

Пример работы доктора, обученного на хорошем датасете “sentenses_for_learning.txt” с n-граммой длины 3.

who are you?
** Nikita
hello, nikita!
what seems to be the trouble?
** i feel depressed
you're important to me.
** thank you doctor
is there anything you want to talk about it?
** it is hard for me to wake up in mornings
you're a good person.
** thank you doctor
i'm here for you.
** i don't have energy for my work
this isn't an ending.
** i don't know what to do with it
when did you begin feeling like this?
** month or two ago...
i'm sure we can figure it out together.
** i feel depressed
that sounds like it's really hard.
** yes it is
this isn't an ending.
** thank you doctor
depression can be defeated.
** i don't think so
your feelings are valid.
** it is hard to wake up in mornings
depression is a serious condition.
**** can you say that i am not alone?**
you're not alone.
** goodbye
Goodbye, nikita!
See you next week!
next!
who are you?
** stop
working day finished

В выделенном жирным курсивом вопросе-ответе срабатывает генерация ответа по n-1-грамме (где n=3). Встретив “not alone”, генератор начал построение предложения в левую и правую стороны от нее. Слева он построил “you're” и остановился, т.к. в качестве предыдущего слова ему выпала точка. Справа он добавил точку, как слово, следующие за “alone”.

Пример работы доктора, обученного на плохом датасете - книге психоаналитика Джексона "jackson.txt" с n-граммой длины 3.

next!
who are you?
** Nikita
hello, nikita!
what seems to be the trouble?
** i feel depressed
he has fed his flock.
** thank you doctor
no, thank you, and in many cases been dropped out of gear.
** i dont have energy for my work
again and is a costly proceeding, an inner and spiritual process very different from all foods
finally reaches the rectum, it is far too compelling to be.
** i dont have energy for my work
all the petting he craves, is capable also of displacement or the energy for creative work.
** i dont know what to do with it
how many times it pays to know what made her play tennis every day.
** month or two ago...
suggestibility is one or two before the period of parental care.
** i feel depressed
i made her a stranger in the past.
** can you say i am not alone?
neither can you wash it out of the physician may be.
** goodbye
Goodbye, nikita!
See you next week!
next!
who are you?
** stop
working day finished

Структура данных, в которую происходит обучение, выглядит следующим образом:

(starts . order),

где **starts** - хэш-таблица, в которой ключами являются n-1-граммы, перед которыми в тексте встречаются конец предложения ".?!", а значениями - количество раз, которые соответствующая n-1-грамма встречается в обучающих текстах.

order - тоже хэш-таблица, где ключом является n-1-грамма.

Значением в таблице order является пара (**prevs . nexts**), где **prevs** и **nexts** - хэш-таблицы с ключами-словами и значениями-вхождениями этих слов до и после n-1-граммы соответственно.

Сохранение результатов

Сохранение результатов происходит при помощи функции **writeFile**, в качестве параметров к которой передаются описанная выше структура данных и название файла, в котором ее следует сохранить.

Чтение сохраненных ранее результатов происходит при помощи функции **readFileAsObject**.

Можно слить две пары (starts . order) при помощи функции (merge pair1 pair2), при этом первый параметр изменится, дополнившись данными из второго.

Нельзя дополнить пару (starts . order), если она была прочитана из файла. Это связано с тем, что хэш-таблицы читаются из файлов как не мутабельные. Если мы всё-таки хотим смержить две пары, обе из которых были прочитаны из файлов, нужно создать новую пару и мержить их в нее:

```
(define newPair (cons (make-hash) (make-hash)))  
(merge newPair oldPair1)  
(merge newPair oldPair2)
```