

УДК 316.35

## МЕТОД ВЫЯВЛЕНИЯ СЕМАНТИЧЕСКОГО КАРКАСА С ИСПОЛЬЗОВАНИЕМ ДАННЫХ ЛОКАЛЬНЫХ СЕТЕВЫХ СООБЩЕСТВ НА ПРИМЕРЕ КУПЧИНО

Курохтин И.В. (ИТМО), Турков Л.А. (ИТМО)

Научный руководитель – д.ф.н., доцент Горнова Г.В.  
(ИТМО)

**Введение.** Направление исследований, связанных с выявлением локальных проблем городских территорий цифровыми методами, активно развивается в урбанистике. Новые методы уменьшают количество времени и ресурсов на определение уникальных характеристик и объектов территорий, которые формируют идентичность жителей. Социальные онлайн-группы, как площадки для коммуникации между местными жителями, являются местом сохранения цифрового следа процесса формирования вернакулярных районов разных масштабов. Изучение и представление наборов данных онлайн-групп представляет интерес для исследования внутренней городской структуры, а также актуально для создания документов территориального планирования.

**Основная часть.** Вернакулярный район - это территория, проживающее на ней население, а также ментальное представление о территории [1]. Последнее свойство отделяет городские вернакулярные районы от других любых районов, поскольку именно внутри вернакулярного района формируется идентичность. Вернакулярный район состоит из отдельных городских объектов, которые жители формируют в ментальном образе района - ориентиров [2]. Рассмотрение вернакулярного района через систему “идентичность - ориентиры - характеристики” возможно графовым способом. Такой способ используется в зарубежных исследованиях для выявления различия, дальнейшего брендинга и развития территорий.

Специфика социальных сетей формирует неструктурированность выгружаемых данных и разнообразие городского взаимодействия внутри сообществ, которое всё больше неотличимо от реального. Два основных режима коммуникации онлайн-групп (соседский и гражданский [3]) по-разному влияют на идентичность, что и отражается в анализируемых данных.

Использование обширных корпусов текстов социальных сетей способно обогатить представления о социальной структуре изучаемого объекта [4]. Учитывая специфику неструктурированных данных для их обработки на сегодняшний день задействуются различные языковые модели для широкого спектра задач.

Предлагаемый нами метод основывается на использовании дообученной языковой модели BERT для извлечения топонимов и ключевых слов из текстов постов и комментариев социальных сетей [5]. При помощи оценки семантической близости между ключевыми словами, относящимися к разным топонимам можно выделять специфические связи между ними, например, при помощи построения графа знаний, который активно используется для отображения связей между семантическими сущностями в текстах [6]. Уникальные сущности, не имеющие связи с другими топонимами или группы топонимов рассматриваются как уникальные аспекты, которые полезны при выявлении семантического каркаса исследуемых территорий.

Выходными данными метода является ориентированный граф, узлами которого являются извлеченные из текста топонимы и сущности в виде ключевых слов, а рёбрами смысловые связи выделенных сущностей.

**Выводы.** Разработанный метод способен выявлять городские сущности, связанные с идентичностью района или города: внутригородские объекты и их народные названия, события и реакцию населения на них, проблемы городской среды. Полученный алгоритм позволяет создавать граф знаний о территориях и их ментальных образах, выделяя наиболее центральные объекты и важные семантические связи. В дальнейшем метод применим для

исследований вернакулярных районов (делимитация, выявление идентичности), учёта и размещения новых точек притяжения района. Полученные данные социальных сетей могут играть подтверждением сильных и слабых сторон района для градостроительного анализа и дальнейшего развития.

#### **Список использованных источников:**

1. Казакова Г. М. Вернакулярный район как условие интенсификации социальных процессов //Социологические исследования. – 2017. – №. 9. – С. 57-65.
2. Tuan Y. F. Space and place: The perspective of experience. – U of Minnesota Press, 1977.
3. Павлов А. В. Локальные городские сообщества в социальных сетях: между «соседской» и «гражданской» коммуникацией //Лабиринт. Журнал социально-гуманитарных исследований. – 2016. – №. 5. – С. 46-57.
4. Erétéo, G., Gandon, F.L., & Buffa, M. (2011). SemTagP: Semantic Community Detection in Folksonomies. 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, 1, 324-331.
5. Nimisha, M., Shamitha, T., & Mishra, G. (2023). Comparative Analysis of Embedding Models for Keyphrase Extraction: A KeyBERT-Based Approach. 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), 1-6.
6. A. Sheth, S. Padhee and A. Gyrard, "Knowledge Graphs and Knowledge Networks: The Story in Brief," in IEEE Internet Computing, vol. 23, no. 4, pp. 67-75, 1 July-Aug. 2019