

Инструменты автоматизированного сбора и анализа социологической информации о территориальной идентичности жителей крупных городов

М. С. Байнова¹, А. М. Соколов^{2}*

¹ Университет «Синергия», Москва, Россия

² Филиал ФГБОУ ВО «Национальный исследовательский университет "МЭИ"» в г. Смоленске, Смоленск, Россия

* ansokol98@mail.ru

Аннотация. В статье предлагается алгоритм автоматизированного поиска и первичного анализа социологической информации для исследования территориальной идентичности жителей районов крупных городов в интернет-источниках. В качестве основного источника информации рассматриваются сообщества в социальных сетях (на примере социальной сети «ВКонтакте»), в качестве вспомогательных – интернет-порталы о топографических объектах, находящихся на исследуемых территориях. Показано, что с точки зрения информационного обеспечения наибольшим потенциалом обладают публичные страницы и группы с открытой и ограниченной «стенкой». Разработанный алгоритм предполагает выделение релевантных для решаемой задачи групп, выявление содержащихся в них записей по районной тематике и определение показателей активности участников сообщества при обсуждении территориальных проблем. Извлечение требуемой информации осуществляется посредством взаимодействия с сервером социальной сети с использованием официального программного интерфейса приложения (API). Для идентификации сообществ и записей предлагается использовать методы морфологического анализа текстовой информации. Описана программная реализация указанного алгоритма на языке Python 3.8.5, которая включает оригинальные функции для получения данных о сообществах по их идентификационным номерам, для формирования набора урбанонимов для заданного района и др. С использованием разработанной программы проведен анализ территориальных групп трех районов г. Москвы. Определена погрешность результатов работы программы относительно результатов, полученных вручную.

Ключевые слова: автоматизированное извлечение данных, анализ текстов, Python, социальные сети, территориальная идентичность, социологические исследования

Для цитирования: Байнова М. С., Соколов А. М. Инструменты автоматизированного сбора и анализа социологической информации о территориальной идентичности жителей крупных городов // Прикладная информатика. 2021. Т. 16. № 2. С. 92–102. DOI: 10.37791/2687-0649-2021-16-2-92-102

Tools for automated collection and analysis of sociological information on the territorial identity of city residents

M. Baynova¹, A. Sokolov^{2*}

¹ Synergy University, Moscow, Russia

² Branch of the National Research University "MPEI" in Smolensk, Smolensk, Russia

*ansokol98@mail.ru

Abstract. The paper proposes an algorithm for automated search and initial analysis of sociological information aimed at studying the territorial identity of city area residents using Internet sources. Communities of social networks, e.g. VKontakte, are considered as the main data source, and websites of topographic objects found in the territories under study are used as auxiliary information sources. It is demonstrated that, in terms of information support, public pages and groups with open or restricted access walls have the greatest potential. The developed algorithm implies selecting relevant groups, finding content concerning area issues, and determining the indices of community activity in discussing territorial problems. The required information is retrieved through the interaction with a social network server with the use of the official Application Programming Interface (API). To identify communities and posts, it is proposed to apply methods of morphological analysis of textual information. The software implementation of the algorithm is described in Python 3.8.5, including original functions for the acquisition of data on communities by their identification numbers, for the formation of a set of urbanonyms for a specified area, and some other ones. The developed program has been used to analyze territorial groups in three areas of Moscow; the results of the analysis enable us to estimate the degree of the territorial identity of their residents. The analysis of the error in the results of automated data collection and processing shows good agreement of these results with manually obtained ones, i.e. the error is 2.6% in the identification of relevant groups and about 3% in the identification of posts on area issues. Therewith, a much higher speed of response and lower labor effort required to perform routine operations allow the algorithm and the implementing computer program to be viewed as an effective tool for sociological research based on data from social networks.

Keywords: automated data mining, text analysis, Python, social networks, territorial identity, sociological research

For citation: Baynova M., Sokolov A. Tools for automated collection and analysis of sociological information on the territorial identity of city residents. *Prikladnaya informatika*=Journal of Applied Informatics, 2021, vol.16, no.2, pp.92-102 (in Russian). DOI: 10.37791/2687-0649-2021-16-2-92-102

Введение

Социологические исследования являются одной из ключевых составляющих при принятии важных решений в области управления и социальной инженерии, поскольку позволяют сформировать представление о текущем состоянии общественного мнения, получить обратную связь

от населения относительно актуальных вопросов, а также проследить динамику протекания социальных процессов.

Одной из современных проблем, затрагивающих как социологию, так и урбанистику и экономическую географию, является оценка территориальной идентичности населения различных административных образований.