

СИСТЕМНЫЙ АНАЛИЗ ГОРОДСКОЙ СРЕДЫ ПОСРЕДСТВОМ МОНИТОРИНГА СОЦИАЛЬНОЙ СЕТИ TWITTER

ВВЕДЕНИЕ

Социальные сети предоставляют большой объём географических данных и могут использоваться для создания модели городского пространства с целью планирования инфраструктуры для общественных мероприятий [1,2]. Для изучения городского пространства Индианаполиса использовался трафик социальной сети Twitter сгенерированный во время Суперкубка 2012 года.

I. АРХИВАЦИЯ И ФИЛЬТРАЦИЯ ТРАФИКА СОЦИАЛЬНОЙ СЕТИ

Для архивации трафика используется инструмент The Archivist (Microsoft, 2010), обеспечивающий доступ к базе данных Twitter и сохраняющий твиты с метками [3]. Сохранение всех твитов с метками Суперкубка теоретически возможно, но практически труднодостижимо [4]. Ограничения, существующие в Twitter, не могут гарантировать, что инструмент архивирования смог вернуть каждый твит с целевым хэштегом в течение определенного периода времени (программное обеспечение позволяет выгрузить не более 1500 твитов за раз [5]). Во время самой игры была захвачена лишь небольшая часть общего трафика, так как во время перерыва Суперкубка трафик Twitter превышал 10 000 твитов в секунду (хотя какая часть этих твитов содержала наши целевые хэштеги, неизвестно). На начальном этапе были выбраны официальные хэштеги мероприятия, а на последующих этапах фильтрации были добавлены неофициальные принятые пользователями сети для маркировки событий не предусмотренных расписанием [6,7]. Например, если знаменитость, посетившая город, написала в твиттере о том, что она находится в игре, люди начинают пользоваться хэштегом этой знаменитости, вместо официальных тегов Суперкубка. Для этого анализа были сохранены только твиты с 23 января 2012 до 9:00 утра 6 февраля 2012. Пользователи Twitter часто применяют несколько хэштегов в одном твите, поэтому многие твиты были записаны несколькими раз. Так как каждому твиту присваивается уникальный идентификационный номер, записи дубликатов легко идентифицируются и удаляются.

II. КОДИРОВАНИЕ ТРАФИКА СОЦИАЛЬНОЙ СЕТИ

Для эффективного анализа полученных данных необходимо выделить основные элементы городской среды, по которым в дальнейшем будет проведено кодирование собранных сообщений [9,10].

Таблица 1 – Базовые элементы городской среды

Название элемента	Описание
Маршруты	Улицы, тропы и другие маршруты, по которым движутся люди
Границы	Края или переходные зоны между двумя зонами. Делятся на природные (речные и горные) и искусственные барьеры (дороги, стены и железнодорожные пути).
Районы	Большие участки города, уникальными свойствами. Например спальный район, центр города или офисный квартал.
Ориентиры	Легко идентифицируемые объекты, которые служат ориентирами, такие как памятники и стадионы.
Узлы	остановки общественного транспорта, площади, перекрестки или вокзалы.

Вручную были соотнесены ключевые слова с пятью элементами городского пространства (табл. 1). Каждое из пространственных слов и словосочетаний, идентифицированных в ходе контент-фильтрации, было отнесено к одной из категорий IC, чтобы обеспечить машинное кодирование контента твита с помощью поиска по шаблону. Из 78 терминов 30 (38,5 процента) были узлами, 18 (10,3 процента) - маршрутами, 16 (20,5 процента) - названными ориентирами, 11 (14,1 процента) - районами и три (3,8 процента) - краями

III. АНАЛИЗ РЕАЛЬНЫХ ДАННЫХ

С 12:01 утра 23 января 2012 г. до 9:00 утра 6 февраля 2012 г. The Archivist зарегистрировал 648 893 твита, содержащих один или несколько целевых хэштегов. Удаление дубликатов сократило количество до 499 756 твитов. Среднее количество твитов в день составило 33 317. Зарегистрированный трафик "Твиттера" был относительно низким в течение большей части иссле-

двумя периода с тенденцией к росту объема в четверг, 2 февраля (рисунок 1), возможно, из-за требования многих отелей в центре города, чтобы посетители Суперкубка резервировали номера как минимум на три или четыре ночи. Пик дневного трафика в 250 000 твитов произошел в день игры 5 февраля. Среднее количество записанных твитов в час составило 1448,6 (рисунок 2).

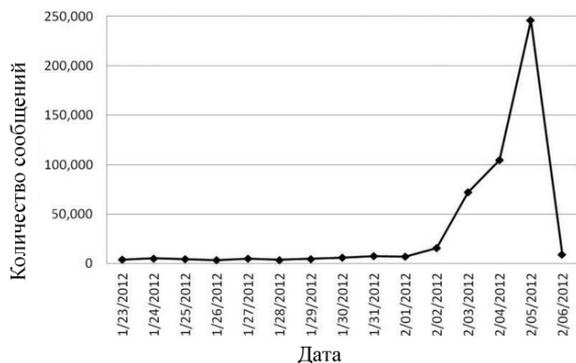


Рис. 1 – Зависимость количества сообщений от дня мероприятия



Рис. 2 – Зависимость количества сообщений в час от времени суток

IV. ЗАКЛЮЧЕНИЕ

Все элементы изображения города были найдены в специфическом для Индианаполиса трафике Twitter. В то время как узловые ссылки были наиболее распространены среди 78 географических идентификаторов. Чаще всего упоминались районы и ориентиры. Ориентиры и узлы обычно представляют собой отдельные объекты и легко запоминаются, служа в качестве ориентиров [8]. Упоминания границ практически отсутствуют и встречаются только в виде упоминания рек; это может указывать на их ограничен-

ную полезность в коротких текстовых описаниях городского пространства или на относительное отсутствие границ городского пространства Индианаполиса. Общая частота трафика в Твиттере и географических "твитов" увеличивалась с течением времени; тем не менее, они имели несколько иную схему распределения. Связанный с событиями трафик достигает пика в день игры, а географические ссылки достигают пика двумя днями ранее, что говорит о том, что посетители и местные жители по-разному используют социальную сеть в предыгровой период и в день игры. Частоты различных типов географических ссылок также различались по дням. Термины "район "узел" и "маршрут" достигли своего пика 3 февраля, в отличие от точек, пик которых пришелся на 5 февраля. Это повышает вероятность того, что туристы чаще задавали общие вопросы о местоположении или описывали общие планы в режиме онлайн, когда они впервые приехали в город, а в день игры внимание переключилось на стадион (ориентир).

V. СПИСОК ЛИТЕРАТУРЫ

1. Гольденберг и М. Леви. Расстояние не умерло: социальное взаимодействие и географическая дистанция в эпоху Интернета. Arxiv, abs / 0906.3202, 2009.
2. Arribas-Bel, K. Kourtit, P. Nijkamp, and J. Steenbruggen, "Cyber Cities: Social Media as a Tool for Understanding Cities," Applied Spatial Analysis and Policy 8: 3 (2015) 231–247.
3. D. Caduff and S. Timpf, "A Framework for Assessing the Salience of Landmarks for Wayfinding Tasks," Cognitive Processing 1: 7 (2006) 23–23.
4. G. Cao, S. Wang, M. Hwang, A. Padmanabhan, Z. Zhang, and K. Soltani, "A Scalable Framework For Spatiotemporal Analysis of Location-Based Social Media Data," Computers, Environment and Urban Systems 51 (2015) 70–82.
5. S. Carr, L. Rodwin, and G. Hack, "Kevin Lynch: Designing the Image of the City," Journal of the American Planning Association 50: 4 (1984) 523
6. C. Chew and G. Eysenbach, "Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak," PLoS ONE 5: 11 (2010) e14118.
7. Eagle, A. Pentland, and D. Lazer. Определение структуры сети дружбы с помощью данных мобильного телефона. PNAS, 2009.
8. Long, P. K. Baren, and R. Moore, "The Role of Space Syntax in Spatial Cognition: Evidence from Urban China," Conference Proceedings of the 6th International Space Syntax Symposium, (Istanbul, Turkey, June 2007).
9. A. Crooks, D. Pfoser, A. Jenkins, A. Croitoru, A. Stefanidis, D. Smith, S. Karagiorgou, A. Efentakis, and G. Lamprianidis, "Crowdsourcing Urban Form and Function," International Journal of Geographical Information Science (2015) 37–41.
10. Lynch, The Image of the City (Cambridge MA: MIT Press, 1960) 20-25

Якубович А.В., Кафедра информационных технологий автоматизированных систем Белорусский государственный университет информатики и радиоэлектроники, yakubovichandrei@gmail.com

Научный руководитель: Шилин Л.Ю., доктор технических наук, профессор, yakubovichandrei@gmail.com