

УДК 004

ТЕКСТОВЫЙ АНАЛИЗ В СОВРЕМЕННОМ МИРЕ

Гурская Д. А., Василевская В. А.

Научный руководитель: к. т. н., доцент Дробыш А. А.

*Белорусский национальный технический университет,
г. Минск, Республика Беларусь*

На сегодняшний день в современном мире информационные технологии играют важную роль в жизни человека. Они присутствуют повсеместно, и даже в анализе текстов. Многие люди хоть раз в жизни после прочтения книги задумывались, сколько слов либо предложений находится в прочтенном тексте – это и есть самый простой анализ текста.

Анализ – метод научного исследования явлений и процессов, в основе которого лежит изучение составных частей, элементов изучаемой системы.

Аналитические методы широко распространены в социуме, поэтому термин «анализ» часто воспринимается как синоним исследования в целом и при решении познавательных задач. Анализ является составной частью любого научного исследования, образуя его первую стадию, когда исследователь выявляет в описании изучаемого объекта его строение, состав, свойства, признаки и т. п.

В современном обществе важным средством оформления, фиксации, сохранения, передачи и обмена информации являются документы.

Анализ документов – это метод сбора первичных данных, при котором документы используются в качестве главного источника информации.

Анализ текста – это процесс получения высококачественной информации из текста на естественном языке. Обычно для этого применяется статистическое обучение на основе шаблонов: входной текст разделяется с помощью шаблонов, затем производится обработка полученных данных.

Такой анализ осуществляет практически каждый человек не задумываясь, что он делает именно это. Например, любой читатель книги анализирует содержание, читая ее по частям. В общем случае всем, особенно трудоспособному населению, постоянно приходится работать с текстовыми и иными видами документов, явно или неявно анализируя их содержание и другие компоненты, например, качество, форму, размер и т. д.

Таким образом, важной задачей практически любых индивидов является нахождение нужных им документов, а также анализ их содержания на предмет подготовки различных документов. При этом нередко возникают ситуации, когда в течение незначительного периода времени необходимо подготовить некоторый материал.

Анализ текста использовался и продолжает использоваться для классификации текстов, содержащихся в них слов и словосочетаний, аннотирования и реферирования текстов, проведения семантически ориентированного поиска текстов по заданным концептам, определения авторского права претендента на соответствующий текст и т. п. Благодаря развитию информационных технологий анализировать текст стало гораздо проще.

Анализ текстовых файлов производился почти с момента создания ЭВМ, и естественно, за все время было создано множество приложений, выполняющих эту функцию. Большая часть таких приложений была усовершенствована и теперь входит в состав текстовых редакторов (MicrosoftWord, Excel и др.).

ЛИТЕРАТУРА

1. Коваль, С. А. Лингвистические проблемы компьютерной морфологии / С. А. Коваль. – СПб. – 2005.
2. Тузов, В. А. Компьютерная семантика русского языка / В. А. Тузов. – СПбГУ. – 2004. – 400 с.