

## **Об управлении избыточной информацией в текстах и данных программных проектов**

*Кознов Д.В., д.т.н., проф. кафедры системного программирования СПбГУ*

В докладе будет рассмотрена задача управления избыточной информацией в программных проектах в общем виде, а также о машинном обучении, NLP и строковых алгоритмах как средствах решения данной задачи. В качестве примеров будет рассмотрена программная документация, а также *stack traces* (стеки вызова приложения в *crash*-отчетах, посылаемых пользователями в компанию-разработчик при падении приложения). Будет рассказано об особенностях задачи поиска нечетких повторов в документации программного обеспечения, о том, в каких случаях дублирование информации является оправданным, а в каких – нет.

Будет рассмотрена *reference* документация и *narrative* документация, а также обсуждены особенности поиска нечетких повторов для этих случаев. Будет представлен подход и программный инструмент *Duplicate Finder* для анализа и рефакторинга документации на основе поиска нечетких повторов. Будет рассмотрен подход для быстрого внесения изменений в повторяющиеся (нечетко) комментарии *JavaDoc* формата для *Java*-приложений (подход реализован в виде плагина к *IntelliJ IDEA*). Будет также рассказано о задаче классификации *stack traces*, а также о задаче назначения разработчика для данного *stack trace*.