

## **ГЕОКОДИРОВАНИЕ ИСТОРИЧЕСКОГО ТЕКСТА**

**Корсукова Е.Г.**

**научный руководитель канд. техн. наук Савельев А.С.**

*Сибирский федеральный университет*

Геокодирование – процесс назначения географических идентификаторов, таких, как географические координаты, выраженные в виде широты и долготы, объектам карты и записям данных. В содержательном плане геокодирование интерпретируется как преобразование обычного адреса в точку на карте.

Любая географическая информация содержит сведения о пространственном положении, будь то привязка к географическим или другим координатам или ссылки на адрес, почтовый индекс, идентификатор земельного или лесного участка, название дороги. Процедура геокодирования автоматически определяет местоположение объекта, используя атрибутивные данные. Это позволяет с ее помощью быстро определить и визуализировать интересующий нас объект на карте.

Обработка текста, выделение ключевых слов в нем – является сложной и в то же время одной из главных задач создания пространственного гипертекста. Основная идея пространственного анализа, воплощенного в технике компьютерного картографирования, заключается в акцентировании внимания, прежде всего на пространственных координатах того или иного исторического события или процесса, на изучении его пространственной локализации и развития. Используя методы геокодирования можно довольно быстро создавать картографические базы данных для информации, имеющей текстовую координатную привязку.

Поэтому ГИС технологии нашли широкое применение для решения прикладных задач в исторических исследованиях, где они используются в следующих целях:

1) визуализация уже имеющихся исторических материалов: нанесение на карту населенных пунктов, дорог, территориально-административных границ, представление распределения численности населения или его плотности и т.д. Составляя тематические исторические карты, можно проследить пространственную динамику в развитии какого-либо явления или объекта. Подобные карты не только являются иллюстрацией к исследованию, но помогают лучше и глубже понять явление или процесс, а также сформулировать вопросы для последующих научных проектов.

2) как база исторических данных, дополненная пространственной информацией, становится основой для дальнейшего пространственного статистического анализа, позволяющего выявить скрытые тенденции и взаимосвязи.

Это приводит к идее карты, но карты не в привычном понимании этого термина как чего-то статичного, а карты как структуры, как системного набора данных с возможностью их разнообразного представления.

Многовариантность создания карт для специфических потребностей пользователя, оперативность, приближающаяся к реальному времени, интерактивность, предоставление новых средств анализа данных, снижает временные и экономические затраты на производство карт, разработку новых видов и типов карт.

Многослойная организация карты позволяет объединить и отобразить не только большее количество информации, чем в обычном формате, но существенно упростить анализ пространственных объектов. Разделение по слоям допускает разбивать данные

на типы, повышать эффективность интерактивной обработки, упрощать решение экспертных задач и процесс хранения информации в базах данных.

Компьютерная карта – это средство моделирования исторических процессов в пространстве и времени. Визуально исследуя с помощью ГИС технологий пространственные модели, можно интегрировать различные, иногда несопоставимые наборы данных.

В то же время цифровая карта наиболее удобна для простой визуальной обработки информации, так как имеет дело с двухмерными образами. Сами карты могут быть дополнены отчетными документами, трехмерными изображениями, графиками и таблицами, фотографиями и другими средствами, например, мультимедийными.

Одной из причин, по которой ГИС технологии не так широко применяются в исторических исследованиях – «привязка» этой технологии к количественным данным, что ограничивает ее применение в большом историческом исследовании, которое больше ориентируется на описательные источники, особенно тексты. Поскольку способ представления пространственных данных в ГИС является количественным, ГИС технологии могут визуализировать только точно расположенные области, описанные при помощи основных графических элементов – точки, линии или многоугольника. Такое представление не подходит для нечетких понятий, как «культурные области» и мест, точное расположение которых не известно.

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:

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

Связь между местоположением объекта и его значением обеспечивается атрибутивными данными, содержащими уникальный номер объекта. Это позволяет создать гипертекст, в котором «уникальные» слова служат ссылками на карту.

Такое построение текста акцентирует внимание пользователя на важной информации, необходимой для лучшего понимания 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.

В прикладном плане ГИС технологии использовались для визуализации исторического текста по книге Л. Н. Гумилева «Древняя Русь и Великая Степь». Основываясь на тематике выбранного текста, были выделены «уникальные» слова, отображающие названия географических объектов, исторических событий, имена людей. Каждому выбранному «уникальному» слову соответствует метка на карте. У тех исторических событий, которые не имели точных координат, привязка осуществлялась к примерной области. В результате проделанной работы был разработан алгоритм геокодирования гипертекста и построен исторический атлас.

В современных условиях, несмотря на широкое использование мультимедиа, текст остается одним из основных видов информации в большинстве электронных хранилищ. И поэтому использование ГИС технологий в прикладных исследованиях будет по-прежнему актуально.