Barcelona vs. #Barcelona:
Semantic Analysis of Twitter Public Data Generated by Tourists

Project Description

1. Methods

- Data source: more than 100 000 tweets produced in the city of Barcelona in March 2015 (tweets geotagged inside the city limit) collected through Twitter API

- Filtering — stage 1: tweets were classified by the origins of their authors. "Locals": people who have been regularly tweeting in the city over the last several months. "Tourists": people who have been tweeting in the city for two weeks or less and regularly geotag their posts in other cities

- Filtering — stage 2: tweets posted by tourists were classified by predominant sentiments. Classification based on Habidatum state-of-the-art semantic analysis algorithms and manual refinement through Habidatum in-house crowdsourcing platform

- Analyzing additional metadata: timestamps and languages. To be included in the visualization on further stages of project development

- Visualization: tweets mapped by their origins — geotags left by the users. Visualization principles: "palimpsest" — overlapping opposite categories for comparative analytics, "dialogue with data" — filters allowing for sampling and selecting particular data layers, "interactive visual analytics" — visualization regimes for "playing" with the maps (while color reflects the density of Twitter activity, the size of "blobs" shows its spatial patterns which vary depending on the generalization levels chosen)

2. Findings

Statistics

- High activity of tourists in Twitter: tweets published by tourists account for 40% of the total amount

- Languages distribution: English — up to 25%, French — 8%, Portuguese — 7%, other languages — less than 1%

- Dominantly positive image: tweets with positive emotions account for 81% of all the posts published by tourists

- Reposts from external platforms (Instagram, Foursquare) account for more than 13% of all the tweets — to be considered (along with Twitter) as self-sufficient data sources for community analytics and digital marketing activities

HABIDATUM

MATHRIOSHKA

Geography

- Geography of the tweets (by their places of origin) produced by tourists differs a lot from the ones produced by locals. The first is represented by a smaller area with much bigger concentrations

- Locals tweet from broader areas, and there can be several factors behind it: (a) they are not linked so much to free WiFi available in cafes and bars predominantly concentrated in the center, (b) locals "consume" broader territory in general with more diverse sets of activities, (c) they are interested in a broader spectrum of location types (not just attractions, shops and cafes, but also sports grounds and gyms, schools and other social infrastructure, grocery stores and markets etc.)

- Tourists tweet most actively from Gothic Quarter, Passeig de Gracia (triangle from the beach to Placa de Catalunya and Passeig de Gracia), as well as from some attractions outside historical center (Sagrada Familia, Camp Nou and surroundings, beach, Avinguda Diagonal)

- Among underused "outsiders" are natural landmarks — parks, beaches et al. – and the most striking example is Montjuic

- Tourists vs. locals — differences in activity: (a) tourists do not actively post from 22@ or Port Olimpic de Barcelona, while locals are active in these local centers, (b) Eixample is more popular among locals than among tourists, (c) locals post from the beach less than tourists while the well-equipped sea-front tends to be actually popular among Barcelona citizens, (d) tourists post more than locals from the airport, but the overall activity in the airport is lower than in the center of the city (probable explanations: free WiFi provided only for 30 minutes, less immediate impressions and attractions — the majority of tweets presents post-trip reflection or, oppositely, expectations, thus predominant sentiments are positive)

- Geography of Twitter activity in Greater Barcelona area repeats the structure of land-use and settlements: huge concentration in the city of Barcelona, activities along major transportation links, concentrations in Barcelona suburbs where the volume of activity correlates with the sizes of the settlements

Content analysis

- "Event economy" works well: a lot of tweets are related to Mobile World Congress (a yearly international event for mobile industry companies) and to sport events at Camp Nou

- Negative tweets are dispersed and scattered, there are no significant concentrations and place-related specific negative senses; negative emotions refer to regular activities and to the most popular places without any specific purports ("routine" negative emotions)

- Positive tweets are in large part place-related and form apparent areas of concentration; large parts of place-related positive tweets contain photos and exact place names (landmarks, cafes etc.)

Plans for further analytics

- Broader time coverage: by extending retrospective analytics up to a full year, or several years we could reveal seasonal patterns, obtain new and validate the existing observations

- Space-time analytics: we could determine overused and underused spatio-temporal niches in the city, which could be a good basis for the time coordination plan of the city. We believe that by employing such a plan the tourist pressure on the overcrowded central historical areas may be dramatically reduced

HABIDATUM

MATHRIOSHKA

- Deeper semantic analysis, through adding new dimensions (languages, key discussion topics) and sources (social networks, online media et al.) of semantic data will help us better understand tourist-related issues, and suggest a roadmap for resolving them

- Overlapping semantic data with other types of spontaneous data (cell phone signals, credit card transactions, apps, etc.) will give us an unprecedented insight into the structure of the urban space including its consumer capacity, proxies for the real estate price geography, etc.

- Expanding regional coverage, international comparisons and city benchmarking could put Barcelona into the broader regional and international context

- Creating analytical tools based on real time spontaneous data to support adaptable tourism infrastructure development, crowd management, flexible space-time planning and coordination

Habidatum Philosophy

1. The Extensions of City

"With the arrival of electric technology, man has extended, or set outside himself, a live model of the central nervous system itself"

Marshall McLuhan, Understanding Media: The Extensions of Man

Internet and ICT allowing for interconnection of all elements of urban habitat and citizens, bridging space and time barriers through continuous data exchange, dramatically shift the way cities operate and should be managed.

Spontaneous (big) data generated by people, objects and environment through various devices, though being a derivative of their regular activities, forms a complete extension of the city. The environment produced by data exchange is not just a footprint of regular "real city" activities or an autonomous "virtual city" layer — it is an extension of the habitat, influencing and influenced by it.

Neither habitat itself, nor its spontaneous data extension is self-sufficient taken separately. They operate in close connection and should be analyzed and managed correspondingly.

2. Digital Jane Jacobs

"The "message" of any medium or technology is the change of scale or pace or pattern that it introduces into human affairs"

Marshall McLuhan, Understanding Media: The Extensions of Man

Habidatum focuses its work on the analytics of urban community behavior in the interconnected environment. We develop analytical tools tailored to help urban administrations and local businesses navigate in hyper-dynamic and unpredictable environment of the city.

Like the famous urban activist Jane Jacobs we support inclusive and responsive urban governance considering community needs — however, the methods are different. We suggest making spontaneous (big) data work for human-scale urban development through analytics of the attitudes, habits, lifestyles and behavioral patterns of micro-communities.

HABIDATUM

MATHRIOSKA

3. Supporting Lean Management

"We need readiness, not plans. We need such readiness as permits to spot the rising challenge when it is still within the agent's power to respond, and own errors when it is still possible to repair them"

Zygmunt Bauman, City of Fears, City of Hopes

Responsiveness, adaptation, readiness and coordination are in the core of the way modern urban habitat can be managed. Control and planning prove their insolvency in the era of self-sufficient and self-tuning communities, adaptive practices and unpredictable patterns of space-time consumption.

Analytical tools developed by Habidatum serve the goals of lean urban governance and lean business development in urban space-time — so-called "grammatical", "dialogical" or "agile" planning. We build a powerful platform capable of collecting and processing large sets of data and its translation to intuitive analytical visualizations open for a dialogue with analysts and ultimate decision-makers in the fields of urban governance, real estate management, retail and retail banking networks development et al.

4. Habidatum Platform

Habidatum platform allows for ongoing analytics of urban community behavior and values based on various types of spontaneous data, such as the senses extracted from social media, cell phone signals or WiFi access data showing mobility patterns, information on urban community activities and transactions. The platform includes technology infrastructure as well as analytical frameworks creating sense out of the chaotic sets of digits.

Habidatum platform will be released for commercial use in 2015. Before the release we are publishing a series of international pilot projects showcasing the capabilities of Habidatum analytical visualization toolkit as well as the unlocked power of spontaneous data.

Credits & Contacts

Alexei Novikov, curator
Arkady Dushatsky, data analytics
Dmitry Demeshkin, web development
Eduard Haiman, design/concept
Katya Serova, analytics/concept
Vadim Smakhtin, software development/data analytics

If you have any questions and thoughts about the project and in case you plan to join our further studies, please ask (at) habidatum.com

ask@habidatum.com

HABIDATUM

MATHRIOSHKA