

УДК 004.62

ПОЛУАВТОМАТИЧЕСКАЯ СИСТЕМА ПАРСИНГА ДЛЯ ООО «ЗАПАДНО-СИБИРСКИЙ НАУЧНЫЙ ЦЕНТР»

Токмагашева Ю.В., студент гр. ПИБ-131, IV курс
Научный руководитель: Сарапулова Т.В., к.т.н., доцент
ФГБОУ ВО «Кузбасский государственный технический университет име-
ни Т.Ф. Горбачева»
г. Кемерово

ООО «Западно-Сибирский научный центр» – это инновационная организация, заинтересованная в развитии научного потенциала молодых ученых, студентов и школьников, а также в повышении эффективности и конкурентоспособности компаний.

Для увеличения производительности работы и повышении эффективности появилась необходимость обеспечить удобный парсинг (поиск и добавление) e-mail адресов целевой аудитории в базу данных для последующей рассылки приглашений.

Ручное добавление адресов позволяет полностью контролировать процесс, но занимает очень много времени. Автоматический парсинг быстр, но не позволяет контролировать добавляемые адреса.

В связи с этим было принято разработать полуавтоматизированную систему парсинга (ПСП), которая позволяет решить данные проблемы, максимально упростив рабочий процесс.

Реализация проекта осуществлялась в среде программирования Microsoft Visual Studio Professional 2013 посредством объектно-ориентированного языка программирования Visual C#. Для хранения данных была выбрана база данных Microsoft Access.

В ходе реализации проекта была разработана полноценная информационная система в виде десктопного приложения.

Приложение реализует следующие задачи:

1. Формирует рабочую область (рис. 1):
 - а) выводит на пользовательский экран посредством встроенного браузера html/pdf страницы с требуемыми данными;
 - б) имеет кнопки добавления (фамилия, имя, отчество, город, e-mail и других данных) информации о клиенте и новой группы, навигации по встроенному веб-браузеру (переход на предыдущую, последующую и домашнюю страницы, имеет адресную строку), сохранения данных клиента, перехода к работе с добавленными данными;
 - в) выводит контрольную информацию.

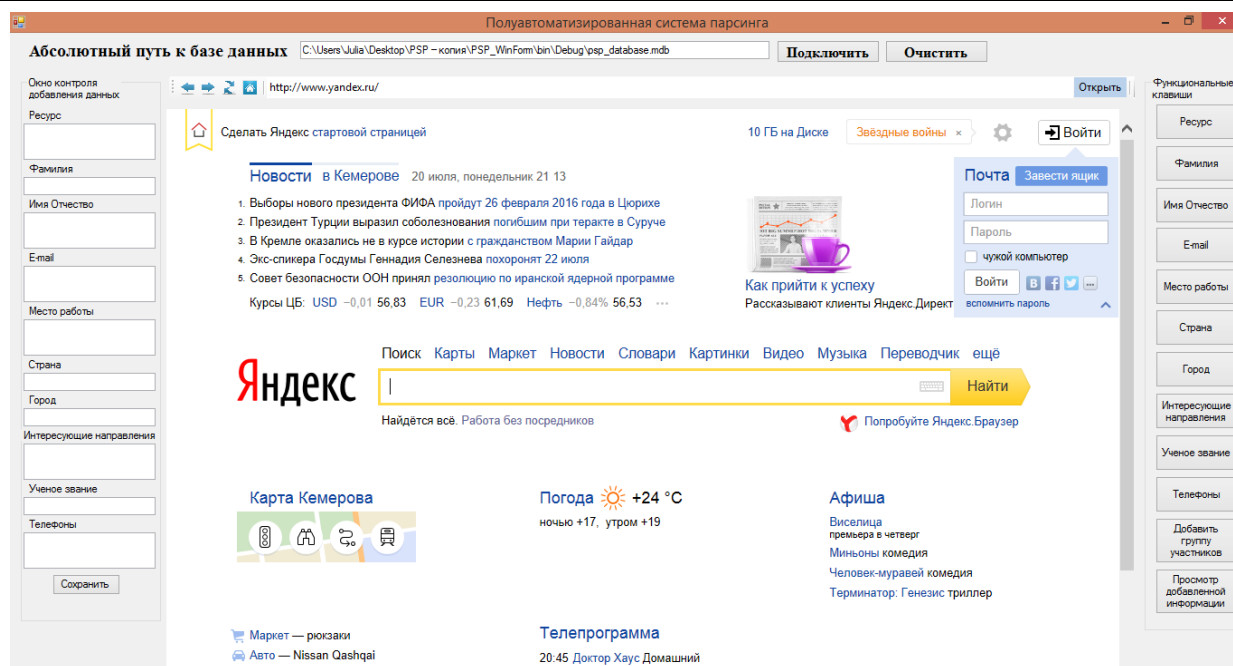


Рис. 1 – Главное окно программы

2. Предоставляет возможность перейти в полноэкранный режим работы с программой, а также свернуть окно программы.
3. Обеспечивает контроль добавляемых данных:
 - а) предупреждает при дублировании данных (рис. 2);

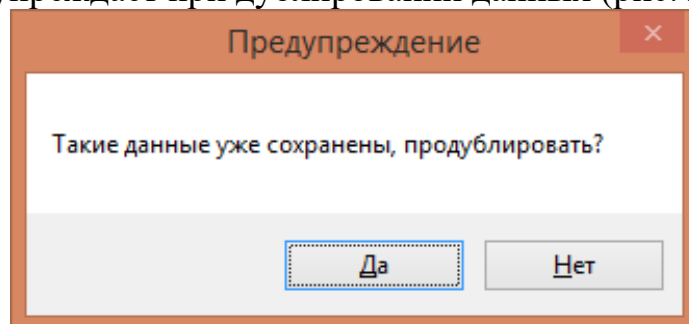


Рис. 2 – Предупреждение

- б) ведет список «посещенных» ресурсов (рис. 3);

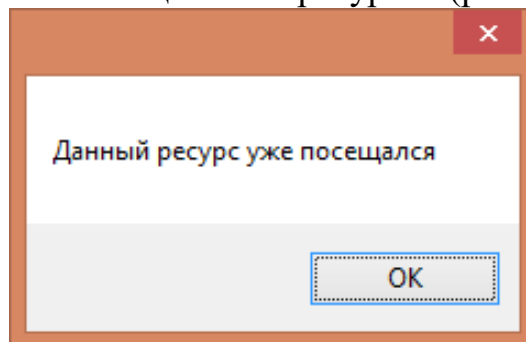


Рис. 3 – Оповещение о посещенных ресурсах

- в) предоставляет визуальный контроль данных в виде окна контроля добавленных данных в главном окне программы (рис. 1).
4. Обеспечивает хранение данных в удобочитаемом формате – в виде реляционной системы управления базами данных Microsoft Access.

5. Предоставляет возможность подключить базу данных, которая хранится в выбранном пользователем месте, а также очистить ее содержимое (рис. 1).

6. Обеспечивает администрирование добавленных данных (рис. 4)

а) двухуровневая сортировка данных по полям фамилия, e-mail, город;

б) группировка по количеству и именуемым папкам, с возможностью изменения группы, просмотра содержащихся в группе данных и возможности добавления новых клиентов в уже существующую группу;

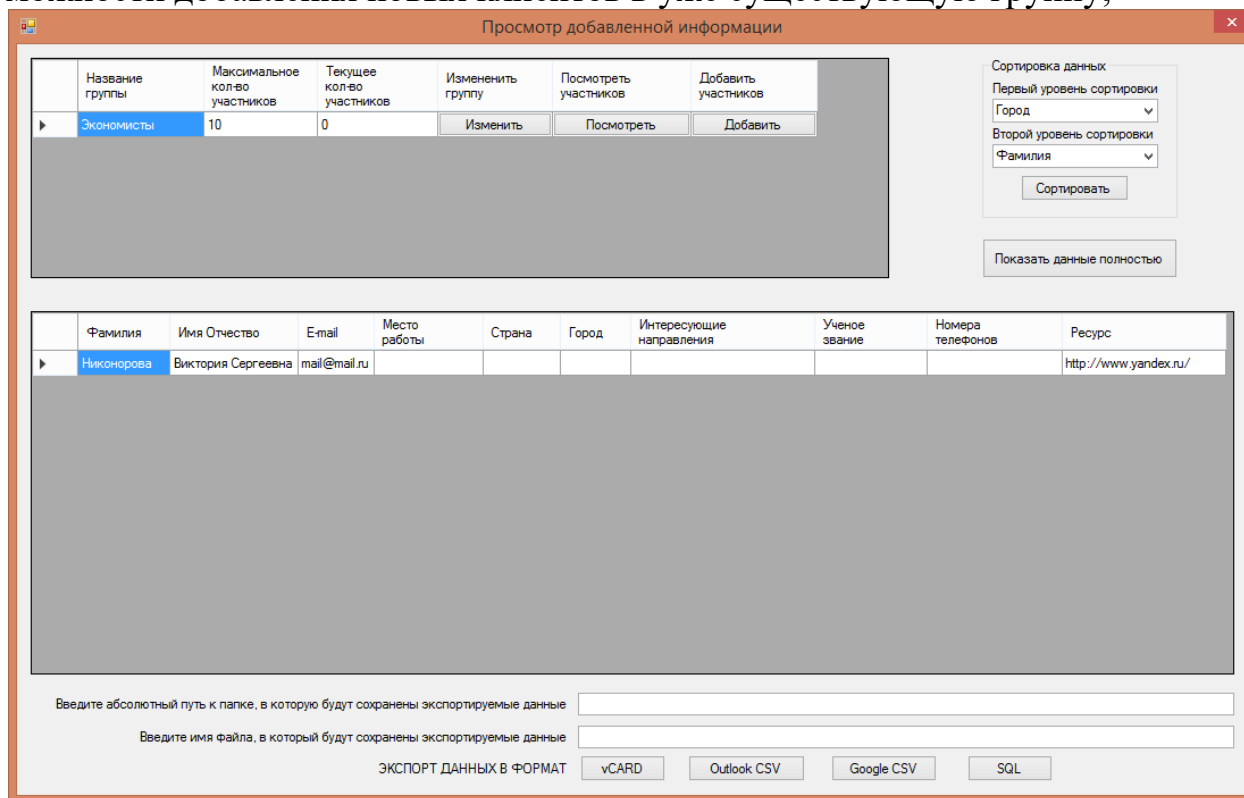


Рис. 4 – Администрирование добавленных данных

7. Поддерживает экспорт данных в популярные форматы баз данных рассылок (vCard (Apple Address Book), Outlook CSV, Google CSV) и баз данных, поддерживающих формат sql (Рис. 4).

8. Имеет модульное построение системы, тем самым обеспечивая масштабируемость системы в дальнейшем.

На данный момент времени программа реализована, введена в эксплуатацию, имеется акт о внедрении.

Список литературы:

1. Общая информация о «ЗапСибНИЦ» // Западно-Сибирский научный центр URL: <http://sibscience.ru/about-us/project/>
2. Пасько, В. П. Эффективная работа в Интернет. – СПб. : Питер, 2005. – 544 с