

УДК 004

ВЕБ-ПРИЛОЖЕНИЕ ДЛЯ СБОРА И ОБРАБОТКИ ИНФОРМАЦИИ
О РЕЙТИНГАХ УНИВЕРСИТЕТОВ

Ю. Д. СТОЛЯРОВ, И. Г. ПЛИСКО, А. И. ВОДОДОХОВ

Белорусско-Российский университет
Могилев, Беларусь

Webometrics Ranking of World Universities – это глобальный рейтинг университетов, публикуемый 2 раза в год (в январе и июле) исследовательской группой Cybermetrics Lab при Высшем совете по научным исследованиям Испании (Consejo Superior de Investigaciones Científicas, CSIC). Webometrics оценивает университеты на основе их присутствия и влияния в интернете, стремясь измерить их веб-активность и открытую научную деятельность. Цель Webometrics – стимулировать университеты и научно-исследовательские институты к увеличению своей веб-активности и открытому распространению знаний. Рейтинг учитывает различные аспекты присутствия университетов в интернете, такие как количество веб-страниц, цитируемость и доступность научных публикаций.

В интернете существуют веб-ресурсы, предоставляющие доступ к различным архивным данным. Самый крупный и популярный из них – Internet Archive. Он предоставляет доступ к различным видам архивной информации: отсканированные копии популярных или редких физических экземпляров книг, исторических газет, журналов, кинолент (в некоторых случаях даже утерянных самой компанией-правообладателем), приложения для различных систем, аудиоконтент разной тематики, изображения, веб-сайты. Архив предоставляет похожий функционал, что и страницы-источники, но может в части случаев не сохранять интерактивные элементы, изображения. Если страницы, на которые указывают ссылки с данной копии, тоже сохранены, возможен переход между ними, как и в актуальной версии сайта.

Разработанное приложение содержит функционал для стандартного взаимодействия с извлекаемой информацией рейтингов университетов, а также ряд функций для удобства работы с ней. При запуске программы пользователю предоставляется окно со всеми необходимыми функциями (рис. 1). Для исследования рейтингов университетов использовались: пакет графического дизайна ggplot2 для построения графиков; broom для преобразования моделей (моделей регрессии, кластерного анализа и др.) в удобные для анализа таблицы данных; psych – пакет для выполнения анализа данных, описательной статистики, визуализации данных, факторного анализа и анализа надежности; rstatix – пакет R для выполнения различных статистических анализов и визуализации данных; dplyr – пакет для манипулирования данными в R, который предлагает функции для фильтрации, трансформации и агрегирования данных.

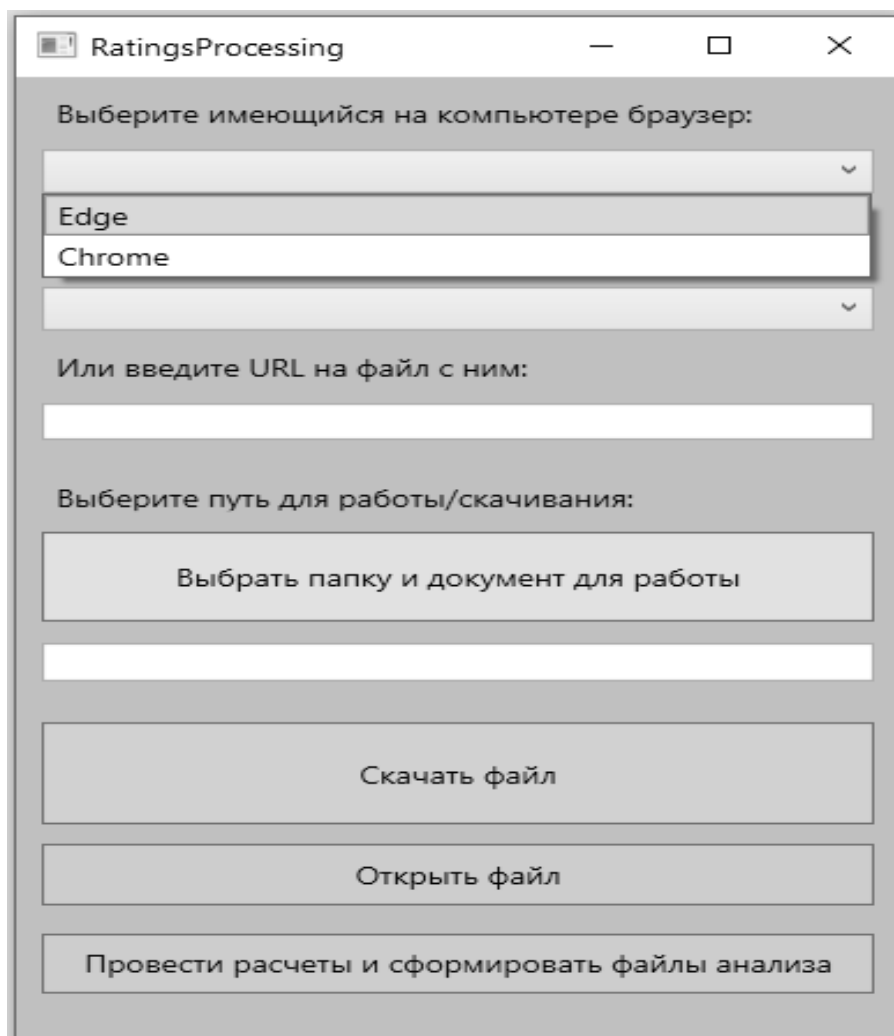


Рис. 1. Главное окно программы с выбором браузера

При всех имеющихся копиях рейтингов в директории программы в папке FormedRankingsData, не указывая пути, нужно выбрать «Провести расчеты и сформировать файлы анализа». В результате будут сформированы книги Excel с выбранными данными рейтингов университетов за 2012–2024 гг., расчеты по копии рейтинга Webometrics первого полугодия 2024 г. и документ Microsoft Office Word с регрессионным анализом, коэффициентами модели, анализом корреляции, описательной статистикой показателей, диагностических графиков моделей для белорусских университетов.

Для извлечения из Excel таблицы сформированных данных рейтингов белорусских вузов (CombinedByYears.xlsx) была использована библиотека readxl. Она поддерживает чтение данных из файлов Excel без необходимости установки Microsoft Excel как в старом (.xls), так и в новом (.xlsx) форматах и формирует таблицу из извлеченного диапазона данных.

Таким образом, использование ресурса Internet Archive позволяет проводить исследование рейтингов университетов за предыдущие периоды, независимо от текущего состояния сайта Webometrics.