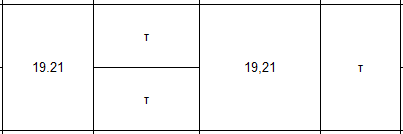
# Сводный статистический отчёт

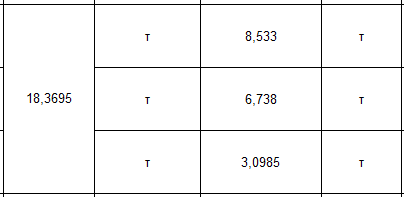
На первый взгляд, в создании сводного статистического отчёта в Excel нет ничего сложного. Исходный бланк формы для заполнения утверждён, поэтому количество столбцов и их порядок в присылаемых файлах формата Excel одинаковы, как и формат представления данных. Достаточно скопировать заполненные данными строки исходных отчётов в сводную таблицу (так называемый копипаст, от английского *copy-paste*) — и сводный статистический отчёт готов.

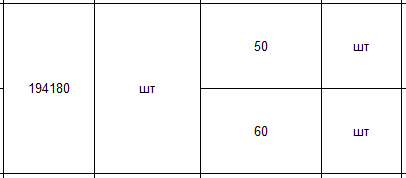
Однако при более детальном рассмотрении возникают определённые сложности. Например, количество продукции может быть указано как в штуках, так и в тоннах. При этом разные организации в своих исходных отчётах используют различные варианты сокращений: одни указывают «*штук*», другие — «*шт.*», а третьи — «*шт*». Аналогичная ситуация наблюдается и с тоннами. Следовательно, для получения корректной статистики необходимо привести все обозначения количества к единому стандарту.

Ещё одна проблема заключается в том, что в исходных отчётах количественные показатели вводятся в таблицу в текстовом формате. Важно помнить, что в Excel данные в ячейках могут быть представлены в трёх форматах: число, текст и формула. Внешне число, записанное как текст, выглядит идентично обычному числу, однако при суммировании такие ячейки не учитываются. Поэтому необходимо тщательно проверить все столбцы с количественными показателями и при необходимости преобразовать текстовые ячейки в числовой формат.

При обработке отчётности может возникнуть следующая проблема: если в одной партии транспортируются различные виды продукции, количественные показатели этой партии часто вносятся в объединённую ячейку исходного отчёта. Такой подход создаёт существенные трудности при последующей обработке и анализе данных, поскольку объединённые ячейки нарушают целостность структуры таблицы и мешают корректной работе с информацией. Примеры:







В Excel при использовании русской кодовой раскладки разделителем между целой и дробной частями числа служит запятая (например, *13,14*). Однако в исходных отчётах нередко встречается ошибочное использование точки в качестве разделителя, что характерно для англоязычных стран. Важно учитывать, что Excel воспринимает числа с точкой как текстовые значения, которые не участвуют в автоматическом суммировании. Это может привести к некорректным результатам при подсчёте итоговых значений.

Обозначение продукции в отчётности часто отличается, что создаёт сложности при анализе данных. Например, для одного и того же продукта (лимона) используются различные варианты записи:

* в одних отчётах указывается просто «*лимон*»;
* в других — «*лимоны*»;
* в третьих — «*лимоны свежие*».

Отсутствие единого подхода к наименованию продукции приводит к тому, что при обработке данных система воспринимает эти записи как разные позиции, хотя фактически это один и тот же вид продукции.

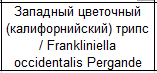
Отсутствие единообразия при указании болезней и вредителей обнаруженных на продукции создаёт серьёзные трудности при обработке данных. В различных исходных отчётах встречаются следующие варианты записи:

* в одних таблицах указываются как латинские, так и русские названия вредителей;
* в других — только латинские наименования;
* в третьих — исключительно русские названия;
* иногда используются синонимы основного латинского названия.

Дополнительные сложности возникают из-за:

* различного порядка написания латинских и русских названий;
* избыточного использования пробелов;
* применения скобок;
* произвольного использования заглавных и строчных букв (что критично при сортировке, так как они имеют разную кодировку);
* различного порядка слов в названиях.

Все эти факторы делают невозможным корректную сортировку и суммирование данных по категориям вредителей. Примеры:

Казалось бы, оптимальным решением было бы издать директиву о единообразном заполнении форм. Однако организации, предоставляющие исходные отчёты, имеют иное организационное подчинение, что существенно затрудняет возможность повлиять на их работу.

Жизнеспособность вредителя может быть указана не только как *да* или *нет*, но и в ином виде, например: “*в живом виде*”, “*жизнеспособные*”, “*живые*”.

Страны также могут указываться по-разному. Например: *Китай* и *КНР*, *Армения* и *Р . Армения*.

Данные о грузоотправителе и грузополучателе содержат не только название фирмы, но и избыточную информацию в виде полного адреса.

Используемые номера документов достаточно длинные, поэтому операторы при наборе их иногда “теряют” одну-две цифры. Поскольку длина номера всегда одинакова (в разных странах она различна), можно проверить нет ли в номере документа пропущенных знаков.

Используемые даты не должны значительно разниться от даты текущей. Соответственно их также можно проверить на корректность.

Итак, перечень выполняемых действий по автоматизации обработки сводного статистического отчёта:

1. Разделение объединённых ячеек количественных показателей
2. Проверка на отсутствие иных объединённых ячеек
3. Замена штучных единиц количества продукции на «шт*.*»
4. Изменение формата представления ячеек штучных единиц на целое число
5. Замена весовых единиц количества продукции на «т»
6. Изменение формата представления ячеек штучных единиц на число с тремя знаками после точки
7. Поиск и преобразование текстовых ячеек количества продукции в числовой формат
8. Поиск ячеек количества продукции где разделитель между целой и дробной частями числа служит точка. Преобразование в числовой формат
9. Удаление избыточных пробелов в тексте
10. Замена наименований видов продукции на стандартное наименование
11. Замена названий вредителей, обнаруженных на продукции на стандартное название из списка
12. Замена названий приславшей отчёт организаций на стандартное название из списка
13. Замена названий стран на стандартное название из списка
14. Замена критериев жизнеспособности на «*да»* или «*нет»*
15. Удаление адреса грузоотправителя и грузополучателя
16. Если вредитель, обнаруженный на продукции, - щитовка, не учитывать её как обнаружение
17. Проверка корректности введённых номеров документов
18. Проверка корректности введённых дат
19. Если грузоотправитель или грузополучатель указан как физическое лицо, замена его на «*фл*»
20. Указание типа склада хранения
21. Указание по возможности название организации, проводившей экспертизу
22. Вычисление и вставка номера недели создания сводного отчёта
23. Вычисление и вставка периода дат создания сводного отчёта
24. Вычисление и вставка заключения о принятых мерах

Приведённый перечень содержит только основные действия по автоматизации.