

## **АВТОМАТИЧЕСКАЯ КЛАССИФИКАЦИЯ ТЕКСТОВ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ**

В современном информационном обществе с активным и повсеместным использованием компьютеров, а также с увеличением объемов хранящейся информации в электронном виде автоматическая классификация текста, т.е. определение принадлежности текста к некой категории, является актуальной и крайне важной задачей.

Классификация применяется в решении многих практических задач, таких как: фильтрация документов; распознавание спама; автоматическое аннотирование; поиск текстовых документов; навигация по большим информационным ресурсам; подбор рекламы; составление интернет-каталогов; классификация новостей; индексация данных в поисковых запросах и т.д.

Если данные не структурировать, не распределять по категориям, то через некоторое время будут накоплены массивы плохо структурированной информации, а отсутствие возможности вовремя и быстро получить необходимую информацию по нужной теме сделает бесполезной большую часть накопленной базы знаний.

Классификация текстов, являющаяся одним из способов структурирования данных, – это сортировка текстовых документов по заранее определенным категориям (тематика, стиль и т.д.). Методы классификации текстовых документов лежат на стыке двух областей: информационного поиска и машинного обучения. Общие части двух этих подходов – способы представления документов и способы оценки качества классификации текстов, а различия состоят только в способах собственно поиска. Как мощный механизм обучения нейронные сети могут широко применяться в различных отраслях. Нейросети должны использоваться для обобщения данных, а не для определения, атрибуты и критерии которого крайне важны при сборе и анализе данных. Обучающиеся машины часто используются при формализации знаний из данных реального мира, но все же сами обучающиеся машины не могут генерировать принципы формализации – здесь не обойтись без помощи человека. Применение автоматических систем классификации текстов позволит сократить трудозатраты человека на рубрикацию и поиск необходимой информации предоставляемой информационными системами, а также позволит сократить участие или вовсе не привлекать специалистов при размещении информации в базе знаний.