprócz tego znajdują się tam statystyki takie jak: prawdopodobieństwo wystąpienia tego zdania, oraz wykres słupkowy ilości wystąpień danego słowa w bazie danych.

Czyli podsumowując, program będzie wykonywał następujące czynności:

1. Analiza tekstu wejściowego
2. Generacja tekstu na podstawie przeanalizowanego tekstu
3. Powiększanie bazy danych o słowa wpisywane przez użytkownika
4. Przy zamykaniu będzie zapisywał bazę by móc jej użyć przy kolejnym uruchomieniu
5. Możliwość zmiany bazy danych na inną
6. Generowanie danych "wewnętrznych" algorytmu generatora (czyli opisany powyżej "Umysł Noobot'a")
7. Analiza wejściowych słów, a w przypadku gdy wpisywane słowo nie będzie występować w bazie to będzie wyświetlać odpowiedź słowa, które ma tyle samo pierwszych n-słów, a gdy jest ich kilka to wybierze ten o najczęstszym występowaniu.

3. Format danych i struktura plików

3.1 Struktura katalogów

Program będzie dostarczany w postaci pliku archiwum *.jar, z podstawową predefiniowaną bazą danych. Czyli będą to pliki z kodem dla maszyny javy i ewentualnie elementy graficzne (i te elementy będą niezmiennie w czasie) oraz plik/pliki binarne z bazą danych, która będzie się zwiększać w miarę korzystania z programu, a także pojedynczy plik tekstowy UTF-8 w której zapisana jest aktualna konfiguracja programu.

4. Scenariusz działania programu

4.1 Scenariusz ogólny

1. Uruchom i załaduj bazę danych z pliku
2. Gdy naciśnięto "Zmień bazę danych" uruchom okno do zmiany bazy danych.
3. Czekaj aż użytkownik zacznie wpisywać tekst
4. Analizuj wpisywany tekst
5. Wygeneruj odpowiedź
6. Wygeneruj statystyki do panelu "Umysł Noobot'a"
7. Wróć do punktu 2
8. Zakończenie programu może nastąpić w każdym momencie, poprzez kliknięcie przez użytkownika przycisku "Zakończ program"

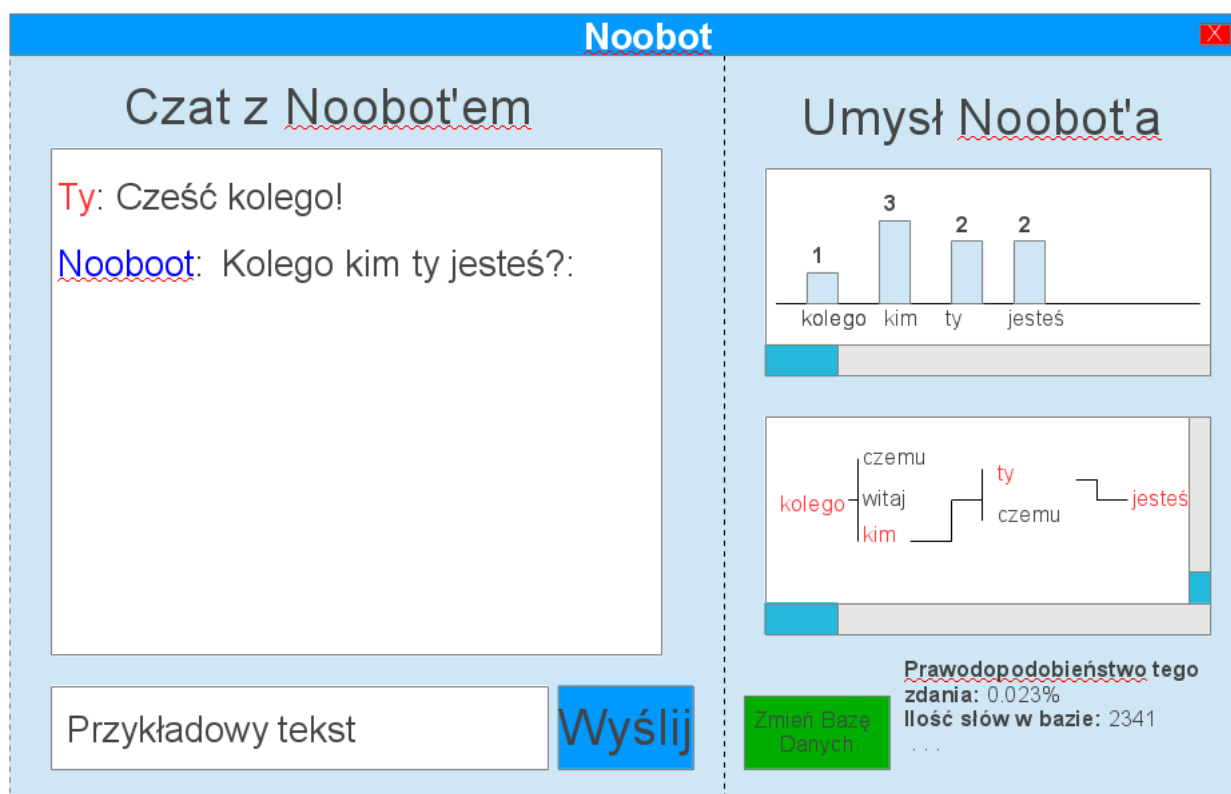
4.2 Scenariusz szczegółowy

1. Uruchom i załaduj bazę danych, której nazwa jest wyspecyfikowana w pliku konfiguracyjnym programu.
2. Gdy naciśnięto "Zmień bazę danych" uruchom okno do zmiany bazy danych.
3. Czekaj aż użytkownik zacznie wpisywać tekst
4. Analizuj wpisywany tekst
 - (a) Dla każdego wpisywanego słowa
 - (b) Gdy wpisano co najmniej 3 litery
 - (c) Sprawdzaj dla każdej nowej litery czy tworzone słowo zgadza się z którymś kolwiek słowem z bazy (wielkość liter nie ma znaczenia), a gdy nie to:
 - (d) Wyszukaj słowa które są identyczne z wpisywanym słowem zmniejszonym o tę jedną literę
 - (e) Wyświetl jako podpowiedź słowo które występowało najczęściej
5. Wygeneruj odpowiedź
 - (a) Wylosuj ngram z tekstu wprowadzonego przez użytkownika i przyjmij je za pierwszy ngram dla generatora.
 - (b) Określ ile słów ma wyprodukować poprzez wylosowanie jakiejś wartości całkowitej z zakresu $\pm 20\%$ długości tekstu wprowadzanego
 - (c) Generuj kolejne ngramy na podstawie łańcuchów Markova, póki nie osiągniesz odpowiedniej długości lub póki nie będzie dalszego przejścia
6. Wygeneruj statystyki do panelu "Umysł Noobot'a"
 - (a) Policz prawdopodobieństwo wystąpienia tego zdania mnożąc ilorazy ilości wystąpień i ilości ogólnej słów dla każdego ngramu. A następnie wypisz je do panelu.
 - (b) Sprawdź ile każde z wygenerowanych słów występuje w bazie, a następnie wygeneruj wykres słupkowy dla tych wszystkich słów

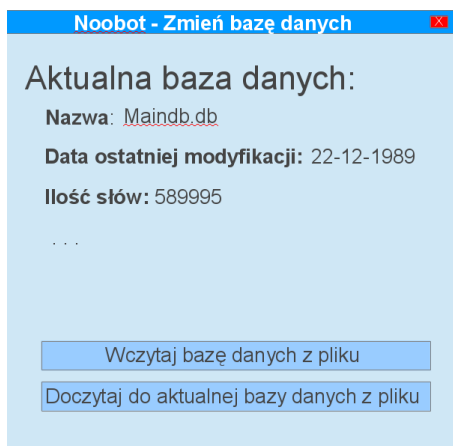
- (c) stwórz graf pokazujący możliwe przejścia z ngramu na ngram, na podstawie ngramów wygenerowanych. Każdy ngram ma pokazywać dostępne sufiksy, oraz ten który został wybrany.
7. Zakończenie programu może nastąpić w każdym momencie, poprzez kliknięcie przez użytkownika przycisku "Zakończ program"

4.3 Ekran działania programu

W tym programie będą dostępne dwa okna: Ekran główny, gdzie będzie w jednym panelu będzie odbywać się konwersacja, a w drugim będą generować się statystyki, oraz z okna dodawania bazy danych, w którym zobaczymy informację o aktualnej bazie danych (jej nazwa, ilość słów,...) oraz przycisk do zmiany bazy danych.



Rys. 1. Okno główne



Rys. 2. Okno zmiany bazy danych

5. Testowanie

Testowanie programu odbędzie się poprzez ręczną obsługę programu dla różnych zestawów danych. Oraz przy pomocy osób zewnętrznych, których zadaniem będzie korzystanie z programu i zgłaszanie ewentualnych uwag i problemów.