

# **Парсерные программы для сайтов с ранжированными списками приемных комиссий ВУЗов**

**Руководство пользователя**

# Содержание

|  |    |
|--|----|
| 1 Введение.....  | 3  |
| 1.1 Область применения.....  | 3  |
| 1.2 Краткое описание возможностей.....   | 3  |
| 1.3 Уровень подготовки пользователя.....   | 3  |
| 1.4 Перечень эксплуатационной документации, с которой<br>необходимо ознакомиться пользователю..... | 3  |
| 2 Назначение и условия применения.....   | 4  |
| 2.1 Виды деятельности, функции.....  | 4  |
| 2.2 Программные и аппаратные требования к системе.....   | 4  |
| 3 Подготовка к работе.....   | 5  |
| 3.1 Входные данные.....  | 5  |
| 3.2 Запуск программ.....   | 5  |
| 3.2.1 Программа «parser».....  | 5  |
| 3.2.1 Программа «parser_all».....  | 9  |
| 3.2.1 Программа «parser_snils».....  | 11 |
| 4 Формирование названий файлов и папок для программ<br>«parser_all» и «parser_snils».....          | 14 |

# **1 Введение**

## **1.1 Область применения**

Данное ПО предназначено для получения csv-таблиц с ранжированными списками приемных комиссий ВУЗов.

В ходе работы создается запрос по заданному url-адресу веб-страницы с таблицей, извлекаются данные и формируется csv-файл с полученными данными в указанной папке.

## **1.2 Краткое описание возможностей**

ПО включает в себя три парсерные программы:

- parser
- parser\_all
- parser\_snils

Программа «parser» работает с одной ссылкой, передает данные в файл (название указывает пользователь) и сохраняет данные в папку, которую указал пользователь.

Программа «parser\_all» получает список ссылок, с которых необходимо выгрузить таблицы (см. п. 3.1), формирует в указанной пользователем директории папки по ВУЗам и загружает в них созданные файлы.

Программа «parser\_snils» получает список ссылок, с которых необходимо выгрузить таблицы (более подробно см. п. 3.1). Для каждой строки таблицы проверяется, имеется ли данный СНИЛС в файле snils.txt (см. 3.1). Если да, то он сохраняется в файл. Программа формирует в указанной пользователем директории папки по ВУЗам и загружает в них созданные файлы.

## **1.3 Уровень подготовки пользователя**

Использование данного ПО не требует специальной подготовки пользователя.

## **1.4 Перечень эксплуатационной документации, с которой необходимо ознакомиться пользователю**

Перед использованием по необходимо ознакомиться с эксплуатационной документацией:

«Парсерные программы для сайтов с ранжированными списками приемных комиссий ВУЗов. Руководство пользователя»

## **2 Назначение и условия применения**

### **2.1 Виды деятельности, функции**

Данное ПО предназначено для автоматизации следующих процессов:

- получение таблиц с заданной веб-страницы;
- загрузка полученной таблицы в файл;
- сохранение файла в заданную пользователем папку;
- получение таблиц с множества ссылок, загрузка в файлы, систематизация файлов по папкам;
- фильтрация данных в таблицах;
- все взаимодействие с программой проходит через простой и понятный интерфейс;
- после успешного завершения работы программы вызывается соответствующее информационное сообщение;
- в случае возникновения ошибки вызывается соответствующее информационное сообщение;
- при попытке закрыть программу вызывается соответствующее уточняющее информационное сообщение.

### **2.2 Программные и аппаратные требования к системе**

Аппаратные требования: процессор x86, монитор, клавиатура, мышь.

Программные требования: программа IDLE Python.3, установленные библиотеки requests, pandas, pathlib, BeautifulSoup, tkinter, messagebox, datetime, os, fake\_useragent, UserAgent.

## 3 Подготовка к работе

### 3.1 Входные данные

Перед запуском программ необходимо загрузить следующие файлы:

- file.txt;
- file1.txt;
- sails.txt.

Считывание и загрузка информации из файлов производится автоматически, поэтому редактировать их не нужно.

Файл «file.txt» сохраняет адрес папки и имя файла для программы «parser».

Файл «file1.txt» содержит перечень ссылок на страницы с ранжированными списками приемных комиссий СГУ, СГТУ, ГУАП, СПбГЭУ, СПбГУ, МЭИ, МТУСИ, РУДН, РТУ МИРЭА, МИСИС, МИФИ, МФТИ.

Файл «sails.txt» содержит список СНИЛС всех абитуриентов СГУ КНиИТ.

Если файлы помещены в нужный каталог и имеют корректный формат, то считывание информации из них и запись данных будет производится автоматически во время работы программы.

### 3.2 Запуск программ

#### 3.2.1 Программа «parser»

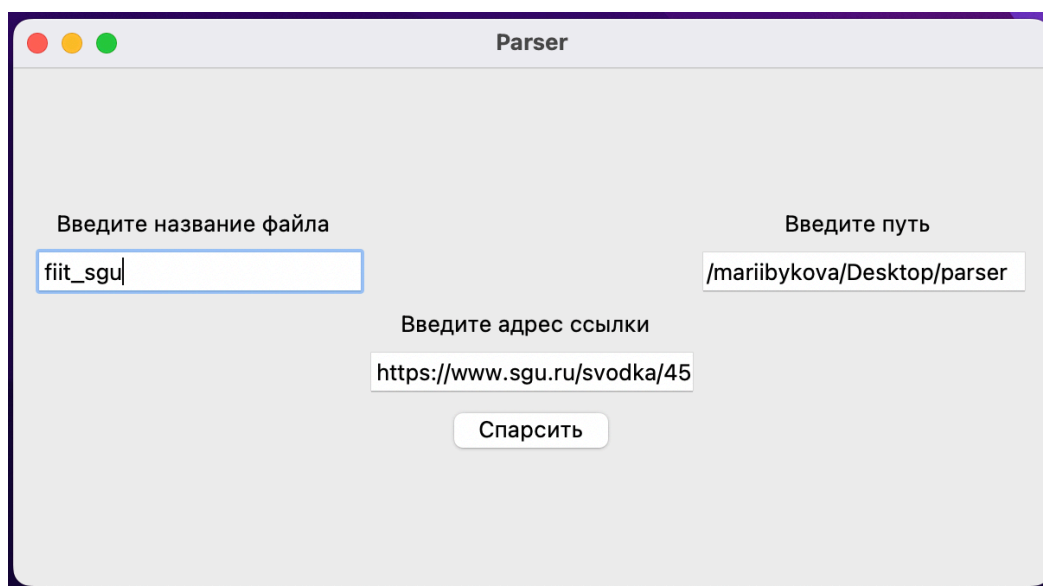
При запуске программы откроется следующее окно:



The image shows a screenshot of a software window titled "Parser". The window has a standard macOS-style title bar with three colored buttons (red, yellow, green) on the left. The main area of the window is light gray and contains three text input fields. The first field is labeled "Введите название файла" (Enter file name) and is located on the left. The second field is labeled "Введите путь" (Enter path) and is located on the right. The third field is labeled "Введите адрес ссылки" (Enter link address) and is located in the center. Below these three fields is a single button labeled "Спарсить" (Parse).

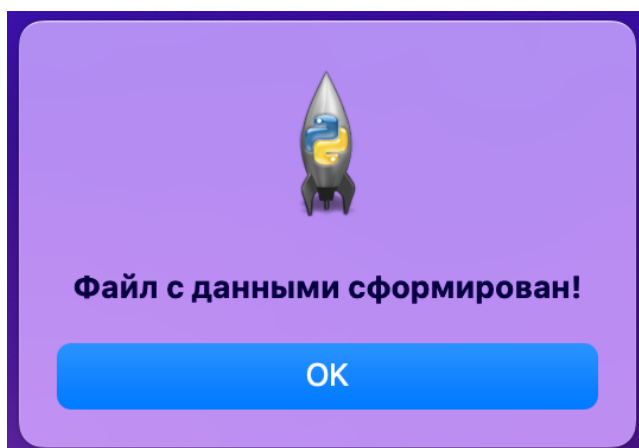
При первом запуске программы необходимо ввести адрес ссылки, с которой нужно спарсить таблицу, путь, куда нужно сохранить сформированный файл, и название, под которым нужно его сохранить. При дальнейшей работе с программой необходимо каждый раз вводить только адрес ссылки. Последние введенные имя файла и путь будут автоматически сохраняться программой в файле «file.txt» (изменять его нельзя). Если нужно обновить эти данные, введите их в соответствующие поля.

После ввода всех данных окно будет иметь вид:

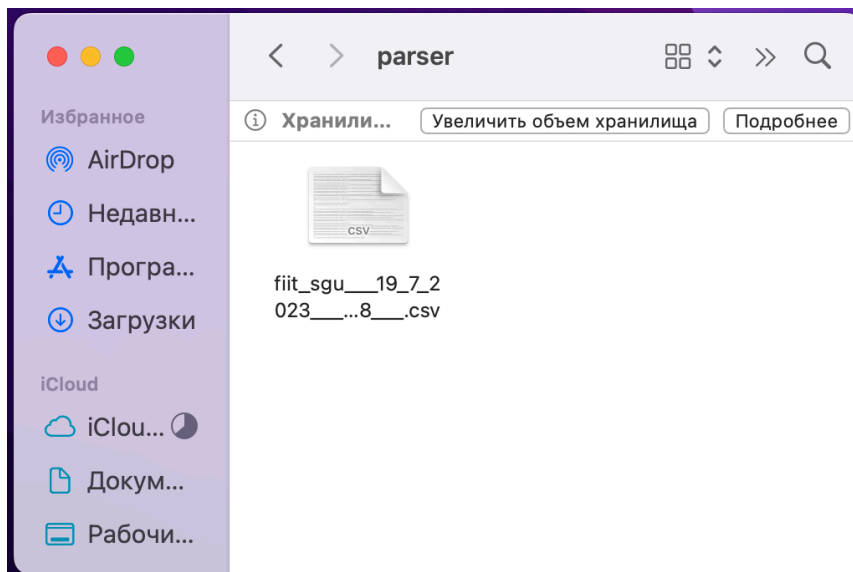


The screenshot shows a window titled "Parser" with a light gray background. It contains three text input fields and one button. The first field, labeled "Введите название файла", contains the text "fiit\_sgu". The second field, labeled "Введите путь", contains the text "/mariibyкова/Desktop/parser". The third field, labeled "Введите адрес ссылки", contains the text "https://www.sgu.ru/svodka/45". Below these fields is a button labeled "Спарсить".

Дальше нужно нажать на кнопку «Спарсить». После этого начнется процесс сбора данных, формирование файла и сохранение его в указанную папку. После успешного завершения работы программы появится следующее информационное окно:

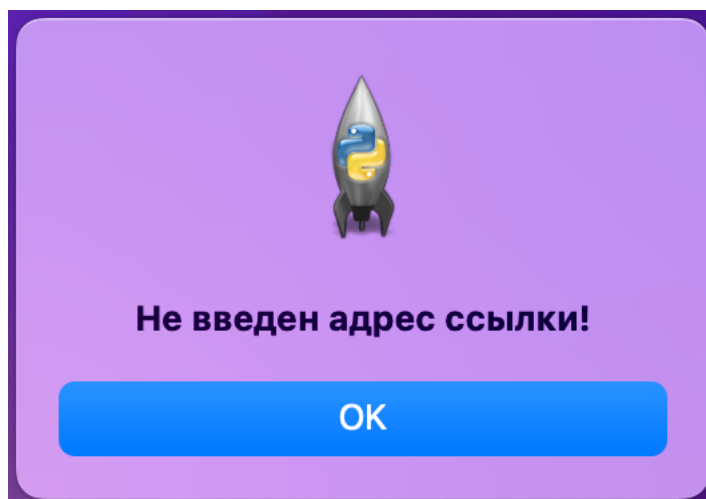


Тогда в указанной папке будет сохранен сформированный файл:

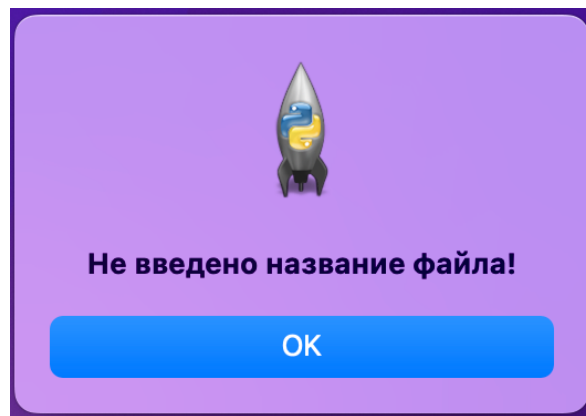


Обратите внимание: к названию файла добавляется дата и время его формирования. Так будет удобнее отслеживать изменения на интересующей странице.

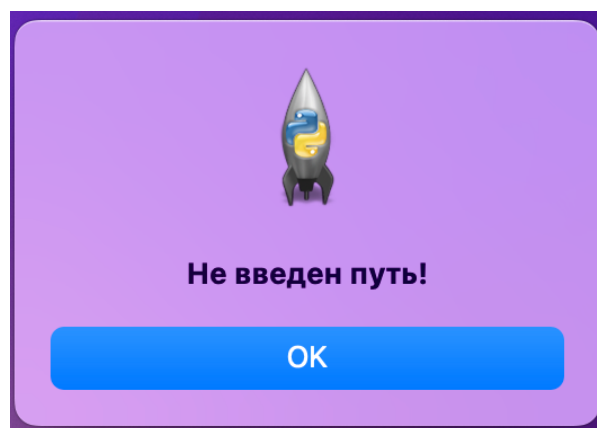
Если не введен адрес ссылки, будет выведено сообщение об ошибке:



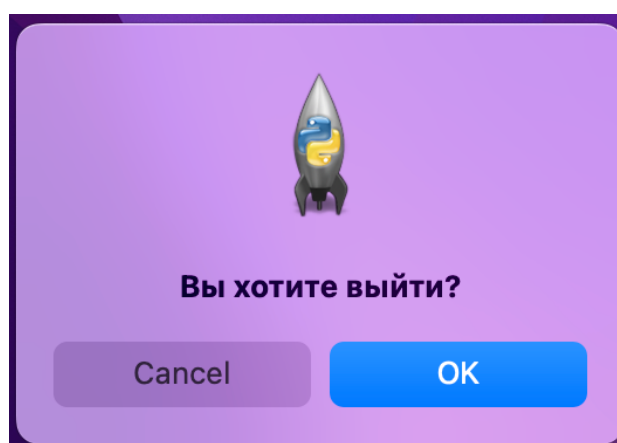
Если при первом запуске не введено имя файла, будет выведено сообщение об ошибке:



Если при первом запуске не введен адрес папки, будет выведено сообщение об ошибке:



При закрытии программы выводится следующее информационное сообщение:

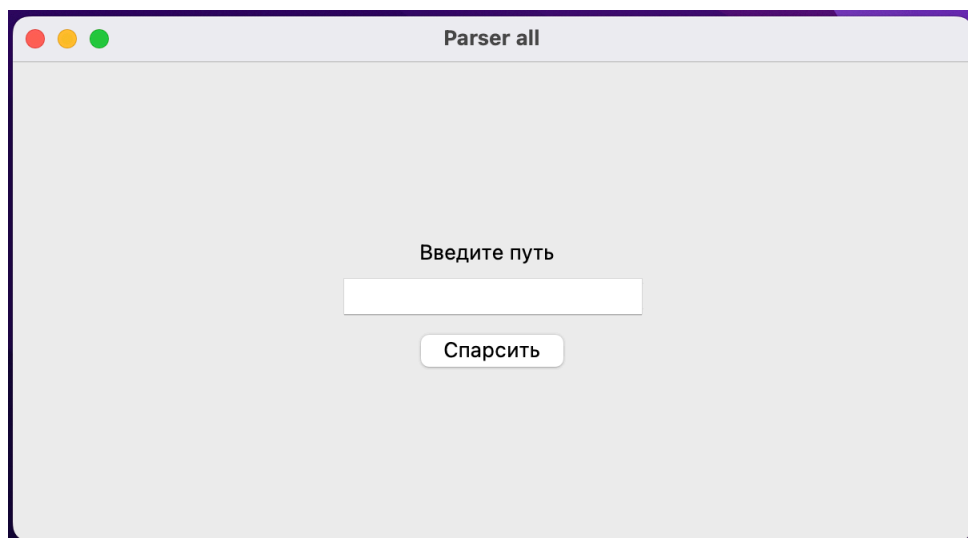


При нажатии кнопки «OK» программа завершит работу, при нажатии кнопки «Cancel» работа программы продолжится.

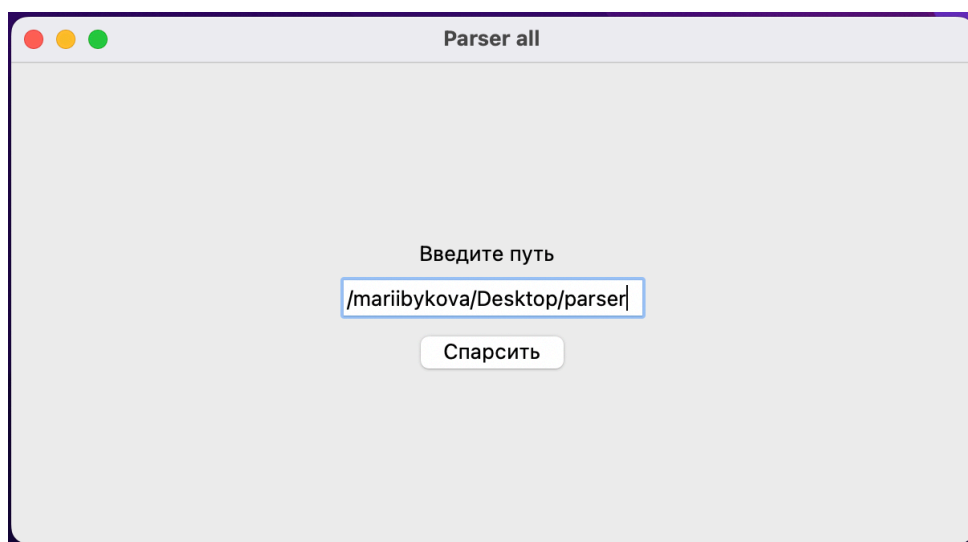


### 3.2.2 Программа «parser\_all»

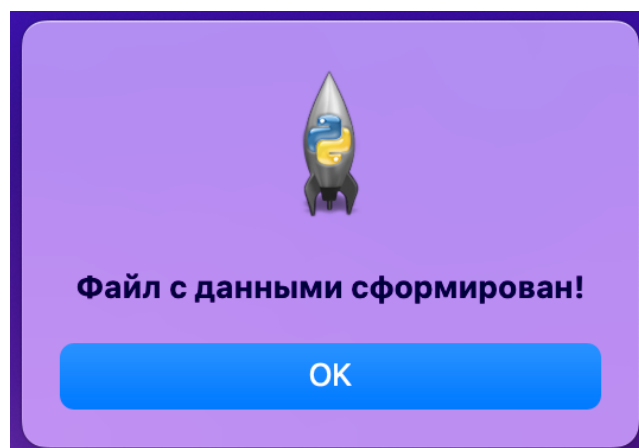
При запуске программы откроется следующее окно:



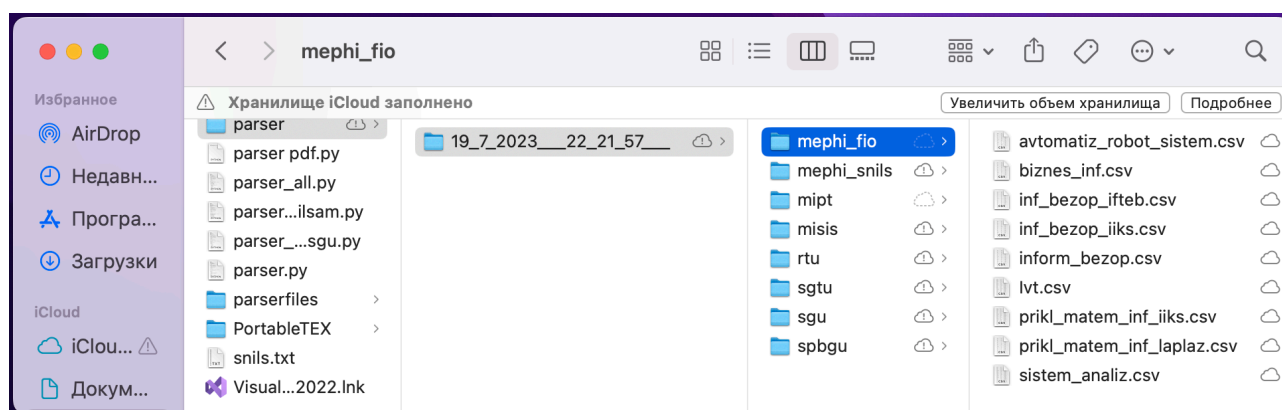
При каждом запуске программы нужно вводить адрес папки, в которую необходимо выгрузить папки с данными. После ввода данных окно будет иметь вид:



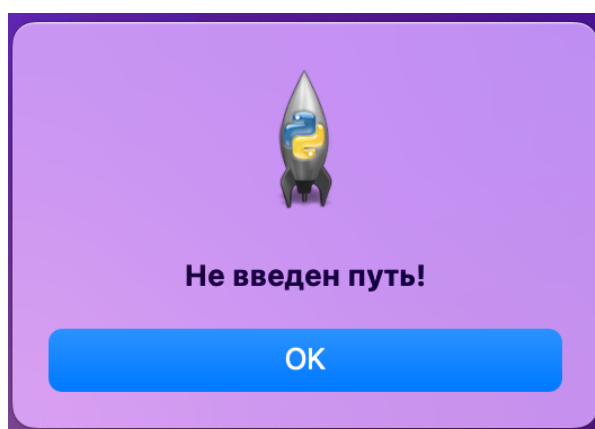
Дальше нужно нажать на кнопку «Спарсить». После этого начнется процесс сбора данных, формирование папок и файлов, сохранение их в указанную директорию. После успешного завершения работы программы появится следующее информационное окно:



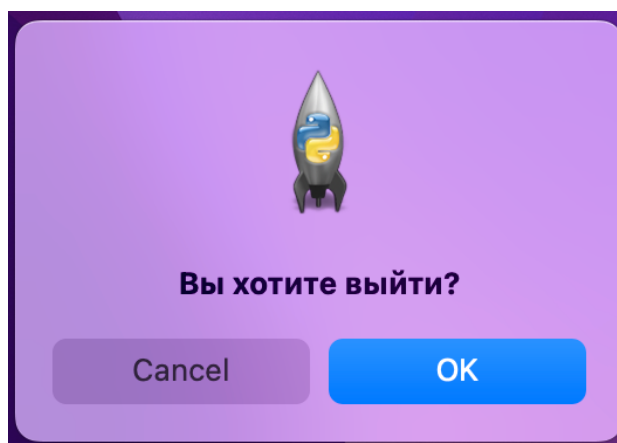
Тогда в указанной директории будет создана папка с текущей датой и временем, содержащая папки с ранжированными списками ВУЗов (см. п. 4):



Если не введен адрес папки, будет выведено сообщение об ошибке:



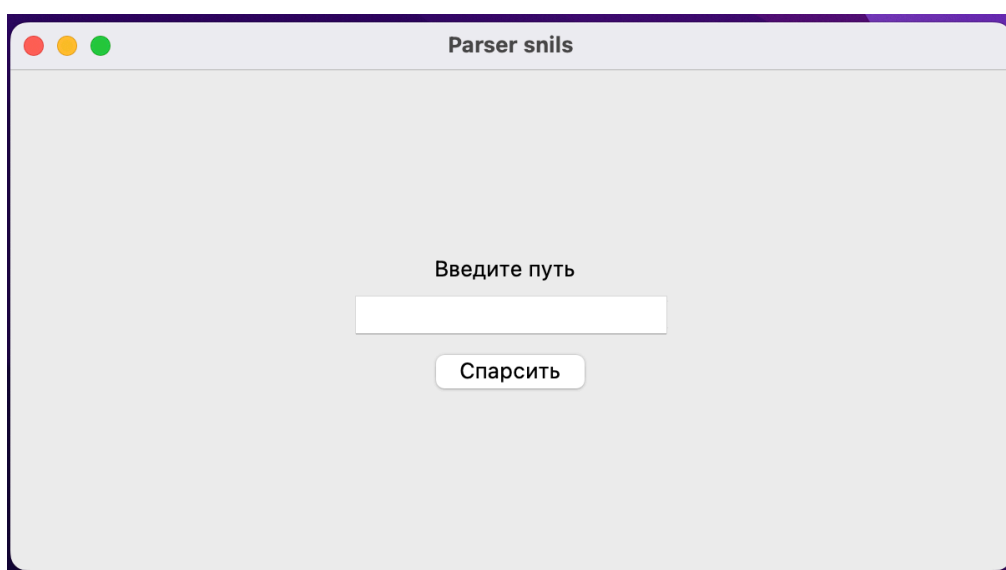
При закрытии программы выводится следующее информационное сообщение:



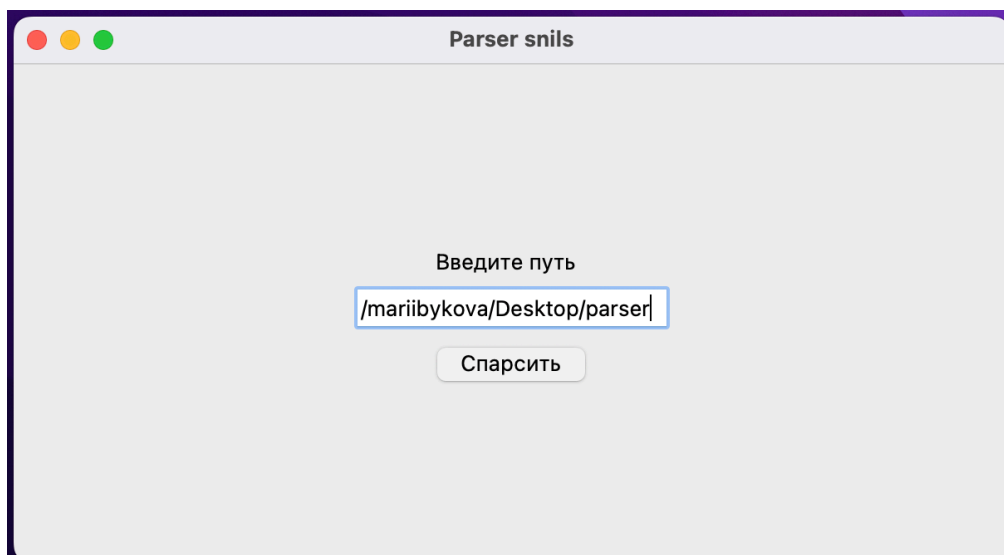
При нажатии кнопки «OK» программа завершит работу, при нажатии кнопки «Cancel» работа программы продолжится.

### 3.2.3 Программа «parser\_snils»

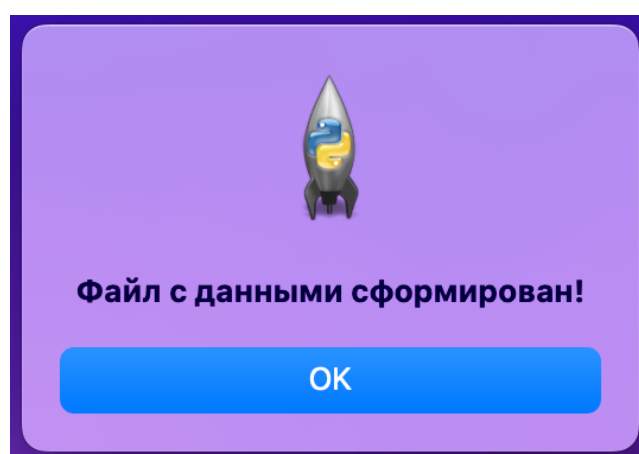
При запуске программы откроется следующее окно:



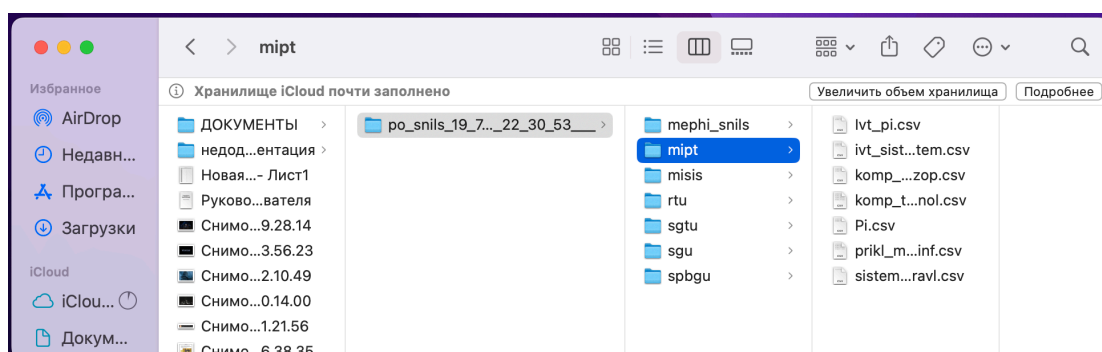
При каждом запуске программы нужно вводить адрес папки, в которую необходимо выгрузить папки с данными. После ввода данных окно будет иметь вид:



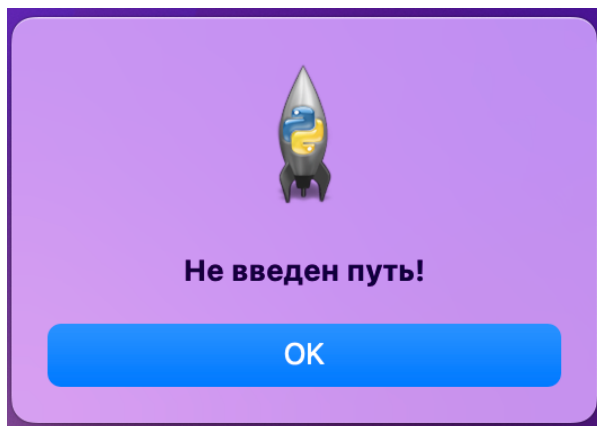
Дальше нужно нажать на кнопку «Спарсить». После этого начнется процесс сбора данных, формирование папок и файлов, сохранение их в указанную директорию. После успешного завершения работы программы появится следующее информационное окно:



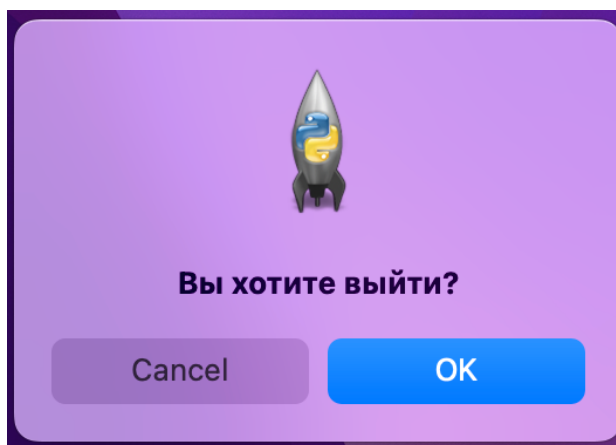
Тогда в указанной директории будет создана папка с текущей датой и временем, содержащая папки с ранжированными списками ВУЗов (см. п. 4):



Если не введен адрес папки, будет выведено сообщение об ошибке:



При закрытии программы выводится следующее информационное сообщение:



При нажатии кнопки «OK» программа завершит работу, при нажатии кнопки «Cancel» работа программы продолжится.

#### **4 Формирование названий файлов и папок для программ «parser\_all» и «parser\_snils»**

Парсерная программа работает для страниц с ранжированными списками приемных комиссий СГУ, СГТУ, СПбГУ, РТУ МИРЭА, МИСИС, МФТИ, МИФИ. Для каждого ВУЗа в файле «file1.txt» хранится его индентификатор и перечень ссылок, с которых нужно спарсить таблицы. Каждая ссылка отдельная конкурсная группа. В соответствии с этим формируется и название файла, куда спарсенная таблица сохраняется.

Элементы в названиях файлов:

- ocho - очная форма обучения;
- za ocho - заочная форма обучения;
- obshchee - общий конкурс;
- ot delnaya - отдельная квота;
- inostr - иностранные абитуриенты;
- platnoe - платная основа;
- selevoj - целевое обучение.

Ниже приведен список названий файлов и соответствующие им конкурсные группы по каждому ВУЗУ:

1) СГУ:

- fiit - фундаментальная информатика и информационные технологии;
- moais - математическое обеспечение и администрирование информационных систем;
- pi - программная инженерия;
- ivy - информатика и вычислительная техника;
- kb - компьютерная безопасность;
- po - педагогическое образование.

2) СГТУ:

- pi - программная инженерия;
- ivt - информатика и вычислительная техника;
- ifst - информационные системы и технологии;
- pinf - прикладная информатика;
- ikts - инфокоммуникационные технологии и системы связи;
- ifbs - информационная безопасность;

— ifs - информационная безопасность автоматизированных систем.

### 3) РТУ МИРЭА:

- prikl\_matem\_inf\_iii - прикладная математика и информатика (ИИИ);
- prikl\_matem\_iit - прикладная математика (ИИТ);
- fiit\_ikb - фундаментальная информатика и информационные технологии (ИКБ);
- ivt\_iii - информатика и вычислительная техника (ИИИ);
- ivt\_iit - информатика и вычислительная техника (ИИТ);
- inform\_sist\_tehn\_ikb - информационные системы и технологии (ИКБ);
- inform\_sist\_tehn\_iri - информационные системы и технологии (ИРИ);
- inform\_sist\_tehn\_komp\_dizayn\_iptip - информационные системы и технологии - компьютерный дизайн (ИПТИП);
- inform\_sist\_tehn\_full\_iptip - информационные системы и технологии - Фулстек разработка (ИПТИП);
- prikl\_inf\_iit - прикладная информатика (ИИТ);
- pi\_iit - программная инженерия (ИИТ);
- inf\_bezop\_ikb - информационная безопасность (ИКБ);
- kb\_iii - компьютерная безопасность (ИИИ);
- inf\_bezop\_avtom\_sist\_ikb - информационная безопасность автоматизированных систем (ИКБ);
- inf\_analit\_sist\_bezop\_ikb - информационно-аналитические системы безопасности (ИКБ);
- bezopasn\_inf\_tehon\_pravoohr\_ikb - безопасность информационных технологий в правоохранительной сфере (ИКБ).

### 4) СПбГУ:

- Math - математика;
- prikl\_matem\_inf - прикладная математика, программирование и искусственный интеллект;
- prikl\_matem\_fund\_inf\_progr - прикладная математика, фундаментальная информатика и программирование;
- sovrem\_program - современное программирование;
- matem\_komp\_nauki - математика и компьютерные науки;

- nauki\_o\_dannyh - математика и компьютерные науки (науки о данных);
- fiit\_bolsh\_dannye - фундаментальная информатика и информационные технологии (большие данные и распределенная цифровая платформа);
- fiit\_progr\_infom\_tehnol - фундаментальная информатика и информационные технологии (программирование и информационные технологии);
- moais\_tehnolog\_programmir - математическое обеспечение и администрирование информационных систем (технологии программирования);
- prikl\_inf\_iskusstv\_intellekt - прикладная информатика (искусственный интеллект и наука о данных);
- prikl\_inf\_iskusstvo - прикладная информатика в области искусств и гуманитарных наук
- Pi - программная инженерия;
- Cau - системный анализ и управление;
- bizness\_inform - бизнес-информатика.

#### 5) МИФИ:

- biznes\_inf - бизнес-информатика;
- avtomatiz\_robot\_sistem - автоматизация и робототехнические системы;
- lvt - информатика и вычислительная техника;
- inf\_bezop\_iiks - информационная безопасность ИИКС;
- inf\_bezop\_ifteb - информационная безопасность ИФТЭБ;
- prikl\_matem\_inf\_iiks - прикладная математика и информатика ИИКС;
- prikl\_matem\_inf\_laplaz - прикладная математика и информатика ЛАПЛАЗ;
- sistem\_analiz - системный анализ и управление;
- inform\_bezop - информационно-аналитические системы безопасности.

#### 6) МФТИ:

- prikl\_matem\_inf - прикладная математика и информатика;
- lvt\_pi - информатика и вычислительная техника (программная инженерия);



- ivt\_sistem\_progr\_prikl\_matem - информатика и вычислительная техника (системное программирование и прикладная математика);
- komp\_tehnot - информатика и вычислительная техника (компьютерные науки);
- sistem\_analiz\_upravl - системный анализ и управление;
- komp\_bezop - компьютерная безопасность.

#### 7) МИСИС:

- prikl\_matem - прикладная математика;
- lvt - информатика и вычислительная техника;
- biznes\_inf - бизнес-информатика.

#### 8) ГУАП:

- prikl\_matem\_inf - прикладная математика и информатика;
- lvt - информатика и вычислительная техника;
- inform\_sistem\_tehnot - информационные системы и технологии;
- prikl\_inf - прикладная информатика;
- Pi - программная инженерия;
- inform\_bezop - информационная безопасность;
- infocomun\_sistem\_tehnot - Инфокоммуникационные технологии и системы связи;
- sistem\_analiz - системный анализ и управление.

#### 9) СПбГЭУ:

- biznes\_inform - бизнес-информатика;
- inform\_sist\_tehnot - информационные системы и технологии;
- prikl\_inf - прикладная информатика;
- prikl\_matem\_inf - прикладная математика и информатика.

#### 10) МЭИ:

- prikl\_matem\_inf - прикладная математика и информатика;
- ivt - информатика и вычислительная техника;
- prikl\_inform - прикладная информатика;
- inform\_bezop - информационная безопасность;
- besprovodn\_tehnot - беспроводные технологии
- biznes\_inform - бизнес-информатика.

#### 11) РУДН:

- prikl\_matem\_inf - прикладная математика и информатика;
- biznes\_inf - бизнес-информатика;
- komp\_infirm\_nauki - компьютерные и информационные науки;
- prikl\_inf - прикладная информатика.

#### 12) МТУСИ:

- prikl\_matem\_inf - прикладная математика и информатика;
- prikl\_matem - прикладная математика;
- Fiit - фундаментальная информатика и информационные технологии;
- lvt - информатика и вычислительная техника;
- inform\_sistem\_tehnol - информационные системы и технологии;
- prikl\_inf - прикладная информатика;
- Pi - программная инженерия;
- inf\_bezop - информационная безопасность;
- inf\_bezop\_telekom\_sist - информационная безопасность телекоммуникационных систем;
- biznes\_inf - бизнес-информатика.