

УДК 004.67

АВТОМАТИЗИРОВАННЫЙ СБОР И АНАЛИЗ ДАННЫХ ИНДИВИДУАЛЬНЫХ ПЛАНОВ ИЗ EXCEL-ФАЙЛОВ

И. И. МАКСИМОВ, А. С. ЛОПАТКО

Научный руководитель А. Д. ОБУХОВ, д-р техн. наук, проф.

Тамбовский государственный технический университет

Тамбов, Россия

Индивидуальные планы сотрудников используются для оценки научной, учебно-методической и организационной деятельности подразделений университета. В рабочем архиве документы, подготовленные по общей форме, различаются именами и датами редакций, заполнением ячеек и формулировками результатов. Ручная обработка требует выбора актуальной редакции, идентификации сотрудника, переноса пунктов в сводную таблицу и их сопоставления с фактическими результатами. Для выполнения такой работы требуется много времени и усилий проверяющих. Поэтому необходимо автоматизировать преобразование архива Excel-файлов в проверяемую структуру данных «план – факт».

Это включено в информационную систему университета, реализованную на Python с использованием Django. Файл XLSX рассматривается как контейнер Office Open XML [1]: парсер читает структуру книги, перечень листов, общие строки и значения ячеек, не изменяя исходный документ [2].

На предварительном этапе файлы группируются по имени плана. Среди отправленных версий для каждого имени пользователя выбирается файл, наиболее поздний по дате. Датированная редакция имеет приоритет перед недатированной. ФИО извлекается из листа «Общие сведения», а при отсутствии реквизита – из титульного листа. Затем оно сопоставляется с учетной записью сотрудника. Отсутствующие и неоднозначные соответствия не принимаются автоматически.

Парсер находит столбец по заголовку «Наименование результата», поэтому извлечение не зависит от фиксированного адреса ячейки. Пустые строки, значения «нет» и «не планируется», заголовки и примечания пропускаются. Точка с запятой и разрыв строки используются как разделители текста. Словари и регулярные выражения относят полученные пункты к категориям статей, докладов, грантов, патентов, программных разработок, учебных изданий и других работ. Для счетных категорий количество определяется по цифрам и словесным конструкциям. Для каждого пункта сохраняются текст, количество, категория и координаты источника: файл, лист и ячейка. На основе координат, номера элемента и категории формируется ключ, поэтому повторный импорт неизмененного файла обновляет запись, а не создает копию. Исправления пользователя помечаются и не перезаписываются автоматически.

Фактические сведения импортируются только из подтвержденных записей университетского реестра. Сотрудник определяется по внешнему кадровому идентификатору. Применена схема сопоставления записей [3]: учитываются

учебный год, категория и сходство нормализованных названий. После сопоставления названий свободная емкость плана заполняется оставшимися результатами того же вида. Один результат может быть использован только один раз. Для записи рассчитываются план, факт, остаток и состояние выполнения, а результаты без автоматической связи сохраняются отдельно.

При апробации ошибки чтения XLSX и разбора XML не зарегистрированы. Из 203 подтвержденных результатов, относящихся к владельцам импортированных планов, 121 результат автоматически связан с плановыми пунктами, а 82 оставлены без связи и доступны для отдельного анализа (табл. 1).

Табл. 1. Результаты апробации на архиве 2025/2026 учебного года

Показатель	Значение
Исходные / отобранные XLSX-документы	92 / 86
Планы с извлеченными пунктами / сотрудники / подразделения	71 / 70 / 7
Исходные ячейки / структурированные записи / плановые единицы	512 / 557 / 611
Ошибки чтения XLSX и разбора XML	0
Подтвержденные результаты владельцев импортированных планов	203
Автоматически связанные / оставленные без связи результаты	121 / 82

Сводное представление содержит сотрудников по строкам, виды деятельности по столбцам и значения «план – факт – остаток» в ячейках.

Практический результат работы состоит в замене повторного ручного переноса сведений воспроизводимым преобразованием данных с сохранением их происхождения и возможностью проверки каждого автоматически сформированного показателя.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. ECMA-376. Office Open XML File Formats. – Geneva: Ecma International, 2016.
2. Разработка информационной системы электронного документооборота управления фундаментальных и прикладных исследований / М. Н. Краснянский, А. Д. Обухов, С. В. Карпушкин, А. В. Остроух // Вестник Тамбовского государственного технического университета. – 2015. – Т. 21, № 2. – С. 216–230.
3. **Краснянский, М. Н.** Математическая модель обработки информации в системе управления электронным документооборотом / М. Н. Краснянский, А. Д. Обухов, И. Л. Коробова // Вестник Тамбовского государственного технического университета. – 2018. – Т. 24, № 3. – С. 382–399.