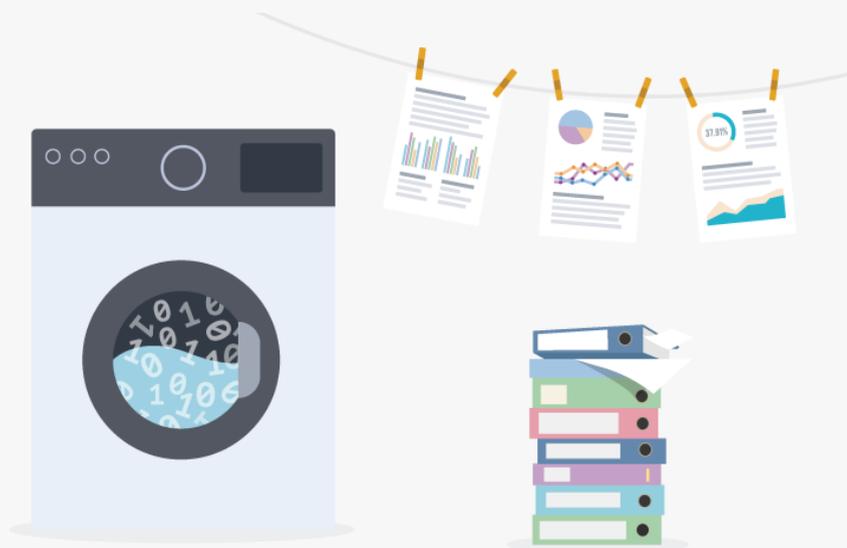




# СТАНДАРТИЗАЦИЯ И ОЧИСТКА НОРМАТИВНО-СПРАВОЧНОЙ ИНФОРМАЦИИ



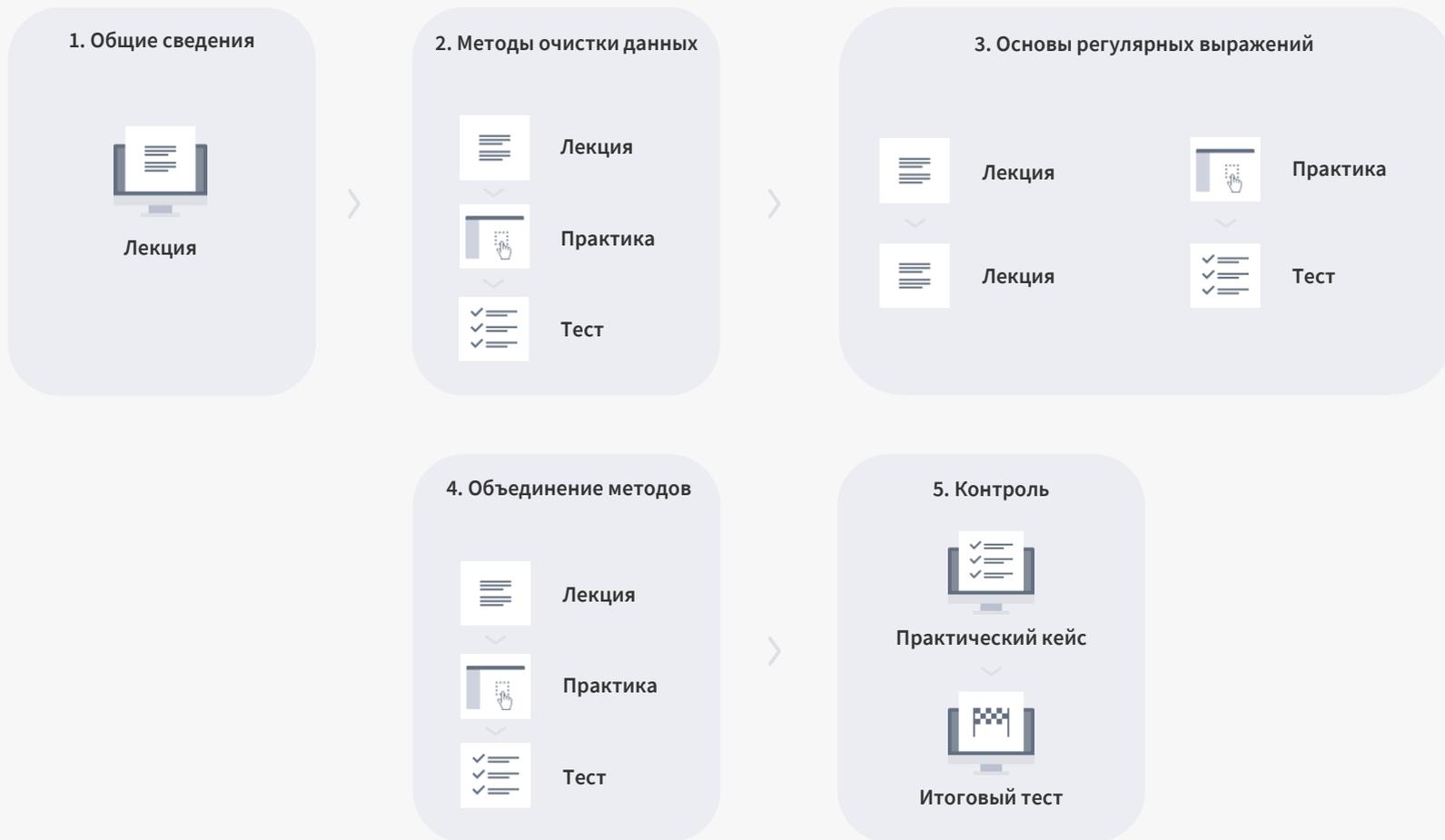
## О курсе

В курсе подробно рассматриваются проблемы, связанные с ведением нормативно-справочной информации организации (НСИ): неполнота, противоречивость, недостоверность или некорректность описаний объектов, несоответствие форматов и другие вопросы.

Обсуждаются основные способы очистки и стандартизации объектов: словари, классификаторы, частотный анализ, контрольные числа, регулярные выражения. Практическая часть построена на решении задач в Loginom.

Для прохождения курса рекомендуется иметь базовые навыки работы с платформой Loginom.

## Карта курса





### Теоретическая часть

Объемы данных. Качество данных. Причины загрязнения данных: ручной ввод данных, не заполнение/некорректное заполнение поля, потеря информации, отсутствие ограничений на вводимые значения, отсутствие стандартов. Виды ошибок в данных: опечатки, пропуски, противоречия, дубликаты, несоответствие форматов, неправдоподобие данных. Понятие очистки данных. Понятие стандартизации данных. Причины необходимости очистки. Показатели качества данных.



### Теоретическая часть

Основные методы очистки. Использование справочников и таблиц замены: виды справочников, понятие классификатора. Общероссийские классификаторы. Классификаторы адресов: КЛАДР, ФИАС. Анализ строк: расстояние Левенштейна, расстояние Дамерау-Левенштейна. Регулярные выражения: понятие, решаемые задачи. Частотный анализ. Контрольные числа: возможности использования, пример: алгоритм формирования контрольного числа СНИЛС.



### Практическая часть

Примеры использования методов очистки в LogiNot: сверка со справочником (на примере БИК), анализ строк (на примере ФИО), частотный анализ (корректировка пола), расчет контрольного числа (на примере СНИЛС).



### Контрольная часть



### Теоретическая часть 1

Понятие регулярного выражения. Понятие метасимволов. История появления регулярных выражений. Диалекты регулярных выражений. Регулярные выражения Perl. Возможности регулярных выражений. Инструменты для проверки.



### Теоретическая часть 2

Виды метасимволов. Метасимволы по умолчанию. Экранирование символов. Диапазоны символов. Метапоследовательности. Мнимые метасимволы. Квантификаторы. Дополнительные опции. Примеры.



### Практическая часть

Компонент Калькулятор. Функции по работе с регулярными выражениями. Описание и примеры использования.



### Контрольная часть

## Блок 4. Объединение методов



### Теоретическая часть

Применимость методов. Использование одного метода. Комбинирование методов. Общий алгоритм очистки данных.



### Практическая часть

Решение задачи по очистке и стандартизации списка регионов. Описание наборов данных. Проведение очистки в четыре этапа: поиск по коду региона, сверка со справочником, использование регулярных выражений, расчет расстояния Дамерау-Левенштейна. Оценка результатов.



### Контрольная часть



### **Практический кейс**

Практический кейс для самостоятельного решения



### **Итоговый тест**

Вопросы с вариантами ответов

## Еще больше курсов и компонентов в [Loginom e-Learning](#)

### Loginom e-Learning

Контент доступен в рамках подписки на тарифы

<b>Всё</b>	<b>Personal</b> для физических лиц 3 из 9	<b>Standard</b> для юридических лиц 5 из 9	<b>Enterprise</b> для юридических лиц 9 из 9
------------	---	--	--



#### Обнаружение и измерение оттока

Методы обнаружения и измерения показателей оттока и активности клиентов, визуализация и интерпретация показателей удержания клиентов.

Электронный курс



#### Базовые методы сегментации клиентов

Методы сегментации, основанные на поведении клиента в прошлом: RF, RFM, LRFM, бинарная сегментация.

Электронный курс



#### Введение в кредитный скоринг

Что такое скоринговая карта, скоринги кредитных бюро, история кредитного скоринга и этапы разработки скоринговой карты.

Электронный курс

