

ПОЛЬЗОВАТЕЛЬСКИЙ ГЕОКОНТЕНТ КАК СРЕДСТВО ДЛЯ АНАЛИЗА И РАЗВИТИЯ ТУРИЗМА

Ахметов Р. Ш., Новиков Д.А

Оренбургский государственный университет, г. Оренбург

Развитие и внедрение современных технологий практически во все сферы нашей жизни породило, в частности, огромный и лавинообразно растущий поток информации. По данным компании IBM более 90 % данных в современном мире были созданы всего за последние два года [1]. Невиданный прежде объем и разнообразие данных поступает из самых разных источников. В их числе системы дистанционного зондирования Земли, информационные устройства с датчиками и мобильные устройства, способные определять местоположение, информация из социальных сетей, цифровые изображения и пр. В результате сформировался феномен и новое понятие «большие данные» (англ.: Big Data). Значительная часть этих данных имеет географическую привязку, т.е. является геоданными, что позволяет использовать их для географического анализа.

Появление и бурное развитие в последние годы картографических интернет-сервисов, основанных на концепции Web 2.0, а также широкое распространение краудсорсинговых проектов в геоинформационной сфере привело к тому, что миллионы обычных пользователей всемирной сети стали не только потребителями, но и активными производителями геоданных.

Производимые пользователями геоданные можно с определенной долей условности разделить на структурированные, т.е. определенным образом организованные и готовые к использованию данные и неструктурированные.

Значительная часть структурированных пользовательских геоданных создается с использованием таких популярных сервисов, как Google Maps, проекты Wikimapia и OpenStreetMap, а также ряда других.

Несмотря на то, что все эти проекты значительно отличаются друг от друга по целям и масштабам, можно говорить о создаваемом в их рамках пользовательском геоконтенте – то есть о географической информации, источником которой являются сами пользователи соответствующего интернет-ресурса. При этом можно выделить следующие общие черты, характеризующие процесс создания структурированного пользовательского геоконтента:

1) *Массовость*. В работе над созданием геоконтента на самых популярных картографических сервисах принимают участие сотни тысяч человек. Такое количество участников позволяет создавать действительно масштабные проекты, поддерживать постоянную актуальность информации и оперативно вносить необходимые изменения и правки.

2) *Непрофессионализм участников*. Подавляющее большинство людей, наполняющих интернет-проект геоконтентом, не имеет географического образования. В связи с этим возникает задача оценки и контроля достоверности создаваемых данных, частично решаемая силами самих участников проекта. Таким образом, пользователи выступают не только в качестве источника данных, они

берут на себя также функцию актуализации и контроля правильности имеющейся информации.

3) *«Волонтерский» характер деятельности участников.* Упомянутые выше проекты (Google Maps частично, остальные полностью) развиваются пользователями на добровольных началах, по модели краудсорсинга. Материальные вложения при этом необходимы лишь для разработки и поддержки сервиса, обеспечивающего взаимодействие пользователей. Само создание геоконтента осуществляется безвозмездно и не требует специальных финансовых затрат.

Большая часть картографических интернет-сервисов предоставляет возможности программного доступа к хранящейся на них информации через интерфейс API, что делает возможным автоматизированную выборку и обработку геоданных с целью дальнейшего их использования, например, в научно-исследовательских целях[2].

Следует заметить, что массив создаваемых пользователями географических данных не ограничивается специализированными сервисами, на которых пользователь создает готовый к использованию структурированный геоконтент. Не менее значительный по объему пласт географической информации может быть извлечен из размещенных в интернете фотоизображений, видеороликов, сообщений в социальных сетях и на тематических форумах, записей в блогах и тому подобного. В этом смысле понятие «пользовательский геоконтент» может трактоваться более широко - как любая доступная информация, производимая интернет-пользователем и имеющая локализацию в пространстве и времени.

Лавинообразному нарастанию количества такого геоконтента во многом способствует широкое распространение мобильных устройств, оснащенных GPS-приемниками, благодаря чему все большее количество создаваемой пользователями информации сегодня имеет географическую привязку. На сегодняшний момент во всемирной сети накоплен огромный и непрерывно пополняемый объем пользовательских данных, имеющих географическую компоненту. В настоящее время эта информация становится важным альтернативным источником данных для исследований в самых разных областях знаний.

В отличие от геоконтента на картографических интернет-сервисах, эта информация неструктурирована с географической точки зрения и ее использование в географическом анализе требует нестандартных подходов и специальной подготовки.

Среди сайтов-агрегаторов подобной информации с точки зрения анализа и развития туристической отрасли особо стоит выделить популярные фотохостинги типа Flickr (www.flickr.com) или инкорпорированного в Google Maps сервиса Panoramio (www.panoramio.com), представляющие собой коллекции фотографий, привязанных пользователем к определенному месту в пространстве.

Значительная часть выкладываемой пользователями в сеть информации (фотографии из отпуска, отзывы об отелях, отчеты о путешествиях) связана с туризмом. Этот факт обуславливает высокую пригодность пользовательского геоконтента как инструмента анализа и развития туризма.

Сбор исходных данных при традиционных исследованиях в области туризма, как правило, включает в себя социологические исследования и прямые наблюдения. Использование этих традиционных методов зачастую осложняется сложностью обеспечения репрезентативной выборки респондентов и достаточно высокой стоимостью исследований.

Использование пользовательского геоконтента как источника данных для исследования туризма позволяют получить результаты на значительной выборке, не требуют больших финансовых и трудовых затрат, при необходимости легко повторяются и актуализируются. Важным преимуществом пользовательского геоконтента является его открытость и доступность.

К недостаткам использования пользовательского геоконтента можно отнести не всегда тривиальную задачу вычленения географической составляющей из общего потока исходящей от пользователей информации, ее структурирования и дальнейшей интерпретации. Кроме того не следует забывать о том, что пользователи всемирной сети представляют только определенную часть туристов, наиболее активную и компьютерно-грамотную.

Можно выделить следующие основные возможности использования пользовательского геоконтента для анализа и развития туризма:

Изучение туристского образа территории. Пользовательский геоконтент, связанный с туризмом, как правило, отражает определенные впечатления отдельного человека о некоторой территории. Обобщая и анализируя подобные впечатления множества пользователей, можно получить представление о некотором туристском образе этой территории в массовом сознании.

Одним из инструментов исследования имиджа той или иной территории в сети интернет может стать анализ пользовательского геоконтента. Его «дилетантский» характер в контексте исследования сложившегося образа территории из недостатка становится преимуществом, позволяя исследовать обобщенное представление туристов о территории, со всеми его упрощениями, ошибками и т.п.

Основным способом исследования пользовательского контента в этом направлении может стать семантический анализ содержания представленной пользователем информации (тексты сообщений на форумах, подписи и тэги изображений) с выявлением положительных или отрицательных коннотаций, связанных с различными географическими объектами или достопримечательностями.

Анализ туристского образа территории, создаваемого пользователями сети Интернет, может найти применение в географической имиджелогии или программах по развитию туризма [3].

Исследование информационной освоенности территории. Предметом исследования в этом случае выступает активность пользователей интернет-ресурсов в конкретной точке пространства-времени. Каждый информационный элемент геоконтента (опубликованная фотография, пост в социальной сети или блоге) рассматривается как универсальное сообщение «я был здесь», мотивированное некоторым интересом пользователя. Смысловое наполнение сообщения

при этом остается за рамками исследования или служит для уточнения и дифференциации его результатов.

Участки территории с максимальной плотностью локализации пользовательского геоконтента позволяют выявить на местности так называемые «точки интереса». Сопоставляя эти точки с известными объектами на географической карте (культурно-историческими или природными достопримечательностями), можно сделать выводы об их популярности и востребованности у туристов.

Работая в этом направлении, исследователи чаще всего обращаются к анализу массива фотоизображений, имеющих пространственную привязку. Примером применения такого подхода могут служить работы эстонского ученого, профессора Т. Таммета (T. Tammet) [4], составившего карту плотности фотографий Panoramio, представленную на сайте <http://sightsmap.com> (см. рис. 1).

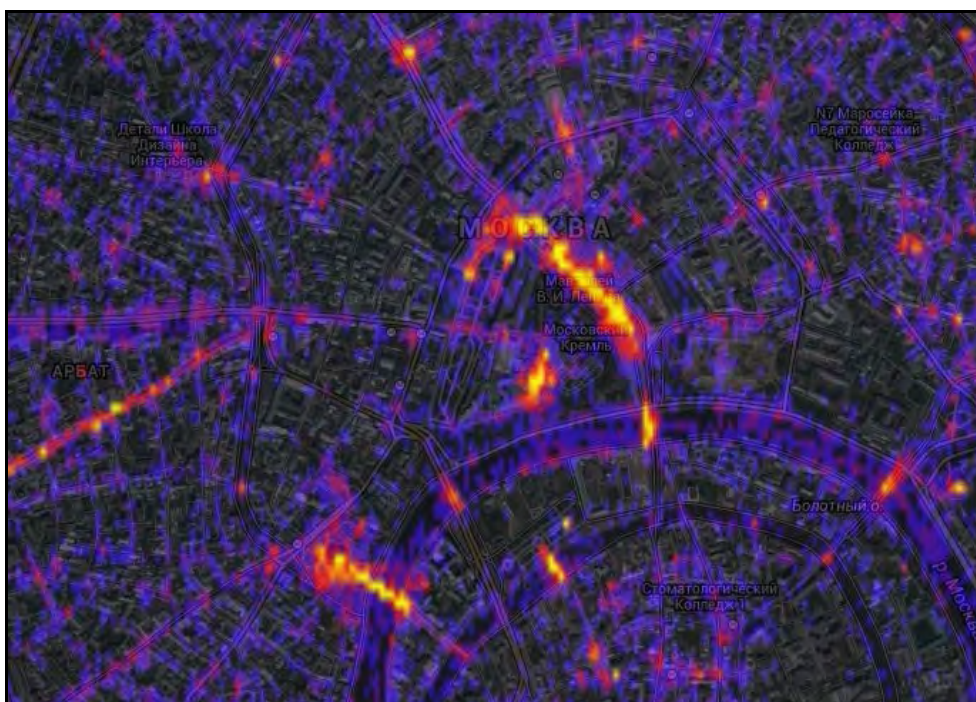


Рис. 1 - Карта плотностей фотографий Panoramio (центр Москвы)

Карта, оформлена в виде тепловой карты - более холодному цвету соответствует низкая плотность фотографий, более теплому - высокая (максимальной плотности соответствует желтый цвет).

Как видно на рисунке, по распределению плотностей фотоизображений можно достоверно выявить основные районы-аттракции туристов. Такая информация может быть полезна как частным лицам при планировании своих путешествий, так и организациям, задействованным в туристическом бизнесе.

Схожий подход был применен В. Фриас-Мартинез, В. Сото и др. (V.Frias-Martinez, V.Soto, H.Hohwald and E.Frias-Martinez) [5]. В своей работе они предлагают использовать сообщения в социальной сети Twitter (т.н. твиты) для районирования городской среды и локализации популярных достопримечательностей, при этом помимо анализа пространственного размещения твитов учитывается также время их публикации.

Рассматривая в качестве примера район Манхэттен в Нью-Йорке, исследователи выделяют кластеры повышенной активности пользователей Twitter в различное время суток в будние и выходные дни. В зависимости от того, какой кластер показывает наибольшую активность в определенное время, делаются выводы о функциональном использовании городской среды в этой зоне (выделяются районы офисов, жилая зона, места досуга населения – парки, музеи и пр., территория «ночной жизни»).



Рис. 2 – Карта функционального использования городской среды Манхэттена (слева направо выделены кластеры: бизнес, досуг, «ночная жизнь», жилье).

Участки территории с наибольшей пиковой активностью твитов относятся авторами исследования к достопримечательностям.

Пользовательский геоконтент в данном подходе может рассматриваться как цифровой информационный след пользователя. Анализ характера его распределения по изучаемой территории позволяет судить о её информационной освоенности, которая согласно А.А. Соколовой хорошо согласуется с интенсивностью туристских потоков [3].

Анализ перемещения туристов. Важной частью исследований, связанных с туризмом, является изучение динамики перемещения людей, определение наиболее популярных направлений путешествий, сезонные изменения и т.п.

В своем исследовании Б. Гавелка (В.Hawelka) [6] использовал данные о пространственном распределении твитов для анализа перемещений их авторов по миру. Авторы исследования обработали около миллиарда твитов, имеющих пространственную привязку. Для каждого пользователя социальной сети они выделили регион проживания (по наибольшей активности в Twitter), а затем проанализировали число, время и расположение в пространстве твитов пользователей за пределами этого региона. Рассматривая полученные результаты для различных стран, были сделаны выводы о мобильности их населения (см. рис. 3), коррелирующие с данными о продаже билетов на международные авиарейсы.

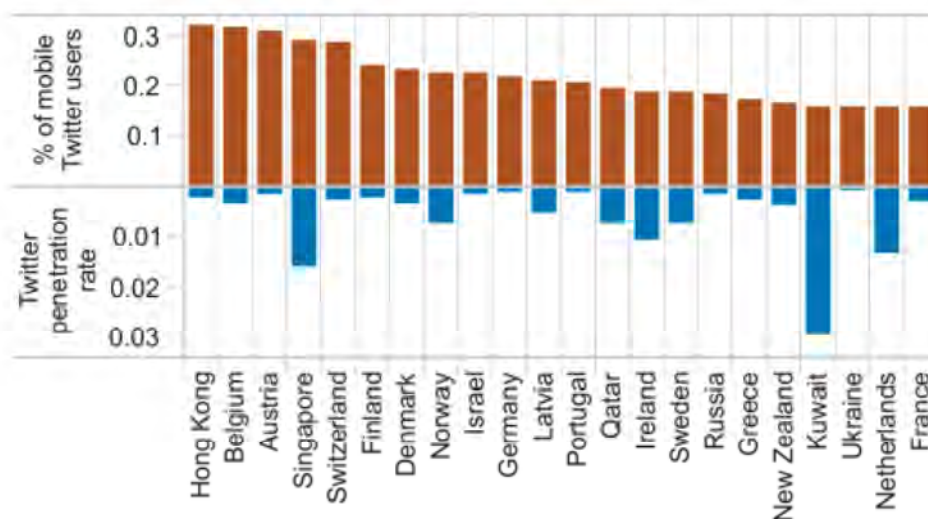


Рис.3 – Страны с максимальным рейтингом активности путешественников в 2012 году (фрагмент).

В верхней части диаграммы - процент пользователей, хотя бы раз оставивших сообщение в Twitter из-за пределов страны. В нижней части – распространенность социальной сети среди населения.

Помимо общей характеристики мобильности населения (в частности туристов) в рамках исследования была получена информация об основных направлениях поездок, изменения количества поездок во времени и многом другом.

Использование описанных выше приемов анализа геоконтента может позволить существенно повысить качество, разнообразие и эффективность предлагаемых туристических продуктов.

Конечно, возможные способы использования пользовательского контента как инструмента анализа и развития туризма не исчерпываются направлениями, рассмотренными выше. Появление новых методик обработки и интерпретации геоданных, созданных пользователями, несомненно, вопрос самого ближайшего времени.

Список литературы

1. Официальный сайт компании IBM [Электронный ресурс]. — Режим доступа: <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>
2. **Goodchild M. F.** Citizens as sensors: web 2.0 and the volunteering of geographic information / M. F. Goodchild // *GeoFocus*. — 2007. — № 7. — С. 8-10.
3. **Соколова А. А.** Виртуальное освоение и виртуальные образы региона (по данным Google Earth и Panoramio) // *Известия РГО*. 2010. Т. 142. Вып. 6. С. 31–40.
4. **Tammet, T.** Sightmap: crowd-sourced popularity of the world places / T. Tammet, A.Luberg, P.Järv // *Information and Communication Technologies in Tourism 2013: proceedings of the International Conference in Innsbruck, Austria, January 22-25, 2013*. — С. 314-325

5. *Characterizing Urban Landscapes Using Geolocated Tweets* / V. Frias-Martinez, V. Soto, H. Hohwald, E. Frias-Martinez // *proceedings of the 4th International Conference on Social Computing (SocialCom'12) in Amsterdam, Netherlands, September 3-5, 2012.* — С. 239-248
6. *Geo-located Twitter as the proxy for global mobility patterns* / B Hawelka, I Sitko, E Beinat и др. // *preprint in arXiv:1311.0680 [Электронный ресурс].* — Режим доступа: <http://arxiv.org/abs/1311.0680>