

Comparing Social Media Topics of Interest associated with places according to user's origin



by **MARIA ATHANASIOU**

The basic theme of this thesis is to compare Social Media Topics of Interest or Terms associated with Points of Interest according to user's origin. It is implemented in the two basic phases. The first phase includes the extraction and visualization of the Topics of Interest from Wikipedia Articles, Flickr and Twitter. The second phase examines the similarity between Flickr, Twitter and Wikipedia Articles.

INTRODUCTION

Social media data are huge data in comparison with data collected by questionnaires or other conventional methods. Recently research findings indicate that social media data are useful for studying the way people perceive and have knowledge of the environments [1].

RESEARCH QUESTIONS

- How similar are the terms extracted from Geosocial Media (Twitter & Flickr) and Wikipedia articles (WA) when associated to the same Points of Interest (POI)?
- Are the same terms present in all Twitter and Flickr?
- How similar are the frequencies of each term according to user's place of origin?
- What proportion of POI present in Twitter and Flickr?

WIKIPEDIA & POI

Firstly, we collect 23 prominent POI of the city of Vienna, based on a list and collect the Wikipedia articles for them in German and English language. Then, we extract terms from Wikipedia with an online wordcloud tool. The wordlist is free from terms that appear only once or from toponyms

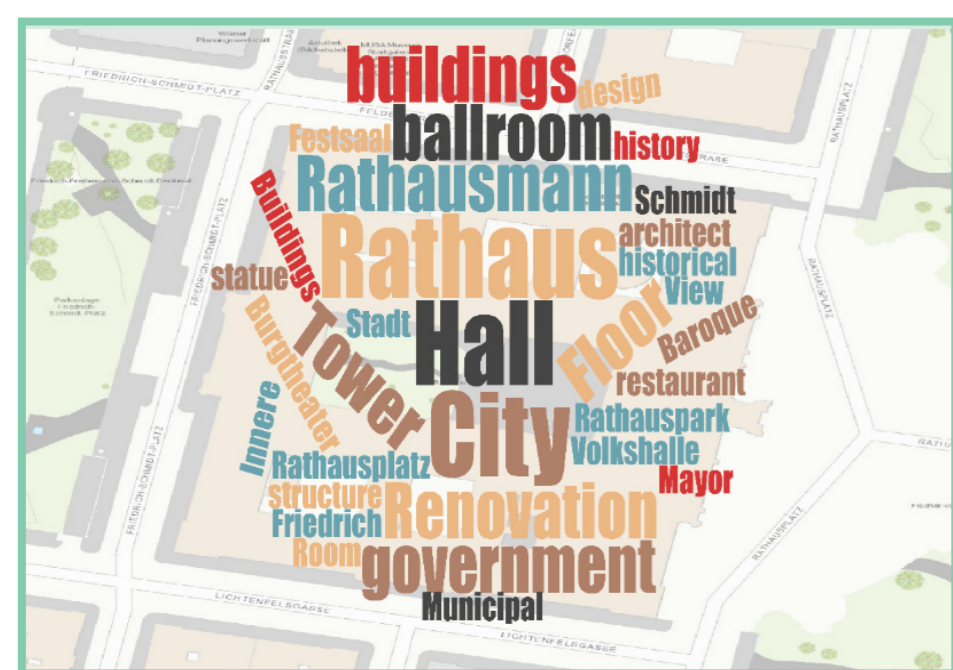


Fig. 1 Wordcloud of POI Vienna City Hall/Rathaus

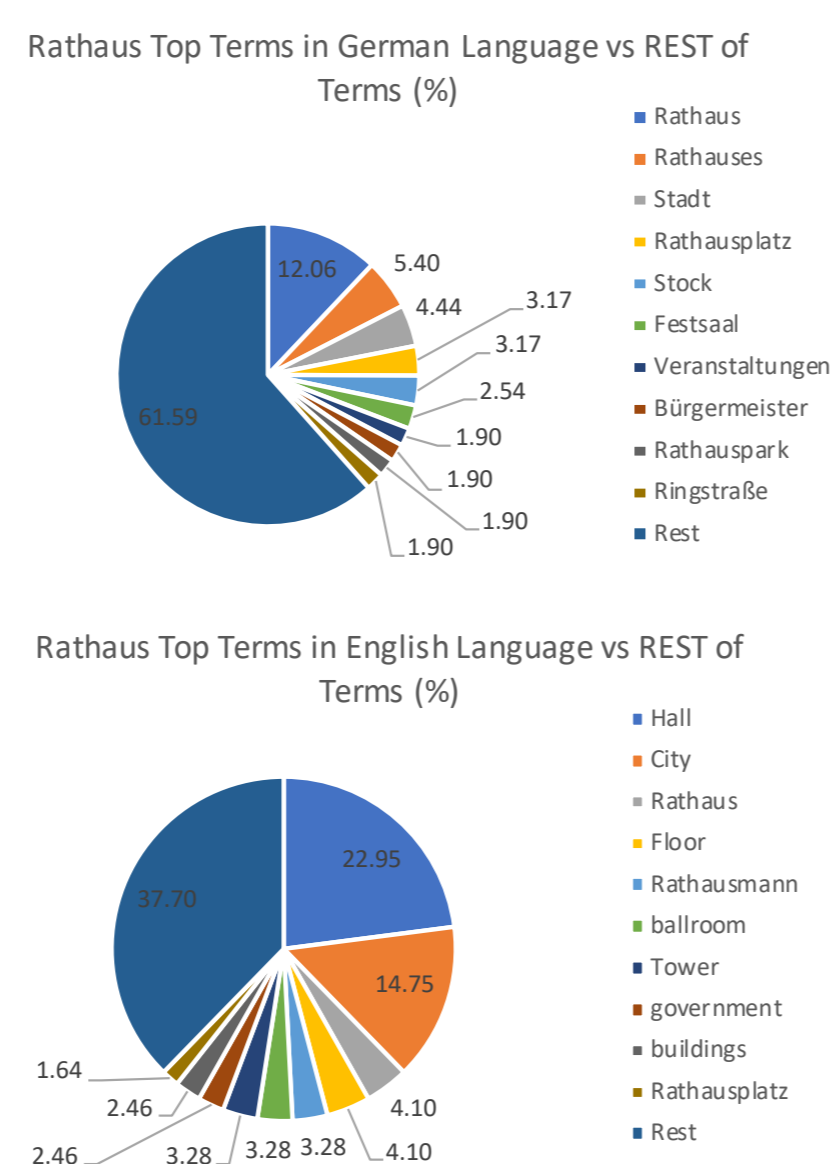


Fig. 2 Top Ten Terms in German and English Language for Vienna City Hall/Rathaus derived from Wikipedia Articles

such as Vienna or Austria. The next step includes the creation of the Wordclouds and lists of Top 10 Terms.

TWITTER DATA

HDBSCAN algorithm is used in order to divide twitter data into clusters. Then we compare POI and clusters by mentioning how many clusters are related to POI and with which POI. Then, it is possible to aggregate tweets within each "related cluster". The final step of the analysis is to obtain the Top 10 Terms for each "related cluster" in English and German version.

FLICKR DATA

HDBSCAN algorithm is used for clustering the flickr data. Similarly to twitter data, we compare Flickr clusters and POI in terms of quantity and name. Aggregate the posts within "related clusters" and obtain the Top 10 Terms.

In order to handle the huge amount of data PgAdmin is used.

RESULTS

Flickr: 15 from 35 clusters are related to POI. Creation of 15 Top 10 Terms lists.

Twitter: 8 from 23 clusters are related to POI. Creation of 16 Top 10 Terms lists (8 in English and 8 in German).

Comparisons between:

- Wikipedia and Twitter: 8 Relations
- Flickr and Twitter: 8 Relations

CONCLUSION

Percentages of WA are higher vs Twitter and Flickr. The use of English words in German version is dominant. Characteristic examples are the terms: "garden", "ticket", "heritage". German words such as toponyms are used in English version such as the term "Schloß". Twitter vs Wikipedia: 8 relations. There are more common terms in English than in German version.

Twitter vs Flickr: 8 relations. Half of the Top 10 Terms are common and the percentages are of similar range. We have many terms with the same meaning. Typical examples are: "Riesenrad", "wheel", "Rode" and "Schloss", "Schloß", "Palace". There are also terms related to activities such as: "shopping", "café", "cinema".

The presence of POI in Flickr: 42.85%. The presence of POI in Twitter is 35%.

Top 10 Terms (EN)	%	Sum(%)	Top 10 Terms (DE)	fr(0%)	Sum(%)
Palace	2.22	6.32	Schönbrunn	8.06	19.00
Schönbrunn	1.12		Schloß	4.10	
Schoenbrunn	0.93		Palace	2.46	
concert	0.37		Schloss	1.91	
rose	0.37		Gardens	0.96	
Schloss	0.37		Tiergarten	0.96	
garden	0.37		chilling	0.14	
environment	0.19		Schönbrunnpalace	0.14	
Sissi	0.19		schlosspark	0.14	
wienerphilharmoniker	0.19		habsburg	0.14	
Rest	99.68	Rest	81.00		

Fig. 3 Top 10 Terms in English and German Language derived from Twitter (Schönbrunn Palace)

Top 10 Terms	Frequency(1%)	Sum(%)
Schönbrunn	3.61	10.72
zoo	1.27	
palace	1.04	
tiergarten	0.85	
schonbrunn	0.79	
schloss	0.69	
schönbrunnpalace	0.69	
park	0.67	
gloriette	0.61	
garden	0.51	
Rest	89.28	

Fig. 4 List of Top 10 Terms derived from Flickr (Schönbrunn Palace)

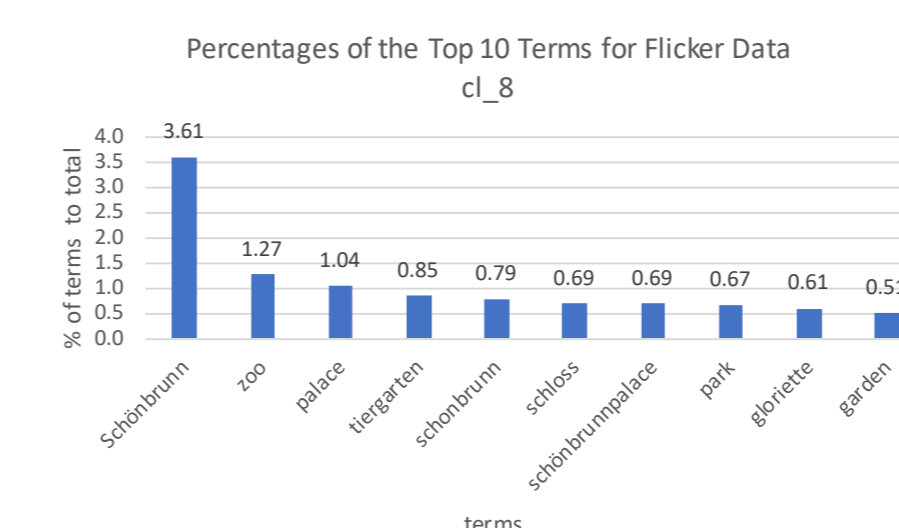


Fig. 5 Top 10 Terms derived from Flickr (Schönbrunn Palace)

THESIS CONDUCTED AT

Research Division Cartography
Department of Geodesy and
Geoinformation
Technische Universität Wien



SUPERVISORS

Univ.Prof. Mag.rer.nat. Dr.rer.nat.
Georg Gartner, TU Wien

MSc. Francisco Porrás-Bernardez, TU
Wien

YEAR

2019

KEYWORDS

Geosocial Media, Twitter, Flickr,
Wikipedia Articles, Points of Interest,
Topics of Interest, Terms, Semantics,
Wordclouds

REFERENCES

- [1] Huang, H., Gartner, G., & Turdean, T. (2013). Social media data as a source for studying people's perception and knowledge of environments. *Mitteilungen Der Österreichischen Geographischen Gesellschaft*, 155, 291–302. <https://doi.org/10.1553/moegg155s291>