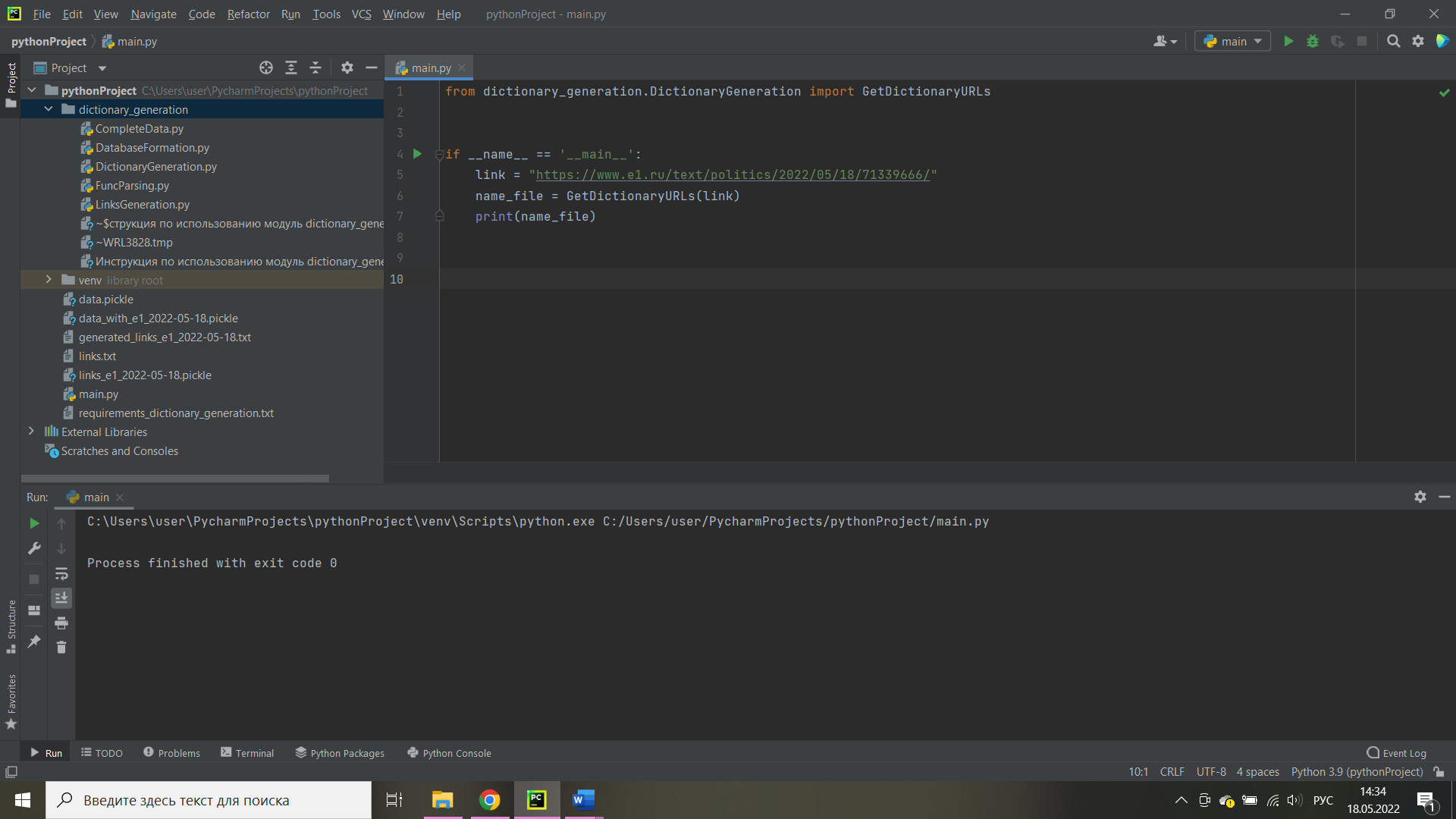
Инструкция по использованию модуль «dictionary\_generation»:

1. Добавьте папку «dictionary\_generation» в папку с вашим кодом и вытащите из неё файлы «requirements\_dictionary\_generation.txt» и «data.pickle»
2. Установите в виртуальную машину все нужны библиотеки из файла «requirements\_dictionary\_generation.txt. Сделать это можно с помощью команды   
   pip install -r requirements\_dictionary\_generation.txt  
   в терминале корневой папки вашего проекта
3. Для того что бы использовать функцию GetDictionaryURLs, которая сгенерируется словарь возможных скрытых фрагментов URL сайта, нужно импортировать её в нужный вам модуль с помощью команды  
   from dictionary\_generation.DictionaryGeneration import GetDictionaryURLs  
   и добавить в код, например, такой строчкой  
   name\_file = GetDictionaryURLs(link)  
   В конечном варианте это должно выглядеть примерно так:



В данном случае мы получим имя файла, где будут храниться ссылки.

Данному модуля можно на вход либо дать полноценную ссылку с названием директорий и прочим, либо главную страницу сайта

В первом случае на выходе будут ссылки, который используют эти токены для генерирования. Например, при подаче такой ссылки: <https://www.e1.ru/text/incidents/2022/05/18/71340008/>, в генерации будут учувствовать такие слова «text», «incidents», «2022», «05», «18», «71340008».

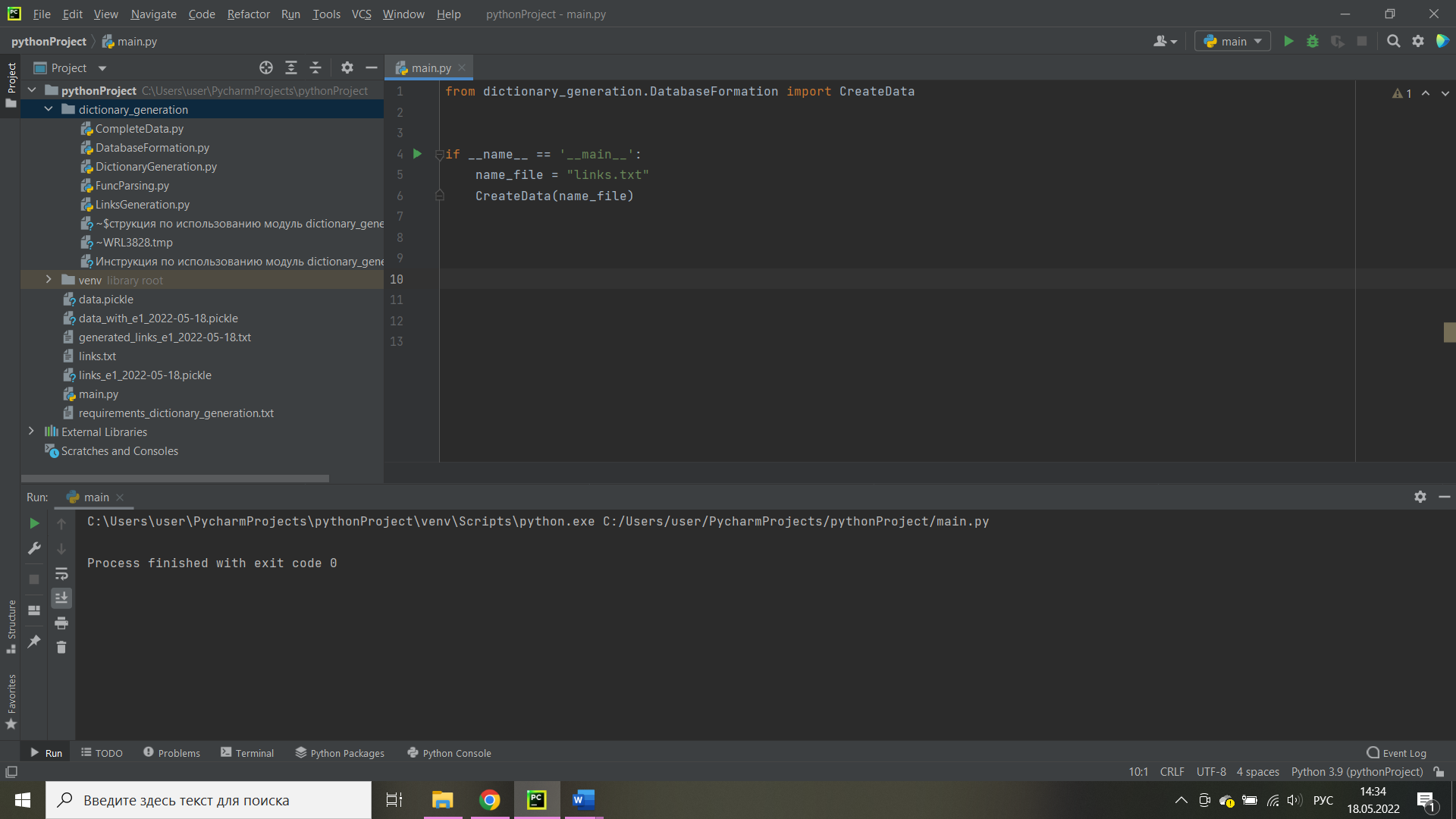
Во втором же случае при подаче главной страницы или страницы, без названий директорий и прочего, например <https://www.e1.ru>, будет сформирован словарь с токенами, который были собраны для обучения модели.

То, что будет сгенерировано моделью, будет находится в файле под именем строящимся следующим образом: «generated\_links», название домена, к которому принадлежит поданная на вход ссылка, текущая дата формирования данного файла. Например, файл с фрагментами URL сайта e1, сформированный 15 мая 2022 будут хранится под именем «generated\_links\_e1\_2022-05-12.txt».

Файлы pickle, начинающиеся со слов «links\_...» и «data\_with\_...» можно при необходимости удалить или написать программу, котороая бы их убирала после использовании функции GetDictionaryURLs.

1. При необходимости модуль предоставляет возможность поменять базу данных для обучения модели. Для этого используется модуль DatabaseFormation. Нужно создать txt файл с новыми ссылкам, поместить её в корневую папку проекта.

В самом коде нужно импортировать функцию CreateData  
from dictionary\_generation.DatabaseFormation import CreateData  
и вызвать её следующим образм   
CreateData(name\_file)  
В итоге должно получится что-то вроде этого:



На вход должен быть подано имя файла, из которого будут считываться ссылки. Новая база данных будет находится в файле «data.pickle». Данный файл стирает предыдущую версию базы данных.

Если нужно будет её вернуть, то файл «links.txt», который находится в папке «dictionary\_generation» нужно извлечь его в корневой.

После создания новой базы данных текстовый файл с ссылками можно убрать.