

Санкт-Петербургский государственный университет
Системное программирование
Группа 24.М71-мм

Автоматический экспорт HTML из облачных электронных таблиц

Малыгин Даниил Александрович

Отчёт по учебной практике в форме «Решение»

Научный руководитель:
доцент кафедры системного программирования, к.ф.-м.н. Д.В. Луцев

Санкт-Петербург
2025

Оглавление

Введение	3
Постановка задачи	4
1. Обзор	5
1.1. Microsoft OneDrive	5
1.2. Google Sheets	7
1.3. CloudConvert	8
2. Описание реализации	11
2.1. Архитектура решения	11
2.2. Требования к сервису	11
2.3. Предлагаемые инструменты	12
Заключение	13
Список литературы	14

Введение

В настоящее время HTML-таблицы играют важную роль при работе с данными. И это не удивительно, так как в настоящее время ими можно воспользоваться везде, где необходимо структурировать информацию и отобразить ее в понятной пользователю форме [1]. Ведь HTML-таблица – отличный способ отобразить такие данные, которые лучше всего смотрелись бы в электронной таблице [2], причём конкретно в среде Интернета. Для этого отображения используется HTML в качестве код разметки.

Может возникать необходимость в получении HTML-таблицы с помощью преобразования уже существующей электронной таблицы. В таких ситуациях достаточно просто сохранить файл в ином формате или использовать стороннее приложение для получения необходимого результата.

Но это работает при нахождении файла на локальном хранилище пользователя. Дело в том, что имеет место ситуации, когда файл хранится на устройстве пользователя, а в облачном хранилище. Безусловно, можно установить необходимый файл и дальше выполнить его преобразование. Но может возникнуть ситуация, когда необходимого для этого объёма памяти может не хватать, что не является

Именно из-за подобных ситуаций и недостатков в решении подобной проблемы существующими сервисами предлагается собственный автоматический экспорт HTML из облачных электронных таблиц.

Постановка задачи

Цель работы заключается в реализации независимого от поставщика сервиса публикации облачных электронных таблиц в вебе.

В рамках данной работы необходимо выполнение следующих задач:

1. Провести анализ функциональности существующих сервисов и возможностей доступных технологий.
2. Выработать требования к сервису.
3. Спроектировать архитектуру сервиса.
4. Реализовать сервис.
5. Провести тестирование и пробу сервиса.

1. Обзор

В рамках данного обзора будут рассмотрены существующие реализации, способные импортировать электронные таблицы из облачного хранилища, а также экспортировать HTML.

Среди имеющихся реализаций можно выделить следующие:

- Microsoft OneDrive;
- Google Sheets;
- CloudConvert;
- Сайт ru.cloud-converter.com;
- Dropbox.

Можно отметить, что последние две реализации хоть и предоставляет необходимый сервис, но требуют за него за плату. По этой причине они не будут рассмотрены в рамках обзора.

1.1. Microsoft OneDrive

OneDrive [3] – это онлайн-сервис от компании Microsoft, который обеспечивает возможность сохранения, синхронизации и обмена различными файлами через Интернет. Данный сервис полностью интегрирован с операционной системой Windows, что позволяет легко работать с файлами через файловый менеджер.

Для получения таблицы от пользователя требуется следующая последовательность действий [4]:

- 1) Открыть веб-сайт OneDrive и перейти к рабочей книге;
- 2) Выбрать опцию «Внедрение» для получения iframe-кода, (рис. 1);
- 3) Скопировать источник iframe (рис. 2);
- 4) Вставить рабочую книгу путем переключения на редактирование блога в формате HTML и вставить код вставки.

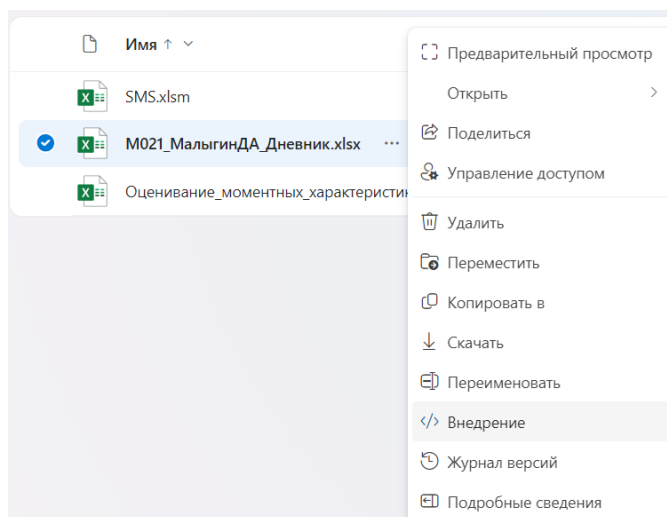


Рис. 1: Выбор опции «Внедрение»

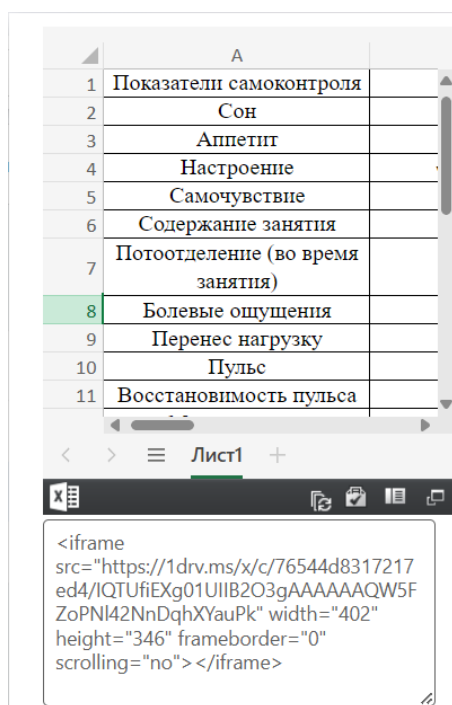


Рис. 2: Получение источника iframe

Проблема данного сервиса заключается в создании HTML-таблицы в виде кода `iframe`, используемого для встраивания в блог пользователя, что не является удобным методом с точки зрения пользовательского опыта. Помимо этого, данный сервис ограничен хранилищем OneDrive, из-за чего использовать файлы из других хранилищ не выйдет.

1.2. Google Sheets

Google Sheets [5] – аналог Excel в режиме онлайн от транснациональной корпорации Google. Это программа, которая помогает систематизировать информацию в табличном виде, считать простые и сложные примеры, решать математические, финансовые и статистические задачи.

Основная функция Google Sheets – создание таблиц и работа с ними совместно с разными пользователями. Главное его преимущество в данном случае – бесплатный доступ.

Чтобы получить HTML-таблицу требуется выполнить следующую последовательность действий:

- Скачать электронную таблицу в формате «Веб-страница» (.html) (рис. 3);
- Открыть полученный архив и разархивировать его;
- Открыть HTML-документ для браузера.

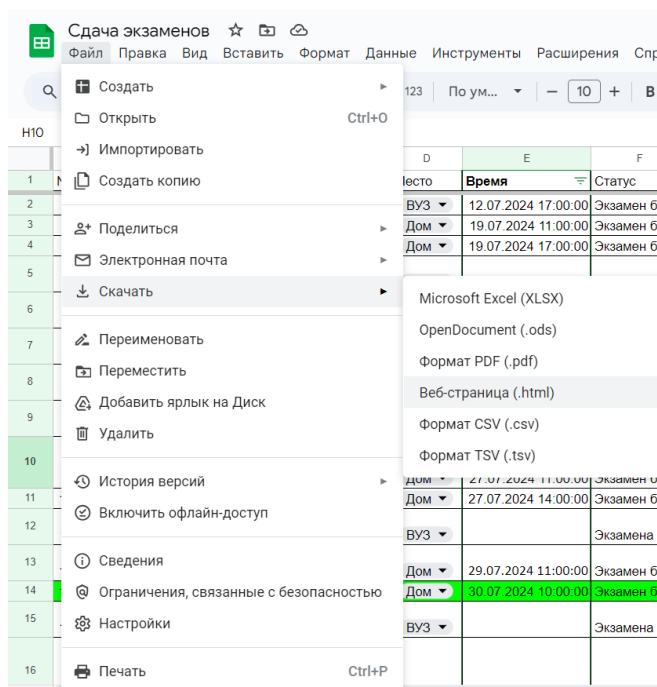


Рис. 3: Скачивание электронной таблицы

	A	B	C	D	E	F	G	H
1	№	ВУЗ	Испытание	Место	Время	Статус	Оценка	Замечание
2	1	РАНХиГС	Бизнес-информатика	ВУЗ	12.07.2024 17:00:00	Экзамен был	74,0	
3	2	ВШЭ	Высшая математика	Дом	19.07.2024 11:00:00	Экзамен был	13,0	
4	3	ГУАП	Информационные науки	Дом	19.07.2024 17:00:00	Экзамен был	74,0	
5	4	ЛЭТИ	Информатика и вычислительная техника	Дом	23.07.2024 11:00:00	Экзамен был	50,0	
6	5	ЛЭТИ	Прикладная математика и информатика	Дом	23.07.2024 11:00:00	Экзамен был	48,0	
7	6	СПбГУ	Технологии искусственного интеллекта и Big Data	Дом	26.07.2024 10:00:00	Экзамен был	49,0	
8	7	ВШЭ	Собеседование по программированию	Дом		Экзамена нет	0,0	
9	8	СПбГУ	Распределенные вычислительные технологии	Дом	26.07.2024 17:00:00	Экзамен был	40,0	
10	9	СПбПУ	Разработка и математическое обеспечение интеллектуальных информационных систем	Дом	27.07.2024 11:00:00	Экзамен был	53,0	
11	10	СПбПУ	Информационная безопасность	Дом	27.07.2024 14:00:00	Экзамен был	20,0	
12	11	СПбГУТ	Автоматизация технологических процессов и производств	ВУЗ		Экзамена нет	0,0	
13	12	ИТМО	Проектирование и разработка систем больших данных	Дом	29.07.2024 11:00:00	Экзамен был	1,0	
14	13	СПбГУ	Программная инженерия	Дом	30.07.2024 10:00:00	Экзамен был	35,0	
15	14	СПбГУТ	Информатика и вычислительная техника	ВУЗ		Экзамена нет	0,0	
16	15	СПбГУТ	Программная инженерия	ВУЗ	01.08.2024 11:00:00	Экзамен был	38,0	38 из 80. С учетом 10 баллов за индивидуальные достижения получается 48 баллов из 100 возможных
17	16	ИТМО	Веб-технологии	Дом	02.08.2024 11:00:00	Экзамен был	50,0	

Рис. 4: HTML-таблица, полученная с помощью Google Sheets

В случае данного сервиса, в сравнение с OneDrive, имеется возможность получить HTML-таблицу в виде отдельного веб-документа.

Но также имеется следующие минусы:

- Результат скачивания оформлен в виде архивированной папки, что может тоже является менее удобным для пользователей;
- Сервис ограничен облачным сервисом корпорации Google

1.3. CloudConvert

CloudConvert [6] позволяет конвертировать более 140 различных форматов в браузере без загрузки приложений на ПК. Удобной возможностью этого приложения является то, что можно выбрать файлы с аккаунта в облаке Google Drive или Dropbox и конвертировать их без сохранения файла на ПК.

В рамках данного сервиса требуется следующая последовательность действий:

- С помощью функции выбора файла указать опцию необходимого облака для загрузки нужного файла с помощью выполнения процедуры авторизации (рис. 5);
- Загрузить файл, выбрать выходной формат для каждого файла;

- Начать процесс преобразования;
- С завершением процесса можно выполнить загрузку на ПК или в облако.

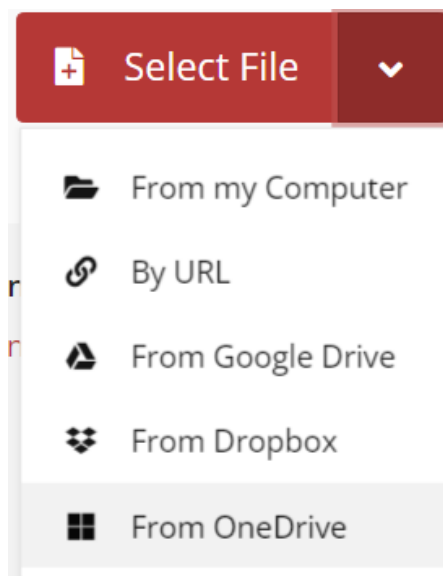


Рис. 5: Выбор облака для загрузки файла

В сравнение с предыдущими реализациями, она реализует работу с файлами, расположенных в разных хранилищах, не только облачных.

Но имеются следующие проблемы:

- Сервис не поддерживает загрузку результата преобразования на локальное хранилище, что может быть критичным моментом в рамках пользовательского опыта;
- При загрузке результата преобразования на Google Drive сохраняется в виде текстового файла с кодом HTML (рис.7), что тоже не делает пользовательский опыт лучше, хотя при загрузке на OneDrive происходит сохранение именно в виде веб-документа (рис. 6), несмотря на проблемы работы со специальными символами.

2. Описание реализации

2.1. Архитектура решения

Предлагаемая архитектура сервиса состоит из двух основных компонентов:

1. Механизма, отвечающего за извлечение необходимой электронной таблицы из облачного хранилища;
2. Преобразователь HTML-таблицы из загруженного файла.

Диаграмма компонентов предлагаемого решения представлена на рис. 8.

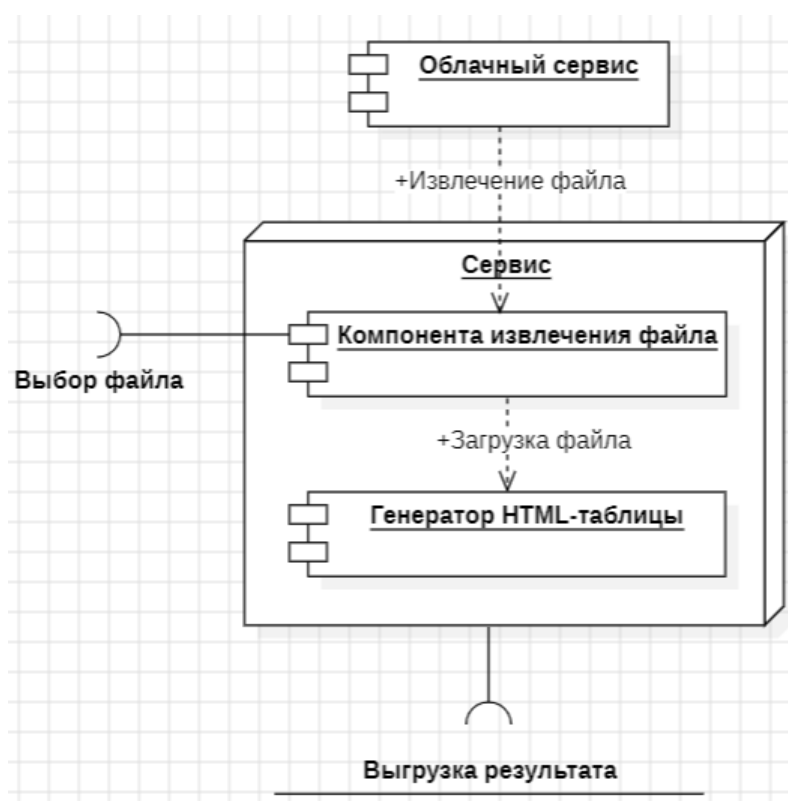


Рис. 8: Диаграмма компонентов

2.2. Требования к сервису

Предлагаемый к реализации сервис должен включать следующие требования:

1. Сервис должен поддерживать возможность извлечение файла из любого облачного хранилища;

2. Сервис должен использовать существующий инструмент генерации HTML-таблицы из файла формата .xlsx и .odb;

3. В процессе преобразования не должно быть использования архивирования;

4. Должна быть предусмотрена возможность сохранить результат преобразования как в облачном хранилище, так и в локальном.

Из нефункциональных требований можно обозначить необходимость сохранения результата в виде HTML-документа, который содержит в себе таблицу.

2.3. Предлагаемые инструменты

Для реализации планируется использование следующего набора инструментов:

- Языки программирования в виде между Python или C. Выбор ограничен выбором конкретно между этими языками обусловлен как легкостью работы с ними, так и доступностью необходимых для них сред разработки;
- Среды разработки в виде PyCharm и VisualCode для работы с языком Python и C соответственно;
- Для взаимодействия как с локальными, так и с облачными хранилищами, рассматриваются Rclone и WebDAV;
- Для тестирования корректности работы сервиса с электронными таблицами планируется использование LibreOffice и Microsoft Office.

Заключение

В текущий момент времени выполнена лишь проработка предлагаемой для реализации идеи, которая будет дополняться в дальнейшем

Из обозначенных изначально задач получилось выполнить следующие:

1. Проведён анализ функциональности существующих сервисов и возможностей доступных технологий;
2. Выработаны требования к сервису;
3. Спроектирована архитектура сервиса.

Список литературы

- [1] HTML и работа с таблицами. – 2023. – URL: <https://otus.ru/journal/html-i-rabota-s-tablicami/?ysclid=m5qzowxkj0923921405> (дата обращения: 2025-01-08).
- [2] Gilmanov A. HTML Tables: All there is to know about them. – 2019. – URL: <https://otus.ru/journal/html-i-rabota-s-tablicami/?ysclid=m5qzowxkj0923921405> (дата обращения: 2025-01-10).
- [3] Умнова А. One Drive — что такое. – 2024. – URL: <https://skyeng.ru/magazine/wiki/it-industriya/chto-takoe-one-drive/> (дата обращения: 2025-01-10).
- [4] Share it: Embed an Excel workbook on your web page or blog from OneDrive. – URL: <https://support.microsoft.com/en-us/office/share-it-embed-an-excel-workbook-on-your-web-page-or-blog-from-onedrive-804e1845-5662-487e-9b38-f96307144081> (дата обращения: 2025-01-09).
- [5] Кирякова А. Гайд по Google Sheets: универсальный инструмент планирования. – 2022. – URL: <https://4brain.ru/blog/gajd-po-google-sheets-universalnyj-instrument-planirovaniya/> (дата обращения: 2025-01-10).
- [6] Leonardo Конвертируем в "облаке" любые файлы в любые другие форматы при помощи CloudConvert. – 2013. – URL: <https://webtun.com/webservices/5443-convert-any-files-to-any-other-format-in-cloud-with-cloudconvert.html> (дата обращения: 2025-01-09).