

**Московский государственный технический  
университет им. Н.Э. Баумана**

**Факультет «Информатика и системы управления»  
Кафедра ИУ5 «Системы обработки информации и управления»**

**Курс «Парадигмы и конструкции языков программирования»**

**Отчет по домашнему заданию**

Выполнил:

студент группы ИУ5-35Б  
Серопегин Я.Р.

Подпись и дата:

Проверил:

Гапанюк Ю.Е.

Подпись и дата:

Москва, 2024 г

**Описание:**

Разработана программа парсер, предназначенная для автоматического сбора обучающего датасета для генеративных нейронных сетей, способных поддерживать и вести диалог. Тип датасета: текстовый, представляет собой диалоги.

**Принцип работы:**

Дан файл с запросами для генерации диалогов. Программа, управляя курсором компьютера, автоматически по очереди вставляет запрос на сайт, ждёт генерацию диалогов и копирует результат, записывая диалоги в файл в необходимом формате обучения.

**Программа:**

```
from datetime import datetime
import random
import time
import pyautogui
from tkinter import Tk
f = open('autoZap.txt', 'r', encoding='utf-8')
zap = f.read().split('\n')
f.close()
root = Tk()
def get_bo():
    root.update()
    return root.clipboard_get()
def set_bo(val):
    root.clipboard_clear()
    root.clipboard_append(val)
    root.update()
    return
def delsk(text):
    new = ''
    f = True
    for i in text:
        if i == '(':
            f = False
        elif i == ')':
            f = True
            continue
    if f:
```

```

        new += i
    return new
def parse(info):
    print('info:\n', info)
    if info[0] == '#':
        info = info.split('\n', 1)[1]
    if info.count('*') > 3:
        info = info.replace('*', '')
    info = info.replace('\n\n', '\n').replace('[model:
toolbaz_v3.5_pro]', '').replace('"', '').replace('--- Конец
диалога ---', '').replace('[Конец диалога]', '')

    info = delsk(info)

    if '\nПервый' in info and '\nВторой':
        info = info.replace('Первый :', 'Первый:
').replace('Второй :', 'Второй: ').split('Первый: ', 1
        )[1].replace('Первый: ',
'\n\\n').replace('Второй: ', '\n*\n').replace('\n\n', '\n')
    else:
        infos = info.split('\n')
        info = ''
        for i in range(len(infos)):
            if i%2==0:
                info += infos[i]+'*\n'
            else:
                info += infos[i]+'\\n\\n'
        while '\n ' in info or '\n:' in info:
            info = info.replace('\n ', '\n').replace('\n:',
'\n')
        f = open('parseAutoDialog.txt', 'a', encoding='utf8')
        f.write(info+'\\n\\n\\n')
        print('itog:\n', info)
        f.close()
def GO():
    global step
    print(zap[step])
    time.sleep(3)
    set_bo(zap[step])
    pyautogui.moveTo(800, 300, duration=0.25)
    pyautogui.click()

```

```

time.sleep(1)
pyautogui.keyDown('ctrl')
pyautogui.press('a')
pyautogui.keyUp('ctrl')
pyautogui.press('backspace')
pyautogui.keyDown('ctrl')
pyautogui.press('v')
pyautogui.keyUp('ctrl')
pyautogui.moveTo(800, 600, duration=0.25)
pyautogui.click()
time.sleep(25)
pyautogui.moveTo(1320, 180, duration=0.25)
pyautogui.click()
pyautogui.scroll(470)
time.sleep(5)
info = get_bo()
if step < len(zap)-1:
    if info != '' and info != zap[step]:
        print('run')
        try:
            parse(info)
            step += 1
        except:
            pass
    else:
        exit()
g = open('autoZapCheckpoint.txt', 'r', encoding='utf-8')
step = int(g.read())#452
g.close()
print('начать с', step)
while True:
    GO()
    g = open('autoZapCheckpoint.txt', 'w', encoding='utf-8')
    g.write(str(step))
    g.close()
    time.sleep(7+0)

```

## Результаты:

Было собрано около 22 МБ диалогов:

parseAutoDialog-1.txt	19.07.2024 16:58	Текстовый докум...	407 КБ
parseAutoDialog-2.txt	26.07.2024 23:49	Текстовый докум...	221 КБ
parseAutoDialog-3.txt	26.07.2024 23:52	Текстовый докум...	412 КБ
parseAutoDialog-4.txt	15.09.2024 18:01	Текстовый докум...	190 КБ
parseAutoDialog-5.txt	16.09.2024 14:47	Текстовый докум...	393 КБ
parseAutoDialog-6.txt	17.09.2024 21:31	Текстовый докум...	539 КБ
parseAutoDialog-7.txt	22.09.2024 17:31	Текстовый докум...	3 314 КБ
parseAutoDialog-7-new-gpt.txt	25.09.2024 21:43	Текстовый докум...	3 284 КБ
parseAutoDialog-8.txt	24.09.2024 23:05	Текстовый докум...	1 570 КБ
parseAutoDialog-8-delstar.txt	25.09.2024 21:15	Текстовый докум...	1 523 КБ
parseAutoDialog-9.txt	26.09.2024 22:51	Текстовый докум...	973 КБ
parseAutoDialog-10 (не очищен).txt	01.10.2024 10:56	Текстовый докум...	1 223 КБ
parseAutoDialog-11.txt	04.10.2024 20:08	Текстовый докум...	1 160 КБ
parseAutoDialog-12.txt	08.10.2024 13:56	Текстовый докум...	953 КБ
parseAutoDialog-13.txt	24.10.2024 22:40	Текстовый докум...	1 272 КБ
parseAutoDialog-14.txt	25.10.2024 19:50	Текстовый докум...	1 248 КБ
parseAutoDialog-14-опир.txt	25.10.2024 19:37	Текстовый докум...	1 253 КБ
parseAutoDialog-15.txt	04.11.2024 18:01	Текстовый докум...	1 021 КБ
parseAutoDialog-16.txt	21.11.2024 19:41	Текстовый докум...	2 169 КБ

19 элем.: 22,5 МБ |

**Каждый файл имеет следующий вид (пример из одного файла):**

\\

Вот, держи, эту штуку прикрути сюда.

\*

Ага, а что это вообще? Крыло от космического корабля?

\

Не, это стабилизатор. Нужен, чтобы самолёт не кувыркался как белка в колесе.

\*

А если белка в колесе будет стабилизатором?

\

Тогда мы получим летающую белку-робота. Запатентуем идею.

\*

Боюсь, патент уже принадлежит профессору Злодею.

\

Не может быть! Он же изобретает злодейские гаджеты, а не летающих белок!

\*

А если это его секретное оружие?

\

Ладно, хватит про белок. Давай лучше эту детальку прикрутим, а то самолёт без неё как курица без головы.

\*

Или как белка без колеса?

\

Прекрати!

\*

Хорошо, хорошо. Вот, прикрутил. Крепко?

\

Надеюсь. А то потом на испытаниях развалится, и мы будем собирать обломки по всему двору.

\*

Будет весело, как в конструкторе LEGO.

\

Только без инструкций.

\*

Ага, полный рандом.

\

Слушай, а ты уверен, что это всё правильно?

\*

Ну, вроде бы да. Вроде бы я в интернете видел такую же конструкцию.

\

В интернете? Там же всё врут!

\*

Ну, не всё. Там есть и полезные вещи. Например, рецепты вкусных пирогов.

\

Или инструкции по сборке самолётов.

\*

Именно!

\

А если мы ошиблись в размерах?

\*

Тогда наш самолёт будет либо слишком маленьким, либо слишком большим.

\

Или слишком квадратным.

\*

Или слишком фиолетовым.

\

Или слишком говорящим.

\*

Не надо. Хватит.

\

Ладно, ладно. Просто шучу.

\*

Вот, теперь осталось только вот эту штуку вот сюда...

\

Осторожно, не сломай!

\*

Да не бойсь, у меня руки золотые.

\

Золотые? Они же в мазуте.

\*

Ну, да. Маслянисто-золотые.

\

Лучше бы они были просто чистые.

\*

Зато блестят!

\

Как новогодняя ёлка.

\*

Ага, а наш самолёт – как игрушка на ней.

\

Главное, чтобы не упал.

\*

Не упадет. Я верю в нашу летающую конструкцию.

\

И я. Хотя, если честно, немного боюсь.

\*

Страх – это нормально. Но мы полетим!

\

Надеюсь, что да. А вдруг он полетит в сторону дома?

\*

Тогда будем учиться управлять самолетом в полете.

\

И куда же он полетит?

\*

В облака!

\

Ну, если только он не разобьется при взлете.

\*

Не разобьется! У нас же есть надежда и супер клея.

\

Супер клей! Реально спасет.

\*

Конечно, спасет! Вот, смотри, готово! Наш самолёт готов к полету!

\

Ух ты! Круто получилось!

\*

Теперь нужно только найти место для взлета и запустить его.

\

Давай запустим его на лужайке, там место побольше.

\*

Отлично! Нас ждет незабываемое приключение!

\\

Ты уверен, что он здесь вообще?

\*

Ну, должен же он где-то быть! Не мог же он просто испариться.

\

Может, в другой мир порталился, ключик-то волшебный?

\*

Ха-ха, очень смешно. Ладно, давай по порядку. Куда ты его последний раз видел?

\

Хм... кажется, в руках держал, когда шел на кухню за чаем...

\*

А потом?

\

Потом... потом... точно! Я же оставил кружку на столе, а ключ положил на подоконник, чтобы...

\*

Чтобы что?

\

Чтобы он зарядился солнечной энергией и открыл портал в другой мир!

\*

Ага, конечно.

\

Ну, а вдруг?



\*

Вряд ли. Ладно, проверяем подоконник.

\

Пусто.

\*

Вот же ж...

\

Может, он упал?

\*

Куда он может упасть с подоконника?

\

На пол, например. Или за батарею.

\*

Давай проверим.

\

Ничего.

\*

Может, ты его потерял еще раньше?

\

Нет, точно на подоконнике клал.

\*

А может, он сам ушел искать приключения?

\

Думаешь, он умеет ходить?

\*

А почему бы и нет? Может, у него маленькие ножки и он ползает по квартире.

\

Ха-ха, ну ты и фантазер!

\*

А что, вдруг он сейчас под диваном прячется?

\

Давай проверим.

\*

Ой, а тут пылица!

\

Давай лучше не будем тут лежать.

\*

Согласен.

\

Может, он под ковром?

\*

Нет, проверял уже.

\

А в тазике с грязным бельем?

\*

Ты издеваешься?

\

А вдруг он решил постираться?

\*

Нет, точно нет.

\

Ну ладно, ладно. Давай по-честному осмотримся. Может, он просто где-то на виду, а мы его не замечаем.

\*

Да, пожалуй, так и сделаем.

\

Ты сегодня какой-то рассеянный.

\*

Сам знаешь, как много дел было.

\

Ну да, ну да. А завтра еще больше будет, если ключ не найдем вовремя.

\*

Опасно, опасно!

\

Ладно, шутки в сторону. Давай еще раз все очень тщательно проверим.

\*

Хорошо.

\

Вот, смотри! Под стулом!

\*

Не может быть!

\

Нашел!

\*

Ура! Спасибо!

\

Не за что. А то я уже начал думать, что он сам сбежал в путешествие.

\*

Да, без него никак. Ну что, пойдём, чай пить?

\

Пойдём, пойдём. Заряженный солнечной энергией чай будет особенно вкусным!

\*

Ага, может, даже портал откроет!

\

Тогда давай быстрее, а то пропустим!

\\