

ИСПОЛЬЗОВАНИЕ ГИС-ТЕХНОЛОГИЙ ДЛЯ СОЗДАНИЯ ПРОСТРАНСТВЕННОГО ГИПЕРТЕКСТА

Корсукова Е.Г.

Научный руководитель – Савельев А.С.

Сибирский федеральный университет

В современных векторных геоинформационных системах информация представлена в виде карты, на которую нанесены изображения объектов с их характеристиками. Атрибутные данные в ГИС обеспечивают связь между местоположением объекта и его значением при помощи уникального номера объекта. Для атрибутных данных, имеющих сложную структуру, предусмотрена возможность хранения таблиц во внешней СУБД. В ГИС выделяется несколько общих блоков: сбора и ввода данных, обработки, моделирования и анализа, распространения и хранения. С помощью ГИС можно преобразовывать геоданные и проводить их пространственный анализ, изучать структуру, взаимосвязи и динамику природных и социально-экономических систем, выводить данные в картографическом виде, что значительно облегчает объяснение и понимание.

Несмотря на широкое использование мультимедиа, текст остается одним из основных видов информации в большинстве электронных хранилищ. Разработка эффективных подходов к обработке текстов с целью фильтрации, формирования смыслового портрета, навигации по базе текстов является одним из наиболее актуальных направлений современных информационных технологий. Визуализация ключевых моментов текста значительно упрощает его понимание. Обработка текста, выделение ключевых слов в нем – является сложной и в то же время одной из главных задач создания пространственного гипертекста. Вручную обработать большой и сложный текст крайне затруднительно и требует длительного времени. Одно из решений этой проблемы – автоматическая обработка текста на основе ГИС – технологий. Данная задача разделяется на следующие этапы:

- 1) написание кода программы для автоматического анализа текста;
- 2) геокодирование уникальных слов текста.

Код программы был написан на языке программирования Python. Python – это интерпретируемый язык высокого уровня, предназначенный для самого широкого круга задач ГИС. С его помощью можно обрабатывать различные пространственные объекты, работать с базами данных, разрабатывать Web-приложения с графическим интерфейсом. Данный язык характеризуется высокой читаемостью и компактностью кода, простотой синтаксиса. Python может быть с легкостью расширен для взаимодействия с другими программными системами и легко встроен в другие программы в качестве компонента. Язык Python является кроссплатформенным языком,

позволяющим создавать программы, которые будут работать во всех операционных системах, и активно развивающимся. Стандартная библиотека включает большой объем полезных функций. Программа на языке Python представляет собой обычный текстовый файл с расширением py (консольная программа). Все инструкции из этого файла выполняются интерпретатором построчно.

На вход разработанной программы автоматического анализа текста поступает исходный текстовый файл в формате ASCII, а на выходе получаются слова и словосочетания, которые являются уникальными. Программа построчно считывает исходный текст и обрабатывает его при помощи регулярных выражений, сравнивая слова со словарем русских слов и удаляя знаки пунктуации. Счетчик считает встречаемость оставшихся слов. Из этих слов формируется словарь, отсортированный по частоте встречаемости в тексте. Словари в Python – это наборы объектов, доступ к которым осуществляется по ключу. В качестве ключа в нашем случае выступает частота повторения слова в тексте. На выходе программы получается словарь встречающихся в тексте географических названий и собственных имен.

Полученные географические названия наносятся на карту путем геокодирования в ГИС. Геокодирование - процесс назначения географических идентификаторов и координат объектам карты и записям данных. Этот механизм позволяет получать пространственные ссылки из текста для определения местоположения на карте. Выделенные в тексте географические названия становятся такими ссылками на карту. Обратное геокодирование связывает географические координаты с фрагментами текста.

Для построения логической системы пространственного гипертекста в качестве знаков используются понятия, способные нести специфическую информацию о географических объектах. В процессе создания пространственного гипертекста решаются следующие задачи:

- 1) выделение объектов, которые будут показаны на карте;
- 2) подготовка структуры пространственной БД;
- 3) выбор инструментальных средств для картографирования, анализа текста;
- 4) подбор литературных источников для создания атласа;
- 5) интерфейс атласа;
- 6) подготовка картографической основы.

На начальном этапе работы геокодирование происходит на мелкомасштабную карту мира. На нее в несколько слоев накладываются векторные данные, которые привязываются к словам в тексте. Первоначально, объекты на карту наносятся схематически, в виде точек. В дальнейшем они будут геокодированы в виде полигонов, линий. При этом появляется возможность перехода с текста на карту и с карты на текст.

С помощью разработанной программы, проведено геокодирование книги Л. Н. Гумилева «Древняя Русь». После автоматической обработки текста было выделено около 4 тыс. слов, которые обозначают собой географические объекты, имена людей. Затем эти объекты были схематически отображены на карте мира.

Работая с полученным таким способом пространственным гипертекстом, можно выводить именно ту информацию, которая необходима пользователю, отфильтровывать все ненужные данные, усложняющие понимание исследуемой области. Можно использовать данные такого рода для поиска искомых объектов на карте. По заданным пользователем параметрам система находит и показывает объекты, которые соответствуют запросу поиска. Вся необходимую информацию можно вывести на печать или же просто сохранить в виде файла, что обеспечивает удобство в дальнейшей работе.

Для многих типов пространственных операций конечным результатом является представление данных в виде карты или графика. С помощью ГИС визуализация самих карт может быть дополнена отчетными документами, трехмерными изображениями, графиками и таблицами, фотографиями и другими средствами, например, мультимедийными.

Пространственный гипертекст можно использовать для различных целей, например, для иллюстрации исторических событий, построения транспортных сетей между различными объектами, отображения местоположения торговых точек, описания достопримечательностей города. Главные моменты текста отображаются на карте, это упрощает поиск по самому тексту и ускоряет работу с ним