

Я. Гуртовой

АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ СОБЫТИЙ В НОВОСТНЫХ СООБЩЕНИЯХ НА АНГЛИЙСКОМ ЯЗЫКЕ

Автоматическое извлечение событий – часть одной из самых быстрорастущих областей компьютерной лингвистики. В последнее время разработка новых языковых ресурсов, принципы их создания, а также теоретическая обоснованность этих принципов стали особенно актуальными. Следует отметить, что в настоящий момент нет общепринятых универсальных критериев для анализа англоязычных текстов для автоматического определения конкретной оценки событий.

Новостные сообщения, которые обладают определенной структурой, являются основными инструментами СМИ в виду их высокой технологичности и по причине их понятности и доступности максимально большому человеческому сообществу.

Событие создается предложением или текстом или, точнее, их интерпретацией.

Любое событие в новостных текстах передается с помощью использования разнообразных синтаксических структур, клишированных, образных выражений, оценочной и экспрессивной лексики.

Автоматическое извлечение событий, которые несут в себе значимую информацию, является приоритетной задачей, которая относится к области обработки текстов. Традиционными методами машинного обучения являются: обучение с учителем; метод Байеса; марковская модель; логический метод.

Метод, основанный на лингвистических шаблонах, на данный момент является наиболее подходящим способом извлечения событий благодаря скорости реализации и возможности расширения алгоритмов, который основывается на составлении и реализации лингвистических шаблонов для выделения необходимых атрибутов и на составлении и применении тематических словарей.

Методы машинного обучения, хотя они все чаще используются для различных задач обработки текста, все еще чрезвычайно сложны и требуют много времени для использования в реальном мире. Это связано не столько со сложностью обучающих алгоритмов, сколько с возможными неудачными методологическими подходами к обучению. Цели обучения применяются по частям на любом конкретном этапе последовательного процесса обработки текста. По этому все еще возникает необходимость управлять ручной маркировкой и не использовать результаты предыдущего обучения системы на последующих и взаимосвязанных этапах обработки.

Несмотря на все существующие разработки и огромное количество людей и компаний, вовлеченных в автоматическое извлечение событий из текстов, до сих пор не решены важные проблемы поиска информации.