

ОГРАНИЧЕНИЯ И ВОЗМОЖНОСТИ ЭЛЕКТРОННЫХ ТАБЛИЦ ПРИ АНАЛИЗЕ ДАННЫХ

Eremina I.M.

LIMITATIONS AND POSSIBILITIES OF SPREADSHEETS IN DATA ANALYSIS

Аннотация. Рассматривается роль электронных таблиц как инструмента разведочного анализа данных в условиях роста объемов информации. Выполнен качественный анализ факторов, влияющих на производительность табличных процессоров при обработке многомерных массивов. Исследуются архитектурные ограничения Excel и Google Sheets, приводящие к деградации скорости работы и снижению точности вычислений. Сформулированы критерии, при достижении которых миграция на специализированные статистические пакеты и языки программирования становится объективной необходимостью.

Ключевые слова: электронные таблицы, Excel, Google Sheets, анализ данных, производительность, ограничения, большие данные, миграция.

Abstract. The role of spreadsheets as a tool for exploratory data analysis in the context of growing information volumes is considered. A qualitative analysis of factors affecting the performance of spreadsheet processors when processing multidimensional arrays is carried out. Architectural limitations of Excel and Google Sheets that lead to degradation of operating speed and reduction in computational accuracy are investigated. Criteria are formulated upon reaching which migration to specialized statistical packages and programming languages becomes an objective necessity.

Keywords: spreadsheets, Excel, Google Sheets, data analysis, performance, limitations, big data, migration.

В современной научной и бизнес-практике электронные таблицы занимают уникальное положение. С одной стороны, они являются наиболее доступным инструментом обработки данных, не требующим от пользователя навыков программирования. С другой стороны, стремительный рост объемов цифровой информации ставит под сомнение универсальность табличных процессоров как средства полноценного статистического анализа [1]. Возникает противоречие между кажущейся простотой работы с таблицами и реальными затратами времени при попытке обработать выборки значительного размера.

Целью настоящей работы является качественное исследование границ применимости электронных таблиц (на примере Microsoft Excel и Google Sheets) [1] и выявление факторов, сигнализирующих о необходимости перехода на специализированные инструменты – статистические пакеты (SPSS, Stata) или среды программирования (R, Python с библиотекой Pandas) [3].

Принцип работы табличного процессора нагляден для малых объемов данных, однако при масштабировании проявляются ее архитектурные ограничения. Ключевым фактором становится способ хранения данных: таблицы вынуждены держать в оперативной памяти не только сами значения, но и информацию о форматировании, стилях и зависимостях между ячейками, что создает

избыточную нагрузку на вычислительную систему.

Особого внимания заслуживает вопрос точности вычислений. Несмотря на то, что современные версии Excel используют математический аппарат двойной точности, визуальное представление данных часто нивелирует эту точность. Такие действия, как округление в ячейках, автоматическое форматирование и ограничения на отображаемые разряды создают иллюзию потери данных.

Другим ограничением является проблема воспроизводимости результатов. Действия пользователя (фильтры, сортировки и др.) не фиксируются. Это затрудняет аудит вычислений и делает невозможным повторение анализа без ручного восстановления всех шагов.

Таким образом на начальном этапе работы, когда объем информации небольшой, электронные таблицы выступают идеальным инструментом, они позволяют быстро визуализировать данные, проверять гипотезы в ручном режиме и качественно подготовить информацию к более глубокому анализу.

Однако по мере накопления данных и усложнения запросов наступает момент, когда инструмент начинает выступать ограничителем и пользователь тратит непропорционально много времени на ожидание отклика программы. На этой стадии работа в таблицах перестает быть эффективной. Возникает объективная потребность в миграции на инструменты, использующие векторные вычисления и не отягощенные графической оболочкой.

Таблица 1. Качественное сравнение эффективности инструментов анализа

Объем данных	Excel / Google Sheets	Python (Pandas) / R
Малый	Высокая эффективность	Избыточная сложность
Средний	Заметные задержки	Высокая эффективность
Крупный	Риск сбоев, низкая скорость	Максимальная эффективность

Таблица 1 иллюстрирует качественную зависимость эффективности работы от объема обрабатываемых данных для различных классов. Однако перспективами данного исследования является разработка методики быстрой оценки данных и автоматического подбора оптимального инструмента анализа.

Критерием для миграции на специализированные пакеты служит совокупность факторов: появление ощутимых задержек при выполнении типовых операций, сложность отслеживания цепочек вычислений и необходимость обеспечения воспроизводимости результатов.

ЛИТЕРАТУРА

1. **Уинстон, У.** Бизнес-моделирование и анализ данных. Решение актуальных задач с помощью Microsoft Excel, 6 издание/ У. Уинстон. – СПб.: Питер, 2021. – 944 с.
2. **Маккинни, У.** Python и анализ данных / У. Маккинни. – М.: ДМК Пресс, 2020. – 540 с.