

Автоматизированное редактирование дат в тексте на естественном языке

Егоров Сергей (Санкт-Петербург, ГБОУ СОШ № 564, 11 класс)

Руководитель: Штукенберг Дмитрий Григорьевич (учитель информатики 564 школы)

Повседневно люди имеют дело с большими объемами документов и рано или поздно сталкиваются с проблемами быстрого их редактирования. В данной работе предпринята попытка создания инструмента, упрощающего работу с текстом, написанном на естественном языке. Так как проблема слишком обширна, мы выбрали только одну из ее частей: работу с датами. В юридических документах часто необходимо менять даты, и при этом несинхронное изменение дат может привести к реальным проблемам.

Подобный инструмент логичнее всего было реализовать в виде плагина для какого-либо текстового редактора, что и было выполнено. Изначально стоял вопрос о выборе редактора и была произведена попытка реализации для редактора Open Office, однако пришлось отказаться от этой идеи из-за большого количества времени, необходимого для освоения системы расширений, которая используется в данном редакторе. В итоге было решено использовать Sublime Text 2 из-за его популярности, простой системы расширений и гибкого интерфейса.

В ходе выполнения работы был разработан небольшой плагин на языке Python, написанный для текстового редактора Sublime Text 2, позволяющий как искать даты в тексте, написанные в числовой и словесной форме, так и автоматически их редактировать (заменять на другие). Программа распознает различные формы записи дат в тексте и предоставляет пользователю список, где все даты записаны в числовой форме. Далее пользователь может выборочно отредактировать даты и применить изменения к исходному файлу (причем изначальная форма записи сохраняется) с помощью различных комбинаций клавиш. Плагин будет выложен в открытом доступе и будет доступен для скачивания.

Из-за гибкости использованного языка, плагин можно улучшать, добавляя новые возможности. Выбор редактора Sublime обоснован исключительно простотой интеграции с ним, поэтому плагин может быть легко перенесен на любой другой редактор.

Список источников

Sublime API - https://www.sublimetext.com/docs/2/api_reference.html

Склонение числительных - <http://numeralonline.ru/>