BeGeo

27-10-2020 www.begeo2020.be

Wordcrowd - A Location-Based Application to Explore the City based on Geo-Social Media and Semantics

enzo Milleville, IDLab – CartoGIS, Ghent University

Steven Verstockt, Nico Van de Weghe, Dilawar Ali – Ghent

Francisco Porras-Bernardez, Georg Gartner - TU Wien



IGN Institut Géographique National

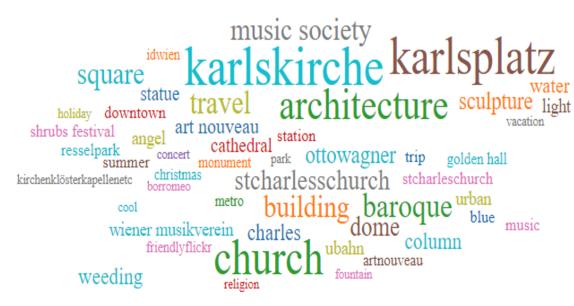


University

co-founding partner

Outline

- Goals and context
- Data collection and preprocessing
- Tourism interest analysis
- Visualization and interface
- Coronavirus sentiment analysis
- Conclusion and future work





Goals and Context

- Wordcrowd focuses on social media data (Flickr)
- Huge amount of spatially-related data
- Automatically determine Areas of Interest (AOIs)
- Compare tourism interest for different nationalities
- Focus on both the location and the content of each post
- Visualize results in an intuitive LBS application



Data Collection

- Geolocated Flickr image tags from 2004-2018 for continental Europe
- Only 33% of total users specified country
- Home determination algorithm: precision of 0.87
- Limit scope to points in Austria (370.000) & Belgium (430.000)



Vienna

Ghent

Brussels



Data Preprocessing

- Multilingual data => translate to English
- Filter irrelevant words (brand names, stop words, etc.)
- Group similar words together with NLP techniques (stemming/lemmatization)
- Try to split joined hashtags (e.g. viennaaustria)
- Filter redundant multilingual city/country names via Wikipedia/Wikidata



Data Preprocessing

- Word cloud of original picture tags near Belvedere palace, Vienna
- In almost every cluster the city name and country name appears in multiple languages
- Many irrelevant tags (Nikon inhoneography inhone8nlus...)

austria garten vienna viennaaustria vienne Wienaustria Österreich museum ^{fountain} statue vienna2018 travel nikon castle baroque internet intern 2018summereuropetrip iphoneography squareformat viena building trip upperbelvedere sculpture iphone8plus geotagged W1en europe wienbelvedere flower autriche botanischergarten plant europa park garden schloss belvedĕre schlossbelvedere

Data Preprocessing

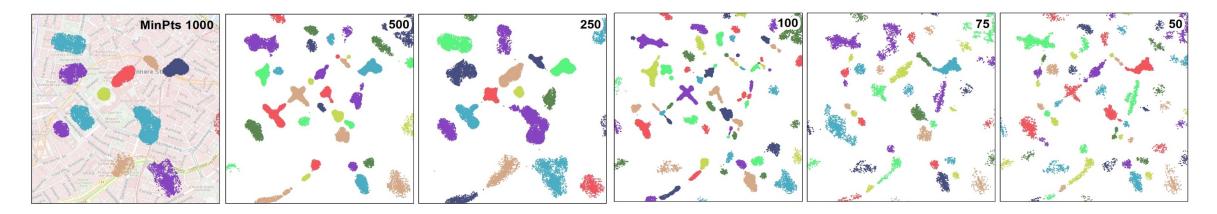
- Preprocessing improves the visualization
- Still some problems with joined hashtags, hard to automatically find a useful split

monochrome tourism park palace palace palace baroque
trip light world blue belvedere palace baroque
tour holiday belvederepalace museum vacation
tourist castle sculpture lowerbelvedere
flower



Clustering

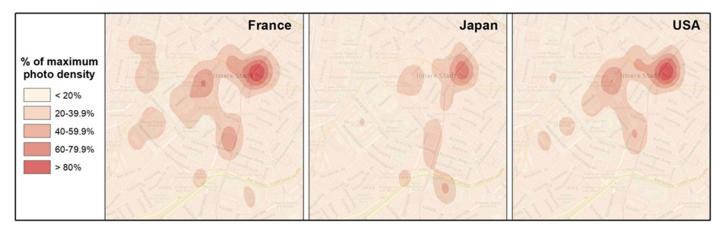
- By spatially clustering the data (HDBSCAN), AOIs can be automatically extracted
- Multi-scale clustering to improve visualization & performance
- Visualize clusters as polygons in the app (convex hull)





Tourism Interest Analysis

- Filter data on nationality
- Generate footprints for each via kernel density estimation (KDE)
- Most popular locations (hotspots) are often shared
- Differences in tourism interest patterns (Center of Vienna)



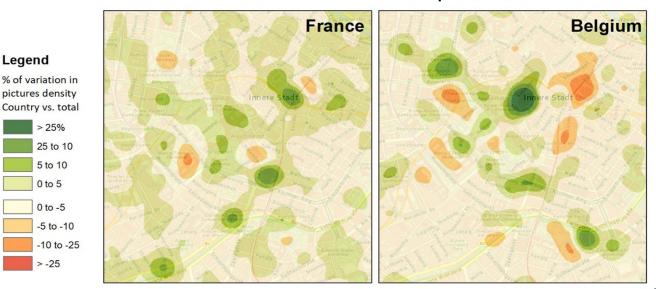
27-10-2020

www.begeo2020.be



Tourism Interest Analysis

- Find relative interest by performing map algebra (Average vs country)
- Green areas indicate more interest than the average user
- Able to determine areas of interest that are specific to certain nationalities





- Wordcrowd: Lightweight JavaScript LBS web application with PostGIS backend
- AOIs were extracted for three different scales
- Most frequently occurring tags saved in the database
- Geolocated AOIs and their descriptions generated from the Flickr picture tags
- Dynamically fetch data near the user at the right scale to reduce network overhead
- As the user explores the city, nearby areas of interest are discovered



- Fetch and visualize nearby clusters in real time as polygons
- Word cloud of related tags is shown when an AOI is clicked
- Dynamic zoom is necessary to improve visibility & user experience
- <u>http://bit.ly/wordcrowd</u>





- KDE rasters were calculated for several nationalities
- Vienna, Brussels, and Ghent
- Exported as GeoJSON to easily import into web application
- Scalable design => more regions/nationalities can always be added
- User can filter data on the fly

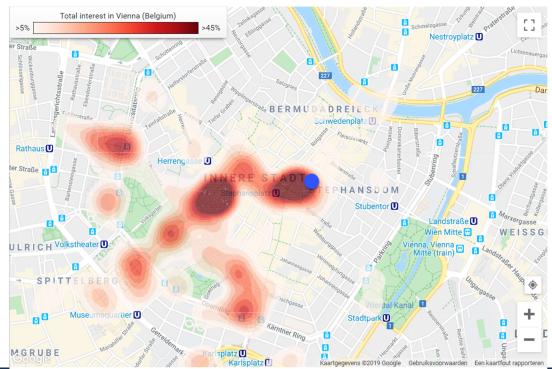
Filter the data	×	
KDE filter: Select a nationality for one of the available regions		
Vienna		
Belgium	√ ▼	
Select a nationality		
All-Austria		
All-Belgium		
All-France		
All-Germany		
All-The Netherlands		
All		
Asia		
Austria		
Belgium		
Czech Republic		
Germany		



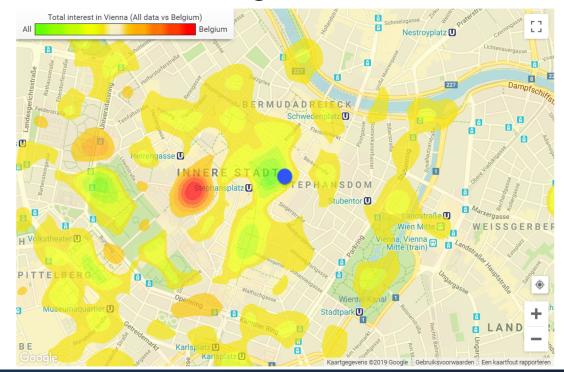
• Total interest in the center of Vienna for Belgian tourists

 $\cup \Theta$

Dev



• Relative interest in the center of Vienna for Belgian tourists



- Specific tags show up for several countries of origin
- E.g. : tags about the Great War show up in data from English tourists in Belgi

bridge localhistory toengu the line dagvandearbeid lievegem nowandthen socialist iseries june channel eastflanders blue spagent urban belfrv west tetramesh building cathedral istation medieval tower architecture square oggle railway memorial boat boat boat december mayday historic army great house electric internationalworkersday

contest

spafrancorchamps

courtneysant parismitton



Coronavirus Sentiment Analysis

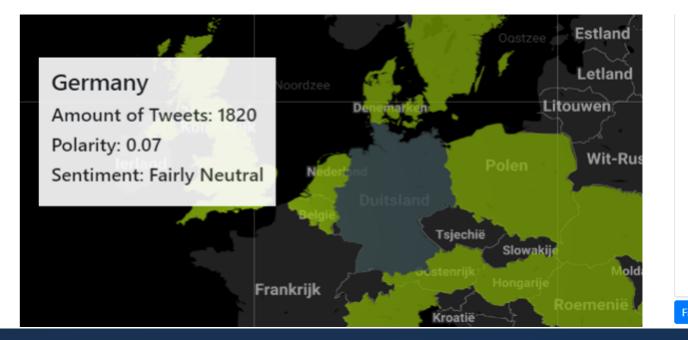
- Applied the Wordcrowd framework to tweets mentioning Coronavirus
- Tweets were preprocessed and dynamically clustered in hexagonal map visualization
- <u>http://bit.ly/wordcrowd-covid</u>





Coronavirus Sentiment Analysis

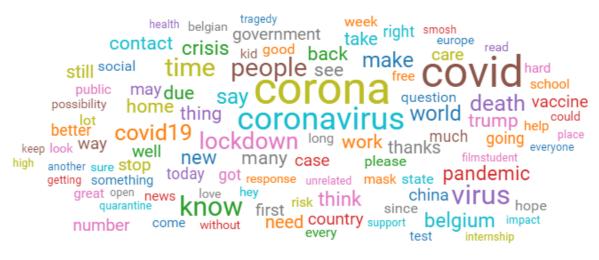
- Alternative visualization per country using GeoJSON
- Basic client-side temporal filtering



Portugal Google	Spanje	reense Z	ee Middelland:
Total tweets	: 68097		
Filter dat	а		
Language:			
en		×	~
Start date:	04/28/2020		
End date:	05/26/2020		
ilter data To	oggle Countries		

Coronavirus Sentiment Analysis

- Wordclouds are generated on the fly with filtered data
- Average sentiment for countries was not informative (always neutral)
- Wordclouds of different countries were very similar





Conclusions

- Social media data requires a large amount of preprocessing
- Developed framework to collect, analyze & visualize geolocated data
- Framework potentially applicable in main urban areas worldwide
- Tourism interest can differ greatly in both location and interest
- Can provide useful insights to various institutions (e.g. museums) about their social media reach and audience
- Can be used to tailor information based on a visitor's nationality



Future Work

- Improve grouping of related tags
- Analyze & integrate other data sources (e.g. full-te
- Add additional temporal filters (evolution over tim
- Automatically link with open data sources (Wikipedia,
- Recommending AOIs that correspond to user interest





BeGeo

27-10-2020 www.begeo2020.be

Wordcrowd - A Location-Based Application to Explore the City based on Geo-Social Media and Semantics

Kenzo Milleville, IDLab – CartoGIS, Ghent University kenzo.milleville@ugent.be Wordcrowd: <u>http://bit.ly/wordcrowd</u>



co-founding partner



Nationaa

Geografisch Instituut