

Е.В. Шарапова
Муромский институт (филиал) Владимирского государственного университета
Россия, Владимирская обл., г. Муром, ул. Орловская, 23
E-mail: sharapovamivlgu@gmail.com

Понятие нечеткого дубликата текста

С развитием и глобальным внедрением в человеческую жизнь информационной сети Internet особо остро проявилась проблема определения первоисточника текста. Является ли документ оригинальным или он копия другого документа, или конфигурация документов? Такая задача стоит в сфере обучения перед преподавателями, а также в сфере авторского права перед правообладателем на текст. Самым важным в этой задаче является определение факта заимствования в одном тексте из другого текста. Эта задача может быть решена проще, если документы систематизировать по типу и мере использованных в нем заимствований. В связи с этим различные копии и дубликаты документов условно можно поделить на 3 группы:

1) Копия документа (полный дубликат документа) – это второй его экземпляр, то есть дубликат в классическом его понимании, с сохранением содержания и всех особенностей форматирования текста.

Чаще всего копия используется для того, чтобы размножить существующий документ. Распознать, что копия – это дубликат другого текста, не представляет труда. Например, когда мы делаем ксерокопию документа, второй получившийся его экземпляр – это копия документа (полный дубликат документа). Или, когда мы копируем файл из одной папки в другую, получившаяся копия во второй папке – это полный дубликат документа.

2) Неполный дубликат – это частичный дубликат текста, то есть его часть, которая взята из текста с сохранением всех особенностей форматирования и содержания, и вставлена в другой текст, возможно, в текст другого автора.

Неполным дубликатом может являться как одна страница текста, так и его абзац, а также половина всего текста или другое его количество. Главное условие неполного дубликата – абсолютное сохранение содержания оригинала. Такой вид дубликата сложнее поддается распознаванию, нахождению в тексте другого автора, потому что не всегда можно определить, откуда взят определенный кусок текста. Например, когда студент составляет определенный реферат по какому-то предмету и вставляет кусок текста из документа, находящегося, например, в сети Internet, без исправлений и редактирования в свой реферат, то таким образом студент создает у себя в реферате неполный дубликат другого текста, кусок из которого он взял. Таких кусков различного размера студент может набрать несколько в своем реферате, тем самым создавая неполный дубликат нескольких различных текстов.

Самый сложный для обнаружения вариант дубликата – это нечеткий дубликат.

3) Нечеткий дубликат – это комбинация различных вариантов переделанных каким-то образом неполных дубликатов текстов.

Нечеткий дубликат текста может быть изначально неполным дубликатом, но переделанным настолько, что невозможно будет понять, что в новом, переделанном тексте есть какие-либо скрытые заимствования из других источников информации. Если воспользоваться предыдущим примером о студенте с рефератом, то в случае, если он переделает свой реферат – неполный дубликат различными способами и вариантами, то его работа уже становится нечетким дубликатом тех заимствованных текстов, которые он использовал в своей работе над рефератом.

Такое условное разделение дубликатов дает возможность систематизировать различного типа дубликаты текстов, а также автоматизировать решение проблемы поиска заимствований в документах.