MASTERARBEIT

Comparing Social Media Topics of Interest associated

with places according to user's origin

Ausgeführt am Department für
Geodäsie und Geoinformation
der Technischen Universität Wien

unter der Anleitung von
Univ.Prof. Mag.rer.nat. Dr.rer.nat. Georg Gartner, TU Wien

und

MSc.Francisco Porras-Bernardez,TU Wien

durch
Maria Athanasiou
Kaisermühlenstrasse 14, Wien

10.09.2019 Unterschrift (Student)

MASTER'S THESIS

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

Conducted at the Department of

Geodesy and Geoinformation

Technical University Vienna

Under the supervision of
Univ.Prof. Mag.rer.nat. Dr.rer.nat. Georg Gartner, TU Vienna
and
MSc.Francisco Porras-Bernardez,TU Wien

by
Maria Athanasiou
Kaisermühlenstrasse 14, Vienna

10.09.2019 Signature
(Student)

Statement of Authorship

Herewith I declare that I am the sole author of the submitted Master’s thesis entitled: **Comparing Social Media Topics of Interest associated with places according to user’s origin**

I have fully referenced the ideas and work of others, whether published or unpublished. Literal or analogous citations are clearly marked as such.

Vienna, 10.09.2019 Maria Athanasiou

Acknowledgements

The completion of this master thesis could not have been possible without the support and assistance of several people. I would like to express my sincere gratitude to Professor Georg Gartner for his advice, ideas and support during this master thesis. His trust in humanity and his boundless creativity in the domain of cartography inspired me throughout my “cartographic life” in Vienna.

Special thanks to Francisco Porras for his help, ideas, technical support and suggestions during the last months. I really appreciate the way he is working in order to overcome problems. His recommendations, encouragement and his flexibility were really important in order to bring my thesis to fruition.

Many thanks to the members of Research Division Cartography of Vienna University of Technology for providing assistance and sharing their knowledge with students. I thank them all for their hard work which often exceeded by far their everyday duties.

I would like to thank Juliane Cron not only for valuable advice during my studies, but also for her dedication to the Cartography Master Program and for her talent to fix any problem that occurred with patience and pleasure.

I would like to thank Corné van Elzakker for organizing the common schedule of four universities during the last semester and for his recommendations during the mid-term presentation.

I am deeply grateful to all people of the 7th intake for their company, their support and their willingness to share precious moments. The multicultural character of our group makes me once more in my life to think that as human beings we have more in common with each other than things that divide us.

Last but not least, I would like to express my gratitude to my family and in particular to my parents for their unceasing affection and support. A lifetime is insufficient to thank them for all they have given me. Also many thanks to my sister, Ioanna for her kindness to share a big amount of her precious time with me.

Abstract

The main theme of the current master thesis is to compare Social Media Topics of Interest 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 deals with the determination of the similarity between Flickr, Twitter and Wikipedia Articles.

The analysis of Wikipedia articles involves the selection of prominent Point of Interests of the city of Vienna and the collection of Wikipedia Articles in German and English language for them. The extraction of the Topics of Interest based on an online tool of creating wordclouds. The next step is to obtain the top ten Topics of Interest or Terms per Point of Interest in both versions (English and German). The visualization of the Topics of Interest or Terms takes place by adding the georeferenced wordclouds to the map. Tables, bar charts, and pie charts depict the most frequent Topics of Interest or Terms.

The analysis of the Geosocial Media data divided into two groups. The first analysis refers to the data derived from Twitter and the second refers to the data derived from Flickr. The analysis of Twitter data contains the clustering of the data by using HDBSCAN algorithm, the extraction of the number of clusters that are related to the Points of Interest and the aggregation of tweets within "related" clusters. The final step of the procedure is to obtain the top ten Topics of Interest or Terms for the related to clusters in English and German language. The most frequent terms depicted with the assistance of tables and bar charts. The analysis of the Flickr data contains the clustering of the data by using HDBSCAN algorithm, the extraction of the number of clusters that are related to the Points of Interest and the aggregation of posts within the "related" clusters. The last step of the process is to obtain the top ten Topics of Interest or Terms for the related clusters. Tables and bar charts are the tools that help us to depict the most frequent terms.

The similarity is "measured" by calculating the frequency of appearance of the Topics of Interest or Terms. The comparison of the percentages of the top ten Terms or Topics of Interest between Twitter, Wikipedia and Flicker is the key to define the similarity. In the current thesis the comparison based into two groups. The first group is Twitter-Wikipedia and the second group is Flicker-Twitter.

Table of Contents

List of Images

List of charts

List of Tables

Abbreviations

GSM: Geosocial Media

POI: Point of Interest

AOI: Area of Interest

HDBSCAN: High Density Based Spatial Clustering of Application with Noise

API/APIs: Application Programming Interfaces

NLP: Natural Language Processing

VGI: Volunteered Geographic Information

DBMS: Data Base Management System

ESRI: Environmental Systems Research Institute

UTF-8: Unicode Transformation Format

CSV: Character Separated Values

SHP: Shapefile (vector data format for GIS software)

CRS: Coordinate Reference System

SRS: Spatial Reference System

SVG: Scalar Vector Graphics

PNG: Portable Network Graphics

GIS: Geographic Information System

GUI: Graphical User Interface

GCP: Ground control Point

1. Introduction

1.1 Motivation and Problem Statement

More and more people use online social network and media-sharing services. So, nowadays scientists who are dealing with the collection and the analysis of data are interested in analyzing user generated data. As a result, these new type of data creates a new and a challenging source of information. In the domain of geosciences user generated data could contain information about user's location. This is really promising because geotagged social media data give us the opportunity to study the interaction between users and places. Tagging is a classical way for users to annotate User Generated Content (UGC) and derives from their need to organize their content and their desire to share with others and be part of the social contribution (Hollenstein & Purves, 2010). It is worthy to mention that social media data are huge data in comparison with data collected by questionnaires or other conventional methods. Social media data could be used in many fields, such as geography, environmental psychology, sociology, computer graphics and decision making. Recently research findings indicate that social media data are useful for studying the way people perceive and have knowledge of the environments (Huang, Gartner, & Turdean, 2013). Social Media Analysis Systems (GIS) offer the possibility to explore spatial, temporal and even affective aspects of users behavior, even for public places (Kovacs-Györi et al., 2018). For example, a recent study of Hollenstein & Purves shows that we can model the city center of a big city according to user perception by using the Flickr tags (Hollenstein & Purves, 2010). A similar research study collects georeferenced postings from Twitter and Instagram in order to define cognitive regions (Gao et al., 2017). Dunkel claims that is important to know how the public perceives the environment, particularly in landscape planning, management, and design (Dunkel, 2015). During his project a visualization of people's perceptual responses to land-scape is demonstrated with crowdsourced photo geodata from Flickr. Geosocial Media data combine spatial and textual information. The existing techniques in order to visualize and synthesize location-specific qualitative data are narrow (Dunkel, 2015). And it is necessary for environment and planning community to investigate new techniques in order to improve the way georeferenced qualitative data are depicted (Li & Zhou, 2017). To conclude, social media data add new information or patterns and constitute a challenging and promising mission in Geosciences. So, the relevance of this master thesis to the current scientific and research work is obvious.

1.2 Research Identification

The central theme of this master thesis is to extract and visualize Topics of Interest associated with Points of Interest based on social media sources. Checking how these Topics of Interest related to user's place of origin is an additional interest of my research. In order to achieve this aim, I will mention the issues that need to be researched by presenting the research question and the objectives.

The research question of the current master thesis is the following:

- How similar are the terms extracted from Geosocial Media (Twitter & Flickr) and Wikipedia articles when associated to the same Points of Interest?

Here, it is worthy to mention that terms or topics of interest referred to semantics, words, tweets, tags, texts, or hashtags. Users describe the same things with different vocabulary depending on many reasons such as their feelings, experience, education, gender, age, place of origin and so on. Generally speaking, semantics refer to the meanings of a word or to the study of a linguistic meaning. According to Oxford dictionary the use of the term semantics refers to the branch of linguistics and logic concerned with meaning. The two main areas are logical semantics, concerned with matters such as sense and reference and presupposition and implication, and lexical semantics, concerned with the analysis of word meanings and relations between them. In the current master thesis "terms of interest" or "topics of interest" are used in order to mention words, texts or any other expression that users post in Twitter and Flickr. It is also essential to point out that in this thesis the term "similarity" is used in order to detect the connections or relations between topics of interest and points of interest. Similarity could be measured by frequency of appearance of the terms. Obtaining the most frequent terms is a significant step in analysis of the semantics. Then the frequency could be visualized by using percentages. The examined place is the city of Vienna. In general, Points of Interests (POI) play an important role in geography in terms of defining important places for people. To define a Point of Interest (POI) is a difficult procedure based on the complication of the environment and on different perceptions of people. In urban places, we can define as Point of Interest (POI) a building due to its specific importance for the city in terms of architectural or historical value. Points of Interest (POI) are basic elements of tour and city guides and in collaboration with social media data tags could be more précised and better defined. Many studies that dealing with the identification of boundaries of vague places and regions use Flickr photos to show a natural link between textual tags and locations (Hu & Janowicz, 2018). Also, in urban places there are Areas of Interest (AOI) that are regions relevant to people due to different reasons. These areas attract the attention of the citizens due to presence of specific infrastructures, services or landmarks in general that determine the functionality of the space. In the city of Vienna, Karlsplazt could be considered as an Area of Interest (AOI). A museum located in the area of Karlsplazt could be considered as a point of interest (POI). In this research the selection of these prominent Points of Interest (POI) based on authoritative and commercial data sources. The list of the fore-mentioned Points of Interest (POI) based on a list provided by Vienna Tourist Board [1].These lists rank places according to their historical, architectural, and cultural value. The amount of visitors play a significant role in the ranking list. That means, the higher the footfall (according to tickets sale statistics for example) for a place, the bigger the possibility to be ranked as a highly visited place.

Geosocial Media (GSM) data are location-based data derived from social media. These data are unstructured data, based on the fact that contain photos, tags, graphic images and so on. Nowadays, examples of Geosocial Media that allow users to interact relative to their locations are Twitter, Flickr and Instagram. Users post their location not only as latitude and longitude but also by tagging the name of the place or their location ("The Essential Guide to Geosocial Data", 2018). The content that user post in their Geosocial media accounts characterized by variety, as it can contain thoughts, feelings, interests and different activities("The Essential Guide to Geosocial Data", 2018). In the current thesis, Geosocial Media (GSM) that offer

[1] https://www.wien.info/en/locations/vienna-tourist-board

access to georeferenced posts through the use of application programming interfaces (APIs) are Twitter and Flickr.

The Research Sub-questions of the current master thesis are the following:

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

The latter sub-question refers to those Points of Interest (POI) that are relevant to a specific place or a specific neighborhood. So, checking the presence of the fore-mentioned points by being inside, outside or near a neighborhood could be a sign of proportional presence in Geosocial Media (GSM).

The research objectives are indicators of the study and give us a clearer view of the expected outcome of the thesis. The research objectives are the following:

- Extract terms from Wikipedia Articles and Geosocial Media (GSM) from Flickr and Twitter
- Determine the similarity between Geosocial Media (Flickr and Twitter) and Wikipedia Articles

1.3 Overview of Contents

This master thesis has 5 chapters.

- The first chapter is the Introduction and describes the motivation and the problem statement.
- The second chapter defines Geosocial Media and gives an overview of useful meanings related to the concept of the thesis. Also, this chapter presents other projects that carried out in the domain of Goesciences and related to our work.
- The third chapter presents the methodology.
- The forth chapter displays the results of the current research.
- The fifth chapter presents the discussion and the conclusion and proposes potential future work.

2. Literature Review

2.1 Geosocial Media useful meanings

Geosocial Media (GSM) data are location-based data derived from social media. These data are unstructured data, based on the fact that contain photos, tags, graphic images and so on. Nowadays, examples of Geosocial Media that allow users to interact relative to their locations are Twitter, Flickr and Instagram. Users post their location not only as latitude and longitude but also by tagging the name of the place or their location ("The Essential Guide to Geosocial Data", 2018). The content that user post in their Geosocial media accounts characterized by variety, as it can contain thoughts, feelings, interests and different activities ("The Essential Guide to Geosocial Data", 2018). Posts could have the form of tags. Tagging is a very common way to annotate User Generated Content (UGC) and established from user's need to have their content in order and their willingness to share with other people and take place in the social contribution (Hollenstein & Purves, 2010). Also the same researchers mentioned that the words used in tagging are completely flat and has been stated that could show the structure of users in terms of conception, lingual ability and their diversity based on geographical and cultural background (Hollenstein & Purves, 2010). To assign a name to a place is an indication of the importance of the moment of space and people's integrated experience and that improves the social construction of a place (Hu & Janowicz, 2018).

Flickr is an image and video hosting service and was created in 2004. The main targets of Flickr is to help people sharing their photos to people who are interested in them and to enable new ways of organizing photos and videos (https://www.Flickr.com/). According to information derived from the official website of Flickr[2], Flickr is one of the best online photo management and sharing application in the world. Flickr photos are geotagged photos and reveal information not only about the user and his comments in form of tags but also about the location. A title of a Flickr photo or other tags that describe the content of the photo are typical examples of tags in this photo sharing service. The presence of a natural link between textual tags and locations in the Flickr photos has been studied many times, particularly in identifying the vagueness of places and regions (Grothe & Schaab, 2009). The allowance of Flickr's users to annotate multiple tags, not only in their own photos but also in the photos uploaded by other uses. These tags facilitate the way users organize, retrieve and explore the data and seems that they appreciate the opportunity to use and create those type of tags. These powerful tools of tagging and annotation are powerful as they give users unparalleled power of mold and influence the way of encountering with the interacted information (Winget, 2006). Current research presents that the majority of Flickr photos were taken in urban environment (Hollenstein & Purves, 2010). That means that the higher the appearance of people, the bigger the amount of photos by giving scientists an extra hint to analyze the relations between human beings and their environment. Twitter is a news and social networking service and was created in 2006 (https://Twitter.com/). Twitter as a social networking site relies on micro-blogging for communication and people are able to share text up to a limit of 140 characters[3].

[2] https://www.Flickr.com/

[3] https://en.wikipedia.org/wiki/Twitter

Wikipedia Articles are derived from Wikipedia[4], the online multilingual encyclopedia that uses a wiki-based editing system and editing is free for everyone. According to Marshall, Wikipedia is the largest and the most popular reference on the World Wide Web (Marshall, 2006).

A Wordcloud is a new way of visual representation of text data. Wordclouds have several names such as tag clouds or weighted list clouds. A wordcloud depicts words with different fonts according to the criteria of frequency[5] ("Create Live World Cloud",n.d.).Words that appear more frequently into a given text are more highlighted in comparison with others that appear in the same text. Wordcloud based on the weighted list of words that appear into the text. The use of wordclouds that have a spatial context could be a helpful tool in order to visualize on an organized way the textual context.

2.2 Related Work

Many studies and researches focus on Geosocial Media data and their impact in people's everyday life. These type of data could be used in many fields, such as geography, environmental psychology, sociology, urban planning, computer graphics and decision making. Recently research findings indicate that social media data are useful for studying the way people perceive and have knowledge of the environments (Huang et al., 2013). This section presents related work in the domain of analysis of data derived from Geosocial Media.

Results of Huang et al. (2013) study reveals the utility of social media data in order to study people's perception and knowledge of the environment. The researches come to this conclusion through three case studies. The aim of the first one was to model the perceived boundaries of Vienna's city center, whereas the aim of the second was to model responses of affection related to environments. The target of the last case study was to identify popular landmarks in the city.

Many attempts have been made in order to investigate affective and sentiments derives from social media data according to specific places. Kovacs- Györi et al.(2018) study focused on findying a new methodology in order to detect spatial, temporal and affective patterns of parks based on their visitors posted tweets. The outcome of the study was to define profile of parks and their visitors of London, and help urban planners and decision makers. Hollestein Purves (2010), used Flickr tags to describe the city cores. Their target was to explore methods and ways of using Flickr tags as authoritative empirical data in order to see how people name places. One of their results was that data derived from Flickr tags could provide information about vague places and place names.

Vague cognitive places were also studied by Gao et al. (2017). The researchers propose an automatable framework that can synthesize datasets that characterized by their variability and heterogeneity. The importance of this study based on the use of different data sources. Examples of them are Flickr, Twitter, Instagram, travel blogs and Wikipedia pages.

Another study that has been performed on the domain of analyzing social media data is the study conducted by Hu & Janowicz (2018). In their analysis they propose an empirical study on place names and how these are changing with the impact of geographic distance.

[4] https://en.wikipedia.org/wiki/Main_Page

[5] https://www.mentimeter.com/features/word-cloud

A way of visualizing the semantics that extracted from Geosocial Media data is the wordcloud or tag cloud. Georeferenced word-cloud shows clearly people’s narrative feelings as they contribute in variable outdoor activities from one side to the other side of city (Li & Zhou, 2017). The study of Li & Zhou (2017) presents the utility of the wordclouds in order to present semantic data. Researchers use GPS location of the users in order to combine location and narrative data. The outcome of their study was that georeferenced word clouds could provide information about narrative data and allow the production of their patterns.

Tag clouds visualization technique is used by Hahmann & Burghardt (2011) study. They present a method of depicting the most significant semantic information on a map by the use of a wordcloud. The creation of their wordcloud based on technique that visualizes feature frequencies on the map.

Tag maps are also used by Dunkel (2015) in order to depict his results after the analysis of Flickr crowdsourced data. The interpretation of geo-crowdsourcing data is a big issue in the domain of geosciences. Dunkel’s study focused on providing information to urban planners and landscape planners in order to be able for exploration, visualization and interpretation of crowdsourcing data.

Many attempts have been made with the purpose of understanding the environment with the assistance of geotagged photos, in order to help urban planners and other scientists in their mission. Prasad et al. (2015) try not only to identify the Areas of Interest (AOI), but also to understand how they formed over time with the assistance of geotagged Flickr photos. The use of the method of clustering (DBSCAN algorithm) was important as it helps to group and identify important places.

Another related project that help us to understand the humans and their interaction with their location is the Jiang (2016) project. The team of researchers study the spatial distribution of city tweets and their densities. They examine a variety of cities in different countries and use different scales in order to examine the presence of the tweets. The study showed social media data enable us to understand their significance and how they distributed in cities that are particularly the places where the amount of data is becoming bigger.

A framework was proposed by Gao, Janowicz, & Couclelis (2017) to study urban functional regions by using data derived from social media and different points of interest. Their technique for finding functional patterns of points of interest based on a model of analysing location based social media data.

3. Methodology

3.1 Workflow

This part presents the methodology of the current master thesis. We have a variety of sources in order to analyze the terms. The used Geosocial Media are Twitter and Flickr. The other source of data derives from Wikipedia Articles.

The following images show us the workflow of the thesis. The first part of the project is to select the most prominent Points of Interest according to a ranking list. The final list consists of 23 Points of Interest (POI) and more details could be find in subchapter 3.3.The next step is the collection of the Wikipedia Articles for Points of Interest (POI) in German and English language. So the total amount of Wikipedia Articles is 46 (each POI corresponds to two Articles). The next processing part is to extract terms from the Wikipedia Articles. An online tool for creation of the wordclouds is useful in order to create a suitable wordlist by removing no relevant terms from the list.

Terms that appear only once and toponyms (name of the city of Vienna or the Austria) removed from the wordlist. After that, it is possible to create a wordcloud by selecting the parameters of our interest. Adding the wordcloud as a map element is possible through the georeference procedure of the wordcloud image. More information about the fore mentioned procedures could be find in the 4th sub chapter of the current chapter. It is possible to obtain the top terms per Point of Interest by analysing the data of the wordlist (see chapter 4.3). A table, a pie chart and a bar chart are used in order to visualize the final results.

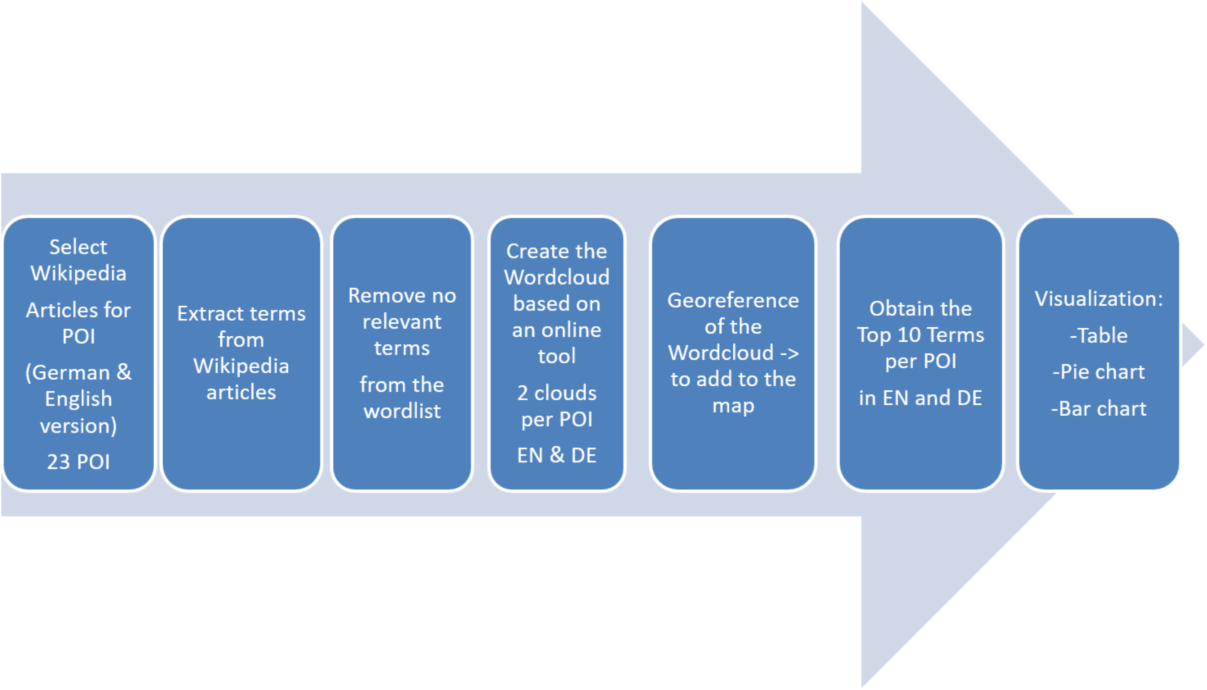

Select Wikipedia Articles for POI
(German & English version)
23 POI
Extract terms from Wikipedia articles
Remove no relevant terms
from the wordlist
Create the Wordcloud based on an online tool
2 clouds per POI
EN & DE
Georeference of the Wordcloud -> to add to the map
Obtain the Top 10 Terms per POI
in EN and DE
Visualization:
-Table
-Pie chart
-Bar chart

Figure 1Workflow for Wikipedia Articles

The next image depicts the workflow for analysing the Twitter data set. The amount of the Twitter data is huge and we use the method of clustering in order to handle these type of data. HDBSCAN algorithm is used in order to divide our data into clusters. The outcome of the HDBSCAN is the creation of different clusters. The next step is to compare the Points of Interest (POI) and the 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 of the Twitter data is to obtain the top 10 terms for each “related” cluster in English and German version. The outcome could be visualized by a table and a bar chart.

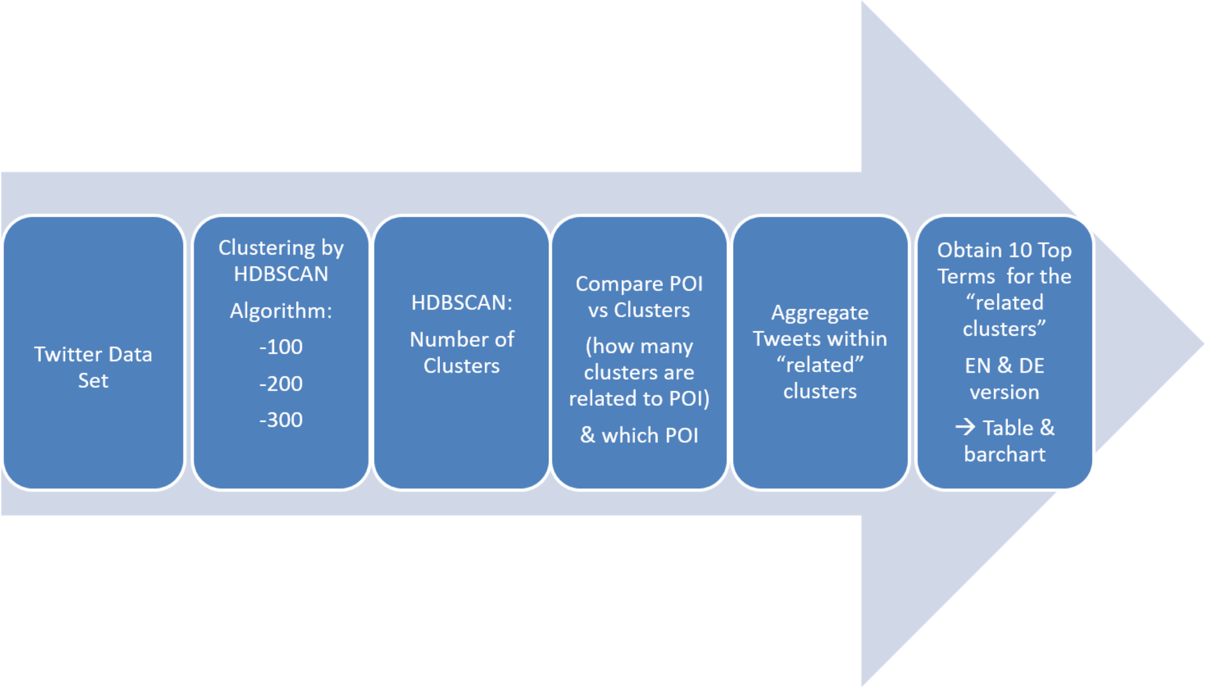

Twitter Data Set
Clustering by HDBSCAN Algorithm: -100 -200 -300
HDBSCAN: Number of Clusters
Compare POI vs Clusters (how many clusters are related to POI) & which POI
Aggregate Tweets within "related" clusters
Obtain 10 Top Terms for the "related clusters" EN & DE version → Table & barchart

Figure 2 Workflow for Twitter Data

The following image presents the workflow for analysing the Flickr data set. The amount of the Flickr data is huge and we use the method of clustering in order to handle these type of data. The same algorithm is used in order to divide our data into clusters. The outcome of the HDBSCAN algorithm is a specific number of clusters. The next step is to compare the Points of Interest (POI) and the created clusters. So, we can mention the number of clusters that are related to Points of Interest (POI). Also, it is possible to define the relationship between a specific cluster and a Point of Interest (POI). The final step is to obtain the ten top terms per related cluster and visualize the results with the assistance of a table and a bar chart.

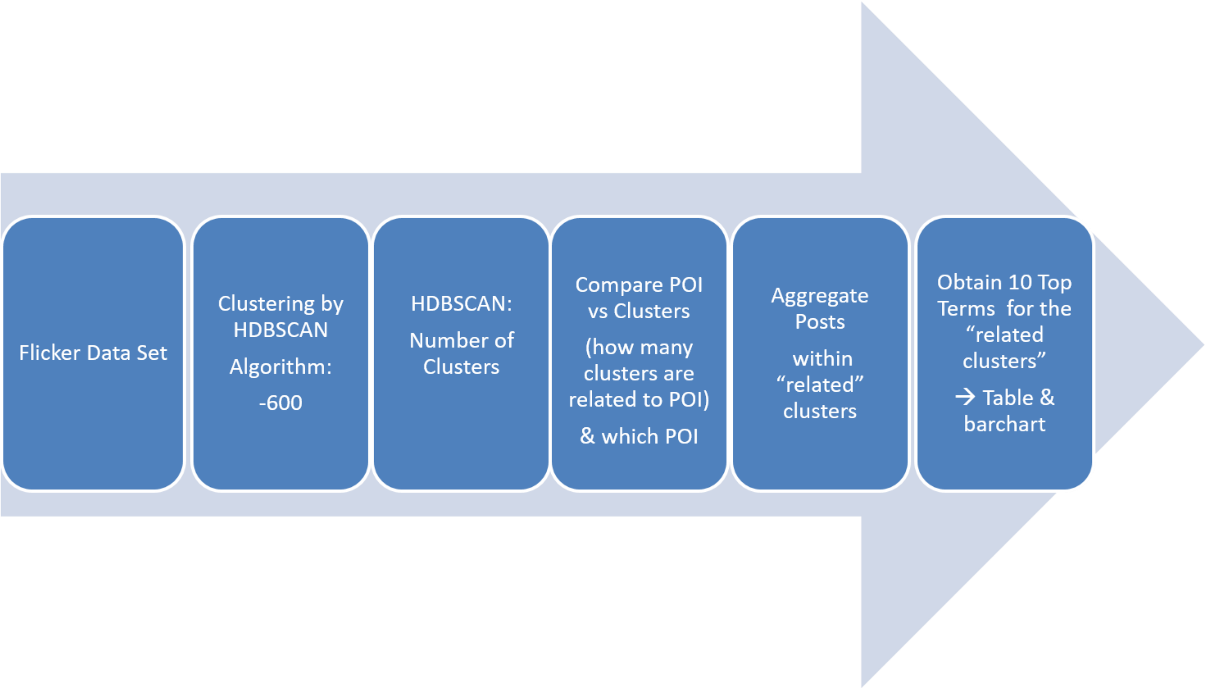


Figure 3 Workflow for Flickr Data

3.2 Data Sets from GSM and Clustering

This subchapter presents some information about the point data sets of Flickr and Twitter. It also describes methods of clustering the point data sets by giving emphasis to the selected method of the HDBSCAN algorithm. In the end gives an overview of the number of different clusters that have been created during the implementation of the HDBSCAN method.

The point data sets derived from Flickr and Twitter APIs. The Flickr data set contains information about the owner of the photo, the tags and their location. The location could be described with the term geolocation due to the fact that the data set consists of the uploaded geotagged photos. The tags referred to texts or words. A typical example of a tag is the title of an uploaded photo or other textual tags that describe the uploaded photo. The Twitter data set contains information about the user, the date of the created tweets, the texts (tweets) and the location. The Twitter data set contains information about the place of origin of the user. The place of origin or the region of origin based on the method developed by EURECA project. In general, the researches based on information that users publish in their own profiles and on information derived after the implementation of a test in order to define the region of origin. This test based on a temporal threshold of six months. The country that the user uploaded the highest number of the total amount of his photographs over a period of six months is selected as the place of origin of the user (Verstockt et al., 2009). In the current master thesis the use of the word "term" or "topics of interest" refers to any kind of textual source. So texts, hashtags, words, tags (Flickr) or tweets (Twitter) are considered as terms. The extraction of the terms from Wikipedia Articles, Twitter and Flickr could be implemented during the analysis of the semantics. Also, it is worthy to point out that the use of the word "language" like German Language or English Language, or the use of the word "users" like German users or English users refers to the region of origin.

One fundamental step of the preprocessing part of the data is the clustering of the data sets. The amount Flickr and Twitter data is quite big and the division of the point data sets into groups is required in terms of identifying the most significant clusters. The software of Arc GIS Pro is used in order to cluster the point data sets of Twitter and Flickr. The software provides the following clustering methods:

- Defined distance (DBSCAN)
- Self-adjusting (HDBSCAN)
- Multi-scale (OPTICS)

DBSCAN: Density Based Spatial Clustering of Applications with Noise is a well-known data clustering algorithm that is commonly used in data mining and machine learning. In ArcGIS Pro we have the Density-based Clustering tool that can detect areas where points are concentrates and where they are separated by areas that are empty or sparse[6]. DBSCAN algorithm configured by two parameters Eps, as a searching radius and MinPts as a minimum number of points within the search radius. The Eps and the MinPts define a threshold of a minimum distance and then the clusters could be defined at the locations whose density of points is larger than the threshold (Hu et al., 2015). One advantage of this tool is that non-clustered points are labeled as noise. According to information provided by ESRI for ArcGIS Pro the fore mentioned tool uses unsupervised machine learning clustering algorithms which automatically detect patterns based purely on spatial location and the distance to a specified number of neighbors (“How Density-based Clustering works”,n.d.).It is worthy to point out that the term “unsupervised” derives from the fact that there is no need for training data before the execution of the process This gives an extra advantage of this algorithm in comparison with other clustering algorithms such as for example K-Means clustering method that requires a pre-determined number of clusters (Hu et al., 2015). For clustering the point data sets from Twitter and Flickr HDBSCAN algorithm is used. The Self-adjusting HDBSCAN tool is based on the DBSCAN algorithm and uses a range of distances to separate clusters of varying densities from sparser noise. Given that, the outcome of the method could give us clusters with different shapes and sizes. The HDBSCAN algorithm is a data-driven clustering method. HDBSCAN requires an input of the point feature data set, a path for the output features and a selection of a value that works as a threshold and represents the minimum number of features required to be considered as a cluster. Any cluster with less features than the threshold could be considered as noise. For Flickr data the name of the data set is fvienni and the name of the output data set is fvienni_hdbscan_600. The number 600 is an indicator for the user to understand easier that for this HDBSCAN processing the minimum number of features that considered as a cluster is 600. Any cluster with fewer features than 600 will be considered as noise. The unnecessary point features depict by a grey color and the rest of the points, the necessary point features depict in different colors. The clustered features have an indicator, whose name is color_id and ranges from 1 to 8. Non-clustered features considered as noise have the value -1 as color_id .Then, it is essential to remove the unnecessary point features, those that clustered as noise according to the threshold of the HDBSCAN tool. The next step is to improve the visualization of the clusters by changing the Symbology (Unique Values). The outcome of the process is the creation of 35 clusters.

[6] https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/how-density-based-clustering-works.htm

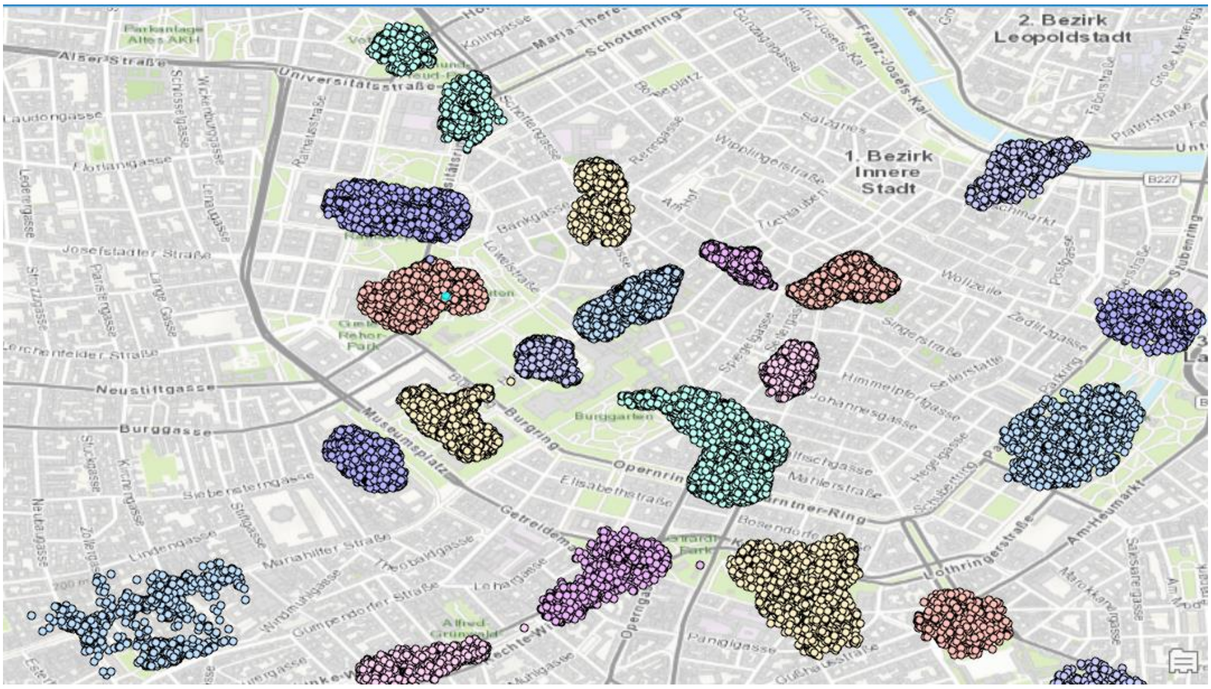


Figure 4 Clusters by HDBSCAN 600 for Flickr

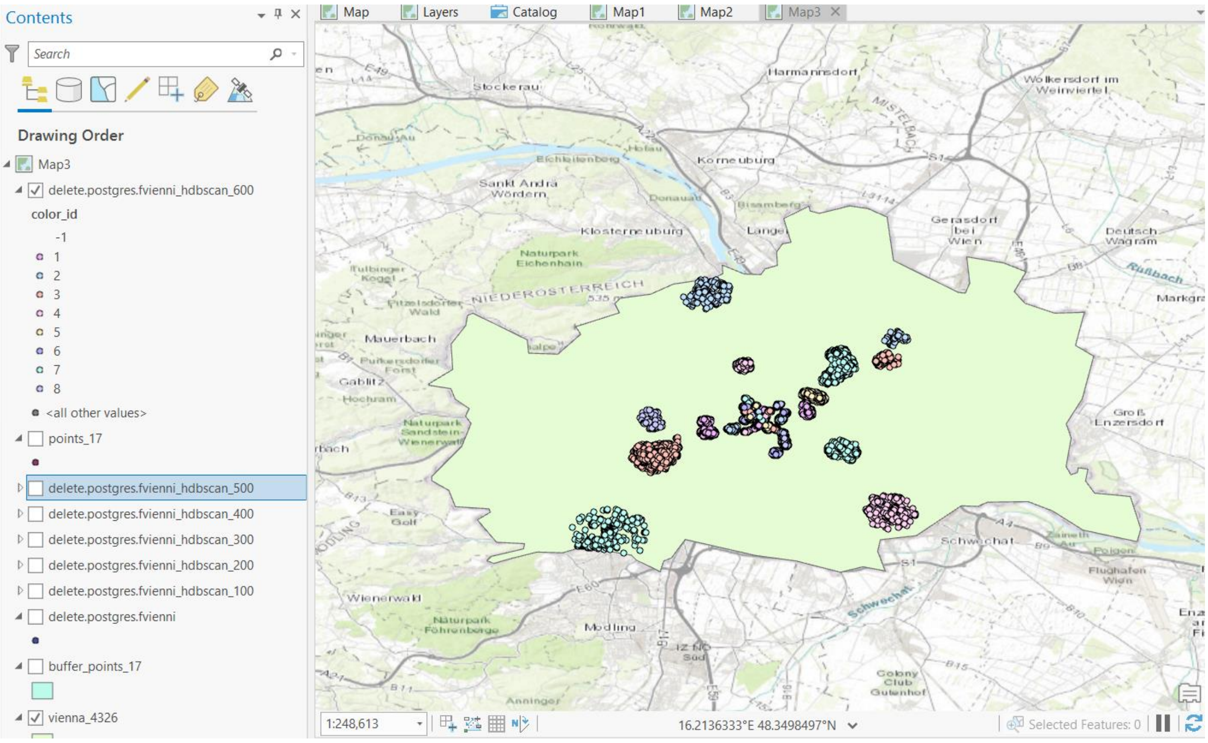
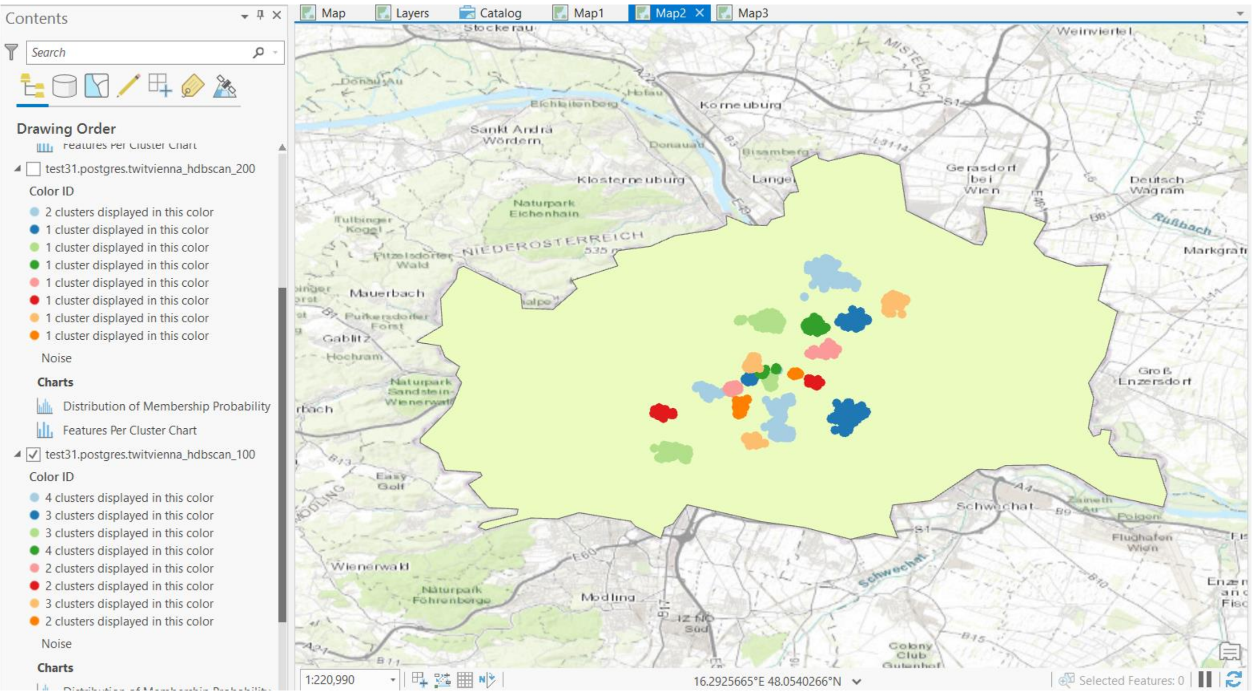


Figure 5 Outcome of HDBSCAN 600 for Flickr Data: 35 clusters

Twitvienna is the name of the Twitter data set and the name data set after the process of the HDBSCAN is twitvienna_hdbscan_100. The value 100 is an indicator for the user to understand that for this HDBSCAN process the minimum number of features that considered as a cluster is 100. Any cluster with fewer features than 100 will be considered as noise. So, twitvienna_hdbscan_200 means that the minimum number of features that considered as a cluster is 200 and so on. The pictures below present the results of the HDBSCAN algorithm for the Twitter data. As we can see, for the twitvienna_hdbscan_300, we have only 3 clusters. In this case each cluster depicted in a different color. The twitvienna_hdbscan_200 algorithm leads us to nine clusters and the twitvienna_hdbscan_100 provides us 23 clusters. For these HDBSCAN algorithms some clusters displayed in the same color. The reason for this is that the use of many different hues reduces the visibility of the user in order to distinguish the clusters. The software has a limit of 8 different colors.



Contents

Search

Drawing Order

Features Per Cluster Chart

◢ □ test31.postgres.twitvienna_hdbscan_200

Color ID

- ● 2 clusters displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color

Noise

Charts

Distribution of Membership Probability

Features Per Cluster Chart

◢ ☑ test31.postgres.twitvienna_hdbscan_100

Color ID

- ● 4 clusters displayed in this color
- ● 3 clusters displayed in this color
- ● 3 clusters displayed in this color
- ● 4 clusters displayed in this color
- ● 2 clusters displayed in this color
- ● 2 clusters displayed in this color
- ● 3 clusters displayed in this color
- ● 2 clusters displayed in this color

Noise

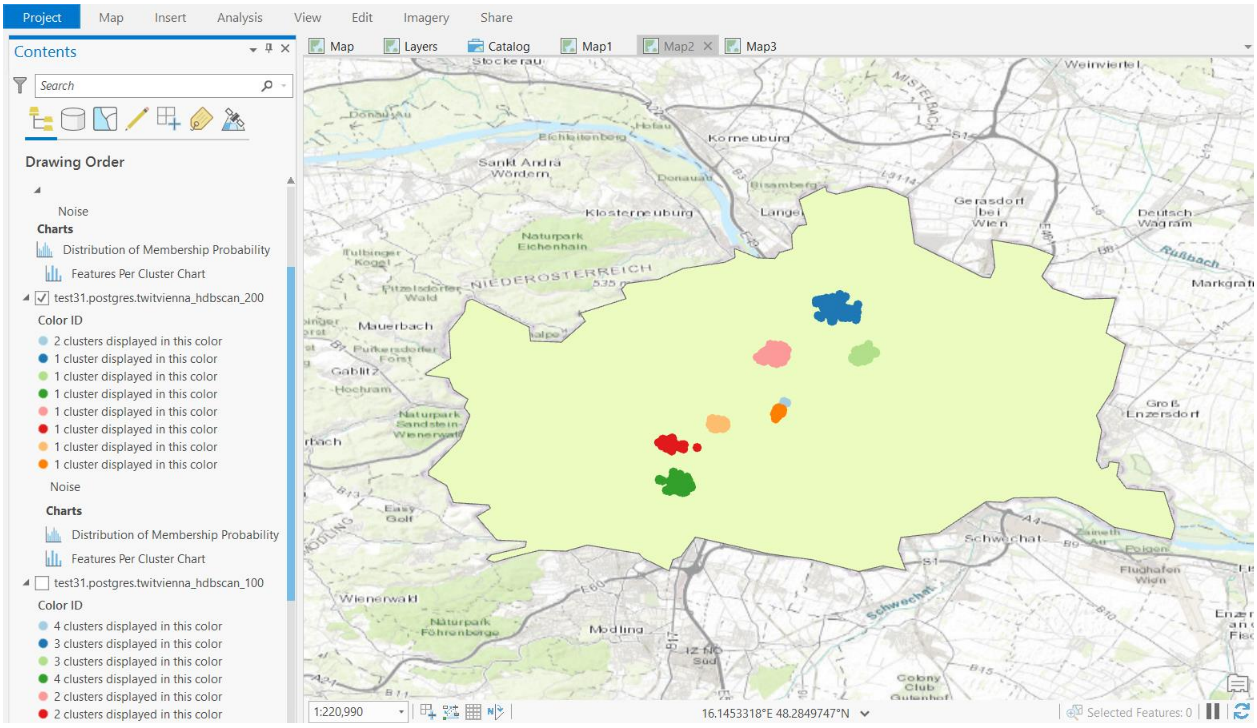
Charts

1:220,990

16.2925665°E 48.0540266°N

Selected Features: 0

Figure 6 Clusters by HDBSCAN 100 for Twitter



Contents

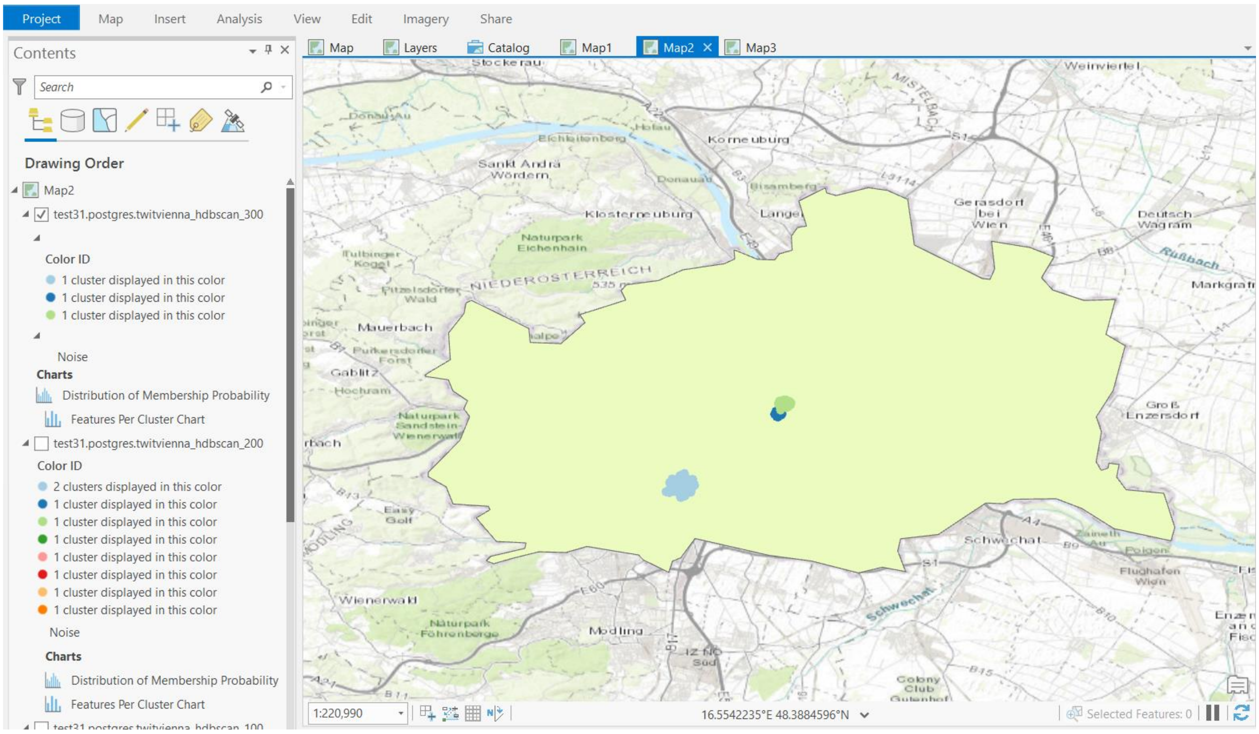
Drawing Order

Noise
Charts
Distribution of Membership Probability
Features Per Cluster Chart
☑ test31.postgres.twitvienna_hdbscan_200
Color ID
- ● 2 clusters displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color

Noise
Charts
Distribution of Membership Probability
Features Per Cluster Chart
☐ test31.postgres.twitvienna_hdbscan_100
Color ID
- ● 4 clusters displayed in this color
- ● 3 clusters displayed in this color
- ● 3 clusters displayed in this color
- ● 4 cluster displayed in this color
- ● 2 clusters displayed in this color
- ● 2 clusters displayed in this color

1:220,990
16.1453318°E 48.2849747°N
Selected Features: 0

Figure 7 Clusters by HDBSCAN 200 for Twitter



Contents

Drawing Order

Map2
☑ test31.postgres.twitvienna_hdbscan_300
Color ID
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color

Noise
Charts
Distribution of Membership Probability
Features Per Cluster Chart
☐ test31.postgres.twitvienna_hdbscan_200
Color ID
- ● 2 clusters displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color
- ● 1 cluster displayed in this color

Noise
Charts
Distribution of Membership Probability
Features Per Cluster Chart
☐ test31.postgres.twitvienna_hdbscan_100

1:220,990
16.5542235°E 48.3884596°N
Selected Features: 0

Figure 8 Clusters by HDBSCAN 300 for Twitter

3.3 Points of Interest

The selection of prominent Points of Interest (POI) is a difficult procedure. For this master thesis the selection of POI (Points of Interest) based on authoritative and commercial data sources, such as the list provided by Vienna Tourist Board [7].The following table presents the selected POI (Point of Interest) and consists of 5 columns providing some information about the fore-mentioned points. Columns present information about the name of the POI (Point of Interest), its location by latitude and longitude and the link for Wikipedia articles in German and English language. It is worthy to point out that Volkstheater/People's Garden, Wiener Prater/Prater Park and Karlsplatz/Charle's square are considered as Points of Interested because of their high importance according to the ranking lists.

Table 1 List of POI

Number	POI	Wikipedia Article German Version	Wikipedia Article English Version	Location (Latitude, Longitude)
1	Albertina Museum	https://de.wikipedia.org/wiki/Albertina_(Wien)	https://en.wikipedia.org/wiki/Albertina	48.204864, 16.368193
2	Peterskirche /St. Peter's Church	https://de.wikipedia.org/wiki/Peterskirche_(Wien)	https://en.wikipedia.org/wiki/Peterskirche,_Vienna	48.209590, 16.370243
3	Kunsthistorisches Museum/Museum of Art History	https://de.wikipedia.org/wiki/Kunsthistorisches_Museum	https://en.wikipedia.org/wiki/Kunsthistorisches_Museum	48.203977, 16.361766
4	Naturhistorisches Museum/Museum of Natural History	https://de.wikipedia.org/wiki/Naturhistorisches_Museum_Wien	https://en.wikipedia.org/wiki/Natural_History_Museum,_Vienna	48.205448, 16.359867
5	Catholic Church Maria am Gestade/St Mary's on the Bank	https://de.wikipedia.org/wiki/Maria_am_Gestade	https://en.wikipedia.org/wiki/Maria_am_Gestade	48.212954, 16.370286
6	Universität Wien/Main Building University of Vienna	https://de.wikipedia.org/wiki/Universit%C3%A4t_Wien	https://en.wikipedia.org/wiki/University_of_Vienna	48.213178, 16.360844
7	Kaisergruft/Imperial Crypt	https://de.wikipedia.org/wiki/Kaisergruft	https://en.wikipedia.org/wiki/Imperial_Crypt	48.205818, 16.370020
8	Parlament/Austrian Parliament Building	https://de.wikipedia.org/wiki/%C3%96sterreichisches_Parlament	https://en.wikipedia.org/wiki/Austrian_Parliament_Building	48.208289, 16.358608
9	Stephansdom"/St. Stephen's Cathedral	https://de.wikipedia.org/wiki/Stephansdom_(Wien)	https://en.wikipedia.org/wiki/St._Stephen%27s_Cathedral,_Vienna	48.208612, 16.373428

[7] https://www.wien.info/en/locations/vienna-tourist-board

10	Schönbrunn Schloss/ Schönbrunn Palace	https://de.wikipedia.org/wiki/Schloss_Sch%C3%B6nbrunn	https://en.wikipedia.org/wiki/Sch%C3%B6nbrunn_Palace	48.184864, 16.312241
11	Österreichische Nationalbibliothek/Austrian National Library	https://de.wikipedia.org/wiki/%C3%96sterreichische_Nationalbibliothek	https://en.wikipedia.org/wiki/Austrian_National_Library	48.206112, 16.366388
12	Hofburg Scloss/Hofburg Palace	https://de.wikipedia.org/wiki/Hofburg	https://en.wikipedia.org/wiki/Hofburg	48.205532, 16.364763
13	Belvedere Schloss/Belvedere Palace	https://de.wikipedia.org/wiki/Schloss_Belvedere	https://en.wikipedia.org/wiki/Belvedere,_Vienna	48.19135, 16.3809
14	Volksgarten /People's Garden	https://de.wikipedia.org/wiki/Volksgarten_(Wien)	https://en.wikipedia.org/wiki/Volksgarten,_Vienna	48.20804, 16.36174
15	Wiener Prater/Prater Park	https://de.wikipedia.org/wiki/Wiener_Prater	https://en.wikipedia.org/wiki/Prater	48.21645, 16.39783
16	Rathaus/Vienna City Hall	https://de.wikipedia.org/wiki/Wiener_Rathaus	https://en.wikipedia.org/wiki/Vienna_City_Hall	48.21094, 16.35734
17	Wiener Staatsoper/ Vienna State Opera	https://de.wikipedia.org/wiki/Wiener_Staatsoper	https://en.wikipedia.org/wiki/Vienna_State_Opera	48.20327, 16.36914
18	Karlsplatzt /Charle's square	https://de.wikipedia.org/wiki/Karlsplatz_(Wien)	https://en.wikipedia.org/wiki/Karlsplatz	48.19965, 16.37031
19	Karlskirche /St Charle's Church	https://de.wikipedia.org/wiki/Wiener_Karlskirche	https://en.wikipedia.org/wiki/Karlskirche	48.19834, 16.37187
20	Hundertwasserhaus/Hundertwasser House	https://de.wikipedia.org/wiki/Hundertwasserhaus_(Wien)	https://en.wikipedia.org/wiki/Hundertwasserhaus	48.20734, 16.39429
21	Musikverein/Viennese Music Association	https://de.wikipedia.org/wiki/Wiener_Musikverein	https://en.wikipedia.org/wiki/Musikverein	48.200382, 16.372789
22	MariaHilfe Kirche/ church of Mariahilf	https://de.wikipedia.org/wiki/Mariahilfer_Kirche	https://en.wikipedia.org/wiki/Church_of_Mariahilf	46.199028, 16.353065
23	Burgtheater/ Imperial CourtTheater	https://de.wikipedia.org/wiki/Burgtheater	https://en.wikipedia.org/wiki/Burgtheater	48.21036, 16.36154

The Points of Interest (POI) are converted to shapefile format in order to insert into GIS software and make modifications. Shapefile format is a simple format for storing information about the location and attributes of geographic features. It is obvious that the 23 Points of Interest can be represented by points in a shapefile format. QGIS is the main GIS software that used in order create the shapefile format of the Points of Interest (POI). The crucial part of this procedure is to select the suitable Reference System (CRS) in order to have point adaptable to the map. The selected CRS is EPSG: 4326 WGS 84.

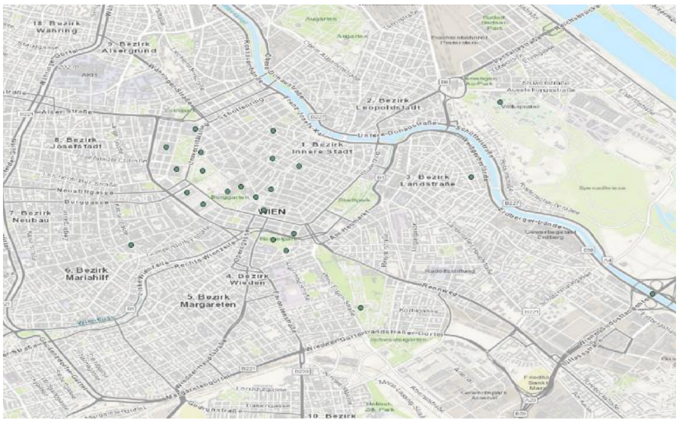


Figure 9 Points of Interest: image from QGIS

Digitizing some important buildings and parks (contained in the examined list of Point of Interest) serve as a helpful tool in order to make comparisons between clusters (created by Twitter and Flickr data sets) and the Points of Interest. In some cases, it is more reasonable to generalize the shape of a building in order to have clear idea about the covering area. For some buildings, that are closed one another it is difficult to refer with certainty their surrounded area. A typical example of this situation is the adjacency of Hofburg Palace and Neue Burg. The following pictures show us some examples of the procedure of digitizing. Areas of interest or polygons are useful in order to make comparisons between the clusters and important places.

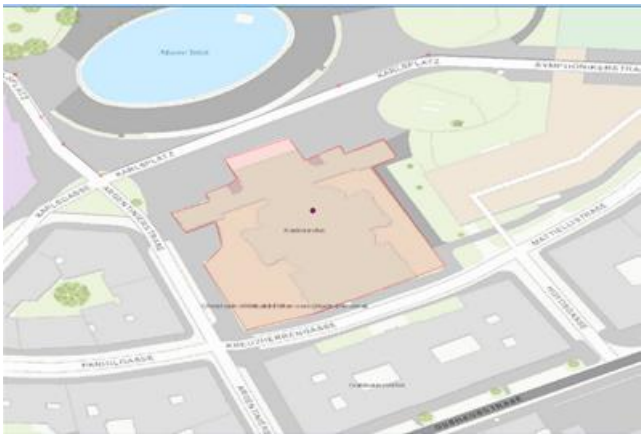


Figure 10 Digitizing of a building: generalized approach

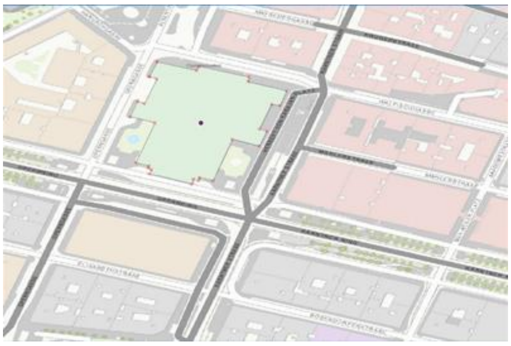


Figure 11 Digitizing a building

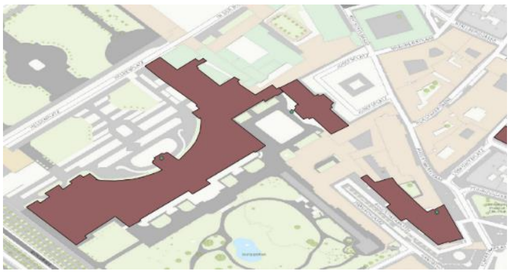


Figure 12 Hofburg Palace and Neue Burg

3.4 Wordclouds

A Wordcloud is a new way of visual representation of text data. Wordclouds have several names such as tag clouds or weighted list clouds. A wordcloud depicts words with different fonts according to the criteria of frequency[8] ("Create Live World Cloud",n.d.).Words that appear more frequently into a given text are more highlighted in comparison with others that appear in the same text. Wordcloud based on the weighted list of words that appear into the text. An online tool [9] is used in order to create the wordclouds of Wikipedia Articles. The process of the creation of the wordcloud of a Point of Interest includes the following steps:

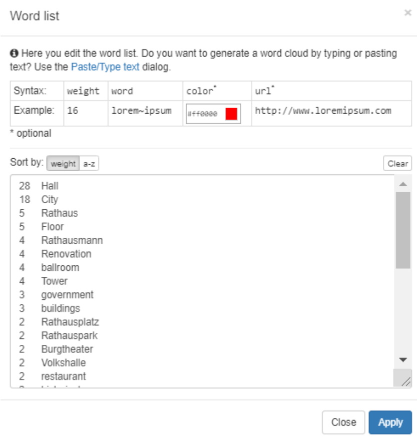
- Insert the text of the Wikipedia article of a specific Point of Interest
- Select the suitable parameters such as font, theme and shape (Font: Impact, Theme: Color band, Shape: Circle)
- Check the wordlist and remove words that are not important for the results. For example, words that appear only once or words that are really common for all Points of Interest such as toponyms or words that are not relevant with the target of the wordcloud
- Save the outcome in SVG format

The wordlist plays an important role for the creation of the Wordcloud. It is really essential to remove words that do not provide valuable information for the Point of Interest. Examples of unnecessary words are toponyms such as Austria, Vienna, Wien, Österreich, linking words or

[8] https://www.mentimeter.com/features/word-cloud

[9] https://www.wordclouds.com/

words like “edit” that appear in every Wikipedia Article. Some names need to be unified in order to make sense for the reader. A typical example is the name of Gustav Mahler.



Word list

Here you edit the word list. Do you want to generate a word cloud by typing or pasting text? Use the Paste/Type text dialog.

Syntax:	weight	word	color*	url*
Example:	16	lorem~ipsum	#ff0000 🟥	http://www.loremipsum.com

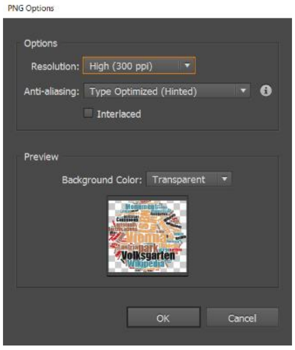
* optional

Sort by: weight a-z Clear

28	Hall
18	City
5	Rathaus
5	Floor
4	Rathausmann
4	Renovation
4	ballroom
4	Tower
3	government
3	buildings
2	Rathausplatz
2	Rathauspark
2	Burgtheater
2	Volkshalle
2	restaurant

Figure 13 Example of a wordlist: Wordcloud tool

The created wordclouds could be fitted into the geographic boundaries of the Points of Interest (POI). Wordclouds should be processed in order to have a suitable format for inserting in a GIS software. The use of Illustrator software transforms the SVG wordcloud in PNG format. It is worthy to mention that the background layer is deleted in order to emphasize the words of the cloud. The selection of parameters like the resolution (>300ppi) and the transparency of the background color could serve the visualization aspect of the Wordcloud.



PNG Options

Options

Resolution: High (300 ppi)

Anti-aliasing: Type Optimized (Hinted)

□ Interlaced

Preview

Background Color: Transparent

Figure 14 Wordcloud preprocess: Illustrator Software

The process of georeference of the wordclouds should take place in order to display them to the map. For this reason, it is important to create some buffers (prepared circles) around the Points of Interest (POI). By using the fixed distance buffer algorithm, it is easy to create some circles around the Points of Interest. The aforementioned algorithm computes a buffer area for all the point feature dataset using a fixed distance.

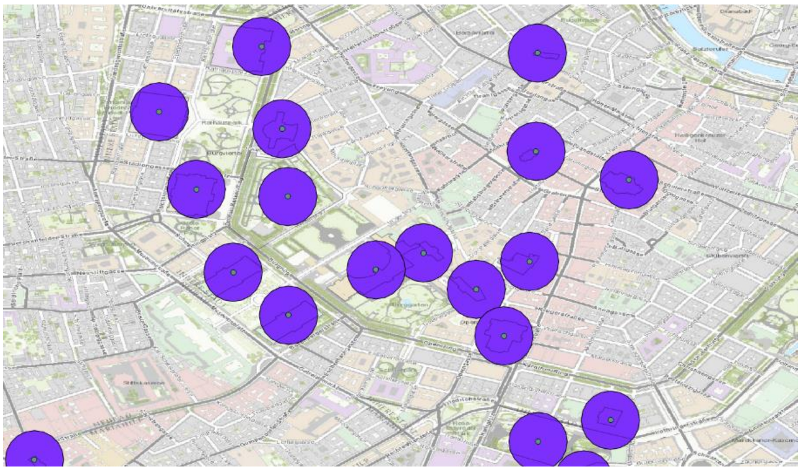
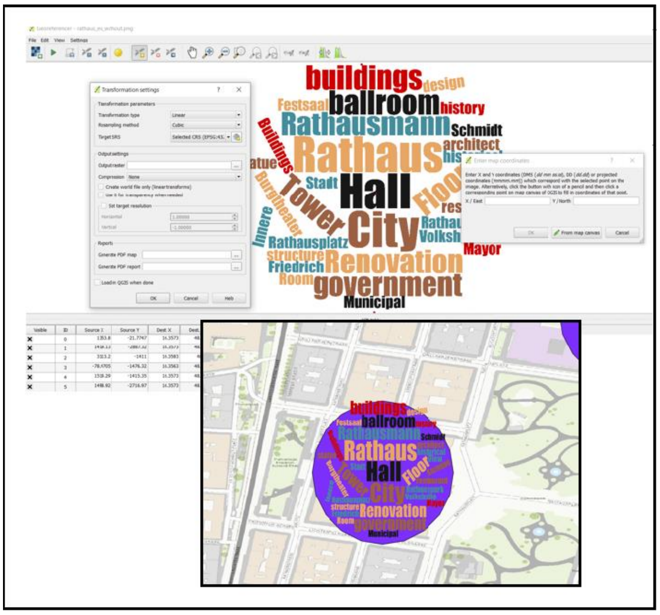


Figure 15 Prepared Circles by using the algorithm fixed distance buffer

The georeference procedure of the Wordclouds performed with the assistance of the fore mentioned prepared circles. Fitting the wordcloud into the prepared circles is possible after the selection of suitable transformation parameters. The selected reference system is CRS: WGS84 EPSG: 4326 and the transformation settings are: (transformation type: linear and resampling method: cubic). The next step is to choose the GCP points. One fundamental step of the procedure is to enter the X and Y coordinates or the projected coordinates which correspond with the selected point on the image. After working with the trial and error method, I concluded that the final number of points is five. That leads us to better results of depiction of the georeferenced wordcloud.



Georeferencer - rathaus_m_without.png

Transformation settings

Transformation parameters

Transformation type: Linear

Resampling method: Cubic

Target SRS: Selected CRS (EPSG:43...

Output settings

Output raster

Compression: None

□ Create world file only (linear transforms)

□ Use 0 for transparency when needed

□ Set target resolution

Horizontal: 1.00000

Vertical: -1.00000

Reports

Generate PDF map

Generate PDF report

□ Load in QGIS when done

OK | Cancel | Help

Enter map coordinates

Enter X and Y coordinates (DMS (*dd mm ss.ss*), DD (*dd.dd*) or projected coordinates (mmmm.mm)) which correspond with the selected point on the image. Alternatively, click the button with icon of a pencil and then click a corresponding point on map canvas of QGIS to fill in coordinates of that point.

X / East | Y / North

OK | From map canvas | Cancel

Visible	ID	Source X	Source Y	Dest. X	Dest.
✕	0	1353.8	-21.7747	16.3573	48.
✕	1	1408.13	-2887.32	16.3573	48.
✕	2	3313.2	-1411	16.3583	4
✕	3	-78.4705	-1476.32	16.3563	48.
✕	4	1319.29	-1415.35	16.3573	48.
✕	5	1488.92	-2716.97	16.3573	48.

Figure 16 Example of the Georefencing the Wordcloud

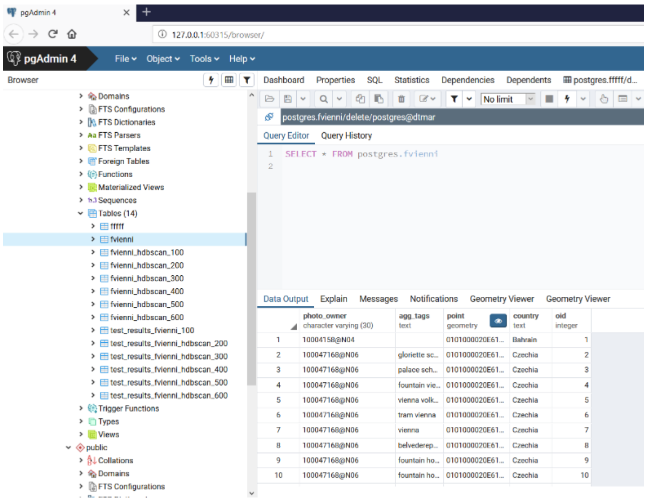
3.5 Handling the data by PgAdmin

The amount of data is huge and it is required to use specific tools in order to handle them. One useful tool is the PgAdmin for managing data sets. PgAdmin is an open source management tool for Postgres, an advanced Open Source database. It provides us a graphical interface (GUI) in order to create, maintain and use database objects[10]. The creation of a database, the storage of our data sets in different files and some other actions of the data like editing, sorting, joining database features are performed by the use of PgAdmin4. The following sections explain the procedure of joining the tables for Flickr and Twitter data.

3.5.1 Joining Tables for Flickr Data

The attribute table of the Flickr data consists of 5 columns whose names are the following: Photo_owner, agg_tags, point, country and oid. The attribute table of the Flickr data set after the implementation of the HDBCAN algorithm (fvienni_hdbscan_600) consists of 9 columns, whose names are the following:

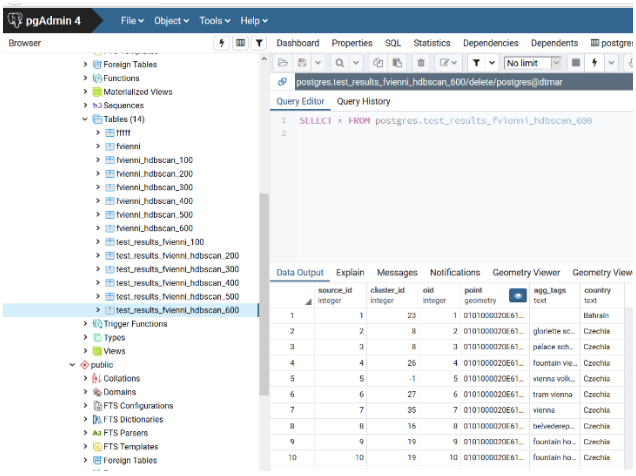
Object_id, source_id,cluster_id,prob,outlier,exemplar,stability,color_id and shape.



```
SELECT * FROM postgres.fvienni
```

	photo_owner character varying (30)	agg_tags text	point geometry	country text	oid integer
1	10004158@N04		0101000020E61...	Bahrain	1
2	100047168@N06	gloriette sc...	0101000020E61...	Czechia	2
3	100047168@N06	palace sch...	0101000020E61...	Czechia	3
4	100047168@N06	fountain vie...	0101000020E61...	Czechia	4
5	100047168@N06	vienna volk...	0101000020E61...	Czechia	5
6	100047168@N06	tram vienna	0101000020E61...	Czechia	6
7	100047168@N06	vienna	0101000020E61...	Czechia	7
8	100047168@N06	belvederep...	0101000020E61...	Czechia	8
9	100047168@N06	fountain ho...	0101000020E61...	Czechia	9
10	100047168@N06	fountain ho...	0101000020E61...	Czechia	10

Figure 17 Attribute table of Flickr data set (fvienni): screenshot of pgAdmin4



```
SELECT * FROM postgres.test_results_fvienni_hdbscan_600
```

	source_id integer	cluster_id integer	oid integer	point geometry	agg_tags text	country text
1	1	23	1	0101000020E61...		Bahrain
2	2	8	2	0101000020E61...	gloriette sc...	Czechia
3	3	8	3	0101000020E61...	palace sch...	Czechia
4	4	26	4	0101000020E61...	fountain vie...	Czechia
5	5	-1	5	0101000020E61...	vienna volk...	Czechia
6	6	27	6	0101000020E61...	tram vienna	Czechia
7	7	35	7	0101000020E61...	vienna	Czechia
8	8	16	8	0101000020E61...	belvederep...	Czechia
9	9	19	9	0101000020E61...	fountain ho...	Czechia
10	10	19	10	0101000020E61...	fountain ho...	Czechia

Figure 18 Attribute table of Flickr data set (test_results_fvienni_hdbscan_600): screenshot of pgAdmin4

[10] https://www.pgadmin.org/docs/pgadmin4/development/

It is essential to join the two tables in order to have all the required information for calculating the frequency of the appearance of the terms in one table. The following SQL Query presents the process of joining the tables. The common column is oid and source_id. In the second line of the query are included the columns (from both tables) that we want to have in the end. The following query refers to the joining of the fvienni_hdbscan_600 and of the fvienni (as the main table). The name of the new table is test_results_fvienni_hdbscan_600.

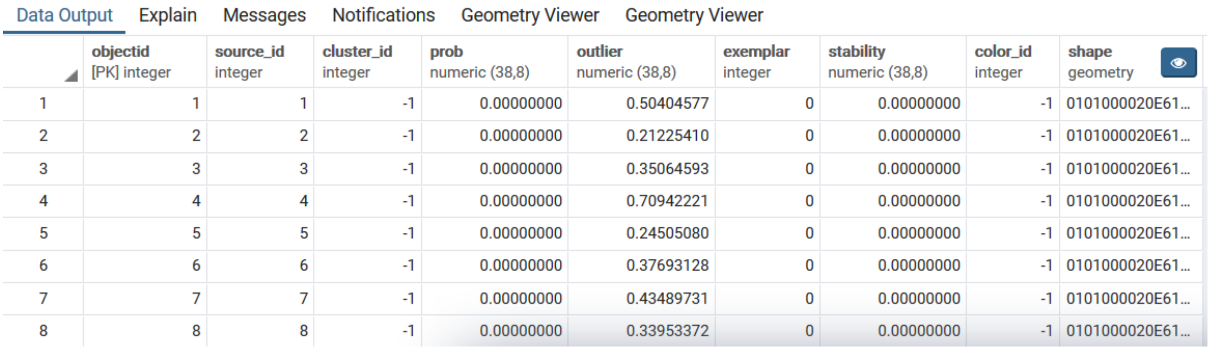
```
CREATE TABLE test_results_fvienni_hdbscan_600 as
SELECT source_id, cluster_id, oid, point, agg_tags, country
from fvienni as t1
LEFT JOIN fvienni_hdbscan_600 as t2
ON t1.oid = t2.source_id
ORDER BY source_id;
```

The new table contains all the information to calculate the frequency of each term. The results exported in csv format by using the UTF8 encoding format in order to represent any character of the Unicode standard.

3.5.2 Joining Tables for Twitter Data

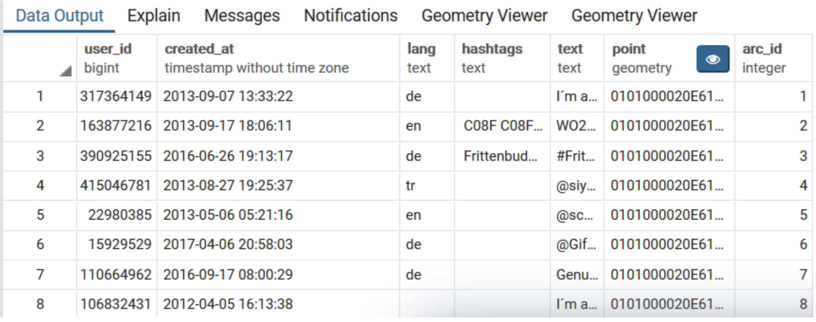
This sub chapter is focusing on the joins of the tables of Twitter data set. The combination of the twitvienna_hdbscan_300 data set, that created by HDBSCAN algorithm and its source, named as twitvienna leads to a new data set that contains all the suitable information with tags. The attribute table of the Twitter data consists of 7 columns whose names are the following: user_id,created_at, lang, hashtags,text,point,arc_id. The attribute table of the Twitter data set after the implementation of the HDBCAN algorithm (twitvienna_hdbscan_300) consists of 9 columns, whose names are the following: object_id, source_id, cluster_id,prob,outlier,exenmplar, stability,color_id and shape. The calculation of the frequency of each term could be possible after the joining of the two tables. The common column is arc_id and source_id. The following SQL Query refers to the process of joining the tables. In the second line of the query included all the columns (from both tables) that we want to have in the final table. The following query refers to the joining of the twitvienna_hdbscan_300 and of the twitvienna (as the main table). The name of the new table is test_results_300_Twitter.

```
CREATE TABLE test_results_300_Twitter as
SELECT
user_id,created_at,lang,hashtags,text,point,arc_id,objectid,source_id,cluster_id,p
rob,outlier,exemplar,stability,color_id,shape
from twitvienna as t1
LEFT JOIN twitvienna_hdbscan_300 as t2
ON t1.arc_id = t2.source_id
ORDER BY source_id;
```



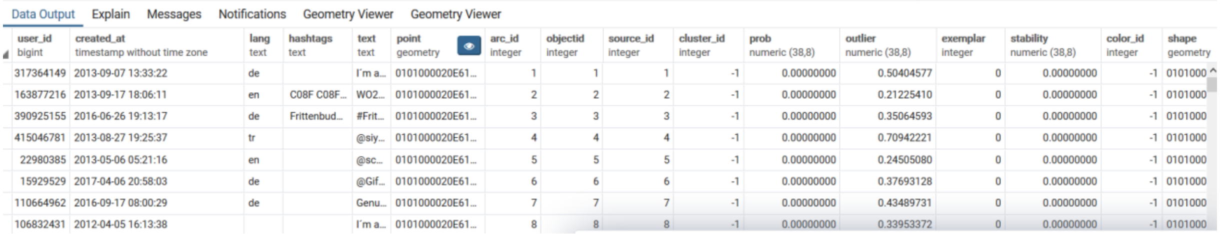
	objectid [PK] integer	source_id integer	cluster_id integer	prob numeric (38,8)	outlier numeric (38,8)	exemplar integer	stability numeric (38,8)	color_id integer	shape geometry
1	1	1	-1	0.00000000	0.50404577	0	0.00000000	-1	0101000020E61...
2	2	2	-1	0.00000000	0.21225410	0	0.00000000	-1	0101000020E61...
3	3	3	-1	0.00000000	0.35064593	0	0.00000000	-1	0101000020E61...
4	4	4	-1	0.00000000	0.70942221	0	0.00000000	-1	0101000020E61...
5	5	5	-1	0.00000000	0.24505080	0	0.00000000	-1	0101000020E61...
6	6	6	-1	0.00000000	0.37693128	0	0.00000000	-1	0101000020E61...
7	7	7	-1	0.00000000	0.43489731	0	0.00000000	-1	0101000020E61...
8	8	8	-1	0.00000000	0.33953372	0	0.00000000	-1	0101000020E61...

Figure 19 Attribute table of the Twitter data set (twitvienna_hdbscan_300)



	user_id bigint	created_at timestamp without time zone	lang text	hashtags text	text text	point geometry	arc_id integer
1	317364149	2013-09-07 13:33:22	de		I´m a...	0101000020E61...	1
2	163877216	2013-09-17 18:06:11	en	C08F C08F...	WO2...	0101000020E61...	2
3	390925155	2016-06-26 19:13:17	de	Frittenbud...	#Frit...	0101000020E61...	3
4	415046781	2013-08-27 19:25:37	tr		@siy...	0101000020E61...	4
5	22980385	2013-05-06 05:21:16	en		@sc...	0101000020E61...	5
6	15929529	2017-04-06 20:58:03	de		@Gif...	0101000020E61...	6
7	110664962	2016-09-17 08:00:29	de		Genu...	0101000020E61...	7
8	106832431	2012-04-05 16:13:38			I´m a...	0101000020E61...	8

Figure 20 Attribute table of the Twitter data set (twitvienna)



user_id bigint	created_at timestamp without time zone	lang text	hashtags text	text text	point geometry	arc_id integer	objectid integer	source_id integer	cluster_id integer	prob numeric (38,8)	outlier numeric (38,8)	exemplar integer	stability numeric (38,8)	color_id integer	shape geometry
317364149	2013-09-07 13:33:22	de		I´m a...	0101000020E61...	1	1	1	-1	0.00000000	0.50404577	0	0.00000000	-1	0101000
163877216	2013-09-17 18:06:11	en	C08F C08F...	WO2...	0101000020E61...	2	2	2	-1	0.00000000	0.21225410	0	0.00000000	-1	0101000
390925155	2016-06-26 19:13:17	de	Frittenbud...	#Frit...	0101000020E61...	3	3	3	-1	0.00000000	0.35064593	0	0.00000000	-1	0101000
415046781	2013-08-27 19:25:37	tr		@siy...	0101000020E61...	4	4	4	-1	0.00000000	0.70942221	0	0.00000000	-1	0101000
22980385	2013-05-06 05:21:16	en		@sc...	0101000020E61...	5	5	5	-1	0.00000000	0.24505080	0	0.00000000	-1	0101000
15929529	2017-04-06 20:58:03	de		@Gif...	0101000020E61...	6	6	6	-1	0.00000000	0.37693128	0	0.00000000	-1	0101000
110664962	2016-09-17 08:00:29	de		Genu...	0101000020E61...	7	7	7	-1	0.00000000	0.43489731	0	0.00000000	-1	0101000
106832431	2012-04-05 16:13:38			I´m a...	0101000020E61...	8	8	8	-1	0.00000000	0.33953372	0	0.00000000	-1	0101000

Figure 21 Attribute table of the final result (test_results_300_Twitter)

The joins were made for all clustering groups of HDBSCAN using the same query. The names of the final joining tables are the following:

- test_results_300_Twitter
- test_results_200_Twitter
- test_results_100_Twitter

3.5.3 Analysis of Twitter Data HDBSCAN 300

In order to find the most frequent terms, it is important to have in the same array all the required information that help the procedure of calculation. Splitting Twitter texts into multiple words could be done by the use of a function. This is essential for calculating the frequency of appearance of each term and not for a group of terms that consist a sentence or phrase. Also it is important to sort the data in a descending order for facilitating the process of determining the most frequent terms. The following part describes the procedure of sorting the clustered Twitter data by HDBSCAN 300.

The number of clusters derived from HDBSCAN algorithm is 3. Each cluster could be distinguished from another cluster by its identification name. In this data base cluster id plays the role of the unique identity of the clusters. So, cluster_id =1 represents the first cluster, cluster_id=2 represents the second cluster and cluster_id=3 represents the third cluster. Data derived from test_results_300_Twitter (joining tables, described in the previous chapter) could be used in order to calculate the frequency of each terms per cluster.

The following query has as outcome two columns: one column with the tags (the terms) and the other with the counts (how frequent they appear in each cluster). This query examines the terms of the first cluster as its id is number 1 (cluster_id=1).

1st

```
SELECT tag, count(*)
FROM (
  SELECT regexp_split_to_table(text, ' ') as tag
  FROM test_results_300_Twitter
  WHERE cluster_id = 1
) t
GROUP BY tag ORDER BY count DESC;
```

An analogous query was executed for the second cluster (cluster_id=2) and the third cluster (cluster_id=3)

2nd

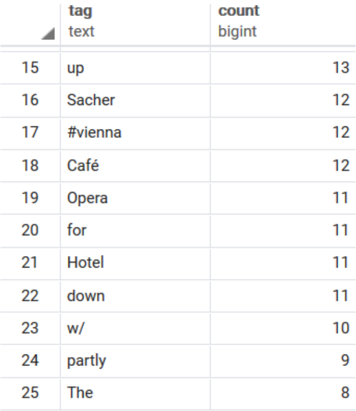
```
SELECT tag, count(*)
FROM (
  SELECT regexp_split_to_table(text, ' ') as tag
  FROM test_results_300_Twitter
  WHERE cluster_id = 2
) t
GROUP BY tag ORDER BY count DESC;
```

3nd

```
SELECT tag, count(*)
FROM (
  SELECT regexp_split_to_table(text, ' ') as tag
  FROM test_results_300_Twitter
  WHERE cluster_id = 3
) t
GROUP BY tag ORDER BY count DESC;
```

Now it is possible to examine the relationship between the POI and the created clusters based on the semantics. From the three created clusters the second cluster (cluster_id =2) and the third cluster (cluster_id=3) are related to Points of Interests (POI). Then, the terms of the related clusters are examined by the user's region of origin. The collection of the Wikipedia Article based on English and German language and this is the reason of comparing semantics between English and German users. The following query shows the sorting of the data based on language perspective in order to calculate the frequency of each term per cluster and per place of origin. The EN is the abbreviation of English and the DE is the abbreviation of German. The result is a table of two columns, one with the counts and the other with the terms.

```
SELECT tag, count(*)
FROM (
  SELECT regexp_split_to_table(text, ' ') as tag
  FROM test_results_300_Twitter
  WHERE cluster_id = 2 AND lang='en'
) t
GROUP BY tag ORDER BY count DESC;
```

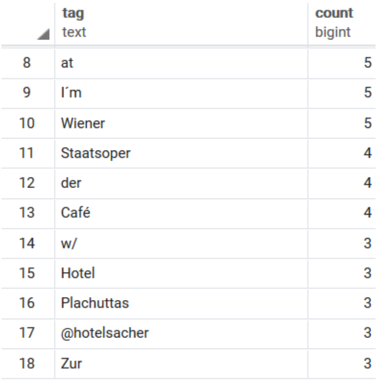


	tag text	count bigint
15	up	13
16	Sacher	12
17	#vienna	12
18	Café	12
19	Opera	11
20	for	11
21	Hotel	11
22	down	11
23	w/	10
24	partly	9
25	The	8

Figure 22 Part of the created table: Outcome for counts and terms based on English users for the 2ndcluster of Twitter Data HDBSCAN 300

The same query for the German language is the following:

```
SELECT tag, count(*)
FROM (
  SELECT regexp_split_to_table(text, ' ') as tag
  FROM test_results_300_Twitter
  WHERE cluster_id = 2 AND lang='de'
) t
GROUP BY tag ORDER BY count DESC;
```



	tag	count
	text	bigint
8	at	5
9	I´m	5
10	Wiener	5
11	Staatsoper	4
12	der	4
13	Café	4
14	w/	3
15	Hotel	3
16	Plachuttas	3
17	@hotelsacher	3
18	Zur	3

Figure 23 Part of the created table: Outcome for counts and terms based on German users for the 2^{nd}cluster of Twitter Data HDBSCAN 300

The analysis of the semantics includes the removal of unnecessary or strange tags and the calculation of the frecuency. Some examples of unnecessary tags are ihpoeography,instagramap, linking words such as and/und or personal pronouns. A typical example of tags with no valuable information for the purpose of the analysis are toponyms such as Wien, Vienna, Austria or Österreich. Removing punctuation symbols and other signs such as the symbol of the hashtag or the symbol "at" sumbol is an important task for the analysis of the semantics. The calculation of frequency based on the following function:

$$f\mathrm{ri}(100) = \frac{\mathrm{fr(i)}}{\mathrm{total}} * 100,$$

where fr(i) is the count of a term and total is the total amount of counts and fri(100) is the final value in form of percentage. The excel software is used for the calculation of the frequency. The results of this analysis could be found in the chapter of the results. The most top ten frequent terms displayed in a table.

3.5.4 Analysis of Twitter Data HDBSCAN 200

The analysis of the Twitter Data with HDBSCAN 200 following the same principles and based on the same steps that described in the previous chapter. The number of clusters derived from HDBSCAN algorithm is 9. Each cluster could be distinguished from another cluster by its identification name. In this data base, cluster id plays the role of the unique identity of the clusters. So, cluster_id =1 represents the first cluster, cluster_id=2 represents the second cluster and so on. Three of them correlated to specific Points of Interest (POI). These clusters are: cluster with id number 5, cluster with id number 7 and cluster with id number 8. The analysis of the semantics performed for three fore-mentioned correlated clusters based on German and English language. The results of this analysis could be found in the chapter of the results. A table and a bar chart is used in order to display the results of the analysis.

3.5.5 Analysis of Twitter Data HDBSCAN 100

The analysis of the Twitter Data with HDBSCAN 100 based on the same steps that described in the previous chapter. The number of clusters derived from HDBSCAN algorithm is 23. Each cluster could be distinguished from another cluster by its identification name. In this data base, cluster id plays the role of the unique identity of the clusters. So, cluster_id =1 represents the first cluster, cluster_id=2 represents the second cluster and so on. Eight of them correlated to specific Points of Interest (POI). These clusters are: cluster with id number 8, cluster with id number 9, cluster with id number 15, cluster with id number 16,cluster with id number 18, cluster with id number 19, cluster with id number 20 and cluster with id number 22. The analysis of the semantics of the 8 correlated clusters based on English and German language performed following the same procedure, whose steps described in the 3.5.3 chapter. The most top ten frequent terms per related cluster displayed in different tables. In general, the analysis procedure leads to 16 tables (2 tables per related cluster). The chapter of results presents these tables.

3.5.6 Analysis of Flickr Data HDBSCAN 600

The analysis of the Flickr Data with HDBSCAN 600 based on the same steps that described in the previous chapters. The only difference is the absence of information of the place of origin of the users. The outcome of the HDBSCAN algorithm is the creation of 35 clusters. Each cluster has a specific code of identification whose name is cluster identification code and its abbreviation is the following: cluster_id=i, i=1,2,..35. The analysis of these data shows that 15 clusters are related to specific Points of Interest (POI). The name of these clusters are the following: cluster with id number 8, cluster with id number 12, cluster with id number 13, cluster with id number 15, cluster with id number 16, cluster with id number 21, cluster with id number 23, cluster with id number 26, cluster with id number 27, cluster with id number 28, cluster with id number 30, cluster with id number 31, cluster with id number 32, cluster with id number 34 and cluster with id number 35. The query of sorting the data per cluster in order to have to counts and the tags is the following (example of the cluster with cluster_id=8). The similar query is executed for the rest correlated clusters.

```
SELECT tag, count(*)
FROM (
  SELECT regexp_split_to_table(agg_tags, ' ') as tag
  FROM test_results_fvienni_hdbscan_600
  WHERE cluster_id = 8
) t
GROUP BY tag ORDER BY count DESC;
```

The analysis leads us to 15 tables, each one per cluster. The results displayed in tables in the chapter of results.

4. Results

4.1 Clusters of Flickr Data

The analysis of these data shows that 15 clusters are related to specific Points of Interest (POI). The total amount of clusters is 35 derived by the execution of HDBCAN algorithm. The following tables display the frequency of appearance of each term per cluster. The first column displays the top 10 terms, the second column presents the percentages of the appearance to the total terms and the third column displays the sum of the top terms.

The 8th cluster is related with the Schönbrunn Palace and the majority of the users use "Schönbrunn" with 3.61% to total terms. The terms "zoo" and "palace" are the 2nd and 3rd most used terms with a percentage of 1.27% and 1.04% to total terms respectively.

Table 2 Top 10 Terms for Flickr Data of cluster_id=8

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

The 12th cluster is related with the Wien Prater and the most frequent term is "Prater" with 5.03% to total terms. The percentages of the other terms are decreasing and are below the 1.60% to total terms. Riesenrad has a value of 1.58% to total terms. Terms that related to Riesenrad, but with less percentages of appearance are the "wheel" with 0.48% "ferriswheel" with 0.74% and "wienerriesenrad" with 0.28%.

Table 3 Top 10 Terms for Flickr Data of cluster_id=12

Top 10 Terms	fr(i)%	Sum(%)
prater	5.03	
riesenrad	1.58	
ferriswheel	0.74	
wheel	0.48	
park	0.47	10.30
wurstelprater	0.47	
amusementpark	0.45	
wienerprater	0.41	
praterstern	0.40	
wienerriesenrad	0.28	
Rest	89.70	

The 13th cluster is related with the Hundertwasserhaus and the term "hundertwasser" has the largest percentage of appearance (3.12% to total terms). The "hundertwasserhaus" has a value of 2.13% to total terms and following terms such as "building", and "art" with 0.68% and 0.51% to total terms correspondingly. "Fountain" is the last one in the top ten list with a percentage of 0.18 to total terms.

Table 4 Top 10 Terms for Flickr Data of cluster_id=13

Top 10 Terms	fr(i)%	Sum(%)
hundertwasser	3.12	
hundertwasserhaus	2.13	
building	0.68	
art	0.51	
city	0.47	8.47
friedensreichhundertwasser	0.45	
kunsthauswien	0.42	
architektur	0.27	
kunsthaus	0.24	
fountain	0.18	
Rest	91.53	

The 15th cluster is related with Mariahilf. The terms “mariahilf” is the first in the top ten list with 0.98% to total terms. “mariahilferstrasse” and “mariahilfestrase” are in the 2nd and 3rd place of the top ten list with 0.63% and 0.37% to total terms correspondingly. The last in this list is the term “architektur” with 0.15% to total terms.

Table 5 Top 10 Terms for Flickr Data of cluster_id=15

Top 10 Terms	fr(i)%	Sum(%)
mariahilf	0.98	
mariahilferstrase	0.63	
mariahilferstrasse	0.37	
building	0.36	
church	0.32	3.73
architecture	0.28	
mariahilferkirche	0.24	
kirche	0.23	
neubaugasse	0.17	
architektur	0.15	
Rest	96.27	

The 16th cluster is related with Belvedere Palace. The term “Belvedere” is the first in the top ten list with 4.12% to total terms. Also, many users use the terms of “palace” and “schloss” with a frequency of appearance 1.18% and 0.82% to total terms respectively. The two last terms of the top ten list are “sculpture” and “statue” with a percentage of 0.44% and 0.41% to total terms respectively.

Table 6 Top 10 Terms for Flickr Data of cluster_id=16

Top 10 Terms	fr(i)%	Sum(%)
belvedere	4.12	
palace	1.18	
schloss	0.82	
architecture	0.76	

museum	0.75	10.37
garden	0.72	
schlossbelvedere	0.61	
park	0.54	
statue	0.44	
sculpture	0.41	
Rest	89.63	

The 21st cluster related to University of Vienna. Many important buildings and points are included in this cluster such as Votiv kirche and the Schottentor. And this is the reason that there are high percentages of the occurrence of these terms in the list.”votivkirche” is the most used term with 2.05% to total terms. But the term that is related to a point of interest of the examined list is “university” with 0.34% to total terms , “universität” with 0.23% to total terms and “universityofvienna” with 0.20% to total terms.

Table 7 Top 10 Terms for Flickr Data of cluster_id=21

Top 10 Terms	fr(i)%	Sum(%)
votivkirche	2.05	
church	1.49	
kirche	0.86	
schottentor	0.72	
architecture	0.69	7.06
university	0.34	
building	0.26	
innerestadt	0.23	
universität	0.23	
universityofvienna	0.20	
Rest	92.94	

The 23rd cluster related to Karlkirche, Karslplatz and Musikverein building. The majority of the users use the term “karlskirche” with 2.79% and then comes the term “karlsplatz” with 2.42%. The 7th place of the top ten list is for the term “musikverein” with 0.37% to total terms.

Table 8 Top 10 Terms for Flickr Data of cluster_id=23

Top 10 Terms	fr(i)%	Sum(%)
karlskirche	2.79	
karlsplatz	2.42	
church	1.58	
architecture	1.00	
kirche	0.59	9.98
baroque	0.48	
musikverein	0.37	
building	0.31	
architektur	0.23	
iglesia	0.22	

Rest	90.02	

The 26th cluster related to Parliament and Volksgarten. Terms of "parlament" and "parliament" appear in the first places of the top ten list with 1.62% and 1.56% to total terms respectively. Other frequent terms are "volksgarten" with 1.10% to total terms and "statue" with 0.75% to total terms.

Table 9 Top 10 Terms for Flickr Data of cluster_id=26

Top 10 Terms	fr(i)%	Sum(%)
parlament	1.62	
parliament	1.56	
volksgarten	1.10	
statue	0.75	
architecture	0.70	7.65
building	0.51	
innerestadt	0.45	
park	0.36	
rathaus	0.34	
garden	0.26	
Rest	92.35	

The 27th cluster related to Rathaus and to Burgtheater. The most frequent term of this cluster is "rathaus" with 2.24% to total terms. Terms that are related to Rathaus are "rathausplatz" with 1.04% to total terms, "townhall" with 0.34% to total terms and "rathauspark" with 0.22% to total terms. The term "burgtheater" appears with a 0.78% to total terms and comes to the 3rd place of the top ten list.

Table 10 Top 10 Terms for Flickr Data of cluster_id=27

Top 10 Terms	fr(i)%	Sum(%)
rathaus	2.24	
rathausplatz	1.04	
burgtheater	0.78	
architecture	0.65	
cityhall	0.50	6.64
townhall	0.34	
innerestadt	0.34	
building	0.30	
hall	0.23	
rathauspark	0.22	
Rest	93.36	

The 28th cluster related to Kaisergruft. The most frequent term is "neuermarkt" with 0.75% to total terms. The term "kaisergruft" comes to the 6th place of the top ten list with 0.25% to total terms. "Kapuzinergruft" follows and has a 0.22% to total terms.

Table 11 Top 10 Terms for Flickr Data of cluster_id=28

Top 10 Terms	fr(i)%	Sum(%)
neuermarkt	0.75	
innerestadt	0.53	
architecture	0.43	
urban	0.35	
innenstadt	0.32	3.40
kaisergruft	0.25	
kapuzinergruft	0.22	
building	0.18	
sculpture	0.18	
church	0.18	
Rest	96.60	

The 30th cluster related to Albertina Museum and to Vienna State Opera. The first term in the top ten list of the frequencies is “albertina” with 1.18% to total terms. The term “opera” is the second most frequent term of this cluster with 1.08% to total terms. Terms related to Opera are “staatoper” and “wienerstaatsoper” and “operahouse” with 0.95%, 0.36% and 0.30% to total terms correspondingly. Terms of “museum” and “art” appear with 0.51% and 0.33% to total terms respectively.

Table 12 Top 10 Terms for Flickr Data of cluster_id=30

Top 10 Terms	fr(i)%	Sum(%)
albertina	1.18	
opera	1.08	
staatsoper	0.95	
architecture	0.71	
museum	0.51	6.09
wienerstaatsoper	0.36	
oper	0.35	
art	0.33	
building	0.32	
operahouse	0.30	
Rest	93.91	

The 31st cluster related to Kunsthistorisches and to Naturhistoriches Museum. The term “museum” has the largest percentage as the most frequent term. In the second place of the top ten list comes the term “kunsthistorischesmuseum” with 1.23% to total terms. In the 9th place of the top ten list comes the “naturhistorischesmuseum” with 0.49% to total terms.

Table 13 Top 10 Terms for Flickr Data of cluster_id=31

Top 10 Terms	fr(i)%	Sum(%)
museum	2.40	
kunsthistorischesmuseum	1.23	
art	0.71	
kunsthistorisches	0.69	

statue	0.60	8.11
mariatheresienplatz	0.58	
architecture	0.58	
history	0.52	
naturhistorischesmuseum	0.49	
sculpture	0.30	
Rest	91.89	

The 32nd cluster related to Hofburg Palace. The most used terms are “hofburg” and “heldenplatz” with 2.57% and 2.27% to total terms respectively. The term “hofburgpalace” comes to the 7th place of the top ten list with 0.46% to total terms.

Table 14 Top 10 Terms for Flickr Data of cluster_id=32

Top 10 Terms	fr(i)%	Sum(%)
hofburg	2.57	
heldenplatz	2.27	
statue	0.90	
palace	0.72	
architecture	0.60	9.11
innerestadt	0.56	
hofburgpalace	0.46	
horse	0.44	
monument	0.33	
imperial	0.25	
Rest	90.89	

The 34th cluster related to Peterskirche. The most used term is “graben” with 2.10% to total terms. The terms “peterskirche”, “church” and “kirche” have the 5th, the 6th and the 8th place of the top ten list.

Table 15 Top 10 Terms for Flickr Data of cluster_id=34

Top 10 Terms	fr(i)%	Sum(%)
graben	2.10	
pestsäule	0.63	
innerestadt	0.58	
architecture	0.56	
peterskirche	0.55	6.00
church	0.53	
statue	0.37	
kirche	0.23	
stpeterschurch	0.23	
monument	0.23	
Rest	94.00	

The 35th cluster related to Stephansdom. The term “stephansdom” has the largest percentage as the most frequent term (2.33% to total terms). The term “cathedral” is the second term of the top ten list with a 1.39% to total terms. Terms related to Stephandom are "stephansplatz", "church" and "ststephenscathedral" and "kirche"with 1.05%, 1.21%, 0.92% and 0.37% to total terms respectively.

Table 16 Top 10 Terms for Flickr Data of cluster_id=35

Top 10 Terms	fr(i)%	Sum(%)
stephansdom	2.33	
cathedral	1.39	
church	1.21	
stephansplatz	1.05	
ststephenscathedral	0.92	8.99
architecture	0.82	
kirche	0.37	
gothic	0.37	
innerestadt	0.28	
ststephens	0.25	
Rest	91.01	

4.2 Clusters of Twitter Data

4.2.1 Twitter Data HDBSCAN 300

The analysis of these data shows that 2 clusters are related to specific Points of Interest (POI). The total amount of clusters is 3 derived by the execution of HDBCAN algorithm. The following tables display the frequency of appearance of each term per cluster. The first column displays the top 10 terms in English language, the second column presents the percentages of the appearance to the total terms, third column displays the sum of the top terms, the forth column displays the top 10 terms in German language, the fifth column presents the percentages of appearance of the terms and the last column displays the sum of the top terms.

Cluster_id=2: This cluster refers to the Point of Interest Vienna State Opera. The analysis of the semantics shows that the majority of the users use terms related to the Vienna State Opera and its surrounded environment. So this is the reason of the appearance of terms such as Opera, Hotel, Sacher and Mozart. The next table displays the Top 10 Terms of this cluster for English and for German Language. The common terms are Sacher, café, Hotel, Staatoper, Mozart. Common terms are the Opera for English, with a percentage of 0.64% to total terms and Oper for German with a percentage of 0.54% to total terms

Table 17 Top 10 terms for EN and DE for Cluster_id=2 for Twitter Data with HDBCAN 300

Top 10 Terms (EN)	fr(i)%	Sum(%)	Top 10 Terms (DE)	fr(i)%	Sum(%)
Sacher	0.7		Sacher	1.63	
Café	0.7		Staatsoper	1.09	
Opera	0.64		Café	1.09	
Hotel	0.64		Hotel	0.82	
cloudy	0.47	4.68	Plachuttas	0.82	8.72
State	0.35		hotelsacher	0.82	
Staatsoper	0.35		Mozart	0.82	

Mozart	0.29		Gasthaus	0.82	
Pub	0.29		Oper	0.54	
Albertina	0.23		torte	0.27	
Rest	95.32		Rest	91.28	

Cluster_id =3: The analysis of the semantics shows that the majority of the users use terms related to the Stephansdom, the Cathedral of the city of Vienna and its surrounded environment. The majority of the terms refer to the St. Stephen's Cathedral and some other activities that are taking place in this neighborhood. This neighborhood is known as a touristic attraction and a place for sightseeing due to many important buildings in terms of architectural and cultural value. The common terms for both languages are Job, Hiring, Stephansplatz, Stephansdom and café. The users of English language mention terms such as hospitality and Graben with 0.06% and 0.04% respectively, where the German users mention Schnitzel, Architect with 0.07% to total terms and Heritage with 0.03% to total terms. More details for the most used terms in English and German language presented in the following table.

Table 18 Top 10 terms for EN and DE for Cluster_id=3 for Twitter Data with HDBCAN 300

Top 10 Terms (DE)	fr(i)%	Sum(%)	Top 10 Terms (EN)	fr(i)%	Sum(%)
Job	0.66		Job	0.49	
Stephansplatz	0.40		Hiring	0.47	
Hiring	0.23		Stephansplatz	0.40	
Praktikant	0.10		Stephen´s	0.28	
Café	0.10		Stephansdom	0.19	2.42
Praktikum	0.07	1.79	Cathedral	0.18	
Schnitzel	0.07		Café	0.11	
Stephansdom	0.07		Cathedral’s	0.18	
Architect	0.07		Hospitality	0.06	
Heritage	0.03		Graben	0.04	
Rest	98.21		Rest	97.58	

The next images depict the relation between the cluster_id=2 and the Points of Interest. The building labeled as 7 is the Vienna State Opera, the building labeled as 4 is the Albertina Museum and the building labeled as 6 is the Kaisergruft. The light blue circles are buffers, the prepared circles that used for the creation of the Wordclouds.

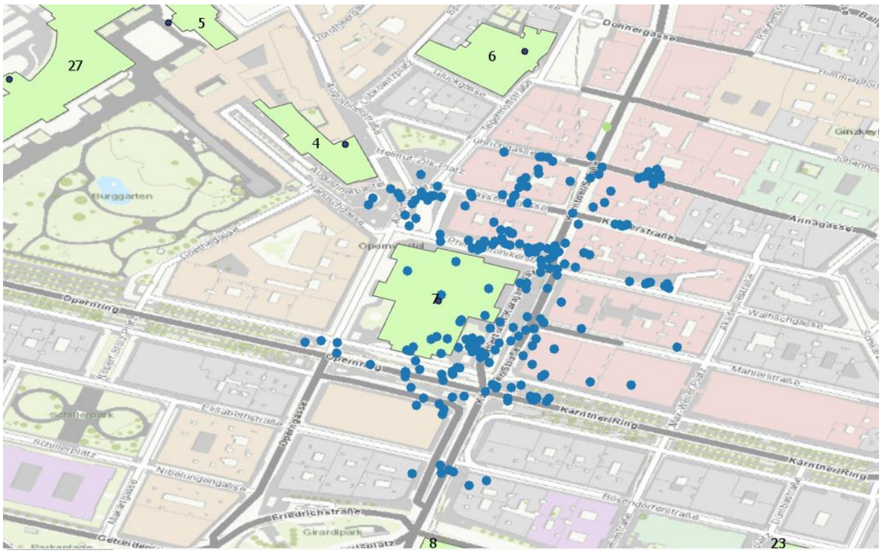


Figure 24 Relation between the cluster_id=2 and POI of Vienna State Opera (7) (HDBCSAN 300)

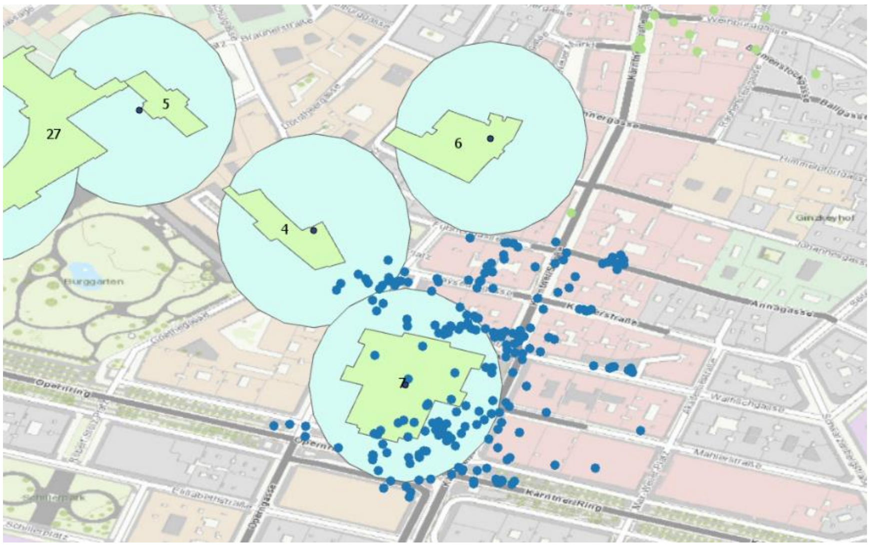


Figure 25 Relation between the cluster_id=2 and POI of Vienna State Opera (7) (HDBCSAN 300): buildings and radius around the POI

4.2.2 Twitter Data HDBSCAN 200

The analysis of these data shows that 3 clusters are related to specific Points of Interest (POI). The total amount of clusters is 9 derived by the execution of HDBCAN algorithm. The following tables display the frequency of appearance of each term per cluster. The first column displays the top 10 terms in English language, the second column presents the percentages of the appearance to the total terms, third column displays the sum of the top terms, the forth column displays the top 10 terms in German language, the fifth column presents the percentages of appearance of the terms and the last column displays the sum of the top terms.

Cluster_id=5: The analysis of the semantics shows that the majority of the users use terms related to the Schonbrunn Palace and its surrounded environment, such as the gardens. The common terms are Schonbrunn, Schloss, Palace and Gardens. It is worthy to mention that the English users use the name of the palace with German characters (Schönbrunn with 0.17% to total terms and Schoenbrunn with 0.86 % to total terms).Also, the term Schloss is a common term for both languages with 1.91% to total terms in German language and with 0.35% to total terms in English language.

Table 19 Top 10 terms for DE and EN for Cluster_id=5 for Twitter Data with HDBCAN 200

Top 10 Terms (DE)	fr(i)%	Sum(%)	Top 10 Terms (EN)	fr(i)%	Sum(%)
Schönbrunn	7.65		Palace	1.38	
Schoß	4.10		Schoenbrunn	0.86	
Palace	2.46		concert	0.35	
Schloss	1.91		Schloss	0.35	
Gardens	0.96		gardens	0.35	4.14
Schönnbrunn	0.55	18.72	Schönbrunn	0.17	
Tiergarten	0.55		environment	0.17	
Bühne	0.27		walk	0.17	
Schönbrunnpalace	0.14		roses	0.17	
schlosspark	0.14		Brauerei	0.17	
Rest	81.28		Rest	95.86	

Cluster_id=7: The analysis of the semantics shows us that the majority of terms referred to Vienna State Opera. The common terms are Sacher, Café, Staatoper, Hotel and Mozart. It is worthy to mention that the English users use the name of Albertina with 0.38% to total terms, whereas the German users not use this term. This occurred because Albertina Museum is adjacent to Vienna State Opera. Another common term is Opera for English, with a percentage of 0.59% to total terms and Oper for German with a percentage of 0.48% to total terms.

Table 20 Top 10 terms for DE and EN for Cluster_id=7 for Twitter Data with HDBCAN 200

Top 10 Terms (DE)	%	Sum(%)	Top 10 Terms (EN)	fr(i)%	Sum(%)
Sacher	1.44		Café	0.64	
café	1.44		Sacher	0.64	
Staatsoper	0.96		Hotel	0.59	
Plachuttas	0.72		Opera	0.59	
hotelsacher	0.72		cloudy	0.43	4.46
Hotel	0.72	8.39	Albertina	0.38	
Mozart	0.72		Staatsoper	0.32	
Gasthaus	0.72		State	0.32	
Oper	0.48		Mozart	0.27	
Ringstrassegalerien	0.48		House	0.27	
Rest	91.61		Rest	95.54	

Cluster_id=8: The analysis of the semantics shows that the majority of terms of this cluster referred to Stephansdom. The terms of Stephansdom and Stephansplatz are common in both English and German users even though the terms are in German language. The percentage of Stephansdom is 0.24% to total terms for German users and 0.12% to total terms for English users. The percentage of Stephansplarz is 0.35% to total terms for German users and 0.41% to total terms for English users. Another common term beautiful for English, with a percentage of 0.23% to total and wunderschön for German with a percentage of 0.12% to total terms.

Table 21 Top 10 terms for DE and EN for Cluster_id=8 for Twitter Data with HDBCAN 200

Top 10 Terms (DE)	%	Sum(%)	Top 10 Terms (EN)	%	Sum(%)
Stephansplatz	0.35		Stephen´s	0.58	
Stephansdom	0.24		Cathedral	0.50	
Stadt	0.23		Stephansplatz	0.41	
building	0.12		beautiful	0.23	
wunderschön	0.12	1.77	city	0.1	2.11
Innenstadt	0.12		Stephansdom	0.12	
stephenplatz	0.12		Coffee	0.04	
Heritage	0.12		building	0.04	
Chor	0.12		UNESCO	0.04	
Mahlzeit	0.12		architecture	0.04	
weihnachtet	0.12		Rest	97.89	

The next images depict the relation between the cluster_buid=7 (orange color) and cluster_id= 8 (light blue color) and cluster_id=5 with the Points of Interest. The building labeled as 7 is the Vienna State Opera, the building labeled as 4 is the Albertina Museum and the building labeled as 6 is the Kaisergruft, the building labeled as 11 is Schönbrunn Palace and the building labeled as 22 is Stephansdom. The light blue circles are buffers, the prepared circles that used for the creation of the Wordclouds.

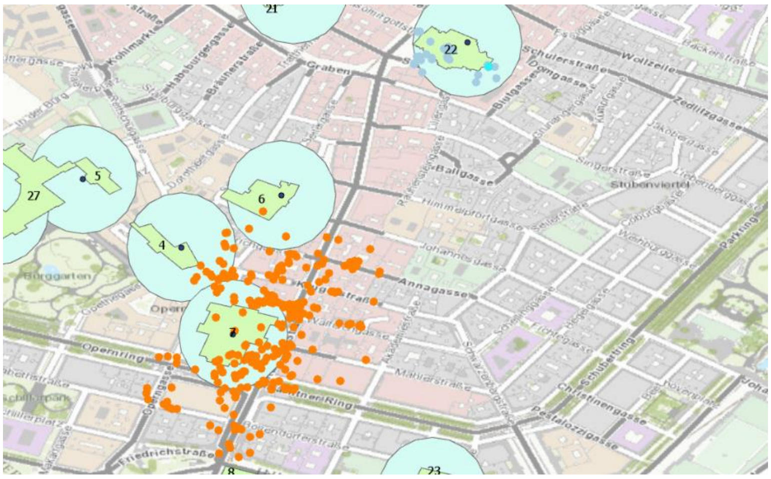


Figure 26 Relation between the cluster_id=7 and Vienna State Opera (7) & cluster_id=8 and Stephansdom (22) (HDBCSAN 200): buildings and radius around the POI

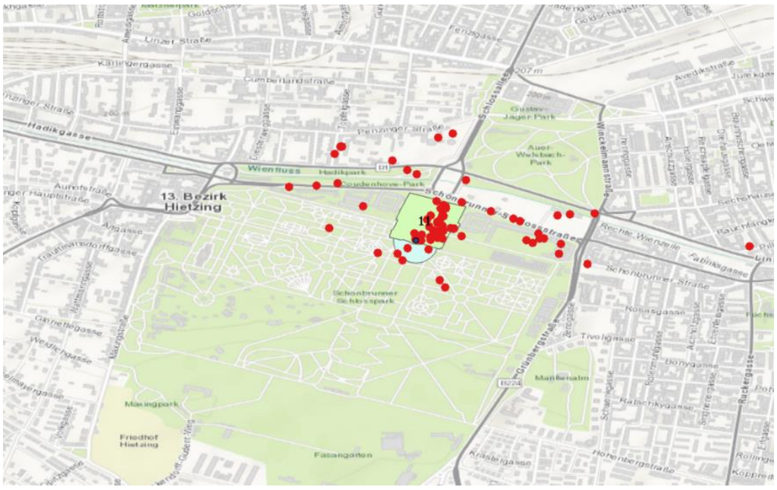


Figure 27 Relation between the cluster_id=5 and Schönbrunn Palace (11) (HDBCSAN 200): buildings and radius around the POI

4.2.3 Twitter Data HDBSCAN 100

The analysis of these data shows that 8 clusters are related to specific Points of Interest (POI). The total amount of clusters is 23 derived by the execution of HDBCAN algorithm. The following tables display the frequency of appearance of each term per cluster. The first column displays the top 10 terms in English language, the second column presents the percentages of the appearance to the total terms, third column displays the sum of the top terms, the forth column displays the top 10 terms in German language, the fifth column presents the percentages of appearance of the terms and the last column displays the sum of the top terms.

Cluster id=8: The analysis of the semantics shows that the majority of the users use terms related to the Schonbrunn Palace and its surrounded environment. The common terms are Schoenbrunn, Schönbrunn, Schloss, Palace and garden. It is worthy to mention that the English users use the name of the palace with German characters (Schönbrunn with 1.12% to total terms and Schoenbrunn with 0.93 % to total terms).Also, the term Schloss is a common term for both languages, with a percentage 0.37% for English and 1.91 for German.

Table 22 Top 10 terms for EN and DE for Cluster_id=8 for Twitter Data with HDBCAN 100

Top 10 Terms (EN)	%	Sum(%)	Top 10 Terms (DE)	fr(i)%	Sum(%)
Palace	2.22		Schönbrunn	8.06	
Schönbrunn	1.12		Schloß	4.10	
Schoenbrunn	0.93		Palace	2.46	
concert	0.37		Schloss	1.91	
rose	0.37	6.32	Gardens	0.96	19.00
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	93.68		Rest	81.00	

Cluster id=9: The analysis of the semantics shows that the majority of the users use terms related to Prater. The common terms are Prater, Pratestern, ferriswheel, Museum and Wurstelprater. Another common term is Wheel in English with 0.20% to total terms and Rode with 0.20% to total terms, whereas Riesenrad in German with 3.47% to total terms.It is worthy to mention that some German users use English words. The term amusement with a percentage of 0.43% to total terms is a typical example of this situation.

Table 23 Top 10 terms for EN and DE for Cluster_id=9 for Twitter Data with HDBCAN 100

Top 10 Terms (EN)	%	Sum(%)	Top 10 Terms (DE)	%	Sum(%)
Prater	3.05		Prater	4.33	
Praterstern	1.22		Riesenrad	3.47	
Wheel	0.20		ferriswheel	0.87	
museum	0.20		Praterstern	0.87	
walking	0.20	5.88	park	0.87	13.01
madametussaudswien	0.20		Museum	0.86	
Wurstelprater	0.20		amusement	0.43	
ferriswheel	0.20		praterriesenrad	0.43	
Rode	0.20		Wurstelprater	0.43	
weinerprater	0.20		Volksprater	0.43	
Rest	94.12		Rest	86.99	

Cluster_id=15: The analysis of the semantics shows that the majority of the users use terms related to Mariahilfe Region. This cluster includes the POI of Mariahilfe Kirche, but there was not any word referered to church, except the name of the street. The terms Mariahilfer and Mariahilferstrasse used by German users with 2.82% and 1.13% to total terms respectively. The next table presents the Top 10 Terms of this cluster for English and for German Language. The common terms are shopping and shop with 0.28% to total terms in German and 0.20% to total terms in English. The appearance of terms such as cinema, lunch, friends, essen and family related to the whole region of this cluster as a characteristic place for shopping, meeting people and other activities related to amusement. The next table presents the Top 10 Terms of this cluster for English and for German Language.

Table 24 Top 10 terms for EN and DE for Cluster_id=15 for Twitter Data with HDBCAN 100

Top 10 Terms (EN)	%	Sum(%)	Top 10 Terms (DE)	%	Sum(%)
Café	0.81		**Mariahilfer**	**2.82**	
Planet	0.61		**Mariahilferstrasse**	**1.13**	
friends	0.60		**Neubaugasse**	**0.85**	
family	0.20		**wienerlinien**	**0.56**	
weekend	0.20	3.83	**Neubau.**	**0.28**	**7.03**
Dancehall	0.40		**Erdbeer**	**0.28**	
Cinema	0.20		**essen**	**0.28**	
lunch	0.20		**shopping**	**0.28**	
Shop	0.20		**Begegnungszone**	**0.28**	
myschnitzel	0.20		**spaziergang**	**0.28**	
Beautiful	0.20		**Rest**	**92.97**	

Cluster_id=16: The analysis of the semantics shows that the majority of the users use terms related to the University of Vienna, Vienna City Hall, Burgtheater and Parliament. The common terms are Rathaus, Rathausplatz, Burgtheater and University. It is worthy to mention that the English users use the name of Rathaus and Rathausplatz, with 1.27% and 0.84% to total terms respectively with German spelling. On the other hand Universität and Rathaus appear to have higher percentages for German language with 2.33% and 1.35% to total terms correspondingly than in English language (0.14% to total terms for University and 1.27% to total terms for Rathaus). The term of Parlament appears with a percentage of 0.24% to total terms in German and does not appear to the top ten list for English language.

Table 25 Top 10 terms for EN and DE for Cluster_id=16 for Twitter Data with HDBCAN 100

Top 10 Terms (EN)	%	Sum(%)	Top 10 Terms (DE)	%	Sum(%)
Rathaus	1.27		Universität	2.33	
Rathausplatz	0.84		Rathaus	1.35	
Café	0.42		univienna	0.98	
viennachristmasmarket	0.28		Rathausplatz	0.86	
University	0.14	3.65	Schottentor	0.74	7.96
burgtheater	0.14		café	0.49	
professors	0.14		Burgtheater	0.37	
campus	0.14		University	0.36	
christmasmarkets	0.14		wienrathaus	0.25	
Filmfestival	0.14		Parlament	0.24	
Rest	96.35		Rest	92.04	

Cluster_id=18: The analysis of the semantics shows that the majority of the users use terms related to Museums quartier. It is obvious that this cluster contains a lot of museums such as Museumsquartier, Kunsthistorisches (POI) and Naturhistorisches Museum (POI). According to the analysis of the semantics the majority of tweets are correlated with the Museums quartier and less tweets are related to the aforementioned POI (Naturhistorisches Museunm and Kunsthistorisches). In German terms Naturhsitorisches has a percentage of 0.51% to total terms and NhM (abbreviation of Naturhistorisches Museum) has a percentage of 0.26% to total terms. In English language, Naturhistorisches appears with a percentage of 0.13% to total terms. The same percentage (0.13%) appears for the following terms: Kunsthistorischesmuseum, exhibition and Kunsthalle. The table below displays the Top 10 Terms of this cluster for English and for German Language.

Table 26 Top 10 terms for EN and DE for Cluster_id=18 for Twitter Data with HDBCAN 100

Top 10 Terms (EN)	%	Sum(%)	Total 10 Terms (DE)	%	Sum(%)
Museum	1.70		Museum	2.57	
mqwien	1.57		MuseumsQuartier	2.05	
MuseumsQuartier	1.44		moderner	1.28	
Leopold	1.18	6.94	Kunst	1.8	
Museumsquartier	0.39		Naturhistorisches	0.51	9.51
culture	0.13		Museumsquartier	0.26	
Naturhistorisches	0.13		museumquarter	0.26	
kunsthistorischesmuseum	0.13		NhM	0.26	

exhibition	0.13		quartier	0.26	
Kunsthalle	0.13		Networking	0.26	
Rest	93.06		Rest	90.49	

Cluster_id=19: The analysis of the semantics shows us that the majority of the users use terms related to Hofburg Palace. The common terms are Hofburg, Hofburg Vienna, Palace, Museum and Imperial. Another common term is the ball with 0.37% total terms in English and Festsaal with 0.58% to total terms in German. The following table presents the Top 10 Terms of this cluster for English and for German Language.

Table 27 Top 10 terms for EN and DE for Cluster_id=19 for Twitter Data with HDBCAN 100

Total 10 Terms (EN)	fr(i)%	Sum(%)	Total 10 Terms (DE)	fr(i)%	Sum(%)
Hofburg	1.87		Hofburg	8.18	
Palace	0.75		Palace	1.75	
Festival	0.56		museum	0.58	
Museum	0.56		Festsaal	0.58	
Sissi	0.38	5.26	Imperial	0.58	14.58
ball	0.37		Burgenland	0.58	
wonderful	0.19		hofburg_vienna	0.58	
Hofburg_vienna	0.19		Strauss	0.58	
Imperial	0.19		Cafe	0.58	
Silvesterball	0.19		Mozart	0.58	
Rest	94.74		Rest	85.42	

Cluster_id=20: The analysis of the semantics shows us that the majority of the users use terms related to Vienna State Opera. Also some users use terms related to the Albertina Museum and to Karlsplatz. The common terms are Sacher, Hotelsacher, Staatoper, and café, Mozart, Albertina and Karlsplatz. The English users use the German version for Karlsplatz and Staatoper, with 0.32% and 0.24% to total terms correspondingly. The same terms appear to have higher percentages for German users with 1.84% and 0.92% to total terms correspondingly.

Table 28 Top 10 terms for EN and DE for Cluster_id=20 for Twitter Data with HDBCAN 100

Top 10 Terms (EN)	fr(i)%	Sum(%)	Top 10 Terms (DE)	fr(i)%	Sum(%)
Sacher	0.67		Karlsplatz	1.84	
Opera	0.59		Sacher	1.11	
café	0.47		Staatsoper	0.92	
Hotel	0.43		café	0.74	
Albertina	0.40	3.52	hotelsacher	0.73	
Karlsplatz	0.32		Plachuttas	0.55	7.35
Staatsoper	0.24		Mozart	0.55	
Mozart	0.20		Albertina	0.37	
hotelsacher	0.12		Torte	0.36	
operahouse	0.08		albertinamuseum	0.18	
Rest	96.48		Rest	92.65	

Cluster_id=22: The analysis of the semantics shows us that the majority of the users use terms related to the St. Stephen's Cathedral and its surrounded environment. The common terms are Stephen's, Cathedral, Stephansplatz and Heritage. It is worthy to point out that the English users include in their tweets terms with German version such as Stephansdom and Stephansplatz with 0.12% and 0.04% correspondingly. There is also important to mention that terms in English (Ticket and Heritage) are published by German users by a percentage of 0.12% to total terms.

Table 29 Top 10 terms for EN and DE for Cluster_id=22 for Twitter Data with HDBCAN 100

Top 10 Terms (EN)	fr(i)%	Sum(%)	Top 10 Terms (DE)	fr(i)%	Sum(%)
Stephen´s	0.65		Stadt	0.24	
Cathedral	0.59		Stephen´s	0.24	
Stephansdom	0.12		stephenplatz	0.12	
architecture	0.06		Innenstadt	0.12	1.45
Stephansplatz	0.04	1.64	Cathedral	0.12	
UNESCO	0.04		Ticket	0.12	
artwork	0.04		klingen	0.12	
City	0.04		Heritage	0.12	
building	0.04		Spaziergang	0.12	
Heritage	0.02		kunst	0.12	
Rest	98.36		Rest	98.55	

The next image illustrates the relationship between clusters and Points of Interest (cluster_id=16 in orange color, the cluster_id=18 in blue color, the cluster_id=19 in green color and the cluster_id=20 in light green). The Point of Interests have the following labels: (Hofburg Palace: 27, Albertina Museum: 4, Vienna State Opera: 7, Kunsthistorische Museum: 2, Naturhistorische Museum:1, University of Vienna:16, Vienna City Hall:17, Burgtheater:15,Parliament:19 and Volksgarten:3).

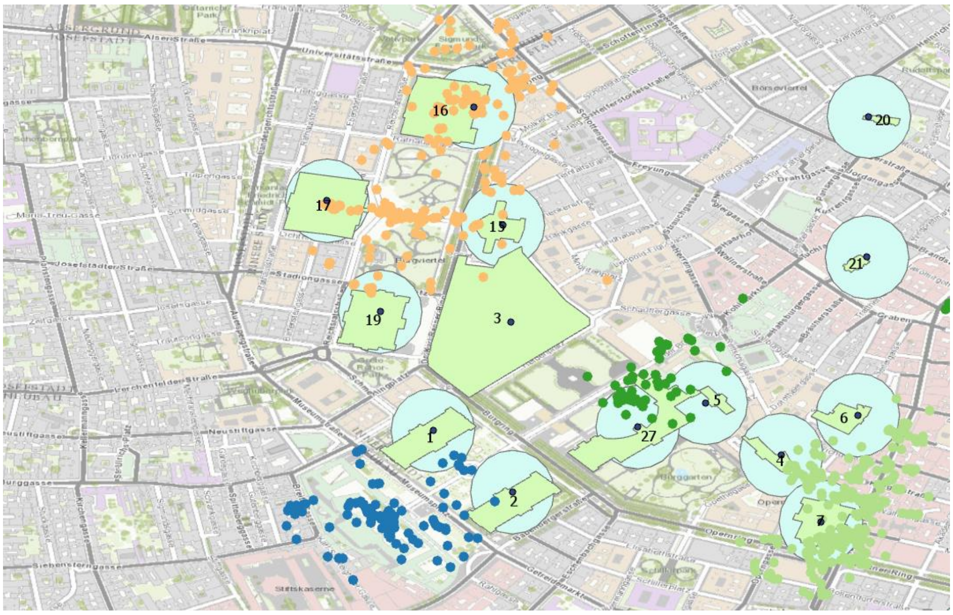


Figure 28 Relation between the clusters and POIs (HDBCSAN 100): buildings and radius around the POI

4.3 Frequent Terms extracted from Wikipedia Articles

This sub chapter presents the results of the Wordclouds for each Point of Interest and depicts the top 10 terms of frequency, obtained from the word list of the each word cloud. A table displays the top 10 terms of frequency by percentages and the percentage of the rest terms. Each term has a different frequency. In order to select the most frequent terms, we calculate the percentage of frequency of each term based on the following function:$f\text{ri}(100) = \frac{\text{fr(i)}}{\text{total}} * 100$, where fr(i) is the count of a term and total is the total amount of counts and fri(100) is the final value in form of percentage. For example the total amount of different terms of Albertina Museum is 136, where the sum of terms is 700. So the most used term for Albertina Museum in German language is the term “Albertina” with 13.29% of the total terms. The second term in this ranking list is the term “Sammlung” with 8.86% of the total terms. The last in the top ten list is the term “Albrecht” with 1.57% of the total terms. The top ten terms of frequency represent the 42.29 % of the total terms and the REST terms represent the 57.71% of the terms. The same visualization ways are used for the depiction of the top ten terms of frequency for Albertina Museum based on Wikipedia articles in English language. According to the following table, the list of the top ten terms of the Albertina Museum is different from the previous top ten list. So the most used term for Albertina Museum in English language is the term “Albertina” with 9.82 % of the total terms. The second term in this ranking list is the term “Collection” with 8.93% of the total terms. The last in the top ten list is the term “Albrecht” with 3.57% of the total terms. The top ten terms of frequency represent the 58.04 % of the total terms and the REST terms represent the 41.96% of the terms.

Table 30 Top 10 Terms for EN and DE based on Wikipedia Articles for Albertina Museum

TOP 10 Terms EN	fr(i) %	Sum (%)	Total	TOP 10 Terms DE	fr (i) (%)	Sum(%)	Total
Albertina	9.82	58.04		Albertina	13.29	42.29	

collection	8.93			Sammlung	8.86		
identifiers	8.04			Albert	4.00		
building	5.36			Direktor	3.14		
Duke	5.36			Palais	2.86		
Archduke	4.46			Herzog	2.71		
museum	4.46			Marie	2.29		
Albrecht	3.57			Christine	1.86		
Art	4.46			Zeichnungen	1.71		
Albert	3.57			Albrecht	1.57		
Rest	41.96	41.96	100%	REST	57.71	57.71	100%

Tow pie charts and one bar chart are used in order to interpret the results of the fore-mentioned calculations. The Pie Chart is a way of visualizing data in percentages or proportions, especially when there are clusters or different categories of data. The first pie chart displays the percentages of the top ten terms, the most frequent terms and the percentage of the rest term as a total value. The "rest" percentage consists of small % of different terms that are not of vital importance to display it separately. The second pie chart shows the distribution of the top ten terms by excluding the proportion of the REST terms. Each slice of the pie chart represents a different term and its size corresponds to the value (numerical value) of the percentage. In general, the bigger the slice, the bigger the percentage of the term. The bar chart consists of rectangular bars whose lengths are proportional to the value of the percentage of the terms. The x-axis displays the top ten terms, whereas the y-axis shows the percentages of the terms to total terms.

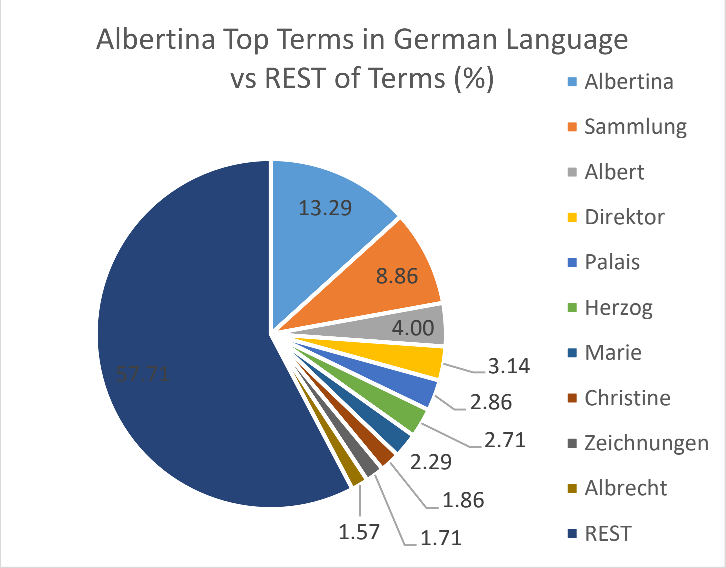


Chart 1 Pie Chart of Top 10 Terms vs Rest for Albertina Museum (POI) in DE

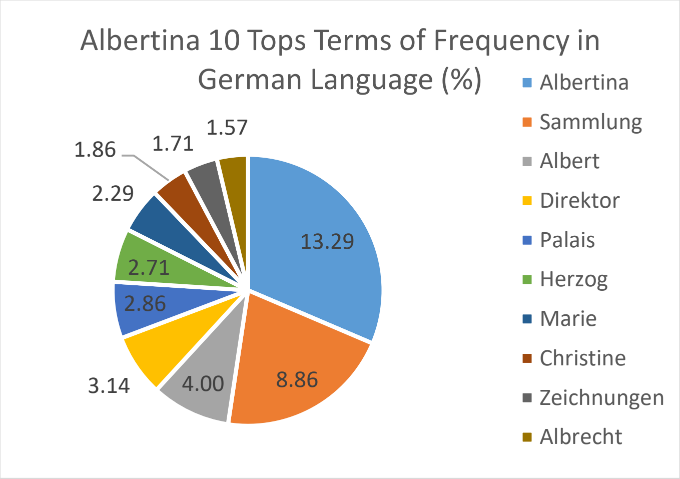


Chart 2 Pie Chart of Top 10 Terms for Albertina Museum (POI) in DE

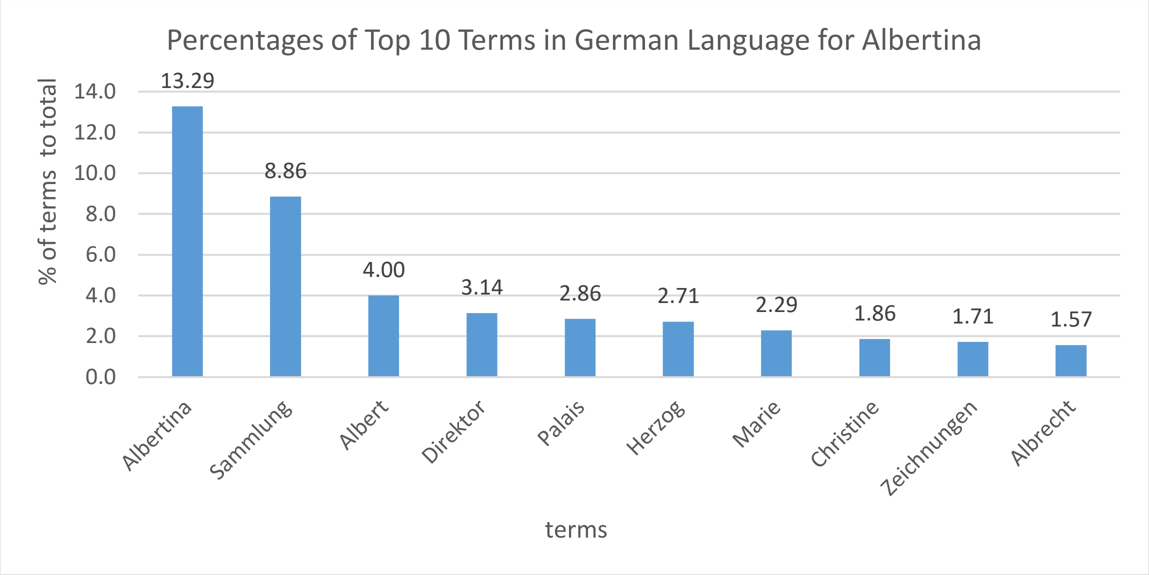


Chart 3 Bar Chart of Top 10 Terms for Albertina Museum (POI) in DE

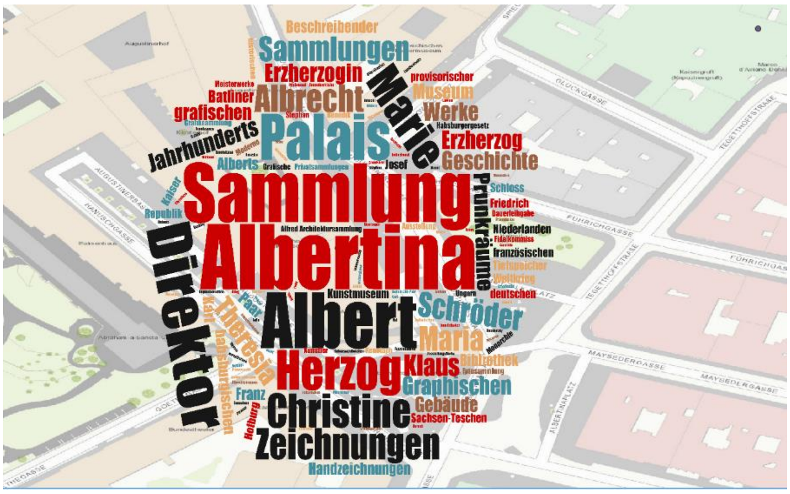


Figure 29 Wordcloud for Albertina Museum in DE

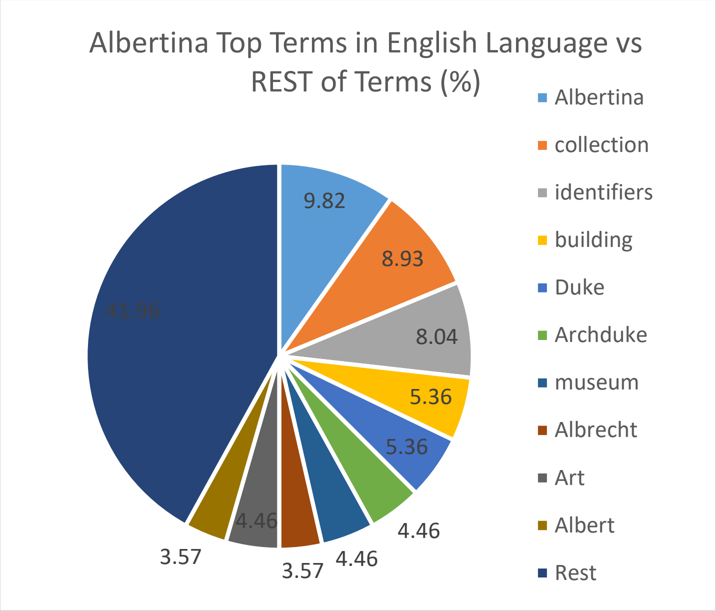


Chart 4 Pie Chart of Top 10 Terms vs Rest for Albertina Museum (POI) in EN

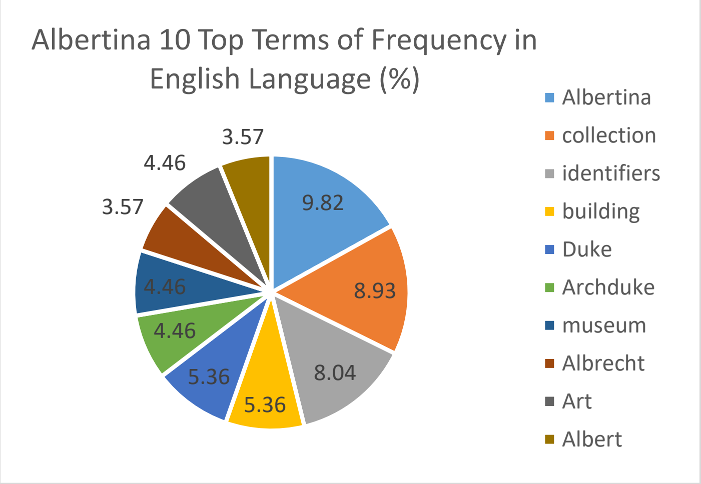


Chart 5 Pie Chart of Top 10 Terms for Albertina Museum (POI) in EN

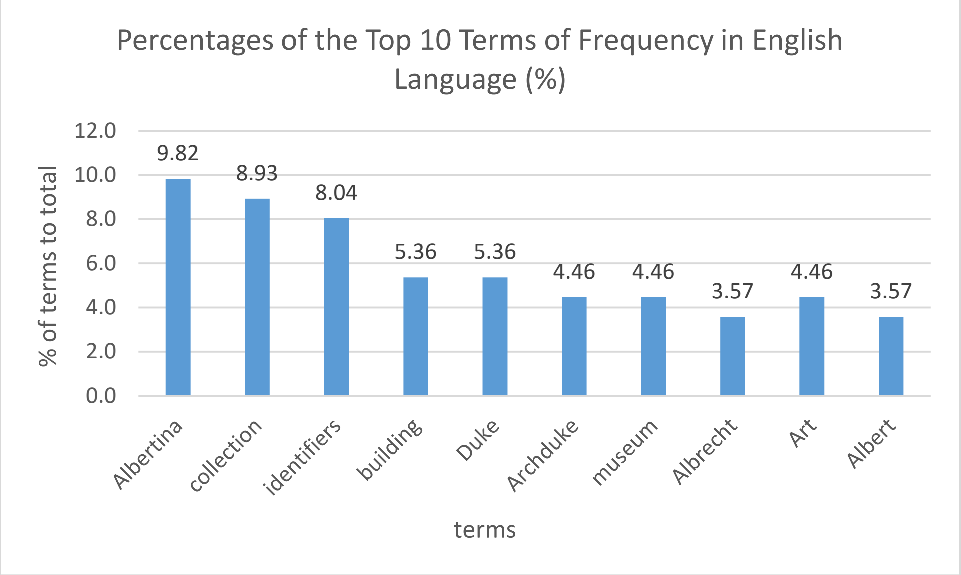


Chart 6 Bar Chart of Top 10 Terms for Albertina Museum (POI) in EN

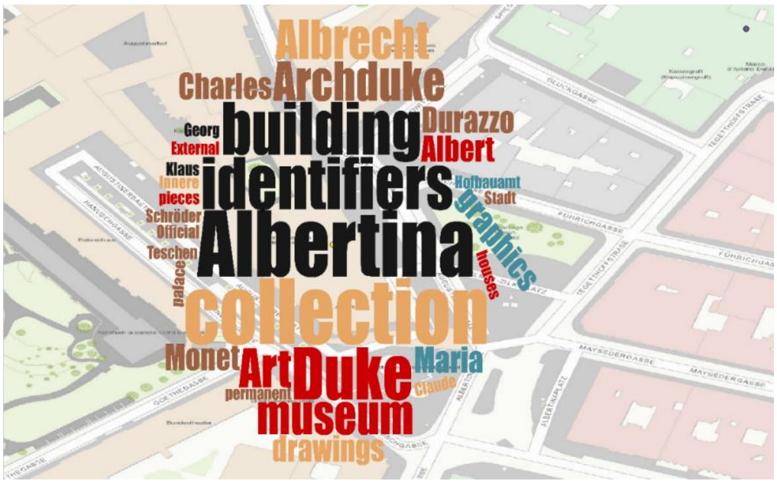


Figure 30 Wordcloud for Albertina Museum in EN

The results for the rest POI could be found in the Appendix. This section displays only the comparing tables of the Top Ten Terms in German and English version.

Table 31 Top 10 Terms for EN and DE based on Wikipedia Articles for Belvedere Palace

Top 10 Terms EN	fr(i) %	Sum(%)	Total	Top 10 of Terms DE	fr(i) %	Sum(%)	Total
Belvedere	12.75	47.98		Belvedere	13.14	43.69	
Palace	6.88			Schloss	7.99		
Lower Belvede	4.45			Eugen	4.44		
imperial	4.05			Franz	4.26		
Garden	3.85			Prinz	4.09		
Museum	3.85			Prinzen	2.13		

Upper Belvedere	3.64			Garten	2.13		
Prince	3.24			Oberen	1.95		
Galerie	2.83			Belvederes	1.78		
collection	2.43			Obere	1.78		
Rest	52.02		100%	Rest	56.31		100%

Table 32 Top 10 Terms for EN and DE based on Wikipedia Articles for Burgtheater

Top 10 Terms DE	fr(i) %	Sum (%)	Total	Top 10 Terms EN	fr(i) %	Sum (%)	Total
Burgtheater	17.58	44.24		Burgtheater	7.81	29.06	
Theater	6.06			theatre	6.25		
Joseph	3.64			Franz	2.50		
Kaiser	3.03			Peter	2.19		
Schauspieler	2.42			Keglevich	1.88		
Ballhaus	2.42			director	1.88		
Klimt	2.42			Joseph	1.88		
Ring	2.42			Theater	1.56		
Burg	2.42			Adolf	1.56		
Michaelerplatz	1.82			Max	1.56		
Rest	55.76		100%	Rest	70.94		100%

Table 33 Top 10 Terms for EN and DE based on Wikipedia Articles for Hofburg Palace

Top 10 Terms DE	fr(i) %	Sum(%)	Total	Top 10 Terms EN	fr(i) %	Sum(%)	Total
Hofburg	7.04	21.87		Wing	9.13	43.46	
Kaiser	2.70			Hofburg	6.14		
Burg	2.70			Imperial	5.35		
Franz	2.02			Castle	4.09		
Schloss	1.87			Joseph	3.46		
Joseph	1.27			Emperor	3.31		
Trakt	1.20			Palace	3.31		
Ferdinand	1.12			Hall	3.31		
Hofbibliothek	1.05			Square	2.68		
Heldenplatz	0.90			Court	2.68		
Rest	78.13		100%	Rest	56.54		100%

Table 34 Top 10 Terms for EN and DE based on Wikipedia Articles for Hundertwasserhaus

Top 10 Terms DE	fr(i) %	Sum(%)	Total	Top 10 Terms EN	fr(i) %	Sum(%)	Total
Hundertwasser	23.73	72.88		Hundertwasser	30.11	84.95	
Krawina	9.32			house	12.90		
Architektur	8.47			Krawina	9.68		

Hundertwasserhaus	7.63			architect	7.53		
Haus	6.78			architecture	5.38		
Hundertwasser-Haus	5.08			Hundertwasserhaus	4.30		
Architekten	3.39			Friedensreich	4.30		
Bauwerke	3.39			German	4.30		
Landstraße	2.54			architectural	3.23		
Architekt	2.54			Foundation	3.23		
Rest	27.12		100%	Rest	15.05		100%

Table 35 Top 10 Terms for EN and DE based on Wikipedia Articles for Kaisergruft

Top Ten Terms DE	**fr(i) %**	**Sum(%)**	**Total**	**Top 10 Terms EN**	**fr(i) %**	**Sum(%)**	**Total**
Kaiser	9.78	54.60		Emperor	7.13	42.62	
Erzherzog	7.55			Family	5.77		
Maria	6.78			Tree	5.74		
Erzherzogin	5.42			Maria	4.78		
Karl	4.84			Archduke	4.31		
Franz	4.45			buried	3.78		
Ferdinand	4.26			Franz	3.05		
Leopold	4.16			Ferdinand	2.75		
Kaiserin	3.97			Empress	2.72		
Gemahlin	3.39			Vault	2.59		
Rest	45.40		100%	Rest	57.38		100%

Table 36 Top 10 Terms for EN and DE based on Wikipedia Articles for Karlskirche

Top 10 Terms DE	**fr(i) %**	**Sum(%)**	**Total**	**Top 10 Terms EN**	**fr(i) %**	**Sum(%)**	**Total**
Karlskirche	10.27	34.86		church	7.38	34.84	
Karl	4.32			Charles	5.33		
Johann	4.05			Fischer	4.10		
Borromäus	2.97			Borromeo	3.69		
Kirche	2.70			Erlach	2.87		
Bass	2.70			Roman	2.87		
Kuppel	2.43			Karlskirche	2.46		
Fischer	2.16			buildings	2.05		
Stadt	1.62			Catholic	2.05		
Bau	1.62			Church	2.05		
Rest	65.14		100%	Rest	65.16		100%

Table 37 Top 10 Terms for EN and DE based on Wikipedia Articles for Karlsplatz

Top 10 Terms DE	**fr(i) %**	**Sum(%)**	**Total**	**Top 10 Terms EN**	**fr(i) %**	**Sum(%)**	**Total**
Karlsplatz	**12.54**	**30.75**		Karlsplatz	14.29	37.36	

Platz	**2.99**			Museum	4.95		
Stadt	**2.69**			Karlskirche	2.75		
Resselpark	**2.39**			building	2.75		
Karlskirche	**2.09**			design	2.20		
Museum	**2.09**			square	2.20		
Künstlerhaus	**1.49**			front	2.20		
Kunstplatz	**1.49**			Park	2.20		
Kärntner	**1.49**			area	2.20		
Projekt	**1.49**			Operngasse	1.65		
Rest	**69.25**		**100%**	Rest	62.64		100%

Table 38 Top 10 Terms for EN and DE based on Wikipedia Articles for Kunsthistorisches Museum

Top 10 Terms DE	**fr(i) %**	**Sum(%)**	**Total**	**Top 10 Terms EN**	**fr(i) %**	**Sum(%)**	**Total**
Museum	**7.77**	**29.38**		Museum	13.94	52.12	
Sammlungen	**3.25**			museums	9.09		
Sammlung	**3.25**			Art	6.67		
Kunsthistorisches	**2.82**			Kunsthistorisches	5.45		
Museums	**2.82**			collections	3.64		
Kunsthistorischen	**2.68**			Collection	3.64		
Direktor	**2.40**			building	2.42		
Kunsthistorische	**1.84**			Portrait	2.42		
Schloss	**1.27**			Madonna	2.42		
Museen	**1.27**			Cellini	2.42		
Rest	**70.62**		**100%**	Rest	45.45		100%

Table 39 Top 10 Terms for EN and DE based on Wikipedia Articles for Maria am Gestade

Top 10 Terms DE	**fr(i) %**	**Sum(%)**	**Total**	**Top 10 Terms EN**	**fr(i) %**	**Sum(%)**	**Total**
Kirche	**16.87**	**53.01**		church	13.08	55.38	
Maria	**10.84**			Maria	7.69		
Gestade	**9.04**			Gestade	6.92		
Langhaus	**3.01**			Gothic	6.15		
Stadt	**2.41**			choir	5.38		
Orgel	**2.41**			nave	3.85		
Chor	**2.41**			Catholic	3.08		
Lage	**2.41**			Church	3.08		
Superoktavkoppel	**1.81**			Danube	3.08		
Geschichte	**1.81**			Art	3.08		
Rest	**46.99**		**100%**	Rest	44.62		100%

Table 40 Top 10 Terms for EN and DE based on Wikipedia Articles for Mariahilf Kirche

Top 10 Terms DE	fr(i) %	Sum(%)	Total	Top 10 Terms EN	fr(i) %	Sum(%)	Total
Kirche	9.46	41.22		Mariahilf	13.10	66.67	
Mariahilfer	7.43			Church	19.05		
Heiligen	5.41			Catholic	5.95		
Mariahilf	4.73			Baroque	5.95		
Geyling	2.70			Roman	4.76		
Heilige	2.70			Paul	4.76		
Joseph	2.70			building	3.57		
Glasmalereien	2.03			Johann	3.57		
Glasmalerei	2.03			Franz	3.57		
Pfarrkirche	2.03			decoration	2.38		
Rest	58.78		100%	Rest	33.33		100%

Table 41 Top 10 Terms for EN and DE based on Wikipedia Articles for Musikverein

Top 10 Terms DE	fr(i) %	Sum(%)	Total	Top 10 Terms EN	fr(i) %	Sum(%)	Total
Saal	6.48	36.44		Hall	9.91	51.35	
Musikverein	5.67			Musikverein	7.21		
Orgel	4.05			seats	6.31		
Rieger-Orgel	3.64			halls	6.31		
Gesellschaft	3.64			Saal	4.50		
Musikfreunde	3.24			acoustics	3.60		
Register	3.24			Symphony	3.60		
Säle	2.43			concert	3.60		
Akustik	2.02			venues	3.60		
Hansen	2.02			Musikvereinssaal	2.70		
Rest	63.56		100%	Rest	48.65		100%

Table 42 Top 10 Terms for EN and DE based on Wikipedia Articles for National Library

Top 10 Terms DE	fr(i) %	Sum(%)	Total	Top 10 Terms EN	fr(i) %	Sum(%)	Total
Nationalbibliothek	**8.06**	**31.51**		Library	14.74	42.88	
Bibliothek	**4.34**			collection	8.54		
Sammlung	**4.34**			National	5.03		
Hofbibliothek	**3.85**			books	3.18		
Bücher	**2.98**			Museum	2.35		
Handschriften	**1.99**			Imperial	2.18		
Werke	**1.74**			Nationalbibliothek	2.01		
Bestände	**1.49**			Papyrus	1.84		
kaiserlichen	**1.36**			Department	1.51		
Kaiser	**1.36**			World	1.51		

Rest	**68.49**		**100%**	Rest	57.12		100%

Table 43 Top 10 Terms for EN and DE based on Wikipedia Articles for Parliament

Top 10 Terms DE	fr(i) %	Sum(%)	Total	Top 10 Terms EN	fr(i) %	Sum(%)	Total
Parlament	12.76	45.27		building	7.80	40.71	
Nationalrat	6.58			Parliament	5.67		
Bundesrat	5.35			Hall	5.11		
Nationalversammlung	3.29			Council	4.96		
Parlamentsgebäude	3.29			Chamber	3.97		
Senat	3.29			House	3.97		
Bundesversammlung	2.88			National	2.70		
Republik	2.88			Federal	2.27		
Abgeordnetenhaus	2.47			Pillars	2.13		
Abgerufen	2.47			marble	2.13		
Rest	54.73		100%	Rest	59.29		100%

Table 44 Top 10 Terms for EN and DE based on Wikipedia Articles for Peterskirche

Top 10 Terms DE	fr(i) %	Sum(%)	Total	Top 10 Terms EN	fr(i) %	Sum(%)	Total
Peterskirche	14.07	48.15		church	17	50.5	
Kirche	13.33			Peterskirche	6.5		
Johann	3.70			Opus	5		
Orgel	2.96			Dei	5		
Stadt	2.96			Holy	3.5		
Dreifaltigkeit	2.22			University	3		
Kirchengebäude	2.22			Baroque	3		
Taufstein	2.22			building	2.5		
Prinzipal	2.22			Catholic	2.5		
Kanzel	2.22			Roman	2.5		
Rest	51.85		100%	Rest	49.5		100%

Table 45 Top 10 Terms for EN and DE based on Wikipedia Articles for Wiener Prater

Top 10 Terms DE	fr(i) %	Sum(%)	Total	Top 10 Terms EN	fr(i) %	Sum(%)	Total
Prater	18.43	40.54		Prater	30.59	76.47	
Hauptallee	4.36			Castle	7.06		
Praterstern	3.35			Palace	7.06		
Wurstelprater	2.68			park	5.88		
Lusthaus	2.68			Wurstelprater	4.71		
Donau	2.01			railway	4.71		
Weltausstellung	1.84			Emperor	4.71		
Stuwer	1.84			area	4.71		

Rotunde	1.68			Liliputbahn	3.53		
Kaiser	1.68			Royal	3.53		
Rest	59.46		100%	Rest	23.53		100%

Table 46 Top 10 Terms for EN and DE based on Wikipedia Articles for Rathaus

Top 10 Terms DE	**fr(i) %**	**Sum(%)**	**Total**	**Top 10 Terms EN**	**fr(i) %**	**Sum(%)**	**Total**
Rathaus	12.06	38.41		Hall	22.95	62.30	
Rathauses	5.40			City	14.75		
Stadt	4.44			Rathaus	4.10		
Rathausplatz	3.17			Floor	4.10		
Stock	3.17			Rathausmann	3.28		
Festsaal	2.54			ballroom	3.28		
Veranstaltungen	1.90			Tower	3.28		
Bürgermeister	1.90			government	2.46		
Rathauspark	1.90			buildings	2.46		
Ringstraße	1.90			Rathausplatz	1.64		
Rest	61.59		100%	Rest	37.70		100%

Table 47 Top 10 Terms for EN and DE based on Wikipedia Articles for Schönbrunn Palace

Top 10 Terms DE	**fr(i) %**	**Sum(%)**	**Total**	**Top 10 Terms EN**	**fr(i) %**	**Sum(%)**	**Total**
Schloss	14.06	46.06		Schönbrunn	15.90	48.25	
Schönbrunn	11.30			Palace	11.86		
Maria	4.08			gardens	4.58		
Kaiser	3.42			Gloriette	3.23		
Franz	2.75			Roman	2.96		
Gloriette	2.28			garden	2.43		
Raum	2.18			Schloss	2.16		
Kultur	1.99			area	1.89		
Garten	1.99			Baroque	1.62		
Joseph	1.99			Castle	1.62		
Rest	53.94		100%	Rest	51.75		100%

Table 48 Top 10 Terms for EN and DE based on Wikipedia Articles for Vienna State Opera

Top 10 Terms DE	**fr(i) %**	**Sum(%)**	**Total**	**Top 10 Terms EN**	**fr(i) %**	**Sum(%)**	**Total**
Staatsoper	**8.88**	**34.91**		Opera	14.68	37.32	
Musik	**4.73**			State	5.45		
Franz	**4.62**			Staatsoper	2.94		
Oper	**4.14**			Franz	2.73		
Libretto	**3.20**			building	2.52		
Richard	**2.37**			performance	2.10		

Karl	2.01			Orchestra	1.89		
Dirigent	1.78			Mahler	1.68		
Rudolf	1.66			Josef	1.68		
Opernhaus	1.54			Karl	1.68		
Rest	65.09		100%	Rest	62.68		100%

Table 49 Top 10 Terms for EN and DE based on Wikipedia Articles for St.Stephen's Cathedral

Top 10 Terms DE	fr(i) %	Sum(%)	Total	Top 10 Terms EN	fr(i) %	Sum(%)	Total
Dom	9.91	37.25		Cathedral	10.13	39.80	
Stephansdom	7.41			Stephen's	4.78		
Stephan	5.00			church	4.78		
Glocke	2.89			German	3.91		
Stadt	2.89			tower	3.76		
gotischen	2.02			Stephansdom	3.47		
Domkirche	1.92			bell	3.18		
Stephansdoms	1.73			Emperor	2.17		
Nordturm	1.73			Roman	1.88		
Chor	1.73			Chapel	1.74		
Rest	62.75		100%	Rest	60.20		100%

Table 50 Top 10 Terms for EN and DE based on Wikipedia Articles for University of Vienna

Top 10 Terms DE	fr(i) %	Sum(%)	Total	Top 10 Terms EN	fr(i) %	Sum(%)	Total
Universität	18.35	42.89		University	24.76	49.05	
Fakultät	6.72			students	3.65		
Karl	3.62			Library	3.65		
Universitäten	2.58			Faculty	3.33		
Hauptgebäude	2.07			Sciences	3.17		
Studenten	2.07			Universities	2.54		
Kurt	2.07			building	2.22		
Fakultäten	1.81			academic	2.06		
University	1.81			Faculties	1.90		
englisch	1.81			Medical	1.75		
Rest	57.11		100%	Rest	50.95		100%

Table 51 Top 10 Terms for EN and DE based on Wikipedia Articles for Volksgarten

Top 10 Terms DE	fr(i) %	Sum(%)	Total	Top 10 Terms EN	fr(i) %	Sum(%)	Total
Volksgarten	15.34	47.73		Monument	8.74	57.28	
Park	5.68			Garden	8.74		
Theseustempel	4.55			park	8.74		
Franz	4.55			Volksgarten	7.77		

Garten	3.41			Empress	5.83		
Stadt	3.41			Elizabeth	3.88		
Rosengarten	2.84			Franz	3.88		
Ringstraße	2.84			rose	3.88		
Platane	2.84			Grillparzer	2.91		
Grillparzerdenkmal	2.27			Rosegarten	2.91		
Rest	52.27		100%	Rest	42.72		100%

Table 52 Top 10 Terms for EN and DE based on Wikipedia Articles for Naturhistorisches Museum

Top 10 Terms DE	fr(i) %	Sum(%)	Total	Top 10 Terms EN	fr(i) %	Sum(%)	Total
Abteilung	10.96	57.11		Museum	17.35	65.30	
Museum	9.79			History	10.73		
Sammlung	8.39			Natural	9.15		
Naturhistorischen	6.99			Hall	7.89		
Museums	5.13			collection	5.36		
Naturhistorisches	4.20			expedition	3.79		
NHM	3.26			Franz	3.15		
Zoologische	3.03			exhibition	2.84		
Saal	2.80			minerals	2.52		
Franz	2.56			objects	2.52		
Rest	42.89		100%	Rest	34.70		100%

5. Discussion and Future Work

5.1 Discussion

The topic of the current Master Thesis is: Comparing Social Media Topics of Interest associated with Points of Interest according to user's origin. The central theme of this master thesis is to extract and visualize topics of interest associated with places based on social media sources. Checking how these topics of interest related to user's place of origin is an additional factor of the current research. The research question of the current master thesis is the following:

- ➢ How similar are the terms extracted from Geosocial Media (Twitter and Flickr) and Wikipedia articles when associated to the Points of Interest?

The Research Question and the following sub-question could be answered by comparing terms from Wikipedia and terms from Flickr. The similarity is defined by mentioning the common terms and the percentages of their appearance based on a top ten list.

- ➢ How similar are the frequencies of each term according to user's place of origin?

According to limitation of the existing data, Twitter is the only Geosocial media data set, which contains information of regions of origin. Those data sets contain the information of the place of origin (in case of Twitter data) and the language (in case of Wikipedia articles). As it mentioned in chapter of 3.2 the word "language" such as German Language or English Language, or the use of the word "users" such as German users or English users refers to the region of origin. The POIs and the clusters that are related with them are the following: (mentioned as relations).These comparisons derived from the tables of the top ten terms that could be found in the part of results (chapter 4.2.3 and 4.3).

1st relation: Schonbrunn Palace

For German version: The common terms are Schloss, Schloß, Schönbrunn. Wikipedia article has a higher percentage of appearance of the term Schloss with 14.06% than Twitter cluster with 1.91%. The term "Schönbrunn" has a higher percentage in Wikipedia than in Twitter with 11.30% in Wikipedia versus 8.06% in Twitter. Users of the Twitter use the term "Tiergarten" with a value of 0.96% to total terms. The same term does not exist in the top ten list of Wikipedia articles. Also it is worthy to point out the Twitter data contain terms in English, such as the term "Gardens" with 0.96% to total terms instead of the use of the term "Garten" that appears in Wikipedia article with 1.99% to total terms.

For the English version: The common terms are Schönbrunn, Palace, garden, Schloss. The common terms appear with higher percentages in data derived from Wikipedia articles than data derived from Twitter data. Similar to the behavior of the German users, the Twitter English users post terms in German. These terms here are: Schönbrunn with a 1.12% to total terms, Schloss with 0.37% to total terms and wienerphilharmoniker with 0.19% to total terms.

2nd relation: Wiener Prater

For German version: The common terms are: Prater, Pratestern and Wurstelprater. Wikipedia article has a higher percentage of appearance of the term Prater with 18.43% than Twitter cluster with 4.33%. The term "Pratestern" frequency has a value of 0.87% to total terms in Twitter and a value of 3.35% to total terms in Wikipedia articles.

For the English version: The common term is only the term "Prater" with 3.05% to total terms in Twitter versus 30.59% to total terms in Wikipedia. In Twitter exist some terms in German alphabet such as: "Pratestern", "Wurstelprater" and "wienerprater" with 1.22%, 0.20% and 0.20% to total terms correspondingly.

3rd relation: Mariahilf Kirche

For German version: The common term is "Mariahilfer". Wikipedia term "Mariahilfer" has a higher percentage of appearance with 7.43% to total terms than Twitter "Mariafilfer" term with 2.82% to total terms. Also, the Wikipedia top ten terms are more related to the church than the terms of the Twitter that related to the street of Mariafilfstasse. In this category belong terms such as "Mariahilferstrasse" and or "shopping" or "spazieregang" have percentages of 1.13%, 0.28% and 0.28% to total terms. Also, here the German users use terms in English such as the term “shopping” with 0.28% to total terms. Terms such as “Heiligen” or “Heilige” occur only in the top ten list of Wikipedia with 5.41% to total terms and 2.70% to total terms correspondingly.

For the English version: There is no coincidence between the top ten terms of Wikipedia list and the top ten list of Twitter cluster. The terms derived from Wikipedia article refer to the church and its architecture, whereas the terms of the Twitter list refer to the activities and more general information. Wikipedia most common term is “Marifilf” with 13.10% to total terms, whereas Twitter most common is “café” with 0.81% to total terms.

4th relation: Rathaus

For German version: The common terms are Rathaus, Rathausplatz. Wikipedia article has a higher percentage of appearance of the term Rathaus with 12.06% than Twitter cluster with 1.35% to total terms. The same occurred for the second common term, “Rathausplatz” that has 3.17% to total terms in Wikipedia versus 0.86% to total terms in Twitter.

For the English version: The common terms are Rathaus and Rathausplatz. So, the common terms are the same with the German version. English uses use the terms of Rathaus and Rathausplatz in German language and not in English version. The common terms appear with higher percentages in Wikipedia than in Twitter list. An example of this is the term “Rathaus” with a value of 4.10% in Wikipedia list versus 1.27% in Twitter list.

5th relation: Naturhistorisches and Kunshistorisches Museum

For German version: The common terms are Museum and Naturhistorisches. Wikipedia article has a higher percentage of appearance of the term Museum with 7.77% to total terms and 9.79% to total terms than Twitter cluster with 2.57% to total terms. The term “Kunsthistorisches” is not in the top ten list of Twitter data, whereas in the Wikipedia article has 2.68% to total terms.

For the English version: The common terms are Museum and exhibition. The Museum 13.94% and 17.35% to total terms whereas the Twitter percentage is 1.70% to total terms. The term “exhibition” occurs in the 9th place of the top ten terms in Twitter list with 0.13% to total terms, whereas in the Wikipedia article comes to the 8th place of the top ten list with 2.84% to total terms.

6th relation: Hofburg

For German version: The common term is Hofburg. The term “Hofburg” appear with similar percentage to total terms based on the top ten list of Wikipedia and on the top ten list of Twitter (7.04% to total terms for wikipedia list and 8.18% to total terms for Twitter list). The “Schloss” appear with 1.87% in Wikipedia whereas the term “palace” is used for German users instead of Schloss with 1.75% to total terms. Also terms such as “museum” and “imperial” share the same percentage of appearance with 0.58% to total terms.

For the English version: The common terms are Hofburg, Palace and Imperial. Wikipedia articles have higher percentages of appearance of the common terms in comparison with the Twitter. The “Hofburg” has 6.14% to total terms in the Wikipedia list, whereas has 1.87% to total terms in Twitter’s list. Palace appears with 3.31% in Wikipedia list, whereas 0.75% in the Twitter top ten list. The term “Imperial” is the third used term based on Wikipedia list with 5.35% to total terms and is the 5th used term on Twitter’s list with 0.19% to total terms.

7th relation: Vienna State Opera

For German version: The common term is Staatsoper. The “Staatsoper” has a higher percentages of appearance (8.88%) in Wikipedia than in Twitter (0.92% to total terms). Wikipedia top ten list based more on terms related with music such as “Musik”, “Libretto” or “Dirigent” with 4.73%, 4.20% and 1.78% to total terms respectively. On the other hand terms from Twitter top ten list related to the surrounding environment of the Vienna State Opera, such as “hotelsacher”, “sacher” and “café” with 0.73%, 1.11% and 0.74% to total terms correspondingly.

For the English version: The common terms are Opera and Staatsoper. Wikipedia articles have higher percentages of appearance of the common terms in comparison with Twitter terms. “Opera” is the first place of the top ten Wikipedia list with 14.68% to total terms, whereas is in the second place of the top ten Twitter list with 0.59% to total terms. Staatsoper has a value of 2.94% to total terms for Wikipedia and 0.24% for Twitter. The terms of the Twitter list are more related to the surrounding environment of the Vienna State Opera. The examples of terms are similar to the German terms: “café” with 0.47% to total terms and “hotelsacher” with 0.12% to total terms. Wikipedia terms are related more to music. Examples are: “performance” with 2.10%, “Orchestra” with 1.89% and “Mahler” with 1.68% to total terms.

8th relation: Stephansdom

For German version: The common term is Stadt. The “Stadt” has a higher percentage of appearance (2.89%) in Wikipedia than in Twitter (0.24% to total terms). Terms related to the cathedral are “Stephansdom” with 7.41% to total terms and comes to the second place of the top ten list and “Stephan” as the third most occurred term in Wikipedia list with 5% to total terms. Terms related to the cathedral from Twitter list are “Stephens” with 4.78% to total terms (the second place of the list) and “Stephensplatz” in the third place of the top ten list with 0.12% to total terms. Terms in English language used by German user. Typical examples of them are the terms “ticket” and “heritage” with 0.12% to total terms.

For the English version: The common terms are Cathedral, Stephen and Stephansdom. Wikipedia articles have higher percentages of appearance of the common terms in comparison with Twitter terms. The term “Cathedral” is the first one in the Wikipedia list with 10.13% whereas in the Twitter list is the second with 0.59% to total terms. The term “Stephens” appear with 4.78% to total terms for Wikipedia and 0.65% to total terms for Twitter gaining the first place in the top ten list. The English version includes the term of “Stephansdom” in German language with 3.47% to total terms for Wikipedia and 0.12% to total terms for Twitter (3rd top ten term).

To sum up, the similarity was defined by mentioning the common terms and the percentages of their appearance based on a top ten list. From the discussed comparison, it is derived that there are eight relations between clusters of Twitter and Points of Interest from Wikipedia Articles.

- Are the same terms present in all the Geosocial Media (Twitter and Flickr)?

Results of the comparison of data from Flickr and Twitter could answer the aforementioned sub-question. Due to data limitations, this comparison does not take into account the factor of place of origin. The POIs and the clusters that are related with them are the following:

(mentioned as relations).These comparisons derived from the tables of the top ten terms that could be found in the part of results (chapter 4.1 and in the appendix).

1st relation: Schonbrunn Palace

The common terms are “Schönbrunn”, “Palace”, “Schloss”, “tiergarten”, “garden”. The first place in the top ten list is “Schönbrunn” in both data sets with 4.71% to total terms for Twitter and 3.61% to total terms for Flickr. The percentages of the top ten comparing lists are in the same ranges (from 0 to 5%). Twitter’s top ten list includes two different types for expressing the palace (“Schloss” with 1.40% and the term “Schloß” with 2.07% to total terms). The term of “Schloss” has 0.69% to total terms according to Flickr top ten list. The term “concert” appear in the 6th place of the Twitter’s list (0.10% to total terms) and not in the list derived from Flickr data. The term “gloriette” comes in the 9th place of the Flickr top ten list with 0.61% to total terms and does not appear in the Twitter top ten list.

2nd relation: Wiener Prater

The common terms are “Prater”, “Riesenrad”, “Pratestern”, “ferriswheel”, “park” and “wurstelprater”. The term “prater” is the most frequent term for both top ten lists (2.55% to total terms for Twitter and 5.03% to total terms for Flickr). The percentages of the top ten comparing lists are in the same ranges (from 0 to 5%). There is a variety of terms referring to the Riesenrad for both data sets. Examples of that are “Riesenrad”, “wheel” and “wienerriesenrad” for Flickr with 1.58%, 0.48% and 0.28% to total terms respectively. The Twitter top ten list contains the terms of “Riesenrad” and “Rode” in the 2nd place and 7th place of the list (0.72% and 0.07% to total terms correspondingly). The term “park” is in the 5th place of the top ten list for both data sets (0.13% to total terms for Twitter and 0.47% to total terms fort Flickr).

3rd relation: Mariahilf kirche

The common terms are “mariafilf” with 0.06% to total terms for Twitter data and 0.98% to total terms for Flickr data. Similar terms to the most common term are “Mariafilfer” (first in the top ten list of Twitter) and “mariahilferstrasse”, “mariahilfestase” (2nd and 3rd place of the Flickr top ten list with 0.37% and 0.63% to total terms). It is worthy to point out that Flickr top ten list contains terms relates to the church such as “church” with 0.32%, “mariahilferkirche” with 0.24%, “kirche” with 0.23% to total terms, whereas those terms are not included in the Twitter’s top ten list. The term related to architecture appears in two different forms such as “architecture” and “architektur” of the Flickr’s list. The top ten list of Twitter data includes more terms related to the external activities such as “cinema”, “dancehall” and “café” than the list of Flickr

4th relation: Rathaus

The common terms are “Rathaus”, “Rathausplatz”, “hall”, “burgtheater” and “rathauspark”. The percentages belong to the same range of (0-3%). Flickr list includes terms related to “Rathaus” with the following terms: “cityhall” with 0.50% to total terms and “townhall” with 0.34% to total terms and “hall” with 0.23% to total terms. The first term in the top ten list for Twitter is “Universität” with 0.96% to total terms, whereas the first term of Flickr list is “rathaus” with 2.24% to total terms.

5th relation: Naturhistorisches and Kunsthistorisches Museum

The common terms are “museum” and “art” and “maria-theresienplatz”.The term “museum” is in the second and first place of the top ten list for Twitter and Flickr respectively. The term “art” has a value of 0.28% to total terms for Twitter and 0.71% to total terms for Flickr. Another common term to the term “art” is the term “kunst” that comes 4th in the Twitters top ten list (0.33% to total terms). The top test list includes terms that are relevant to Museumsquartier such as “Museumsquartier” whereas the top ten list of Flickr does not include those kind of terms.

6th relation: Hofburg Palace

The common terms are “Hofburg” and “Palace” and “imperial” and “horses”. The term “Hofburg” is the first term in the both top ten lists (3.18% to total terms for Twitter and 2.57% to total terms for Flickr). The term “Palace” is one of the most frequent terms for both sources (the second place for Twitter with 1.03% to total terms and the 4th place for Flickr with 0.72% to total terms). The percentages belong to the same range of (0-4%). Twitter list contains terms related to ball such as “ball” with 0.26% to total terms or “hofburgsilvesterball” with 0.09% to total terms. Flickr list contains terms related to architectural crafts such as “architecture”, “monument” and “statue”.

7th relation: Vienna State Opera

The common terms are “Staatsoper” and “Oper” and “albertina” and “opera” and “museum”. The term “Staatsoper” is the first term in Twitter’s top then list and third in Flickr’s list (with 0.57% to total terms and 0.95% to total terms correspondingly). The Twitter top ten list includes terms related to the surrounding environment such as the term “café” and “Sacher” (third and second place of the list respectively). On the other hand Flickr list includes terms related to the buildings of the surrounding environment. These terms are “building”, “operahouse” and “museum” with 0.32%, 0.30% and 0.51% to total terms respectively

8th relation: Stephansdom

The common terms are “Cathedral” and “Stephansdom” and “architecture” and “Stephansplatz”. In both top ten lists “Cathedral” is the second most frequent term with 0.41% to total terms for Twitter and 1.39% to total terms for Flickr. The term “Stephansdom” takes the first place in Flickr list with 2.33% to total terms, whereas the term “Stephens” takes the first place in Twitter’s list with 0.45% to total terms.

To sum up, the similarity was defined by mentioning the common terms and the percentages of their appearance based on a top ten list. From the discussed comparison, it is derived that there are eight relations between clusters of Twitter and cluster of Flickr.

The last sub-question of the thesis is:

- What proportion of POI present in Geosocial Media (GSM)?

The analysis of semantics of Flickr data shows that 15 clusters are related to specific Points of Interest (POI). The total amount of clusters is 35 derived by the execution of HDBCAN algorithm. So, the presence of POI in Flickr is 15/35 and the percentage is 42.85%.The analysis of the semantics of Twitter divides into three categories:

- The total amount of clusters is 3 derived by the execution of HDBCAN 300 algorithm. The analysis of these data shows that 2 clusters are related to specific Points of Interest (POI). So, the presence of POI in Flickr is 2/3 and the percentage is 67%.
- The total amount of clusters is 9 derived by the execution of HDBCAN 200 algorithm. The analysis of these data shows that 3 clusters are related to specific Points of Interest (POI). So, the presence of POI in Flickr is 3/9 and the percentage is 33%.
- The total amount of clusters is 23 derived by the execution of HDBCAN 100 algorithm. The analysis of these data shows that 8 clusters are related to specific Points of Interest (POI). So, the presence of POI in Flickr is 8/23 and the percentage is 35%.

To sum up, the current thesis answered the Research Question and the sub-questions. It is worthy to mention the following limitations and acceptances

- In the current master thesis the use of the word "term" refers to any kind of textual source. So texts, hashtags, words, tags (Flickr) or tweets (Twitter) are considered as terms. The extraction of the terms from Wikipedia Articles, Twitter and Flickr could be implemented during the analysis of the semantics.
- Also, it is worthy to point out that the use of the word "language" like German Language or English Language, or the use of the word "users" like German users or English users refers to the region of origin.
- It is worthy to point out that Volkstheater/People's Garden, Wiener Prater/Prater Park and Karlsplatz/Charle's square are considered as Points of Interested because of their high importance according to the ranking lists.
- The selection of POI (Points of Interest) based on authoritative and commercial data sources, list provided by Vienna Tourist Board
- Only Twitter contains information of regions of origin
- Wikipedia articles in German and English language
- The selection of the location of the POI and the size of its surrounded area
- Unpredictable errors of creating the top ten list due to the process of editing the wordlist are related to the occurrence of deficits of analysis of the semantics, case that could also influence the quality of the percentages of the terms

5.2 Future Work

This sub-chapter gives an overview of the possible future work. Further research is needed in order to examine the influence of people from different places of origin. It would be interesting to compare top ten terms from different places of origin. Different places of origin could be clustered into groups and new researches could examine the differences between people from North, South or Central Europe. Another indicator of clustering the group of people could be the age. That gives the opportunity to future researchers to study the differences of terms according to user's age.

Further studies should take into account more Geosocial Media such as Instagram or Foursquare or Panoramio and observe the impact of terms derived from this new source. That give researches the possibility to make comparisons between the most frequent terms based on different sources.

Further research should take into account the perspective of time and examine if the same terms vary over time. This could provide valuable information about the occurrence of the semantics in terms of temporal analysis.

Additional work is needed in order to include more Points of Interest and examine their influence with the terms. Future researches should establish a new approach in case of overlapping. That occurs when two Points of Interest are too close (in terms of adjacency) and the area around them sharing the same terms.

5.3 Conclusion

The comparison of the results for Twitter and Wikipedia data shows that the percentages of terms based on Wikipedia Articles are higher than the percentages of terms based on Twitter data.

Another general outcome of the comparison is the fact that the ten top list based on German perspective includes English words. A typical example of this is the use of the term “Gardens” instead of “Garten”. Other similar examples are “ticket” and “heritage”. It is also worthy to mention that for the majority of the terms that refers to a name of a Point of Interest, English users use the term in German language. Some characteristic examples are the following terms: “Schloss” or “Schloß” instead of “Palace” and “Schönbrunn” instead of Schoenbrunn.

The comparison between Twitter and Wikipedia terms shows that in most cases the number of the common terms based on English top ten lists is bigger than the number of common terms derived from German top ten lists. The only exceptions are Mariahilfe kirche that has no common terms in English language and Wien Prater that has more common terms is German than in English version.

The comparison between Flickr and Twitter data set terms based only on terms. It was impossible to take into account the place of origin because there is not that type of information in the examined data set. In most cases the common terms between Flickr and Twitter top ten lists are around 4 or 5 words. That means that half of the terms of the list are the same in both top ten lists. The only exception is the case of Mariahilf Kirche that has only one common term in the examined top ten lists. It is worthy to point out that the range of the percentages for both data set is similar. For example, the Twitter and Flickr range of percentages is from 0 to 5 (%). Both top ten lists contain terms that has the same meaning but in different language. Examples of these occurrences are: “Riesenrad”, “wheel”, “Rode” and “wienerriesenrad” or “Schloss”, “Schloß” and “Palace”. Another similarity between these two different type of data sources is that the in their top ten list occurred terms related to activities such as “shopping”, “café”, “cinema”.

6. References

Dunkel, A. (2015). Visualizing the perceived environment using crowdsourced photo geodata. *Landscape and Urban Planning*, *142*, 173–186. https://doi.org/10.1016/j.landurbplan.2015.02.022

Gao, S., Janowicz, K., Montello, D. R., Hu, Y., Yang, J. A., McKenzie, G., … Yan, B. (2017). A data-synthesis-driven method for detecting and extracting vague cognitive regions. *International Journal of Geographical Information Science*, *31*(6), 1245–1271. https://doi.org/10.1080/13658816.2016.1273357

Grothe, C., & Schaab, J. (2009). Automated footprint generation from geotags with kernel density estimation and support vector machines. *Spatial Cognition and Computation*, *9*(3), 195–211. https://doi.org/10.1080/13875860903118307

Hollenstein, L., & Purves, R. (2010). Exploring place through user-generated content: Using Flickr to describe city cores. *Journal of Spatial Information Science*, *1*(1), 21–48. https://doi.org/10.5311/josis.2010.1.3

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems*, *54*, 240–254. https://doi.org/10.1016/j.compenvurbsys.2015.09.001

Hu, Y., & Janowicz, K. (2018). *An empirical study on the names of points of interest and their changes with geographic distance*. (GIScience). Retrieved from http://arxiv.org/abs/1806.08040

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

Kovacs-Györi, A., Ristea, A., Kolcsar, R., Resch, B., Crivellari, A., & Blaschke, T. (2018). Beyond Spatial Proximity—Classifying Parks and Their Visitors in London Based on Spatiotemporal and Sentiment Analysis of Twitter Data. *ISPRS International Journal of Geo-Information*, *7*(9), 378. https://doi.org/10.3390/ijgi7090378

Li, D., & Zhou, X. (2017). “Leave Your Footprints in My Words”—A Georeferenced Word-Cloud Approach. *Environment and Planning A*, *49*(3), 489–492. https://doi.org/10.1177/0308518X16662273

Verstockt, S., Milleville, K., Ali, D., Porras-bernardez, F., Gartner, G., & Weghe, N. Van De. (2009). EURECA - EUropean Region Enrichment in City Archives and collections. *EURECA-EUropean Region Enrichment in City Archives and Collections*, 1–9. Retrieved from https://biblio.ugent.be/publication/8623994/file/8623996

Winget, M. (2006). User-defined classification on the online photo sharing site Flickr...or, how i learned to stop worrying and love the million typing monkeys. *Advances in Classification Research Online*, *17*, 1–16. https://doi.org/10.7152/acro.v17i1.12496

Create Live World Clouds (n.d.) Mrntimeter. Retrieved from https://www.mentimeter.com/features/word-cloud

Kiessling, W. (2018, September 05). The Essential Guide to Geosocial Data. Spatial. Retrieved 02.09.2019 from https://blog.spatial.ai/the-essential-guide-to-geosocial-data

Marshall, Poe. (2006, September). The Hive.The Atlantic. Retrieved 20.08.2019 from https://www.theatlantic.com/magazine/archive/2006/09/the-hive/305118/

How Density-based Clustering works (n.d.). ESRI. Retrieved from https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/how-density-based-clustering-works.htm

Wikipedia contributors. (2019, June 10). Albertina. In *Wikipedia, The Free Encyclopedia*. Retrieved 01:50, June 12, 2019, from https://en.wikipedia.org/w/index.php?title=Albertina&oldid=913253115

Seite „Albertina (Wien) “. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 10. May 2019, 19:46 UTC. URL: https://de.wikipedia.org/w/index.php?title=Albertina_(Wien)&oldid=192180006 (Abgerufen: 12. June 2019, 01:53 UTC)

Wikipedia contributors. (2019, May 7). Peterskirche, Vienna. In *Wikipedia, The Free Encyclopedia*. Retrieved 01:55, May 12, 2019, from https://en.wikipedia.org/w/index.php?title=Peterskirche,_Vienna&oldid=914414040

Seite „Peterskirche (Wien)“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 10. Mai 2019, 13:31 UTC. URL: https://de.wikipedia.org/w/index.php?title=Peterskirche_(Wien)&oldid=188526787 (Abgerufen: 12. May 2019, 01:56 UTC)

Wikipedia contributors. (2019, May 5). Kunsthistorisches Museum. In *Wikipedia, The Free Encyclopedia*. Retrieved 01:57, May 7, 2019, from https://en.wikipedia.org/w/index.php?title=Kunsthistorisches_Museum&oldid=914930480

Seite, Kunsthistorisches Museum“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 2. Juli 2019, 19:11 UTC. URL: https://de.wikipedia.org/w/index.php?title=Kunsthistorisches_Museum&oldid=190073013 (Abgerufen: 6. July, 2019, 01:50 UTC)

Wikipedia contributors. (2019, May 9). Natural History Museum, Vienna. In *Wikipedia, The Free Encyclopedia*. Retrieved 01:58, May 12, 2019, from https://en.wikipedia.org/w/index.php?title=Natural_History_Museum,_Vienna&oldid=914717685

Seite „Naturhistorisches Museum Wien“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 9. May 2019, 14:14 UTC. URL: https://de.wikipedia.org/w/index.php?title=Naturhistorisches_Museum_Wien&oldid=190127822 (Abgerufen: 12. May 2019, 01:59 UTC)

Wikipedia contributors. (2019, May 7). Maria am Gestade. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:00, May 12, 2019, from https://en.wikipedia.org/w/index.php?title=Maria_am_Gestade&oldid=914468331

Seite „Maria am Gestade“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 12. May 2019, 15:23 UTC. URL: https://de.wikipedia.org/w/index.php?title=Maria_am_Gestade&oldid=192173855 *(Abgerufen: 12. May 2019, 02:00 UTC)*

Wikipedia contributors. (2019, August 23). University of Vienna. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:01, September 9, 2019, from https://en.wikipedia.org/w/index.php?title=University_of_Vienna&oldid=912070242

Seite „Universität Wien“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 9. September 2019, 15:49 UTC. URL: https://de.wikipedia.org/w/index.php?title=Universit%C3%A4t_Wien&oldid=192146327 (Abgerufen: 9. September 2019, 02:01 UTC)

Wikipedia contributors. (2019, August 15). Imperial Crypt. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:03, August 20, 2019, from https://en.wikipedia.org/w/index.php?title=Imperial_Crypt&oldid=910916611

Seite „Kaisergruft“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 10. August 2019, 19:29 UTC. URL: https://de.wikipedia.org/w/index.php?title=Kaisergruft&oldid=191798671 (Abgerufen: 20. August 2019, 02:03 UTC)

Wikipedia contributors. (2019, August 24). Austrian Parliament. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:04, August 20, 2019, from https://en.wikipedia.org/w/index.php?title=Austrian_Parliament&oldid=912335974

Seite „Österreichisches Parlament“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 4. Juli 2019, 10:03 UTC. URL: https://de.wikipedia.org/w/index.php?title=%C3%96sterreichisches_Parlament&oldid=190115449 (Abgerufen: 20. August 2019, 02:04 UTC)

Wikipedia contributors. (2019, August 8). St. Stephen's Cathedral, Vienna. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:05, August 20, 2019, from https://en.wikipedia.org/w/index.php?title=St._Stephen%27s_Cathedral,_Vienna&oldid=914697575

Seite „Stephansdom (Wien)“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 8. August 2019, 15:59 UTC. URL: https://de.wikipedia.org/w/index.php?title=Stephansdom_(Wien)&oldid=191583210 (Abgerufen: 30. August 2019, 02:05 UTC)

Wikipedia contributors. (2019, August 15). Schönbrunn Palace. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:06, August 20, 2019, from https://en.wikipedia.org/w/index.php?title=Sch%C3%B6nbrunn_Palace&oldid=910990322

Seite „Schloss Schönbrunn“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 9. August 2019, 10:24 UTC.
URL: https://de.wikipedia.org/w/index.php?title=Schloss_Sch%C3%B6nbrunn&oldid=192106877 (Abgerufen: 20. August 2019, 02:07 UTC)

Wikipedia contributors. (2019, January 10). Austrian National Library. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:07, May 12, 2019,
from https://en.wikipedia.org/w/index.php?title=Austrian_National_Library&oldid=877758029

Seite „Österreichische Nationalbibliothek“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 11. Mai 2019, 03:44 UTC.
URL: https://de.wikipedia.org/w/index.php?title=%C3%96sterreichische_Nationalbibliothek&oldid=188443459 (Abgerufen: 12. May 2019, 02:07 UTC)

Wikipedia contributors. (2019, August 15). Hofburg. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:08, August 18, 2019,
from https://en.wikipedia.org/w/index.php?title=Hofburg&oldid=910892896

Seite „Hofburg“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 7. Juli 2019, 11:32 UTC.
URL: https://de.wikipedia.org/w/index.php?title=Hofburg&oldid=190202978 *(Abgerufen: 18. August 2019, 02:09 UTC)*

Wikipedia contributors. (2019, May 7). Belvedere, Vienna. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:09, May 12, 2019,
from https://en.wikipedia.org/w/index.php?title=Belvedere,_Vienna&oldid=914413852

Seite „Schloss Belvedere“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 5. June 2019, 09:55 UTC.
URL: https://de.wikipedia.org/w/index.php?title=Schloss_Belvedere&oldid=192166150 (Abgerufen: 20. June 2019, 02:09 UTC)

Wikipedia contributors. (2019, August 7). Volksgarten, Vienna. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:10, August 12, 2019,
from https://en.wikipedia.org/w/index.php?title=Volksgarten,_Vienna&oldid=914526275

Seite „Volksgarten (Wien)“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 6. Juli 2019, 12:42 UTC.
URL: https://de.wikipedia.org/w/index.php?title=Volksgarten_(Wien)&oldid=190178825 (Abgerufen: 10. August 2019, 02:10 UTC)

Wikipedia contributors. (2019, August 26). Prater. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:11, August 30, 2019,
from https://en.wikipedia.org/w/index.php?title=Prater&oldid=912645371

Seite „Wiener Prater“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 11. May 2019, 22:11 UTC.
URL: https://de.wikipedia.org/w/index.php?title=Wiener_Prater&oldid=192183112 (Abgerufen: 12. May 2019, 02:11 UTC)

Wikipedia contributors. (2019, August 24). Vienna City Hall. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:12, August 30, 2019, from https://en.wikipedia.org/w/index.php?title=Vienna_City_Hall&oldid=912326694

Seite „Wiener Rathaus“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 2. September 2019, 06:59 UTC. URL: https://de.wikipedia.org/w/index.php?title=Wiener_Rathaus&oldid=191895795 (Abgerufen: 3. September 2019, 02:12 UTC)

Wikipedia contributors. (2019, May 10). Vienna State Opera. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:13, June 18, 2019, from https://en.wikipedia.org/w/index.php?title=Vienna_State_Opera&oldid=915008003

Seite „Wiener Staatsoper“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 30. May 2019, 07:42 UTC. URL: https://de.wikipedia.org/w/index.php?title=Wiener_Staatsoper&oldid=191808609 (Abgerufen: 18. June 2019, 02:13 UTC)

Wikipedia contributors. (2019, January 14). Karlsplatz. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:14, June 18, 2019, from https://en.wikipedia.org/w/index.php?title=Karlsplatz&oldid=878397150

Seite „Karlsplatz (Wien) “. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 14. Juni 2019, 13:28 UTC. URL: https://de.wikipedia.org/w/index.php?title=Karlsplatz_(Wien)&oldid=189537216 (Abgerufen: 18. June 2019, 02:14 UTC)

Wikipedia contributors. (2019, May 8). Karlskirche. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:15, June 18, 2019, from https://en.wikipedia.org/w/index.php?title=Karlskirche&oldid=914636878

Seite „Wiener Karlskirche“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 20. May 2019, 18:37 UTC. URL: https://de.wikipedia.org/w/index.php?title=Wiener_Karlskirche&oldid=191528932 (Abgerufen: 18. June 2019, 02:15 UTC)

Wikipedia contributors. (2019, June 10). Hundertwasserhaus. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:16, June 18, 2019, from https://en.wikipedia.org/w/index.php?title=Hundertwasserhaus&oldid=913882977

Seite „Hundertwasserhaus (Wien)“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 17. März 2019, 22:48 UTC. URL: https://de.wikipedia.org/w/index.php?title=Hundertwasserhaus_(Wien)&oldid=186691397 *(Abgerufen: 18. June 2019, 02:16 UTC)*

Wikipedia contributors. (2019, January 17). Musikverein. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:16, June 18, 2019, from https://en.wikipedia.org/w/index.php?title=Musikverein&oldid=878939542

Seite „Wiener Musikverein“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 29. May 2019, 10:54 UTC.

URL: https://de.wikipedia.org/w/index.php?title=Wiener_Musikverein&oldid=190849991 (A bgerufen: 18. June 2019, 02:17 UTC)

Wikipedia contributors. (2019, February 21). Church of Mariahilf. In *Wikipedia*, *The Free Encyclopedia*. Retrieved 02:17, June 20, 2019, from https://en.wikipedia.org/w/index.php?title=Church_of_Mariahilf&oldid=884384696

Seite „Mariahilfer Kirche". In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 21. Februar 2019, 15:16 UTC. URL: https://de.wikipedia.org/w/index.php?title=Mariahilfer_Kirche&oldid=185907858 (Ab gerufen: 20. June 2019, 02:18 UTC)

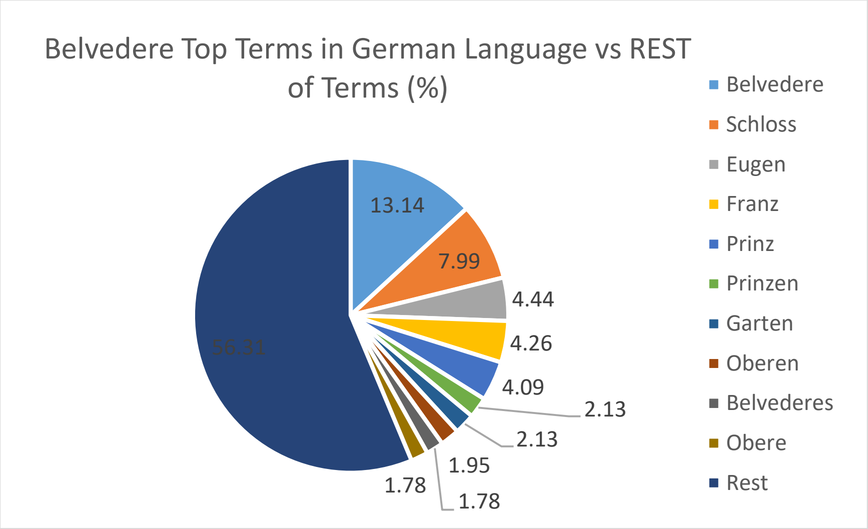
Wikipedia contributors. (2019, May 4). Burgtheater. In *Wikipedia*, *The Free Encyclopedia*. Retrieved 02:18, June 20, 2019, from https://en.wikipedia.org/w/index.php?title=Burgtheater&oldid=913966301

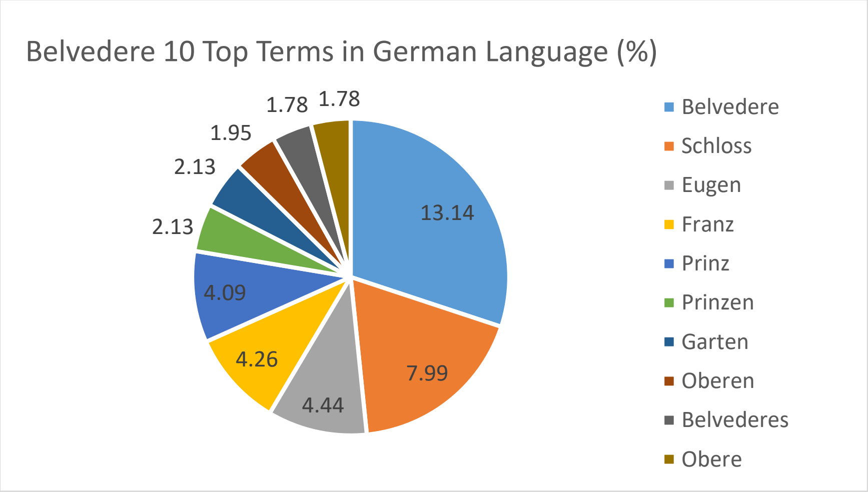
Seite „Burgtheater". In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 14. May 2019, 12:11 UTC. URL: https://de.wikipedia.org/w/index.php?title=Burgtheater&oldid=191903297 (Abgerufen: 12. June 2019, 02:19 UTC)

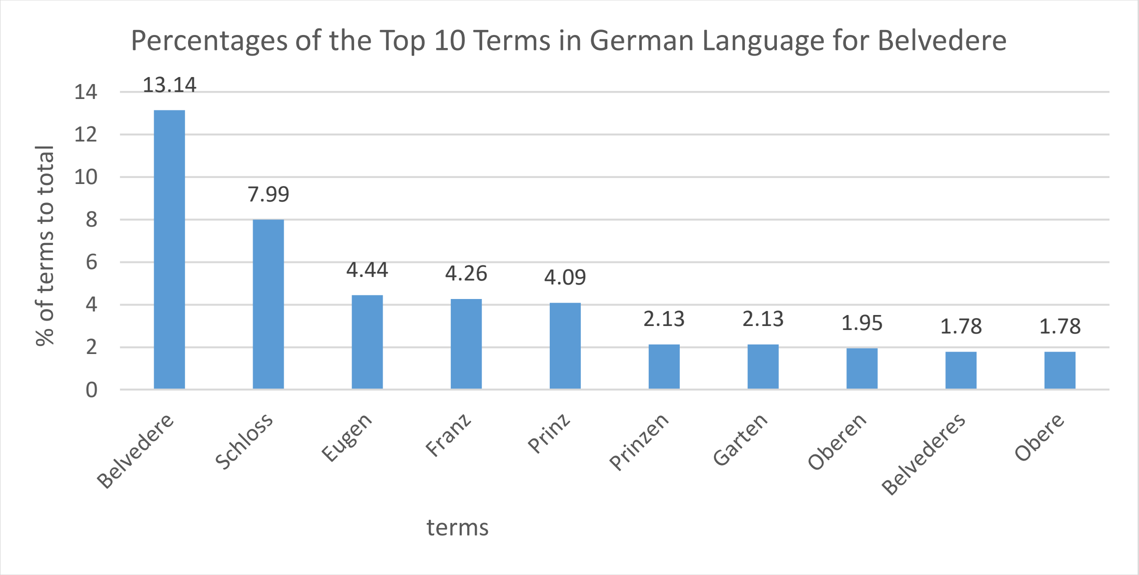
Appendix:

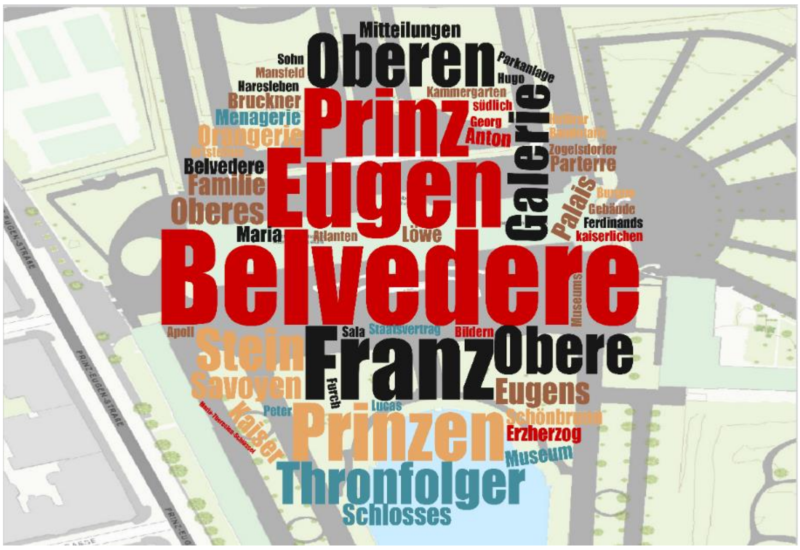
Wikipedia: Wordclouds, Pie Charts and the Bar Charts for chosen POI

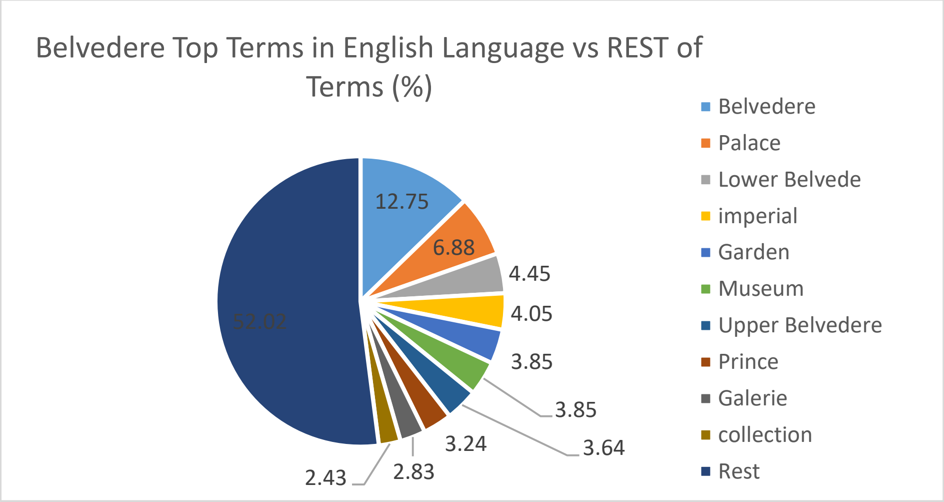
Belvedere Palace:

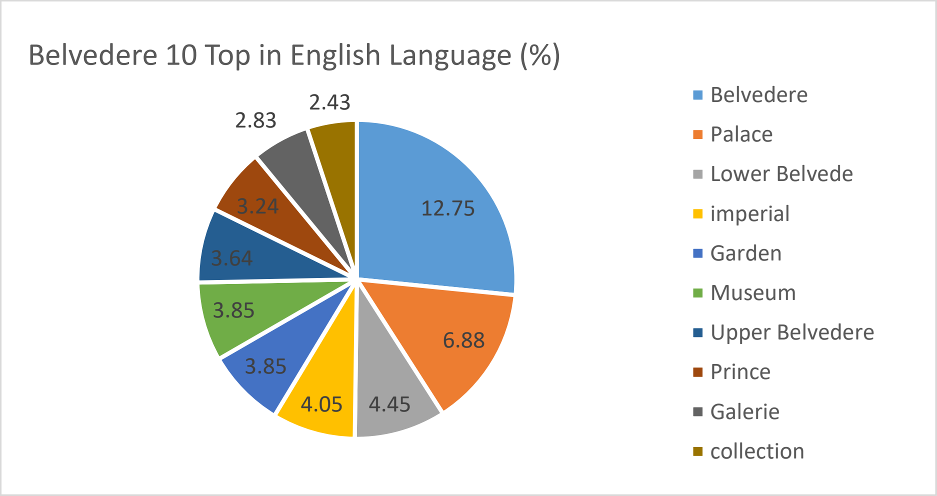


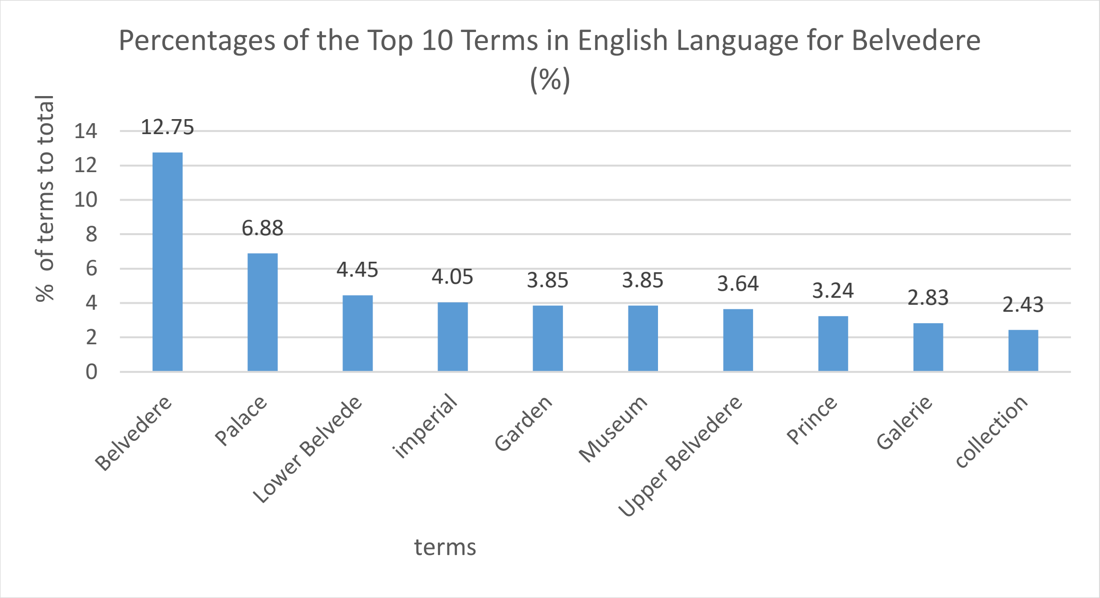






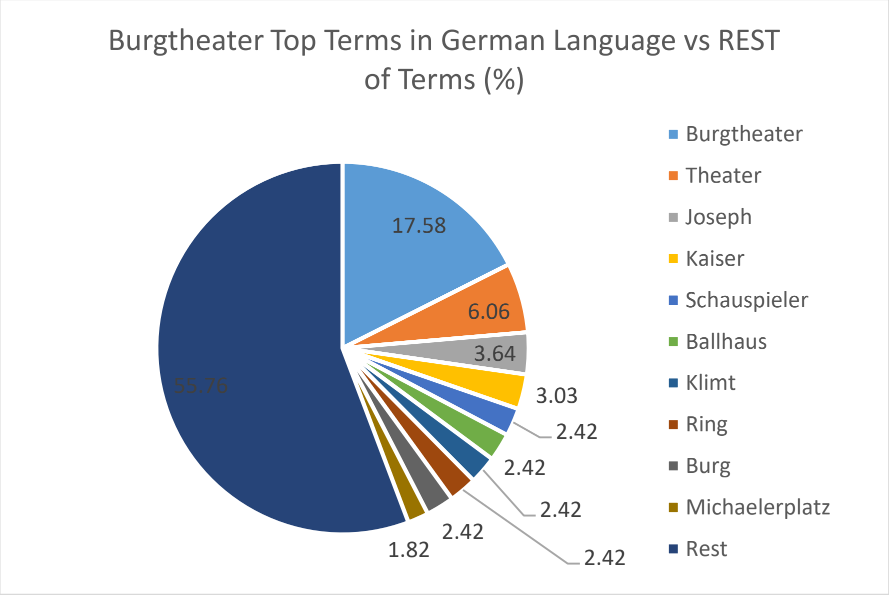

Belvedere Top Terms in English Language vs REST of Terms (%)

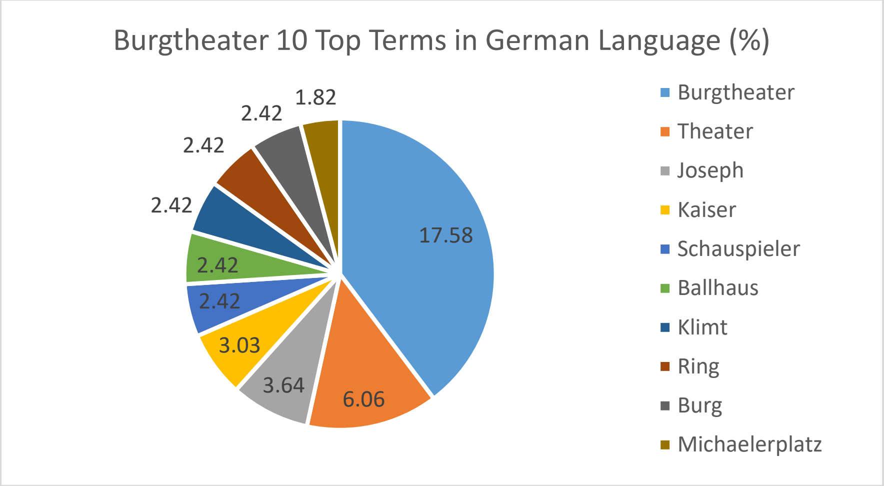

Belvedere 10 Top in English Language (%)

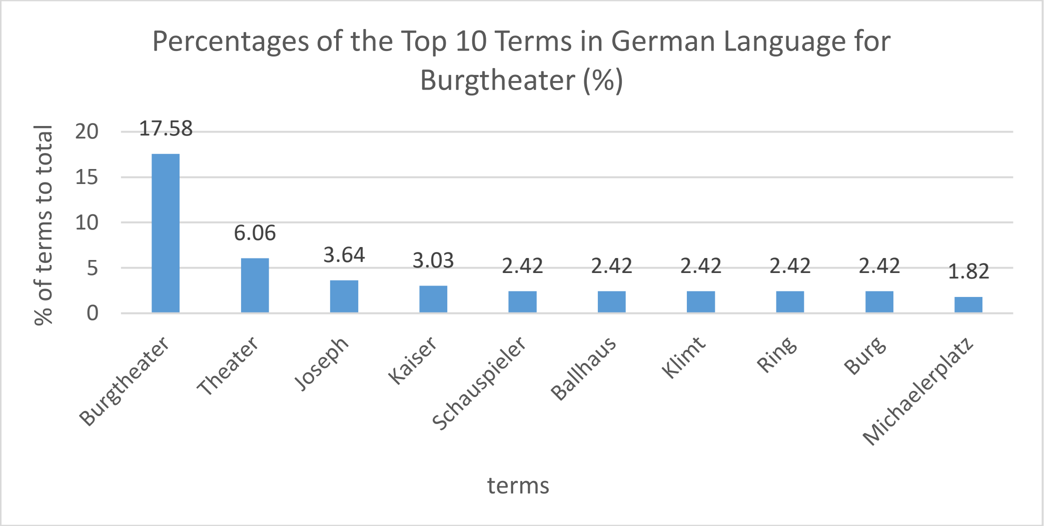


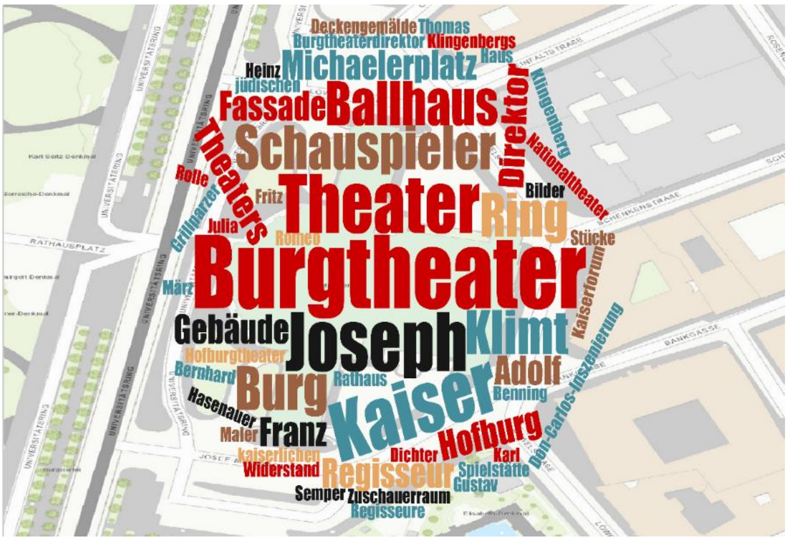
collection
UpperBelvedere
Franz
imperial
Art
residence
Hall
Marble
LowerBelvede
Habsburg
Joseph
Maria
French
building
Belvedere
Baroque
Palace
Prince
Eugene
Ferdinand
Castle
Garden
Museum
Galerie
architect
Construction
Princess Royal

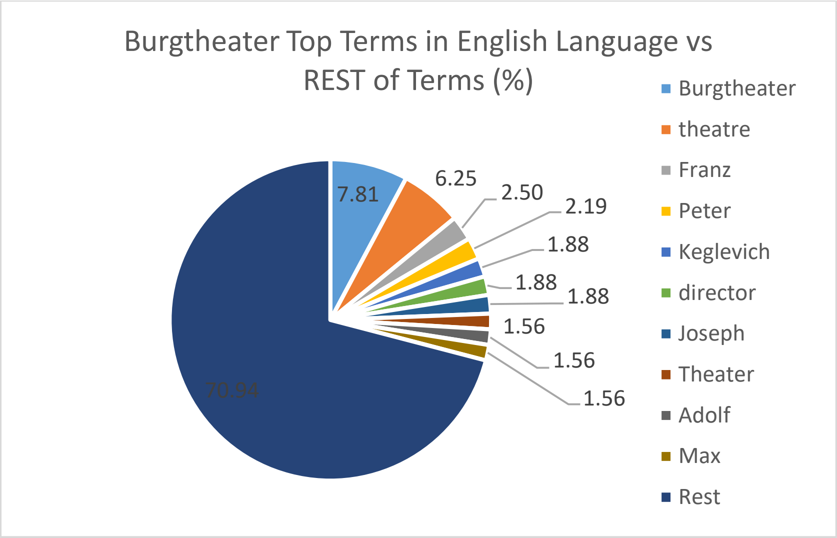
Burgtheater:

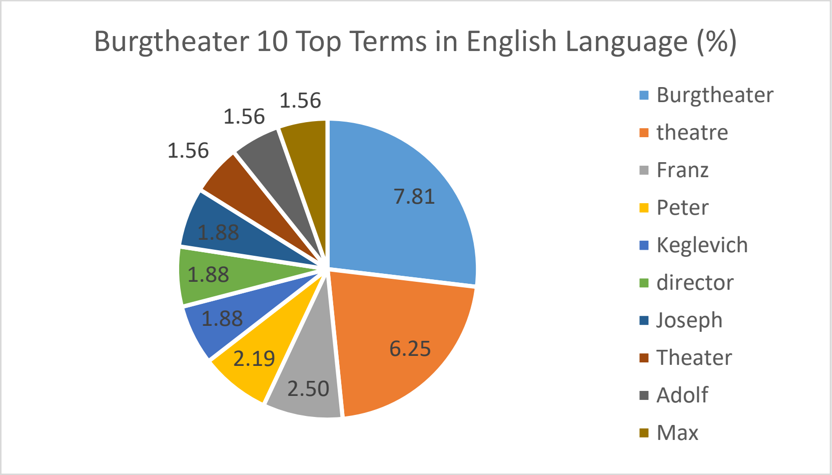


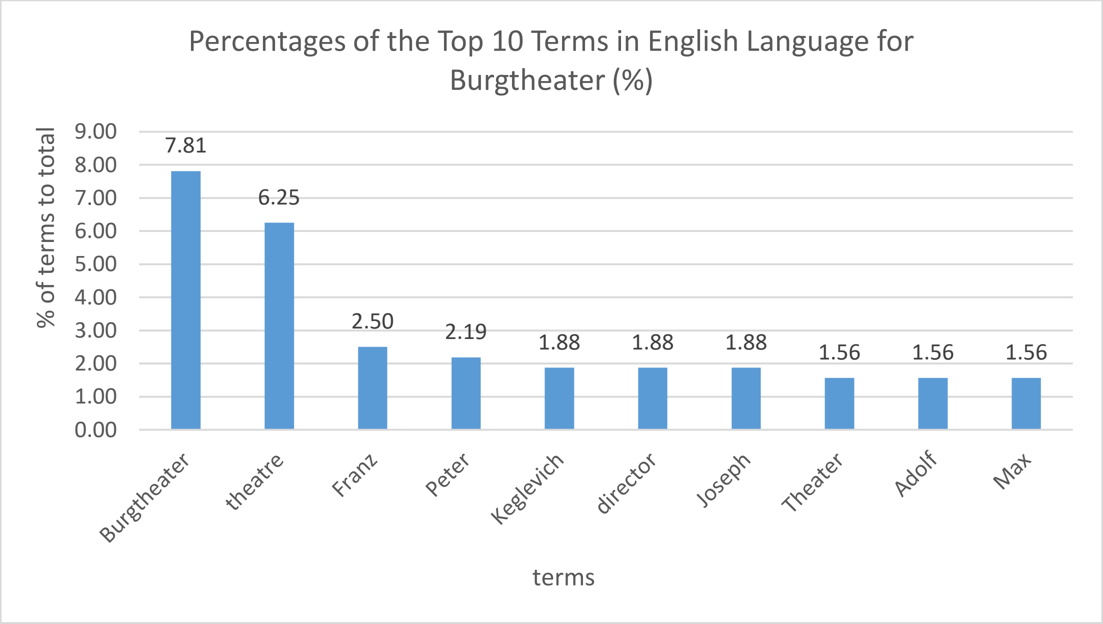

Burgtheater 10 Top Terms in German Language (%)
17.58
6.06
3.64
3.03
2.42
2.42
2.42
2.42
2.42
1.82
Burgtheater
Theater
Joseph
Kaiser
Schauspieler
Ballhaus
Klimt
Ring
Burg
Michaelerplatz


Percentages of the Top 10 Terms in German Language for Burgtheater (%)
% of terms to total
20
15
10
5
0
17.58
6.06
3.64
3.03
2.42
2.42
2.42
2.42
2.42
1.82
Burgtheater
Theater
Joseph
Kaiser
Schauspieler
Ballhaus
Klimt
Ring
Burg
Michaelerplatz
terms


Deckengemälde
Thomas
Burgtheaterdirektor
Klingenbergs
Heinz
Michaelerplatz
Haus
jüdischen
Fassade
Ballhaus
Direktor
Klingenberg
Schauspieler
Nationaltheater
Theaters
Rolle
Fritz
Theater
Ring
Bilder
Grillparzer
Julia
Romeo
Stücke
Burgtheater
März
Kaiserforum
Gebäude
Joseph
Klimt
Hofburgtheater
Bernhard
Adolf
Rathaus
Burg
Benning
Don-Carlos-Inszenierung
Hasenauer
Kaiser
Maler
Franz
Hofburg
kaiserlichen
Dichter
Karl
Widerstand
Regisseur
Spielstätte
Gustav
Semper
Zuschauerraum
Regisseure

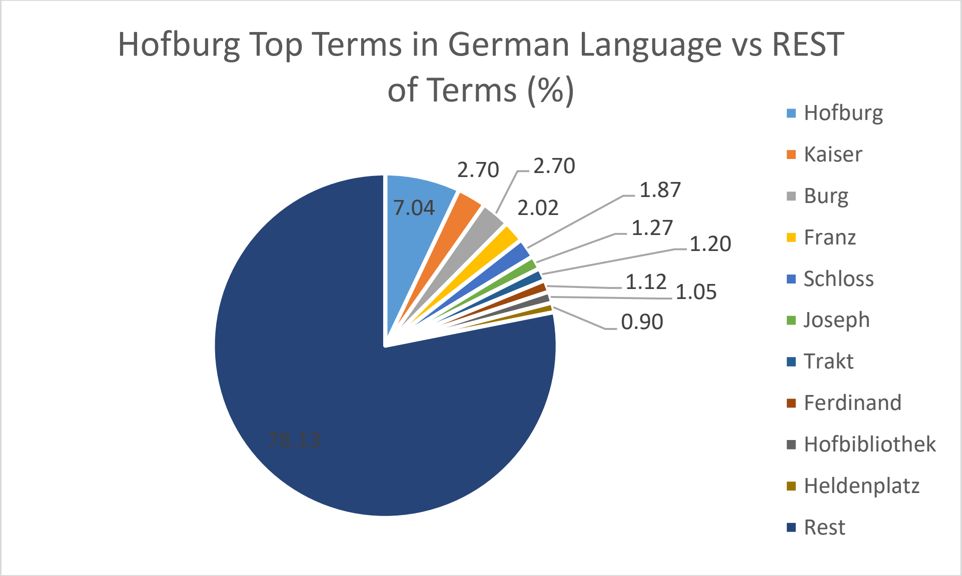

Burgtheater Top Terms in English Language vs REST of Terms (%)

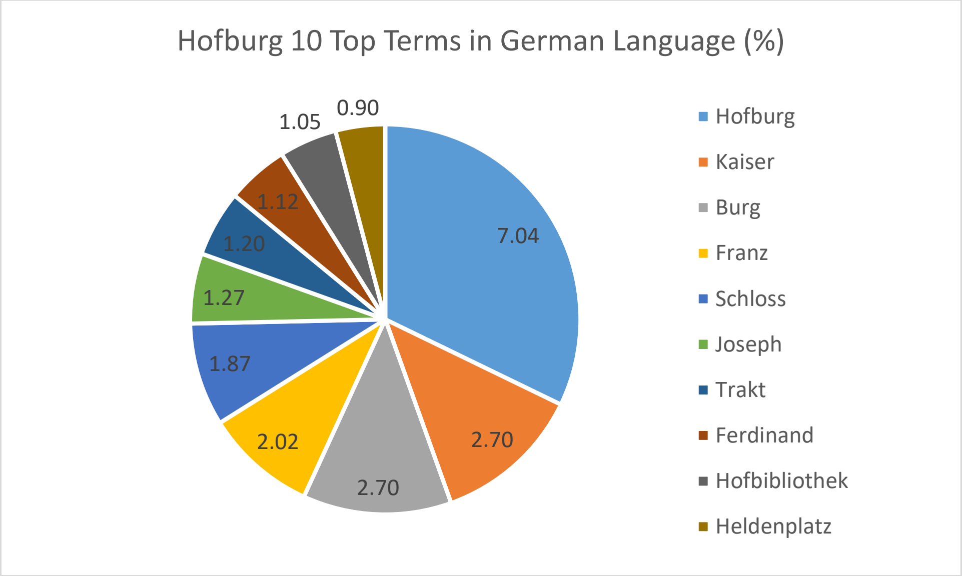

Burgtheater 10 Top Terms in English Language (%)

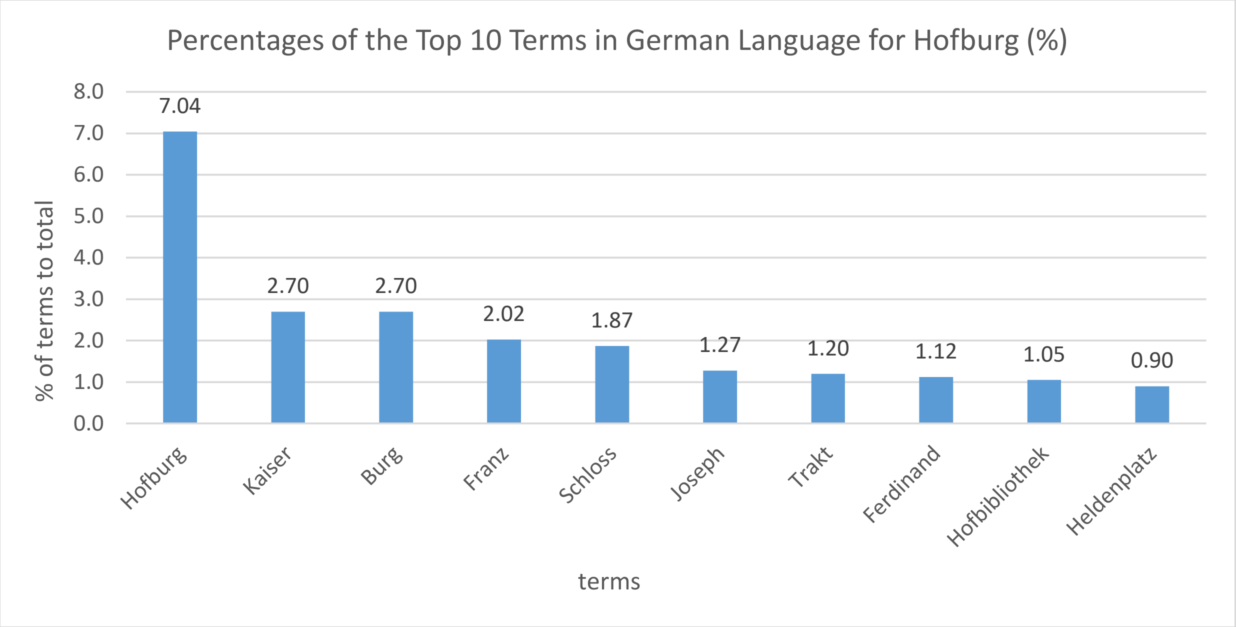

Percentages of the Top 10 Terms in English Language for Burgtheater (%)

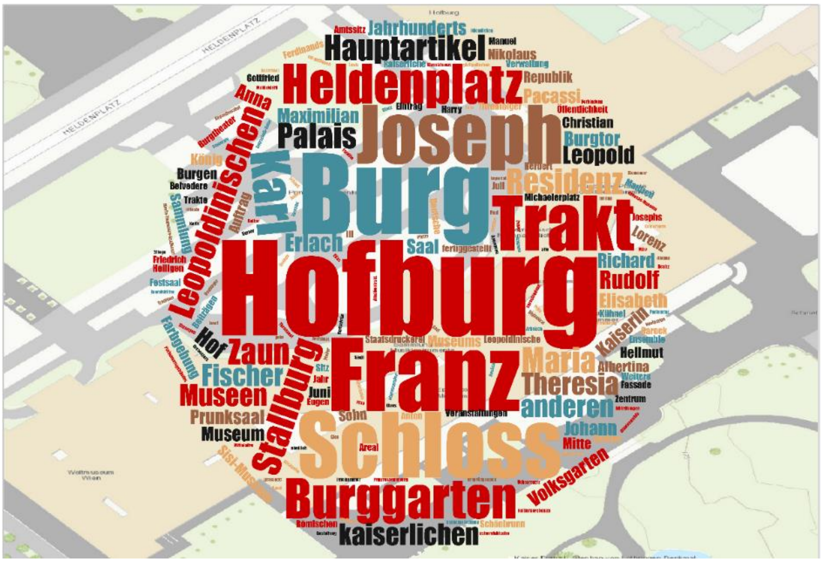
ensemble
Theresa
Burg
Michael's
MusicBrainz
Claus
important
August
film
Karl
director
Sonnenthal
directors
financial
April
Freiherr
Peter
University
Theatre
tenant
External
Franz
Joseph
Adolf
Josef
Fritz
Alfred
committee
financing
internationally
Burgtheater
Magyar
Maria
theatre
Max
Sporck
direction
Navigation
style
Joint
Heinrich
Johann
members
Directors
language
Keglevich
therefore
staged
Paul
stage
performances
Contact
Christoph
Marie
National
senior
Michael
October
Anton
Theater
German
Klaus
Hoftheater
designers
chairman
Wildgans
Ludwig
Hofburg
Empress
Hermann
March

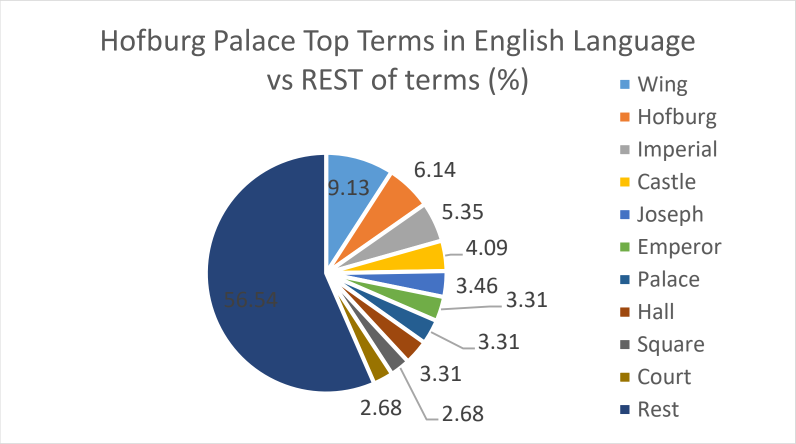
Hofburg Palace:

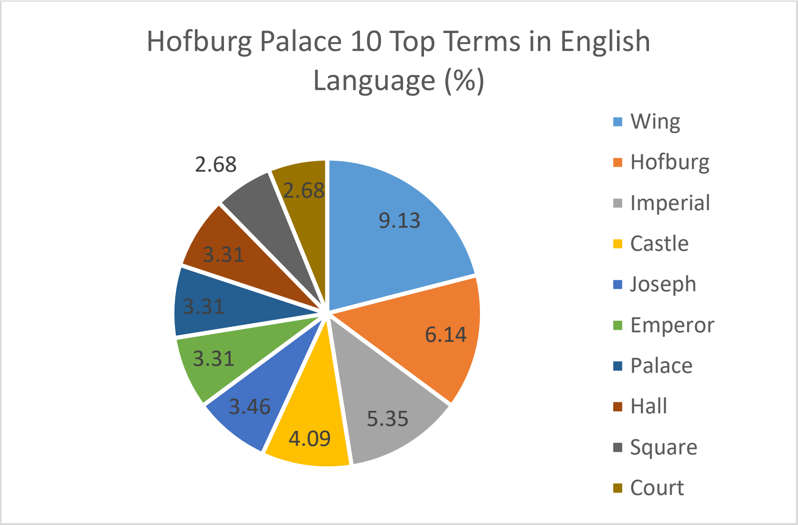


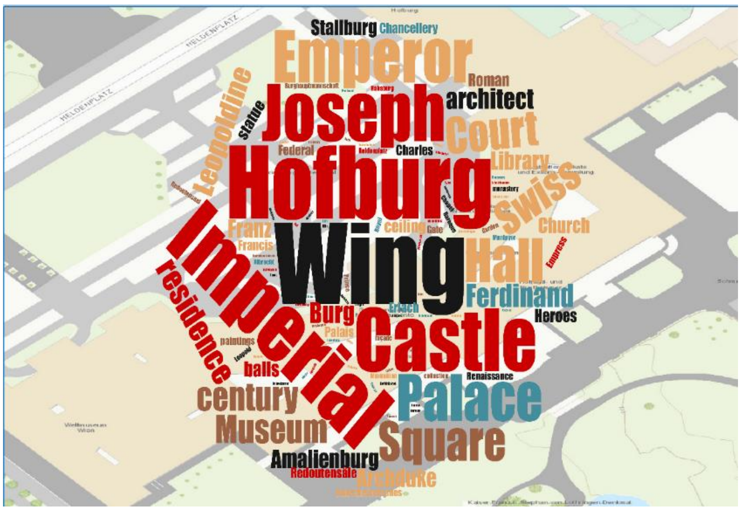



Percentages of the Top 10 Terms in German Language for Hofburg (%)
% of terms to total
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
7.04
2.70
2.70
2.02
1.87
1.27
1.20
1.12
1.05
0.90
Hofburg
Kaiser
Burg
Franz
Schloss
Joseph
Trakt
Ferdinand
Hofbibliothek
Heldenplatz
terms

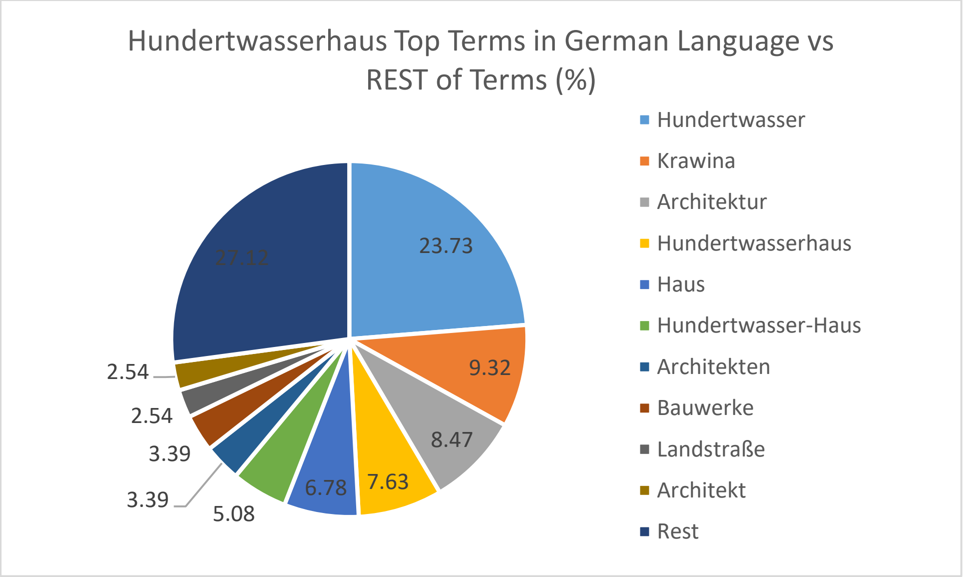

Jahrhunderts
Hauptartikel
Heldenplatz
Joseph
Palais
Maximilian
Christian
Burgtor
Leopold
Burg
Karl
Residenz
Trakt
Leopoldinischen
Anna
Erlach
Saal
Hofburg
Richard
Rudolf
Elisabeth
Kaiserin
Franz
Zaun
Fischer
Museen
Prunksaal
Museum
Stallburg
Maria
Theresia
anderen
Johann
Schloss
Burggarten
Volksgarten
kaiserlichen
Republik
Pacassi
Burgen
Belvedere
Hellmut
Albertina
Nikolaus

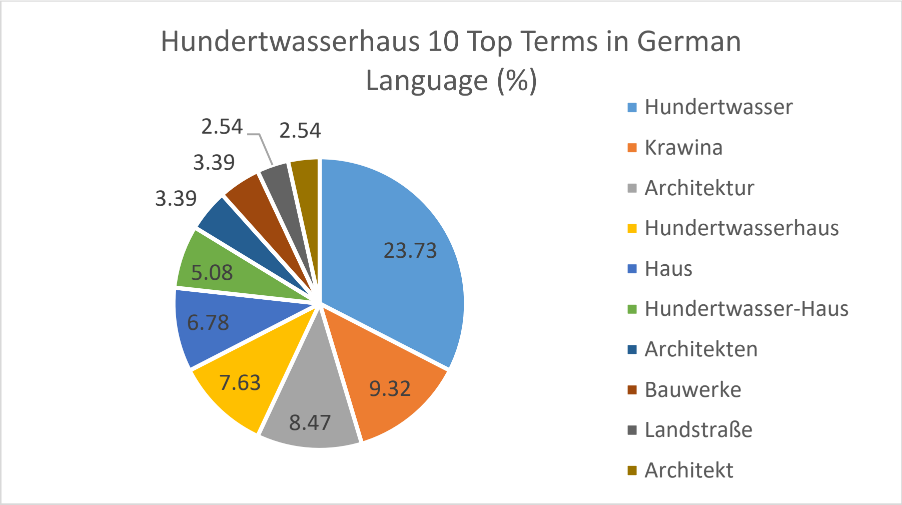

Hofburg Palace Top Terms in English Language vs REST of terms (%)
9.13
6.14
5.35
4.09
3.46
3.31
3.31
3.31
2.68
2.68
56.54
Wing
Hofburg
Imperial
Castle
Joseph
Emperor
Palace
Hall
Square
Court
Rest

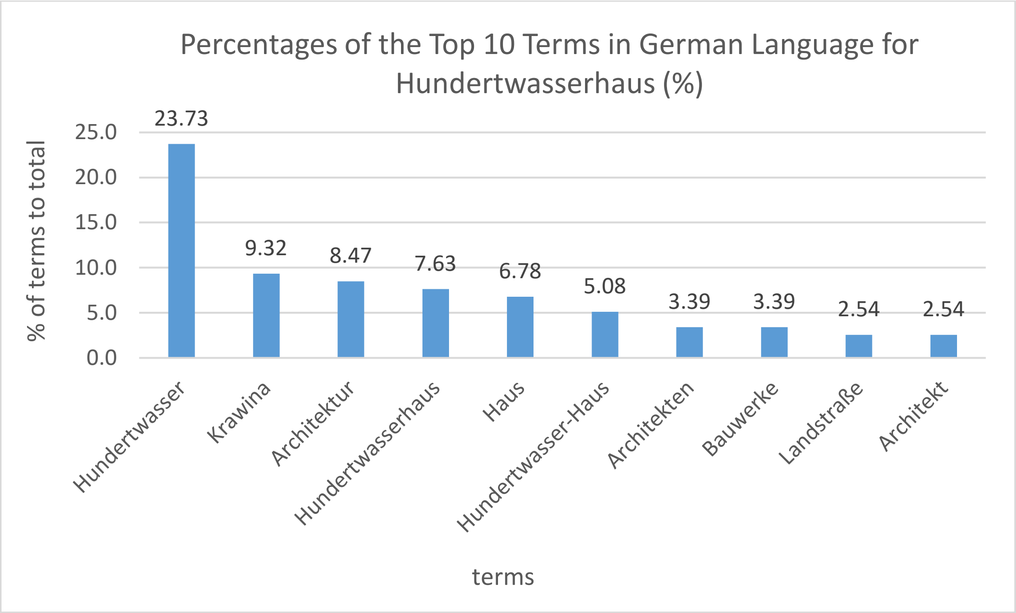




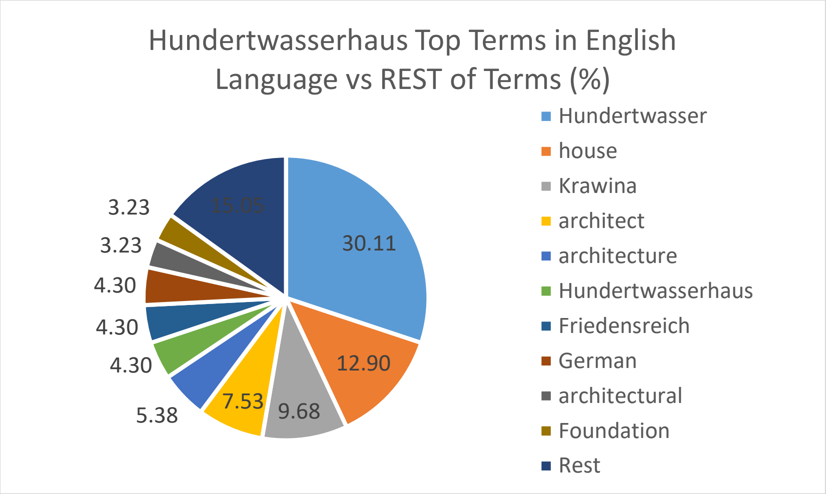
Hundertwasserhaus:

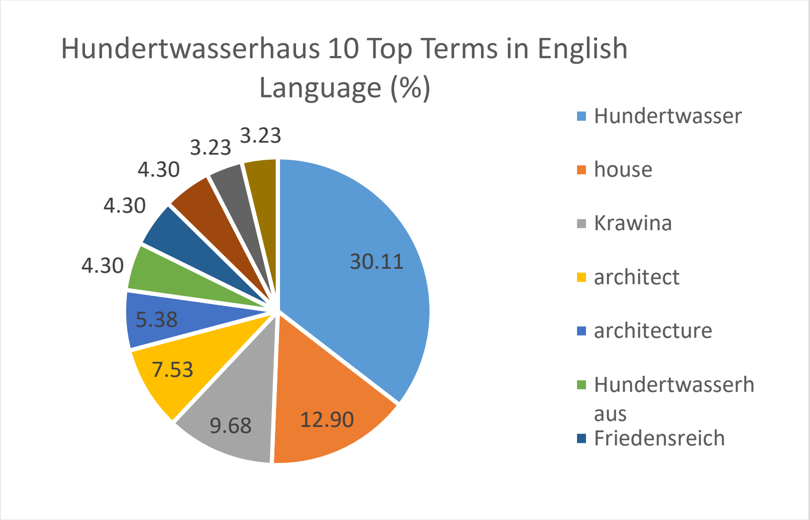


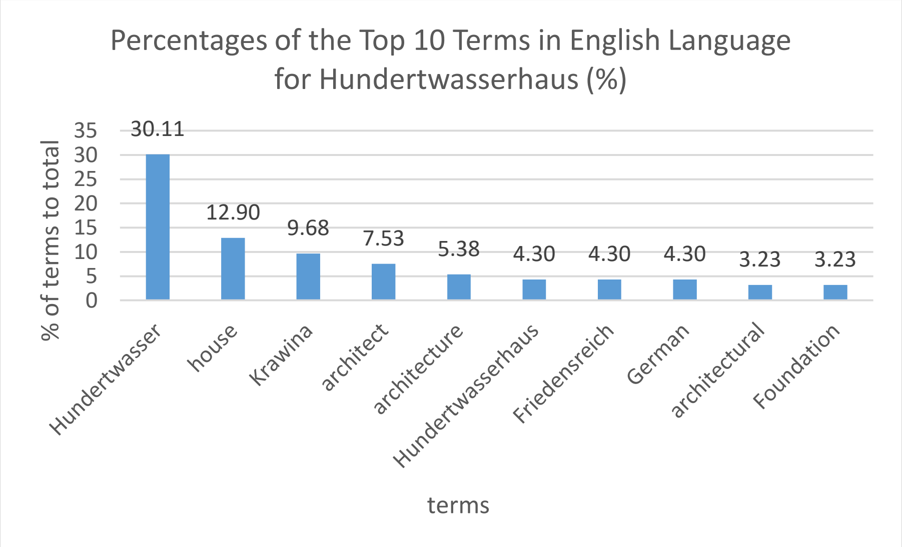

Hundertwasserhaus 10 Top Terms in German Language (%)
23.73
9.32
8.47
7.63
6.78
5.08
3.39
3.39
2.54
2.54
Hundertwasser
Krawina
Architektur
Hundertwasserhaus
Haus
Hundertwasser-Haus
Architekten
Bauwerke
Landstraße
Architekt

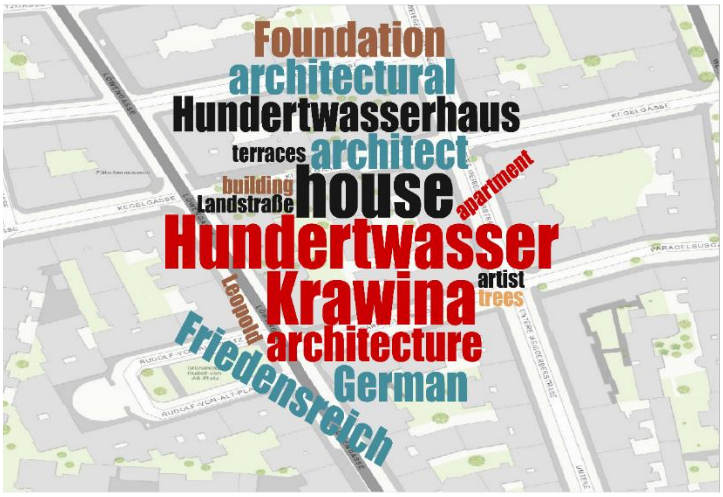

Percentages of the Top 10 Terms in German Language for Hundertwasserhaus (%)
% of terms to total
25.0
20.0
15.0
10.0
5.0
0.0
23.73
9.32
8.47
7.63
6.78
5.08
3.39
3.39
2.54
2.54
Hundertwasser
Krawina
Architektur
Hundertwasserhaus
Haus
Hundertwasser-Haus
Architekten
Bauwerke
Landstraße
Architekt
terms

Magistratsabteilung
Gratz
Gebäude
Bauwerke
Hundertwasser-Haus
Hundertwasserhaus
Landstraße
Krawina
Magdeburg
Hauses
Hundertwasser
Baukörper
Fassade
Bäume
Architektur
deutsche
Bauwerk
Künstler
Haus
Architekt
Vorstellungen
Hundertwassers
Blumau
Architekten
Hundertwasserentscheidung

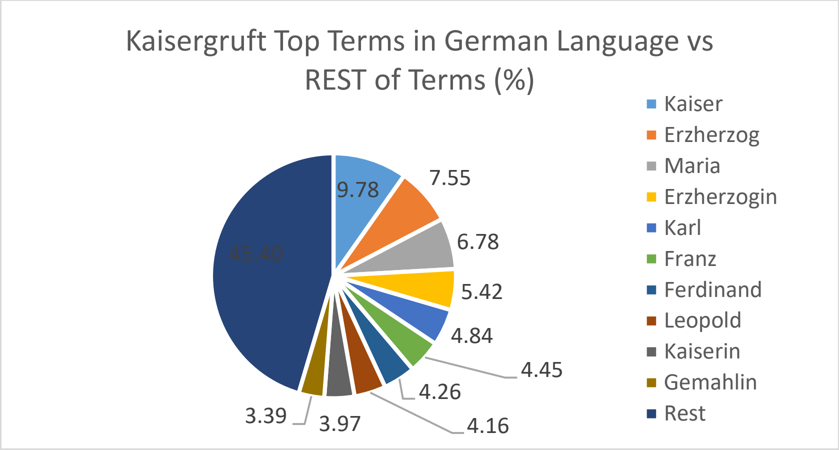

Hundertwasserhaus Top Terms in English Language vs REST of Terms (%)
30.11
12.90
9.68
7.53
5.38
4.30
4.30
4.30
3.23
3.23
15.05
■ Hundertwasser
■ house
■ Krawina
■ architect
■ architecture
■ Hundertwasserhaus
■ Friedensreich
■ German
■ architectural
■ Foundation
■ Rest

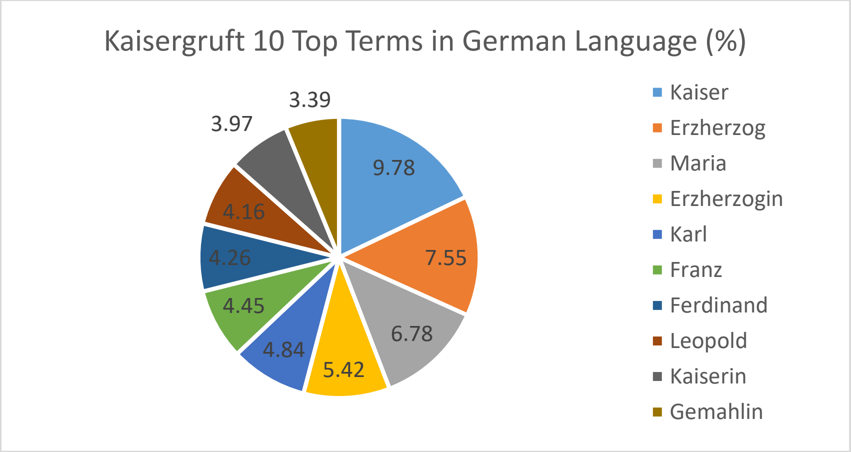

Hundertwasserhaus 10 Top Terms in English Language (%)
30.11
12.90
9.68
7.53
5.38
4.30
4.30
4.30
3.23
3.23
■ Hundertwasser
■ house
■ Krawina
■ architect
■ architecture
■ Hundertwasserhaus
■ Friedensreich

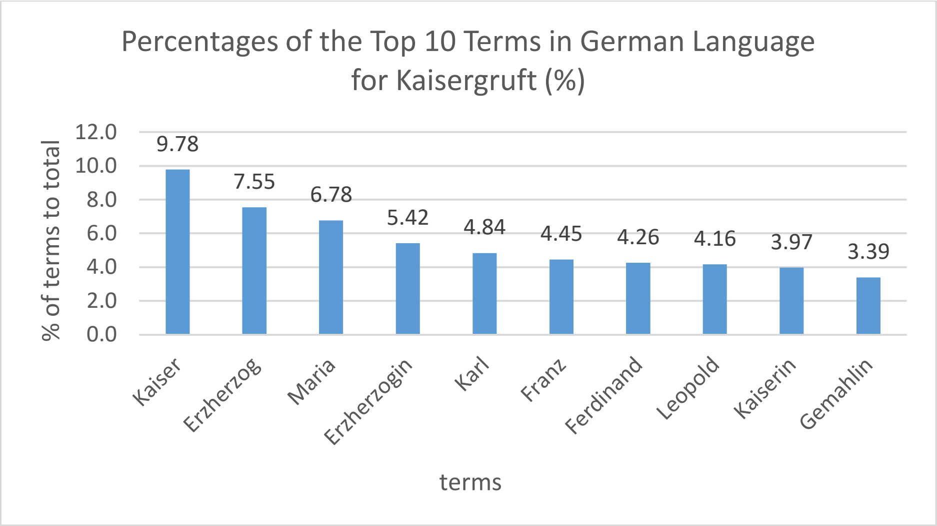

Percentages of the Top 10 Terms in English Language for Hundertwasserhaus (%)
% of terms to total
35
30
25
20
15
10
5
0
30.11
12.90
9.68
7.53
5.38
4.30
4.30
4.30
3.23
3.23
Hundertwasser
house
Krawina
architect
architecture
Hundertwasserhaus
Friedensreich
German
architectural
Foundation
terms

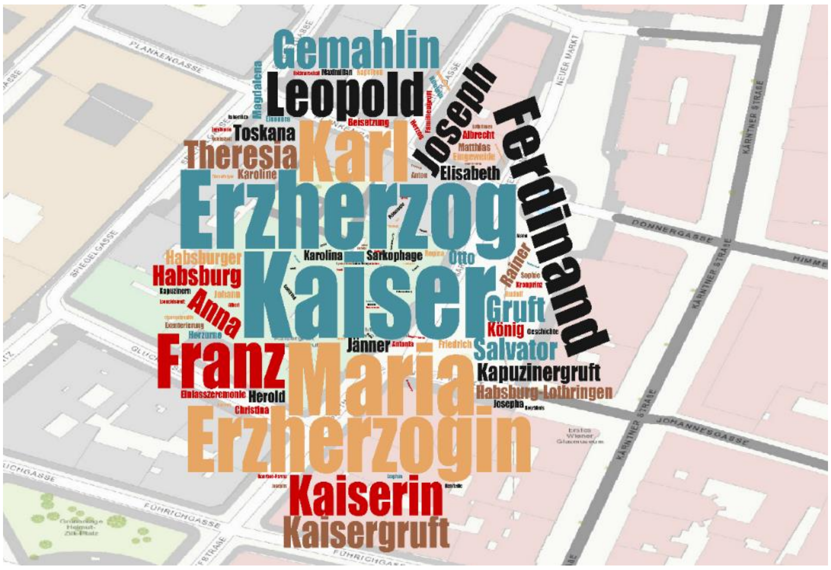


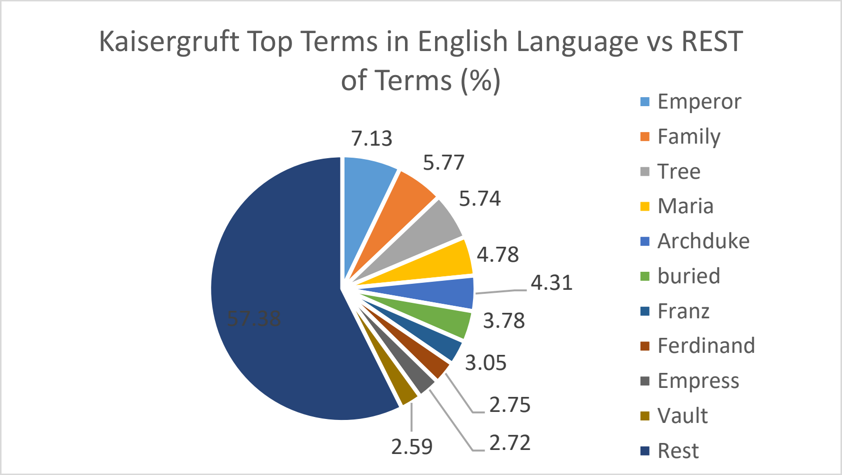
Kaisergruft:

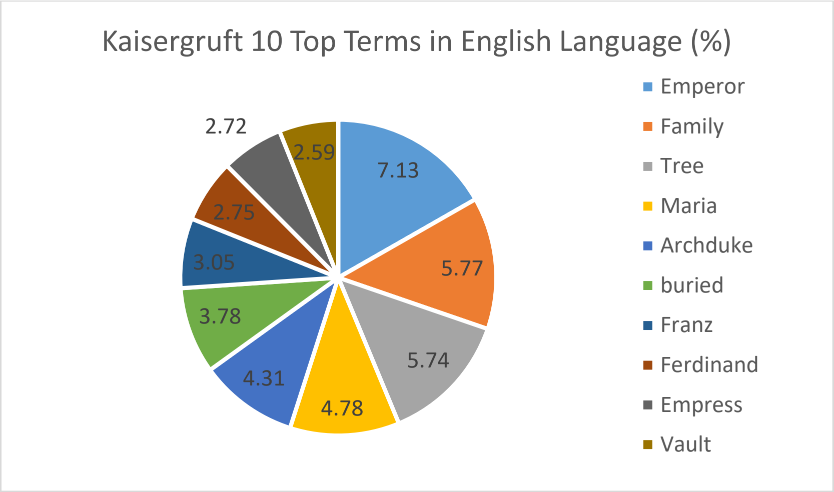

Kaisergruft Top Terms in German Language vs REST of Terms (%)

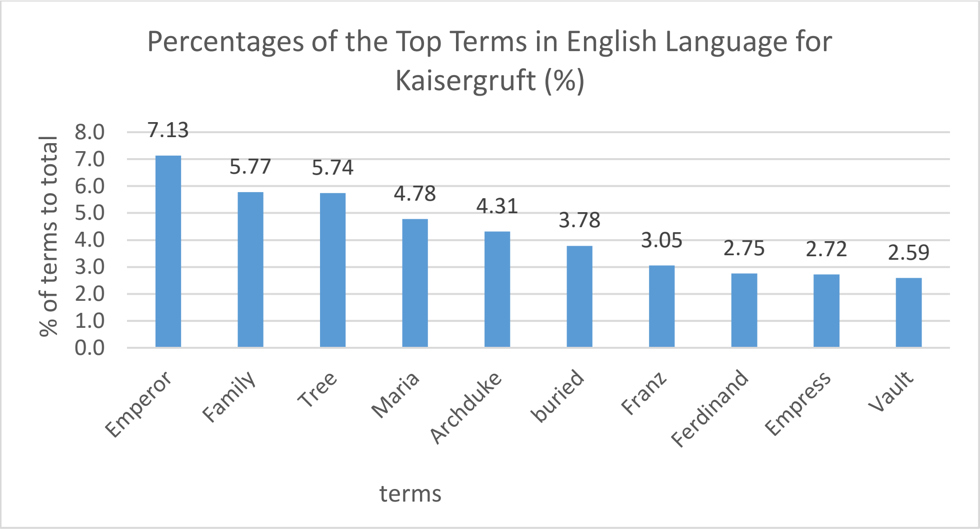

Kaisergruft 10 Top Terms in German Language (%)

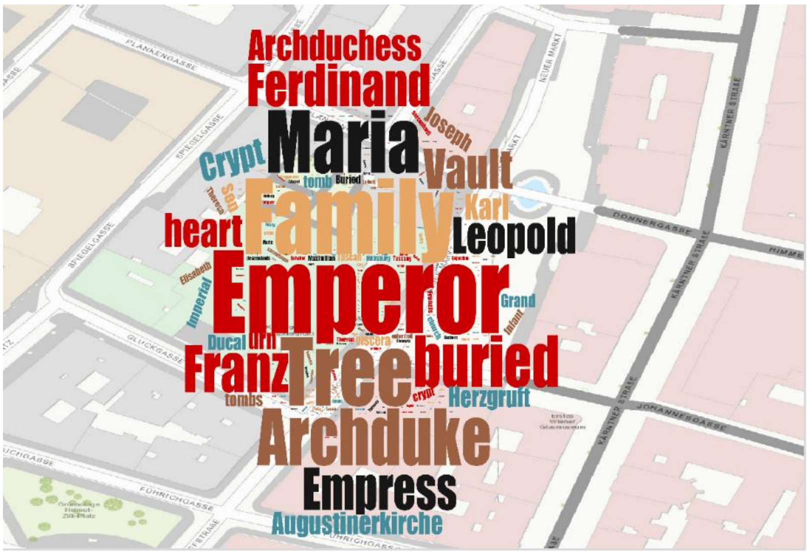

Percentages of the Top 10 Terms in German Language for Kaisergruft (%)
% of terms to total
12.0
10.0
8.0
6.0
4.0
2.0
0.0
9.78
7.55
6.78
5.42
4.84
4.45
4.26
4.16
3.97
3.39
Kaiser
Erzherzog
Maria
Erzherzogin
Karl
Franz
Ferdinand
Leopold
Kaiserin
Gemahlin
terms


Gemahlin
Leopold
Joseph
Ferdinand
Toskana
Theresia
Karl
Albrecht
Matthias
Elisabeth
Karoline
Erzherzog
Habsburger
Habsburg
Karolina
Sarkophage
Otto
Kaiser
Anna
Johann
Gruft
Rainer
König
Salvator
Franz
Jänner
Friedrich
Maria
Kapuzinergruft
Habsburg-Lothringen
Josepha
Herold
Christina
Erzherzogin
Kaiserin
Kaisergruft
Magdalena
Beisetzung
PLANKENGASSE
SPIEGELGASSE
NEUER MARKT
KÄRNTNER STRASSE
DONNERGASSE
JOHANNESGASSE
FÜHRICHGASSE

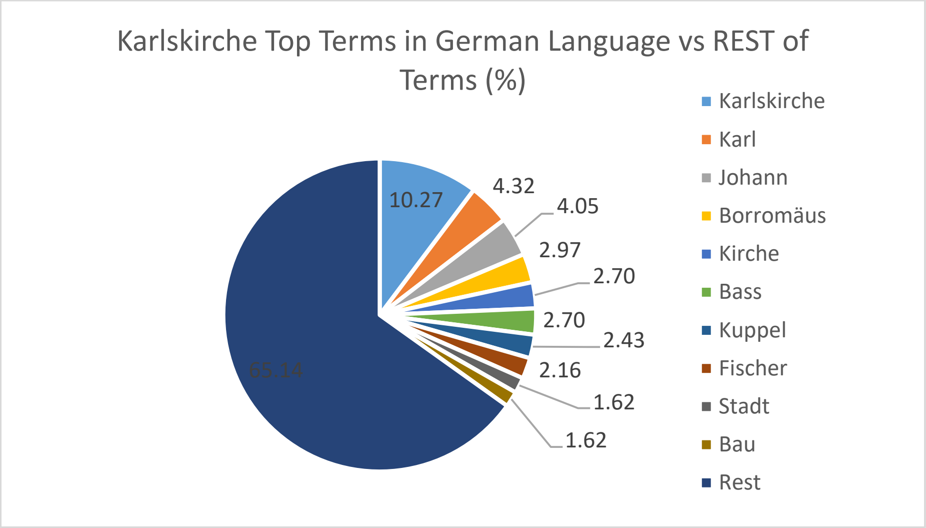

Kaisergruft Top Terms in English Language vs REST of Terms (%)
7.13
5.77
5.74
4.78
4.31
3.78
3.05
2.75
2.72
2.59
57.38
■ Emperor
■ Family
■ Tree
■ Maria
■ Archduke
■ buried
■ Franz
■ Ferdinand
■ Empress
■ Vault
■ Rest

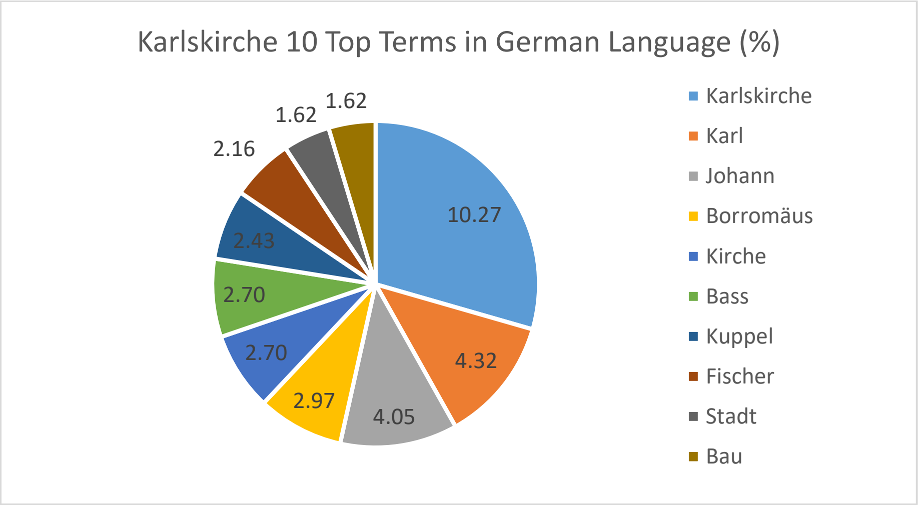

Kaisergruft 10 Top Terms in English Language (%)
7.13
5.77
5.74
4.78
4.31
3.78
3.05
2.75
2.72
2.59
Emperor
Family
Tree
Maria
Archduke
buried
Franz
Ferdinand
Empress
Vault

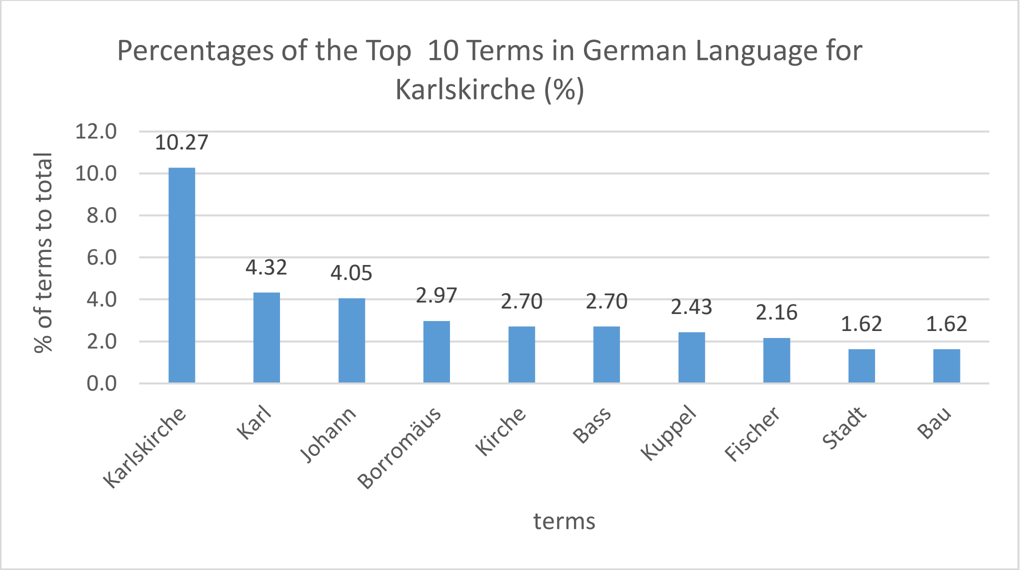

Percentages of the Top Terms in English Language for Kaisergruft (%)
% of terms to total
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
7.13
5.77
5.74
4.78
4.31
3.78
3.05
2.75
2.72
2.59
Emperor
Family
Tree
Maria
Archduke
buried
Franz
Ferdinand
Empress
Vault
terms

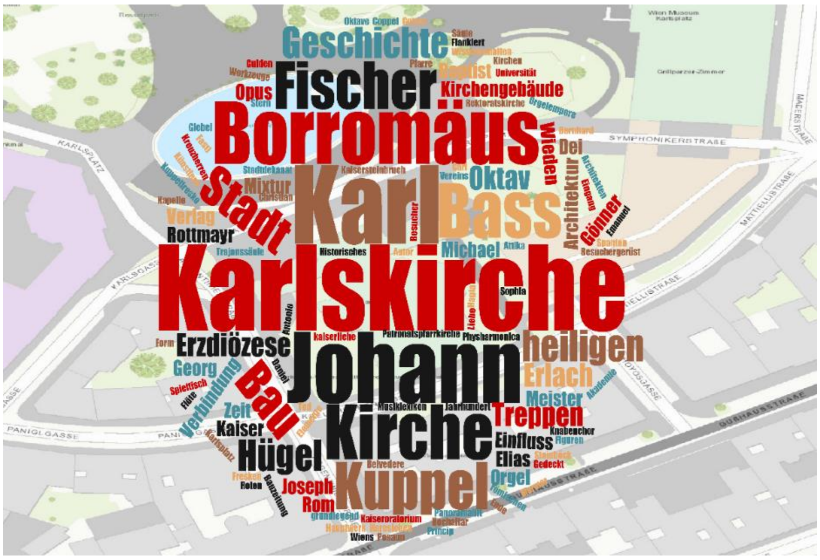

Archduchess
Ferdinand
Maria
Joseph
Crypt
Vault
tomb
Buried
Theresa
Son
Family
Karl
heart
Leopold
Elisabeth
Imperial
Emperor
Grand
Ducal
urn
Franz
Tree
buried
tombs
crypt
Herzgruft
Archduke
Empress
Augustinerkirche

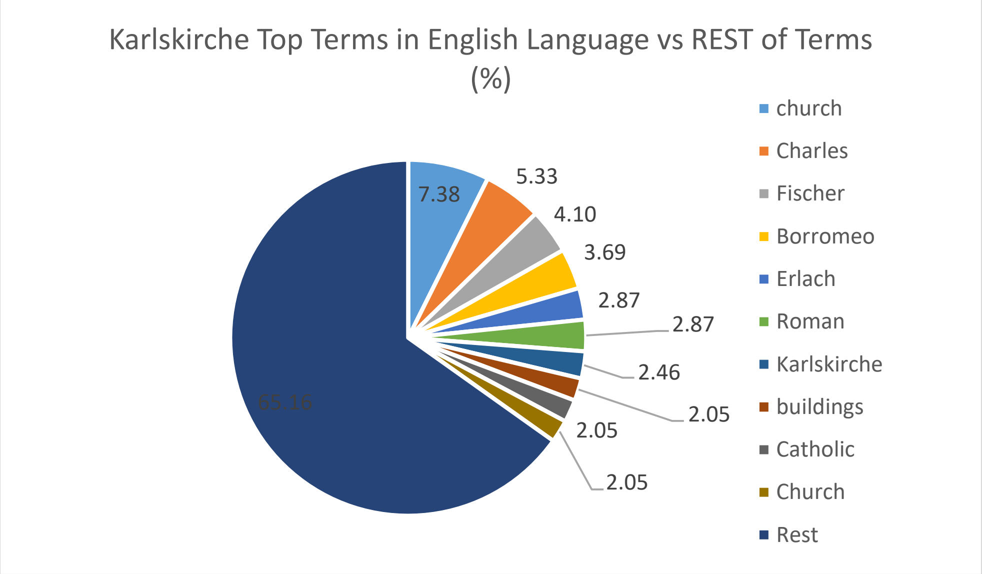
Karlskirche:

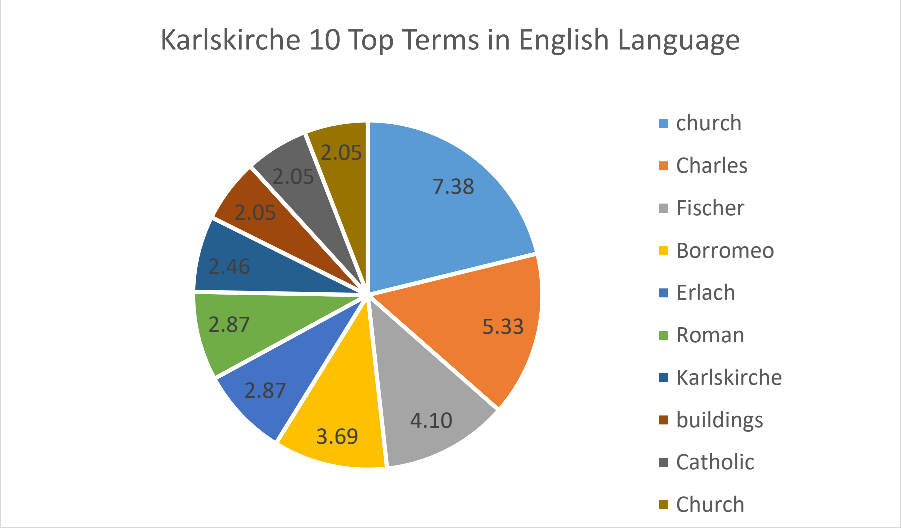


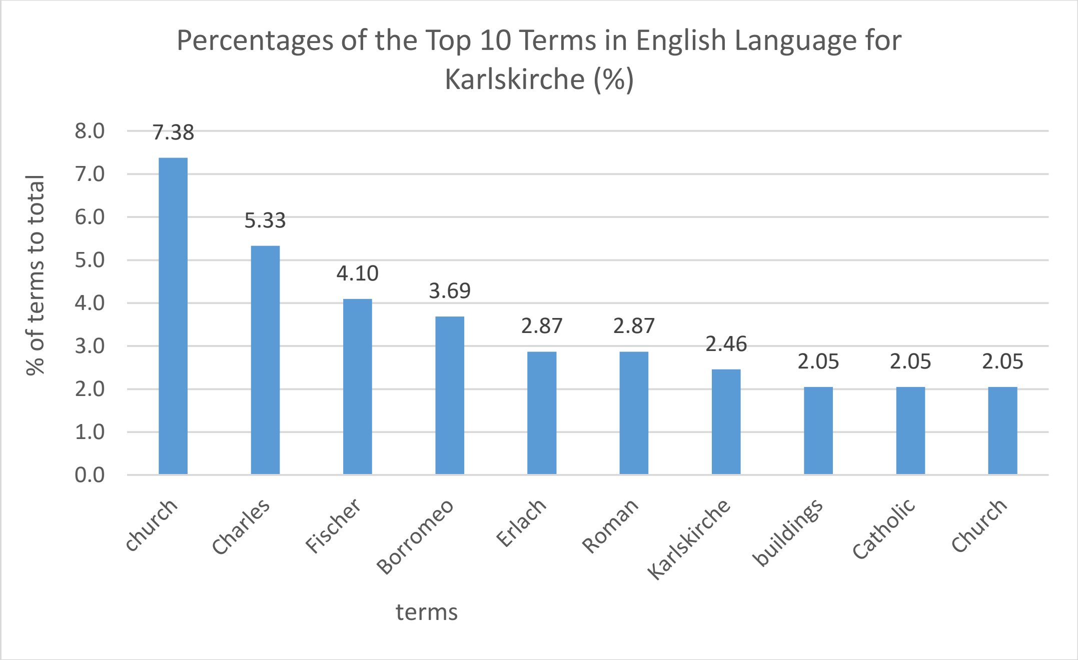


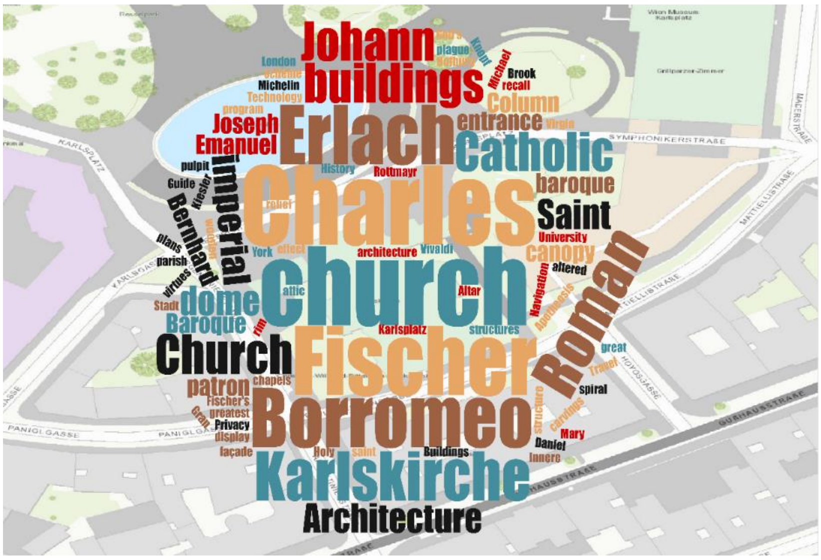



Geschichte
Fischer
Kirchengebäude
Borromäus
Wieden
Oktav
Karl
Bass
Stadt
Mixtur
Verlag
Rottmayr
Architektur
Gönner
Michael
Karlskirche
Erzdiözese
Johann
heiligen
Georg
Erlach
Bau
Verbindung
Meister
Zeit
Kaiser
Kirche
Treppen
Einfluss
Hügel
Elias
Joseph
Kuppel
Orgel
Rom

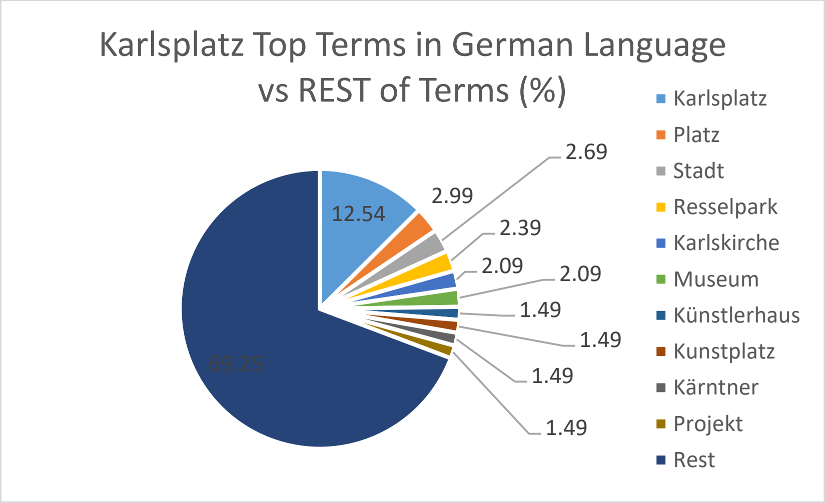

Karlskirche Top Terms in English Language vs REST of Terms (%)
7.38
5.33
4.10
3.69
2.87
2.87
2.46
2.05
2.05
2.05
65.16
church
Charles
Fischer
Borromeo
Erlach
Roman
Karlskirche
buildings
Catholic
Church
Rest

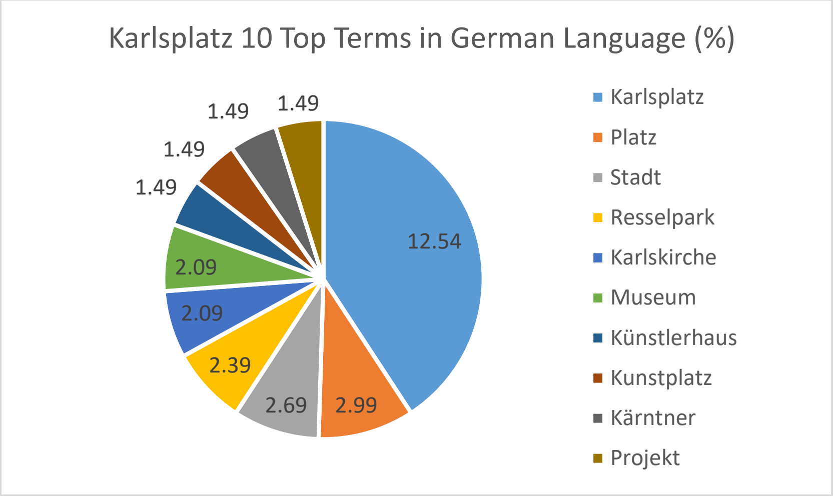

Karlskirche 10 Top Terms in English Language
7.38
5.33
4.10
3.69
2.87
2.87
2.46
2.05
2.05
2.05
church
Charles
Fischer
Borromeo
Erlach
Roman
Karlskirche
buildings
Catholic
Church

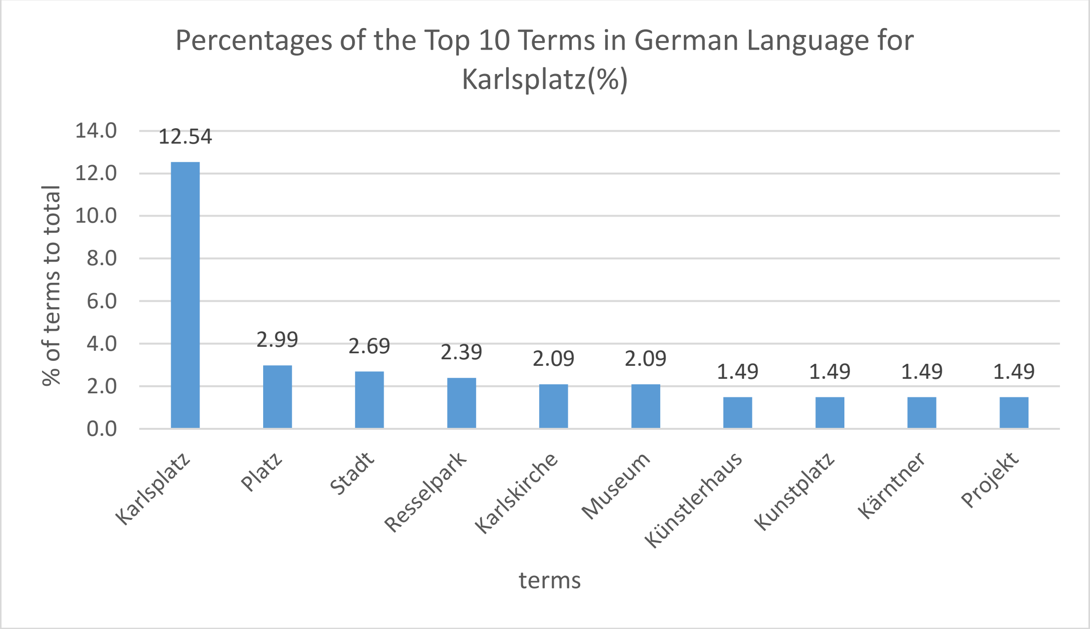

Percentages of the Top 10 Terms in English Language for Karlskirche (%)
% of terms to total
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
7.38
5.33
4.10
3.69
2.87
2.87
2.46
2.05
2.05
2.05
church
Charles
Fischer
Borromeo
Erlach
Roman
Karlskirche
buildings
Catholic
Church
terms

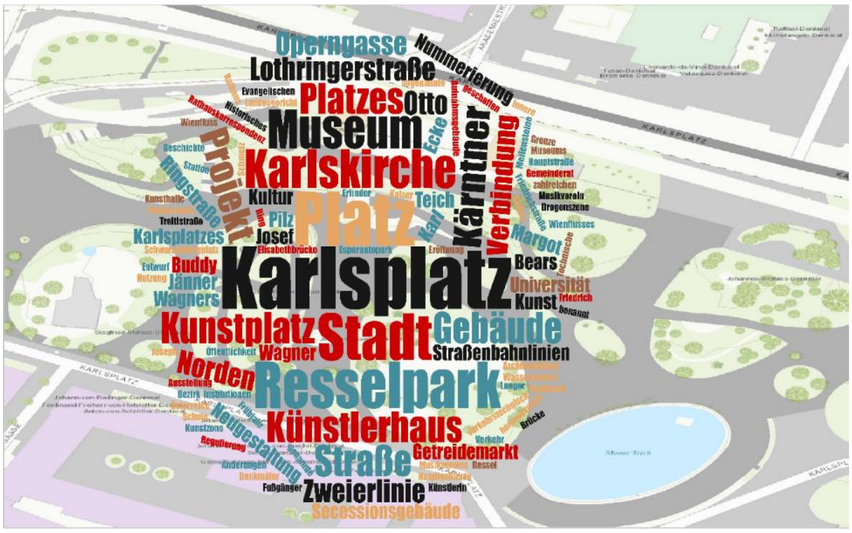

Johann
buildings
Erlach
Joseph
Emanuel
Catholic
Column
entrance
Charles
baroque
Saint
Imperial
Bernhard
church
dome
Baroque
Church
Fischer
Roman
Borromeo
Karlskirche
Architecture
patron
Michelin
London
Technology
Brook
recall
Michael
History
Rottmayr
University
architecture
Vivaldi
canopy
altered
York
attic
Altar
Navigation
Karlsplatz
structures
great
travel
spiral
chapels
Fischer's
greatest
Privacy
display
façade
Holy
saint
Buildings
Mary
Daniel
Innere
pulpit
Guide
plans
parish
virtues
Stadt

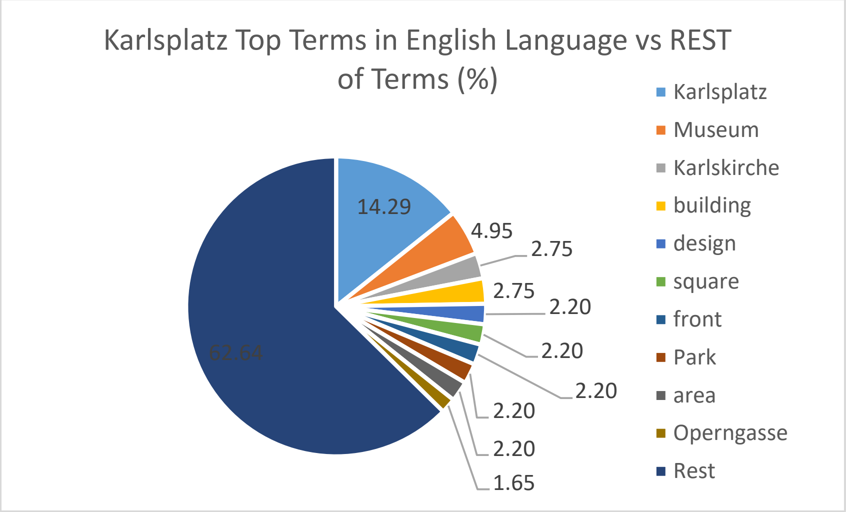
Karlsplatz:

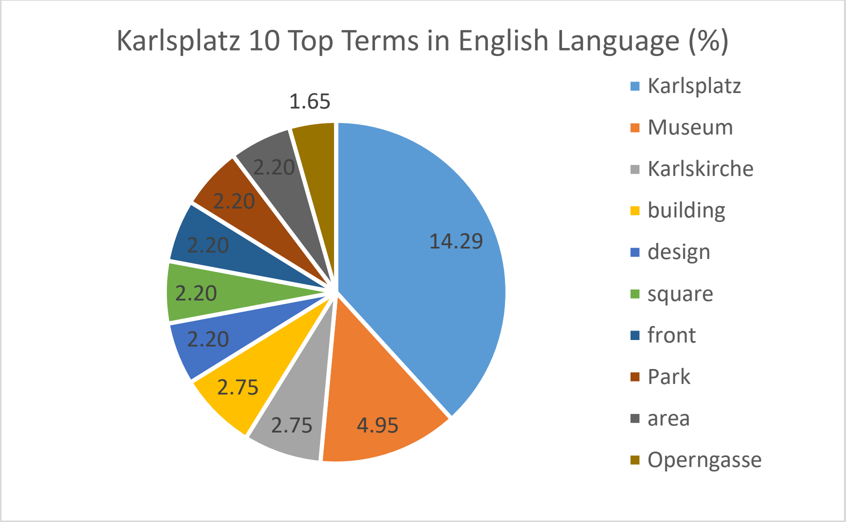

Karlsplatz Top Terms in German Language vs REST of Terms (%)
12.54
2.99
2.69
2.39
2.09
2.09
1.49
1.49
1.49
1.49
69.25
Karlsplatz
Platz
Stadt
Resselpark
Karlskirche
Museum
Künstlerhaus
Kunstplatz
Kärntner
Projekt
Rest

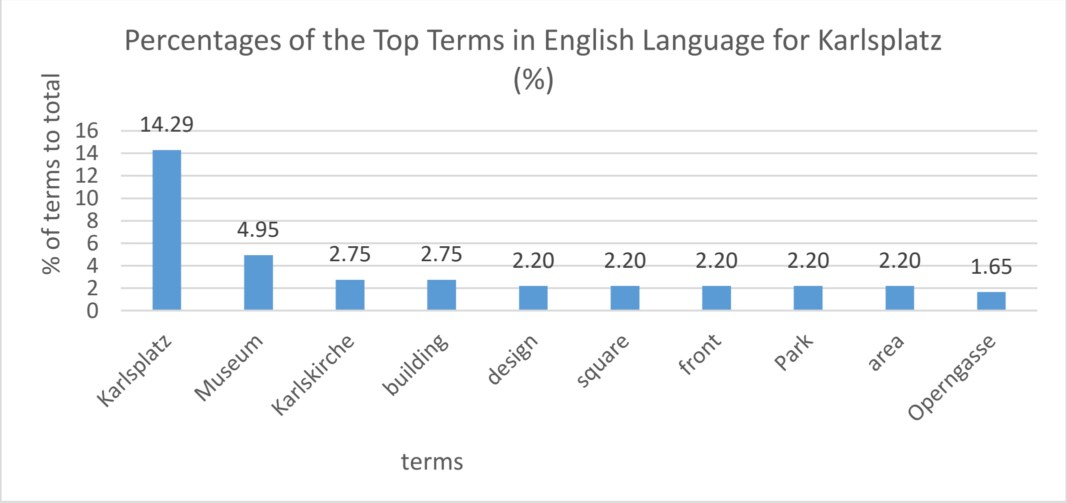

Karlsplatz 10 Top Terms in German Language (%)
12.54
2.99
2.69
2.39
2.09
2.09
1.49
1.49
1.49
1.49
Karlsplatz
Platz
Stadt
Resselpark
Karlskirche
Museum
Künstlerhaus
Kunstplatz
Kärntner
Projekt

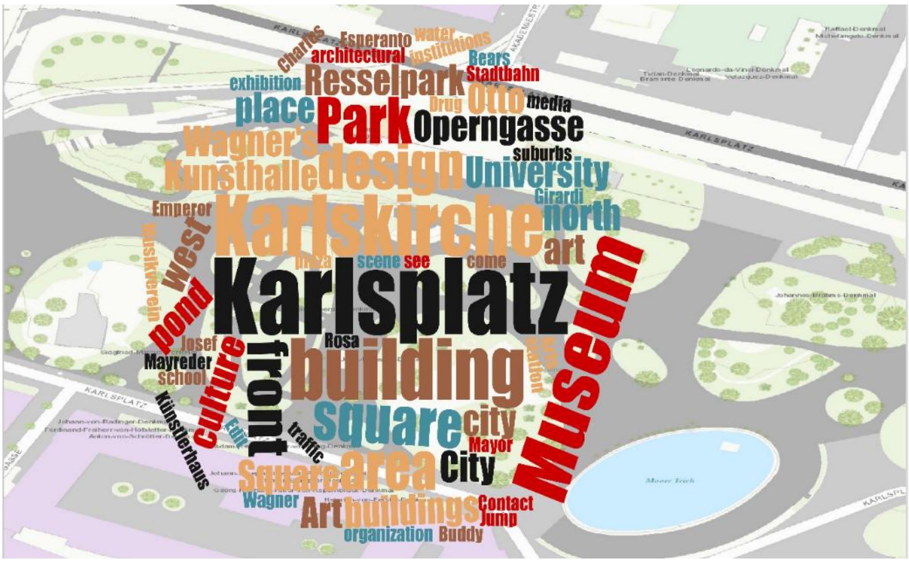

Percentages of the Top 10 Terms in German Language for Karlsplatz(%)
% of terms to total
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
12.54
2.99
2.69
2.39
2.09
2.09
1.49
1.49
1.49
1.49
Karlsplatz
Platz
Stadt
Resselpark
Karlskirche
Museum
Künstlerhaus
Kunstplatz
Kärntner
Projekt
terms


Operngasse
Nummerierung
Lothringerstraße
Platzes
Otto
Museum
Karlskirche
Projekt
Kärntner
Verbindung
Ringstraße
Platz
Kultur
Teich
Pilz
Josef
Karl
Karlsplatzes
Margot
Karlsplatz
Buddy
Jänner
Wagners
Bears
Universität
Kunst
Friedrich
Kunstplatz
Stadt
Gebäude
Wagner
Straßenbahnlinien
Norden
Resselpark
Künstlerhaus
Neugestaltung
Straße
Getreidemarkt
Zweierlinie
Secessionsgebäude

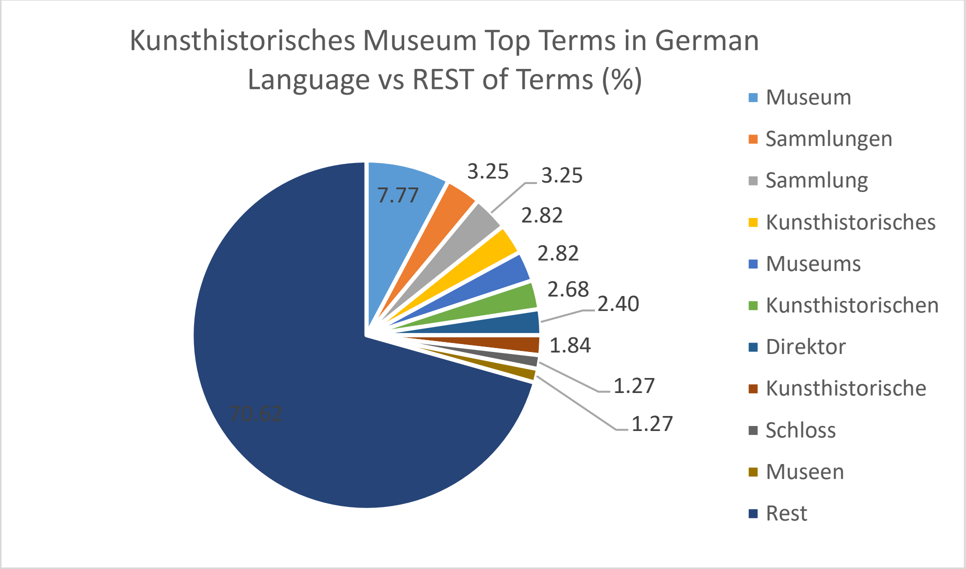

Karlsplatz Top Terms in English Language vs REST of Terms (%)
14.29
4.95
2.75
2.75
2.20
2.20
2.20
2.20
2.20
1.65
62.64
■ Karlsplatz
■ Museum
■ Karlskirche
■ building
■ design
■ square
■ front
■ Park
■ area
■ Operngasse
■ Rest

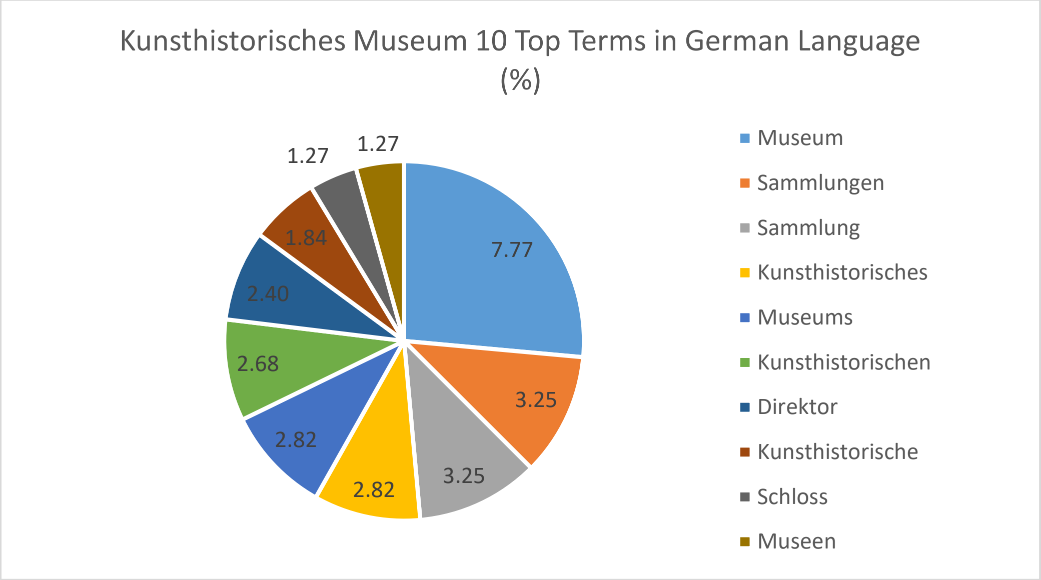

Karlsplatz 10 Top Terms in English Language (%)
1.65
2.20
2.20
2.20
2.20
2.20
2.75
2.75
4.95
14.29
■ Karlsplatz
■ Museum
■ Karlskirche
■ building
■ design
■ square
■ front
■ Park
■ area
■ Operngasse

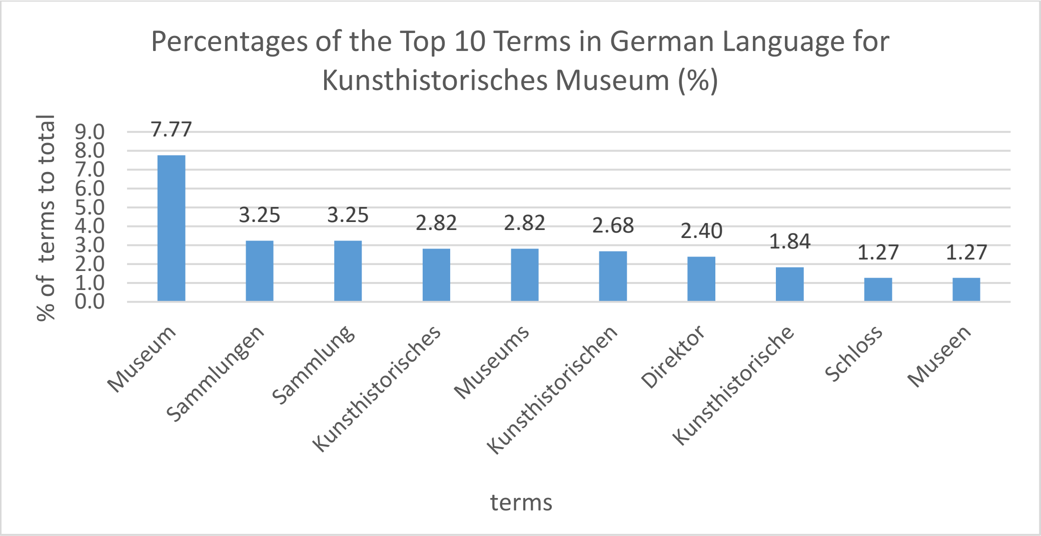

Percentages of the Top Terms in English Language for Karlsplatz (%)
% of terms to total
16
14
12
10
8
6
4
2
0
14.29
4.95
2.75
2.75
2.20
2.20
2.20
2.20
2.20
1.65
Karlsplatz
Museum
Karlskirche
building
design
square
front
Park
area
Operngasse
terms



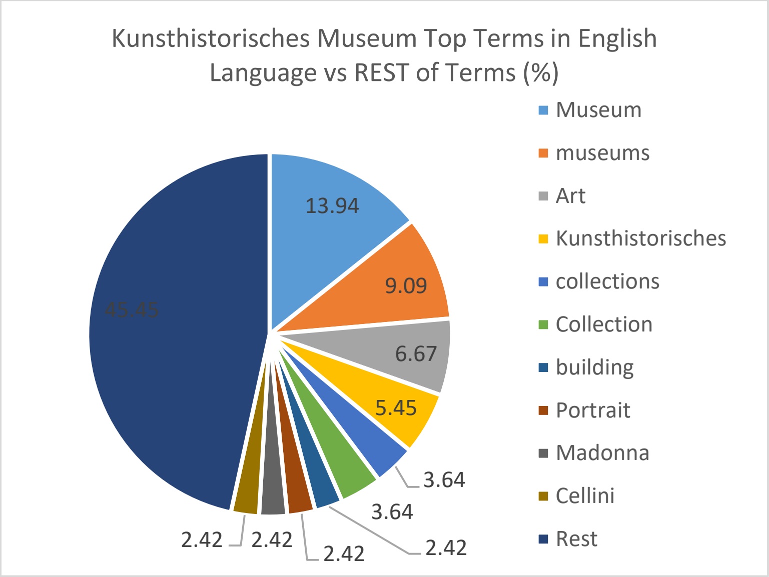
Kunsthistorisches Museum:

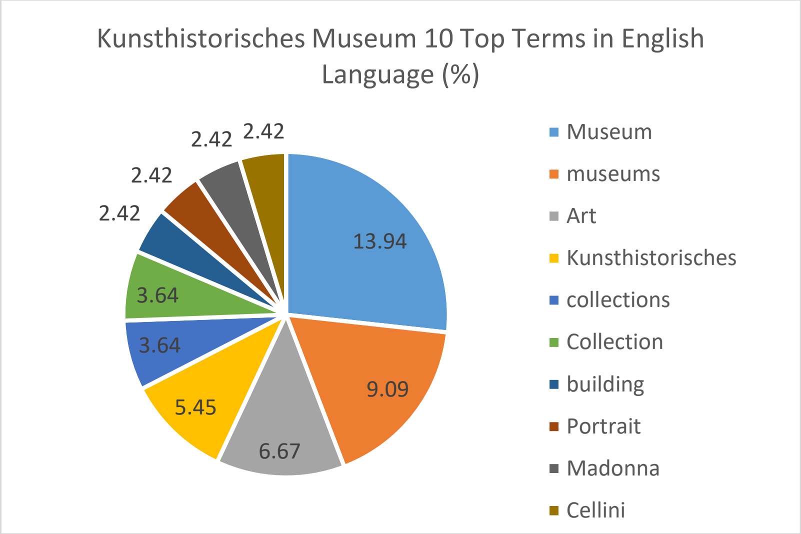


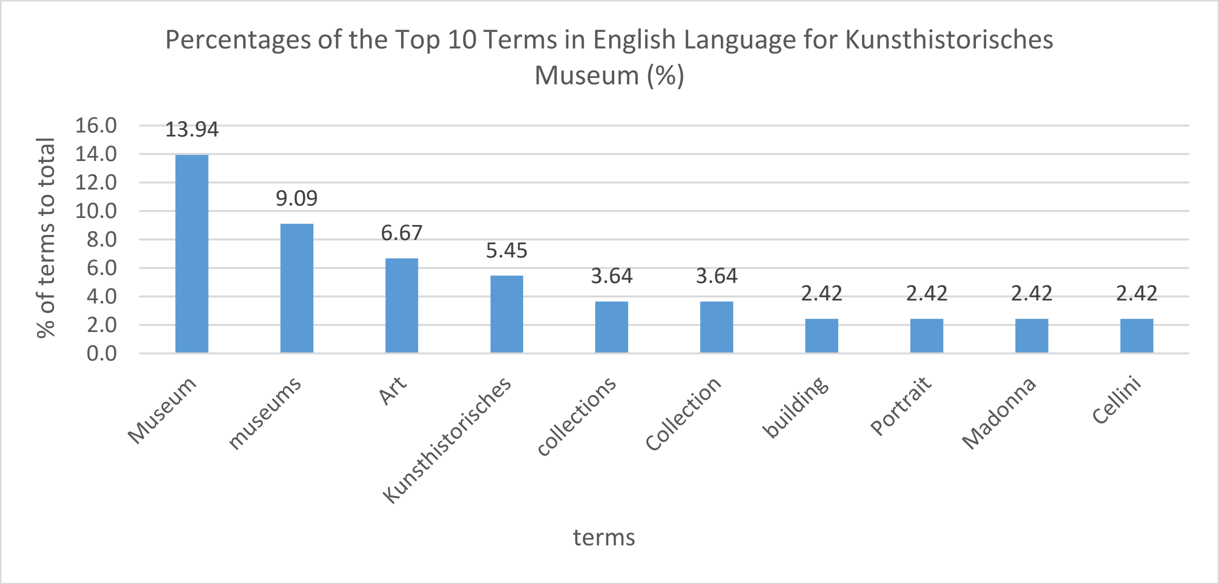


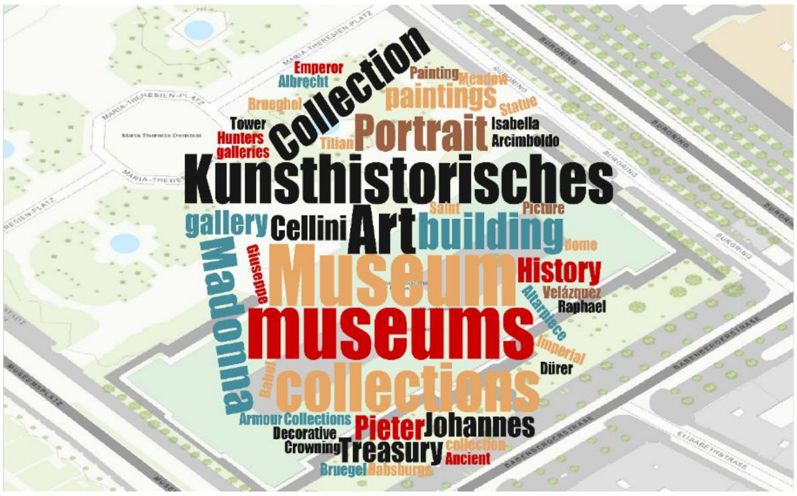

Percentages of the Top 10 Terms in German Language for Kunsthistorisches Museum (%)
% of terms to total
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
7.77
3.25
3.25
2.82
2.82
2.68
2.40
1.84
1.27
1.27
Museum
Sammlungen
Sammlung
Kunsthistorisches
Museums
Kunsthistorischen
Direktor
Kunsthistorische
Schloss
Museen
terms

Waffensammlung
Theatermuseum
Musikinstrumente
Familie
Gemäldegalerie
Hofmusem
Schmidt
Provenienzforschung
Kunsthistorischen
Kunsthistorisches
Sammlung
Bildnis
Museen
Haag
Peter
Joseph
Ambras
Johannes
Haupt
Museum
Kaiser
Franz
Vorsitzender
Hermann
Geschichte
Eike
Sammlungen
Hasenauer
Sabine
Architektur
KHM
Hofburg
Burg
Münzen
Museums
Schloss
Republik
Direktor
kaiserlichen
Rothschild
Schatzkammer
Hauses
Kunsthistorische
Kunstsammlungen
Antikensammlung
Weltmuseum
MARIA-THERESIEN-PLATZ
BURGRING

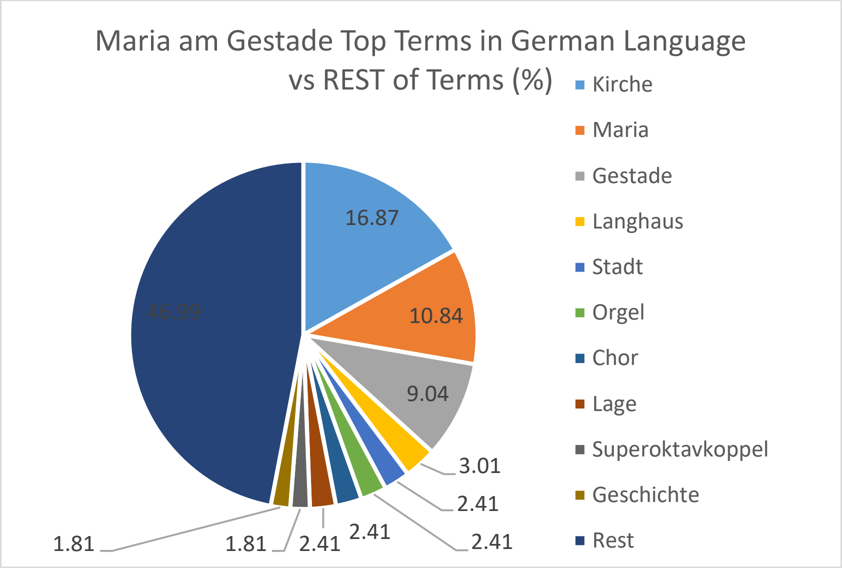

Kunsthistorisches Museum Top Terms in English Language vs REST of Terms (%)
13.94
9.09
6.67
5.45
3.64
3.64
2.42
2.42
2.42
2.42
45.45
■ Museum
■ museums
■ Art
■ Kunsthistorisches
■ collections
■ Collection
■ building
■ Portrait
■ Madonna
■ Cellini
■ Rest

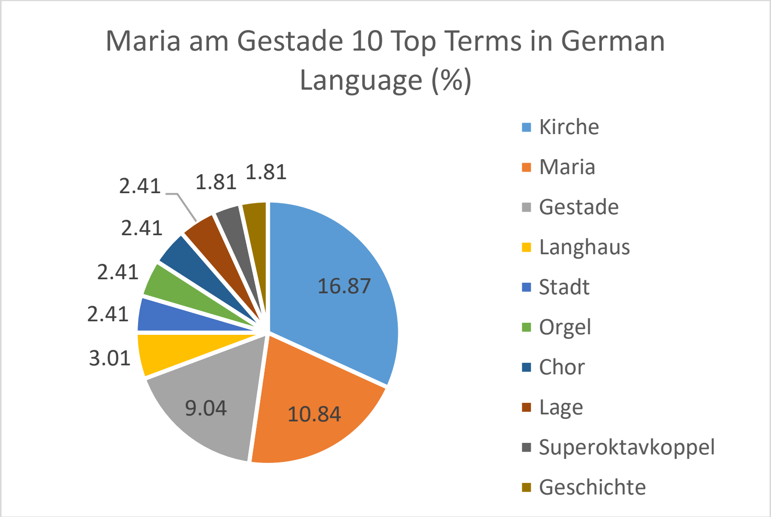

Kunsthistorisches Museum 10 Top Terms in English Language (%)
13.94
9.09
6.67
5.45
3.64
3.64
2.42
2.42
2.42
2.42
Museum
museums
Art
Kunsthistorisches
collections
Collection
building
Portrait
Madonna
Cellini

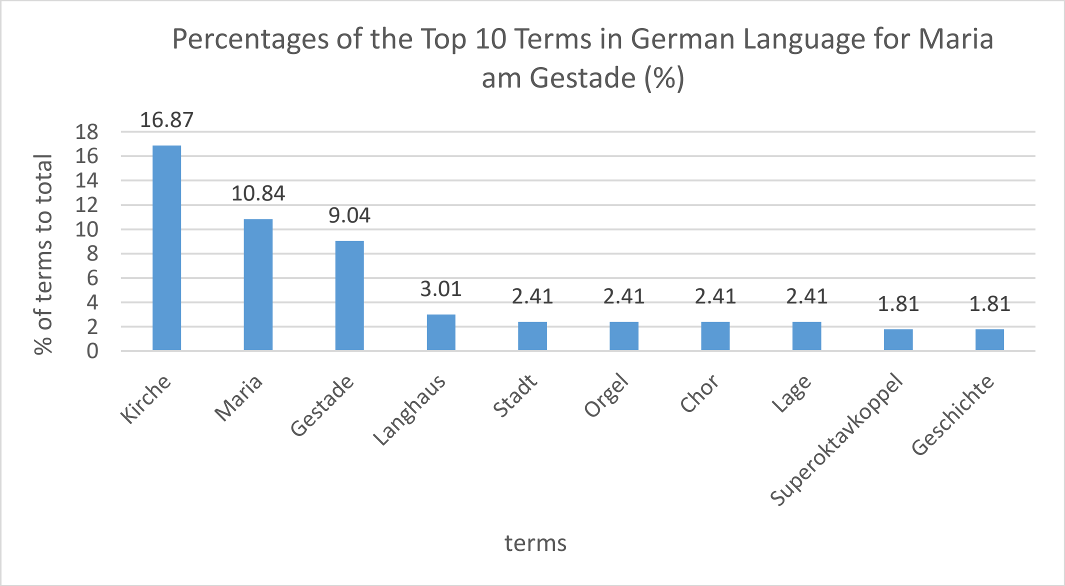

Percentages of the Top 10 Terms in English Language for Kunsthistorisches Museum (%)
% of terms to total
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
13.94
9.09
6.67
5.45
3.64
3.64
2.42
2.42
2.42
2.42
Museum
museums
Art
Kunsthistorisches
collections
Collection
building
Portrait
Madonna
Cellini
terms

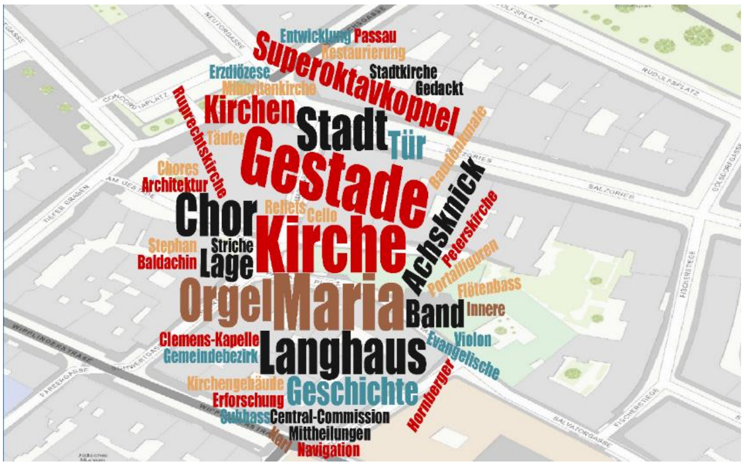

Collection
Emperor
Albrecht
Painting
Meadow
paintings
Brueghel
Statue
Tower
Hunters
galleries
Titian
Portrait
Isabella
Arcimboldo
Kunsthistorisches
Saint
Picture
gallery
Cellini
Art
building
dome
Madonna
Giuseppe
Museum
History
Velázquez
Altarpiece
Raphael
museums
Imperial
Babel
Dürer
collections
Armour
Collections
Pieter
Johannes
Decorative
Crowning
Treasury
collection
Ancient
Bruegel
Habsburgs

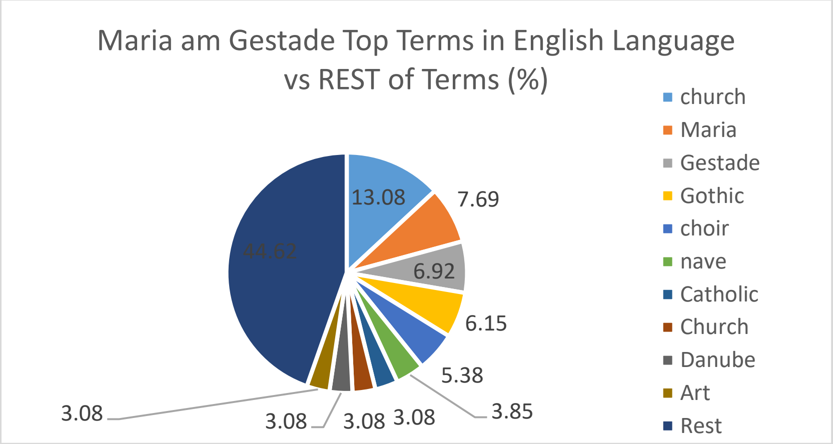
Maria am Gestade:

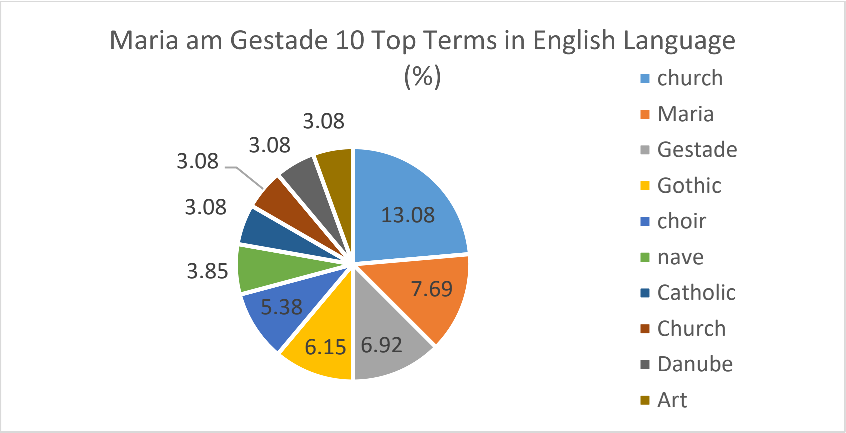

Maria am Gestade Top Terms in German Language vs REST of Terms (%)

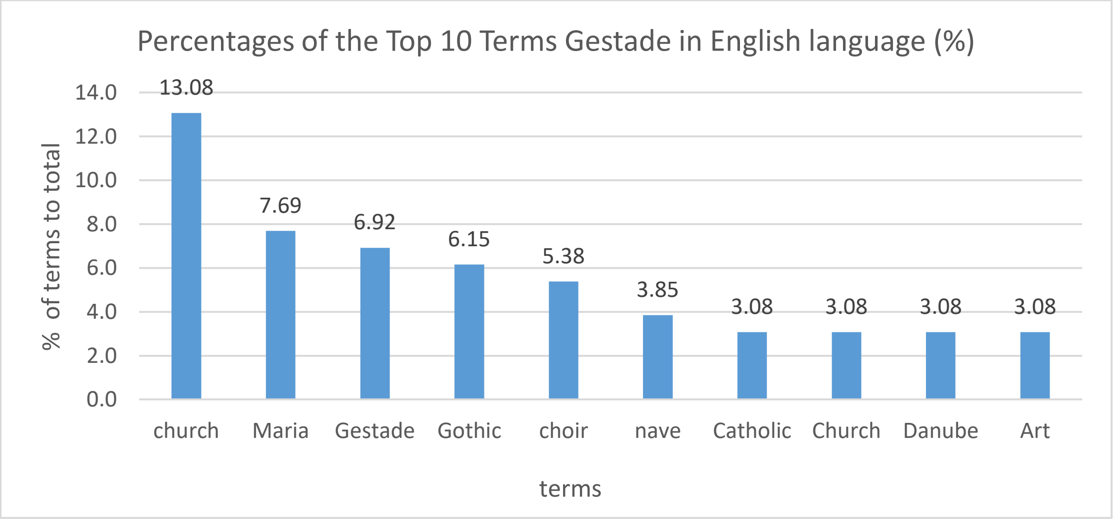

Maria am Gestade 10 Top Terms in German Language (%)

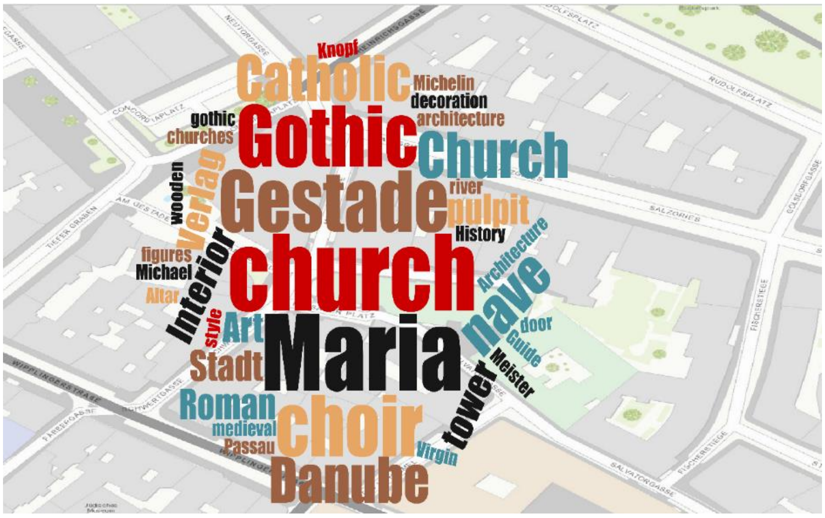

Percentages of the Top 10 Terms in German Language for Maria am Gestade (%)


Entwicklung
Passau
Superoktavkoppel
Restaurierung
Erzdiözese
Stadtkirche
Gedackt
Minoritenkirche
Kirchen
Stadt
Ruprechtskirche
Täufer
Tür
Gestade
Baudenkmale
Chores
Architektur
Achsknick
Chor
Reliefs
Cello
Peterskirche
Stephan
Striche
Kirche
Baldachin
Lage
Portalfiguren
Flötenbass
Orgel
Maria
Band
Innere
Clemens-Kapelle
Violon
Gemeindebezirk
Langhaus
Evangelische
Kirchengebäude
Geschichte
Hornberger
Erforschung
Subbass
Central-Commission
Mittheilungen
Navigation

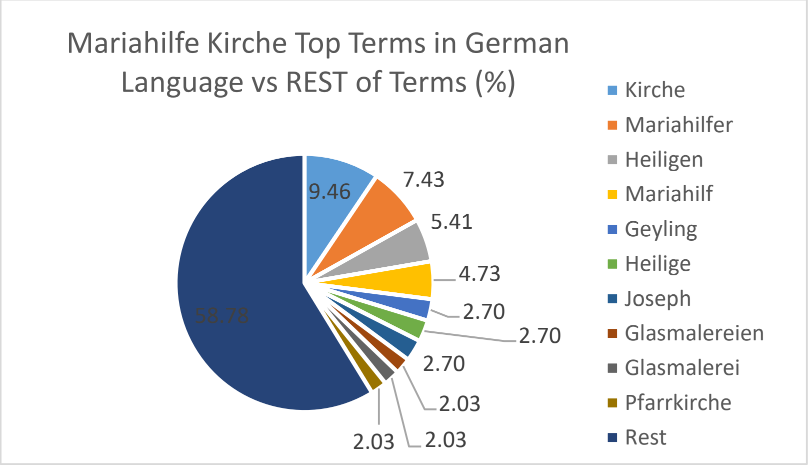

Maria am Gestade Top Terms in English Language vs REST of Terms (%)
13.08
7.69
6.92
6.15
5.38
3.85
3.08
3.08
3.08
3.08
44.62
■ church
■ Maria
■ Gestade
■ Gothic
■ choir
■ nave
■ Catholic
■ Church
■ Danube
■ Art
■ Rest

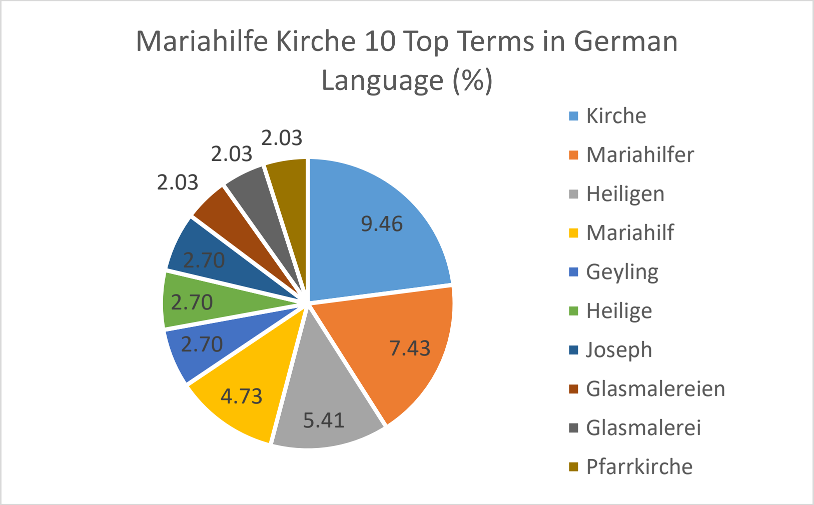

Maria am Gestade 10 Top Terms in English Language (%)
13.08
7.69
6.92
6.15
5.38
3.85
3.08
3.08
3.08
3.08
■ church
■ Maria
■ Gestade
■ Gothic
■ choir
■ nave
■ Catholic
■ Church
■ Danube
■ Art

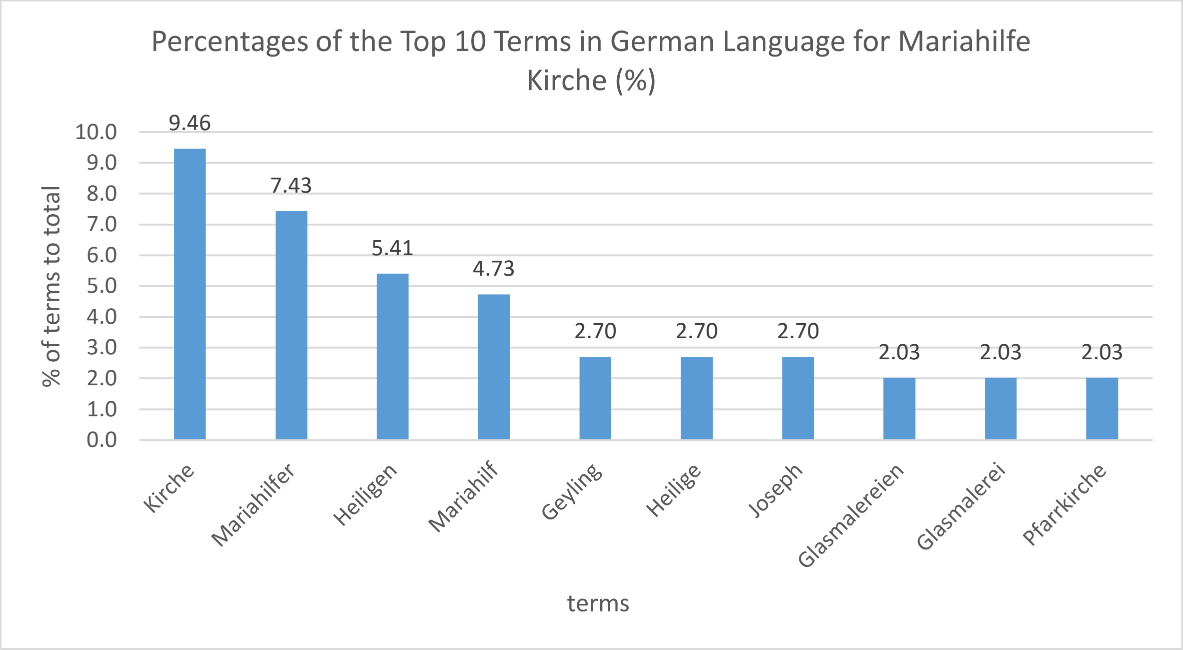

Percentages of the Top 10 Terms Gestade in English language (%)
% of terms to total
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
13.08
7.69
6.92
6.15
5.38
3.85
3.08
3.08
3.08
3.08
church
Maria
Gestade
Gothic
choir
nave
Catholic
Church
Danube
Art
terms



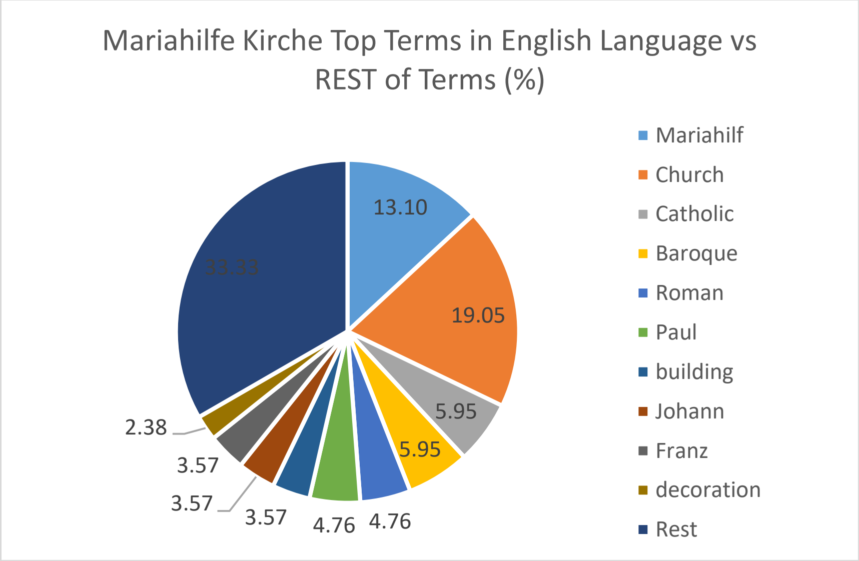
Mariahilfe Kirche:

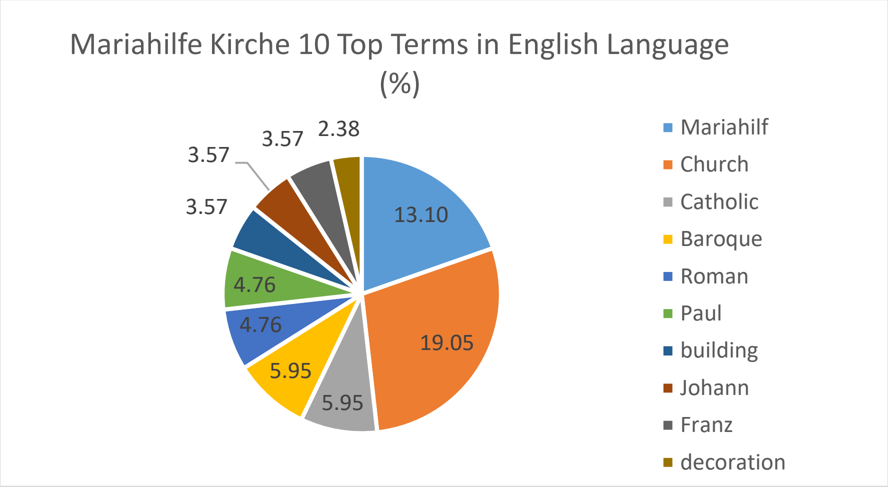


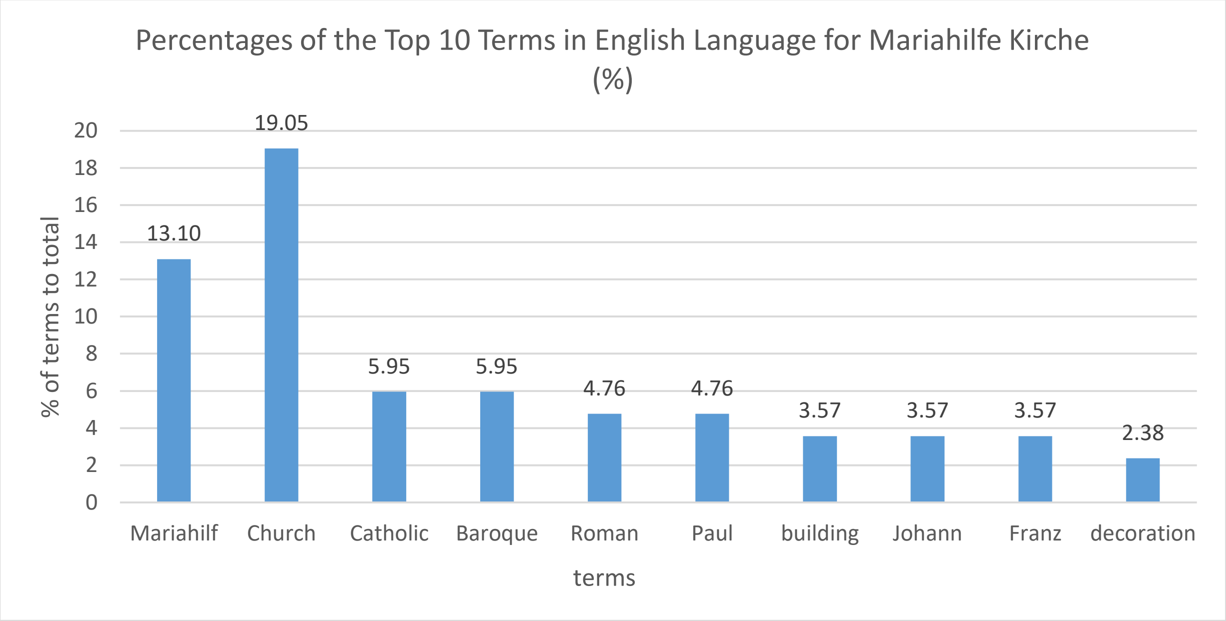


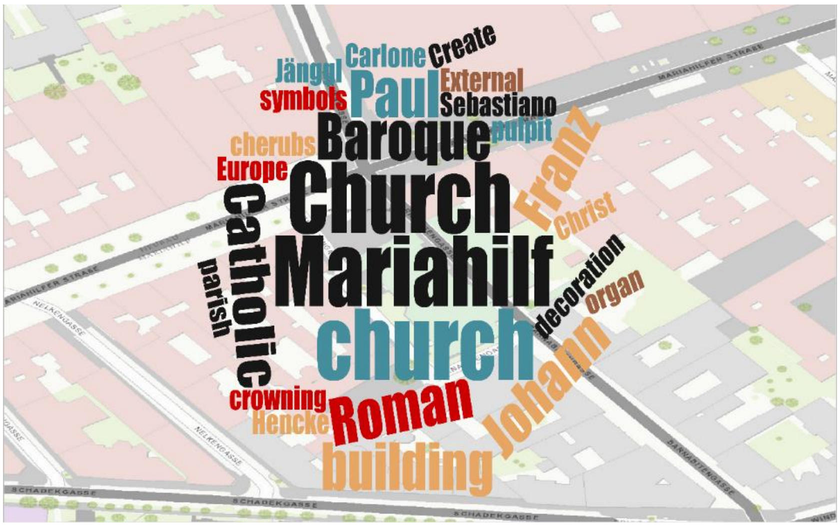

Percentages of the Top 10 Terms in German Language for Mariahilfe Kirche (%)
% of terms to total
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
9.46
7.43
5.41
4.73
2.70
2.70
2.70
2.03
2.03
2.03
Kirche
Mariahilfer
Heiligen
Mariahilf
Geyling
Heilige
Joseph
Glasmalereien
Glasmalerei
Pfarrkirche
terms

Ägypten
Barnabiten
Entwürfe
Geburt
Glasmalerei
Hochaltar
Erzdiözese
Theresa
Heiliger
Heilige
Pfarrkirche
ausgeführt
Mollardgasse
Mariahilf
Wallfahrtskirche
Dreifaltigkeit
Straße
Langhaus
Turm
Mariahilfer
Carl
Mariahilf
Evangelische
Kirche
Maria
Glasmalereien
Flucht
Anna
Leopold
Rudolf
Jahren
Westfassade
Waldemar
restauriert
Augustinus
Heiligen
zwischen
Ignatius
Kirchen
Kind
Geyling
Sakralbauten
Sebastian
Gemeindebezirk
Änderungen
Fenster
Joseph
Barbara
Christi
Geschichte
Amalia

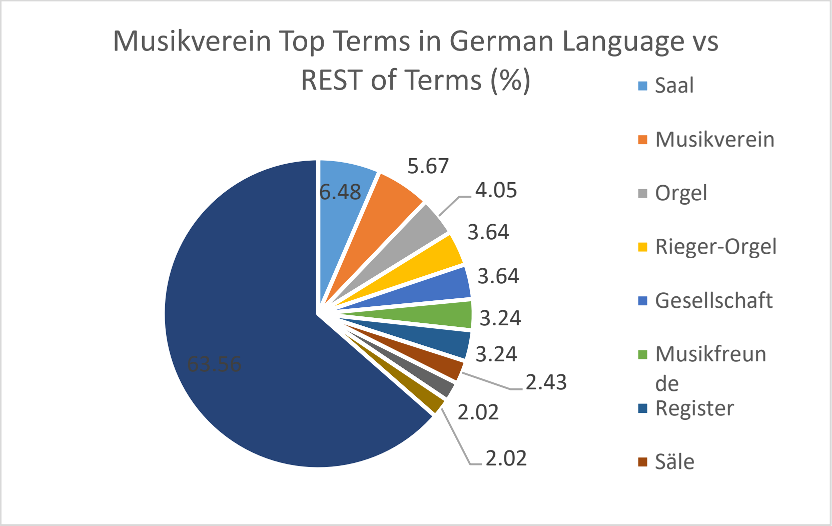

Mariahilfe Kirche Top Terms in English Language vs REST of Terms (%)
13.10
19.05
5.95
5.95
4.76
4.76
3.57
3.57
3.57
2.38
33.33
Mariahilf
Church
Catholic
Baroque
Roman
Paul
building
Johann
Franz
decoration
Rest

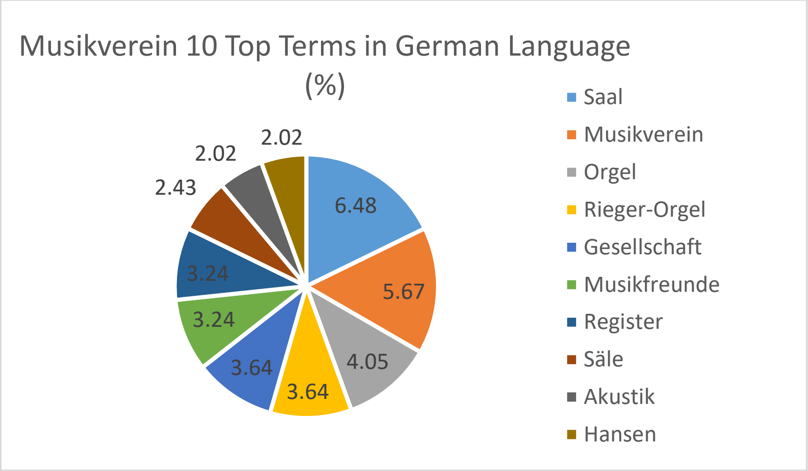

Mariahilfe Kirche 10 Top Terms in English Language (%)
13.10
19.05
5.95
5.95
4.76
4.76
3.57
3.57
3.57
2.38
Mariahilf
Church
Catholic
Baroque
Roman
Paul
building
Johann
Franz
decoration

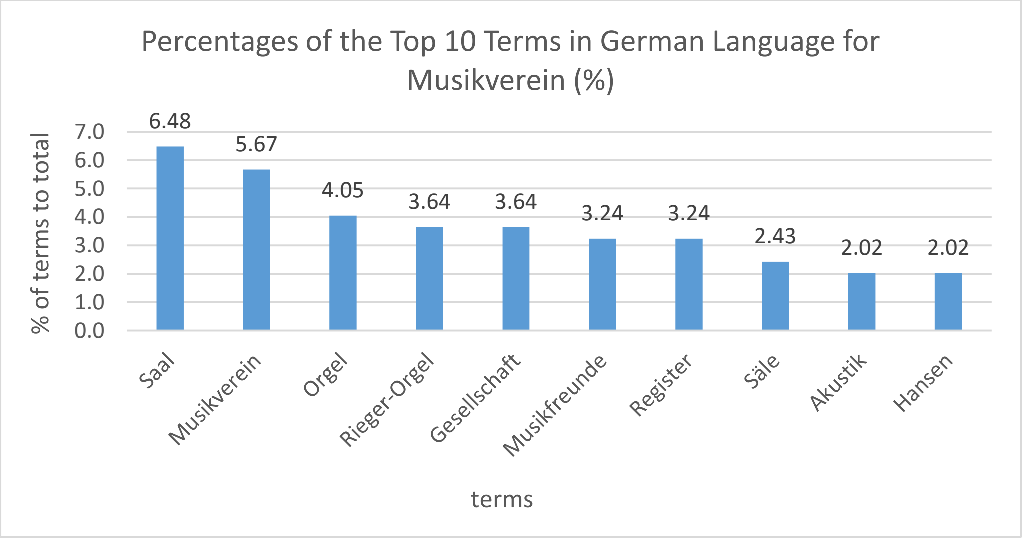

Percentages of the Top 10 Terms in English Language for Mariahilfe Kirche (%)
% of terms to total
0
2
4
6
8
10
12
14
16
18
20
13.10
19.05
5.95
5.95
4.76
4.76
3.57
3.57
3.57
2.38
Mariahilf
Church
Catholic
Baroque
Roman
Paul
building
Johann
Franz
decoration
terms

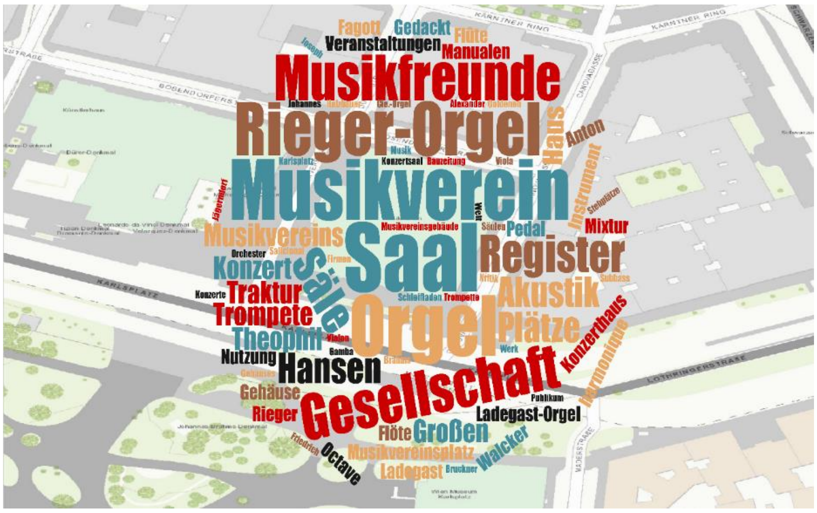

Carlone
Create
Jänggl
Paul
External
symbols
Sebastiano
cherubs
Baroque
pulpit
Europe
Church
Franz
Christ
Catholic
Mariahilf
parish
decoration
organ
church
crowning
Hencke
Roman
Johann
building

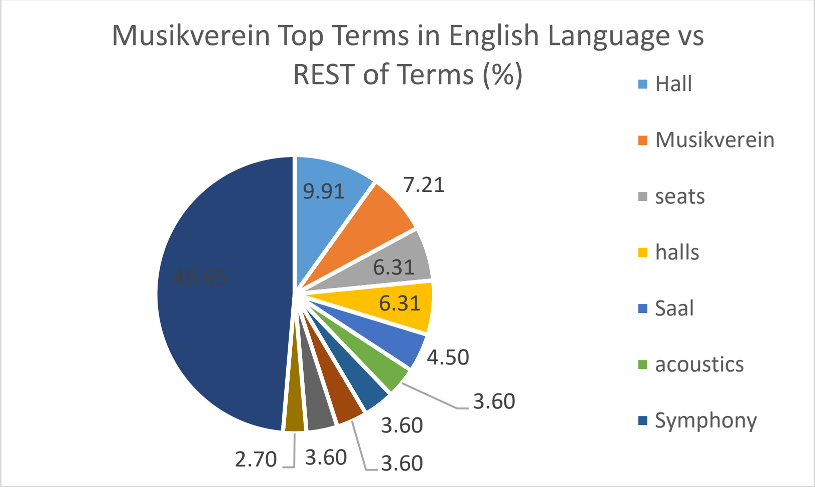
Musikverein:

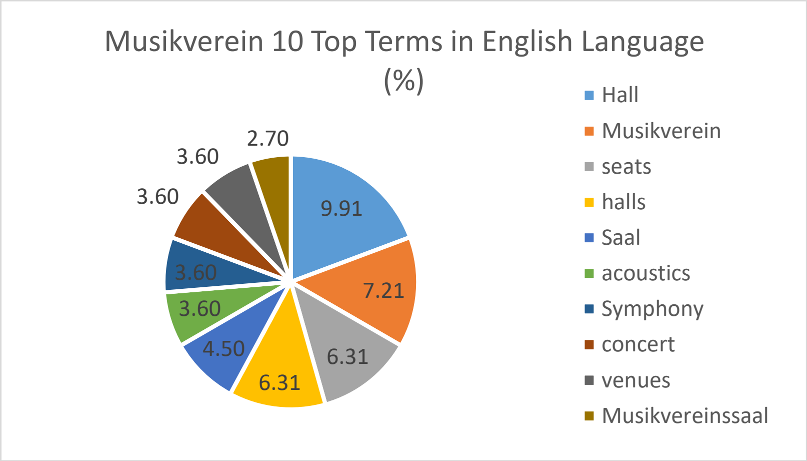


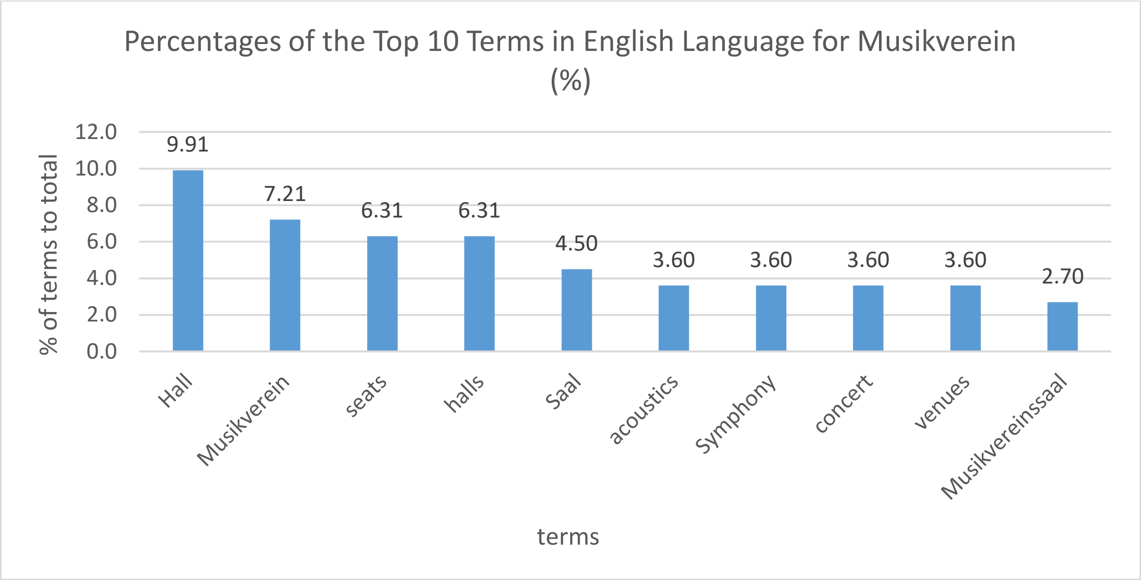


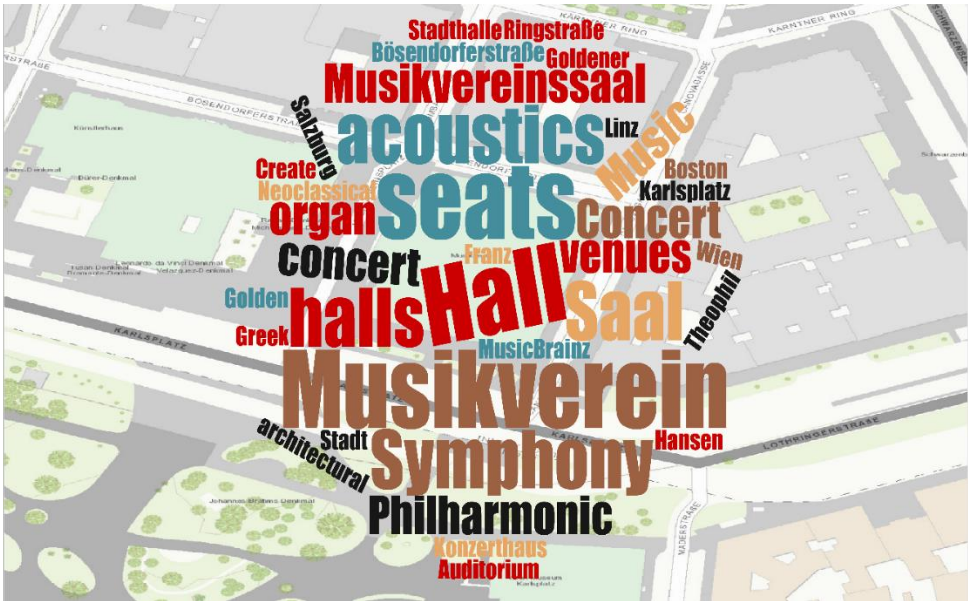



Fagott
Gedackt
Flöte
Veranstaltungen
Manualen
Musikfreunde
Rieger-Orgel
Haus
Anton
Musikverein
Instrument
Mixtur
Musikvereins
Saal
Register
Konzert
Säle
Pedal
Traktur
Akustik
Trompete
Orgel
Plätze
Konzerthaus
Theophil
Nutzung
Hansen
Gesellschaft
Harmonie
Gehäuse
Rieger
Ladegast-Orgel
Flöte
Großen
Musikvereinsplatz
Walcker
Octave
Ladegast

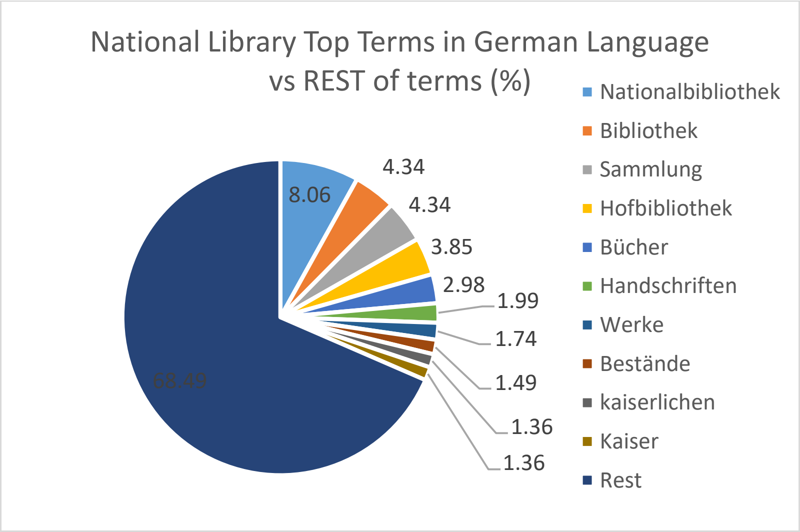

Musikverein Top Terms in English Language vs REST of Terms (%)
9.91
7.21
6.31
6.31
4.50
3.60
3.60
3.60
3.60
2.70
48.65
Hall
Musikverein
seats
halls
Saal
acoustics
Symphony

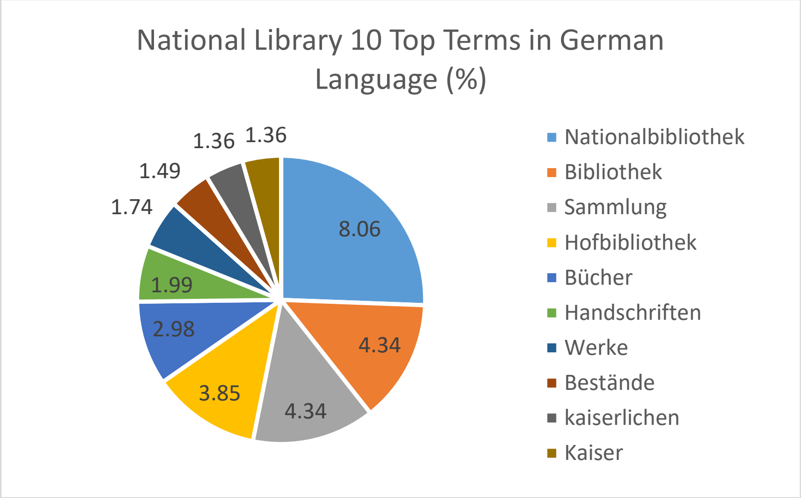

Musikverein 10 Top Terms in English Language (%)
9.91
7.21
6.31
6.31
4.50
3.60
3.60
3.60
3.60
2.70
Hall
Musikverein
seats
halls
Saal
acoustics
Symphony
concert
venues
Musikvereinssaal

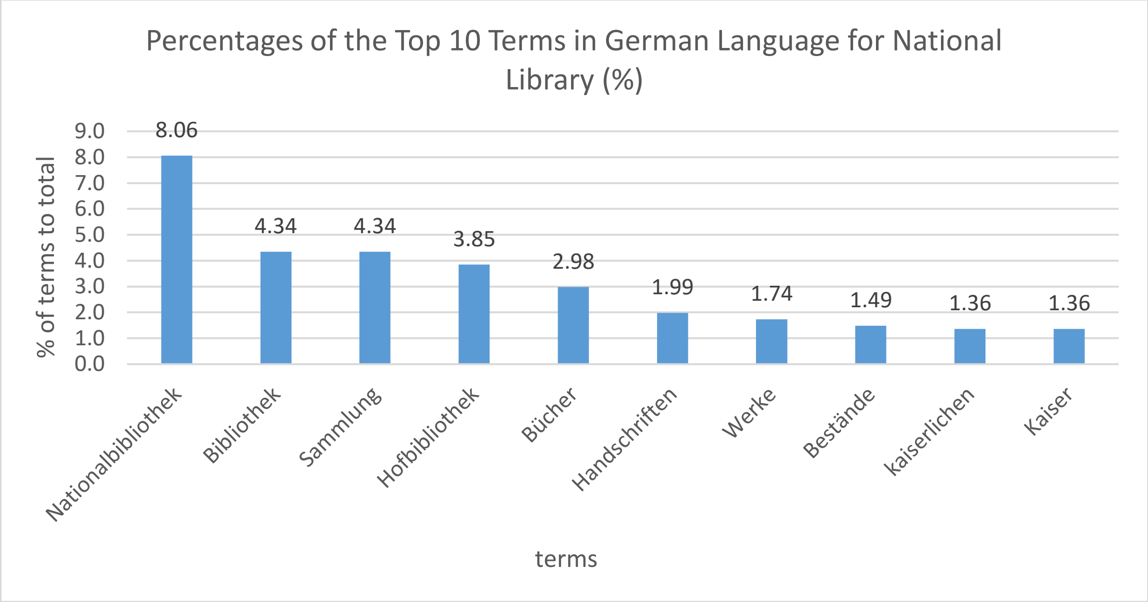




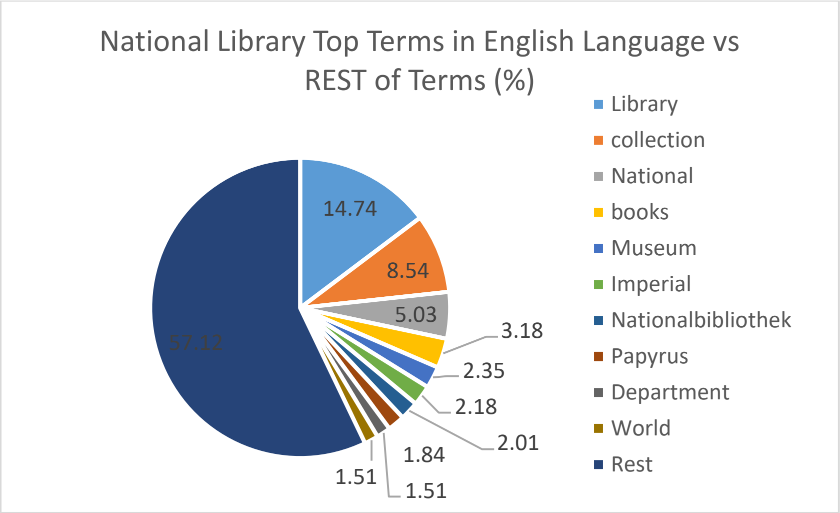
National Library:



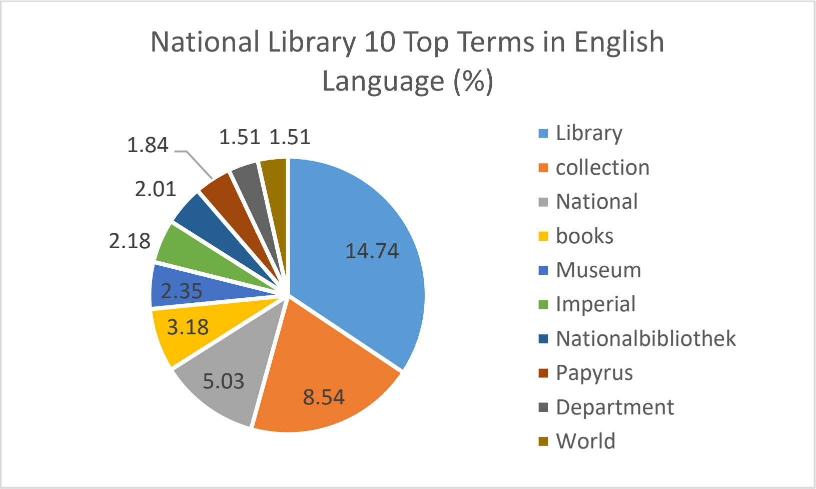

National Library 10 Top Terms in German Language (%)
8.06
4.34
4.34
3.85
2.98
1.99
1.74
1.49
1.36
1.36
■ Nationalbibliothek
■ Bibliothek
■ Sammlung
■ Hofbibliothek
■ Bücher
■ Handschriften
■ Werke
■ Bestände
■ kaiserlichen
■ Kaiser


Percentages of the Top 10 Terms in German Language for National Library (%)
% of terms to total
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
8.06
4.34
4.34
3.85
2.98
1.99
1.74
1.49
1.36
1.36
Nationalbibliothek
Bibliothek
Sammlung
Hofbibliothek
Bücher
Handschriften
Werke
Bestände
kaiserlichen
Kaiser
terms

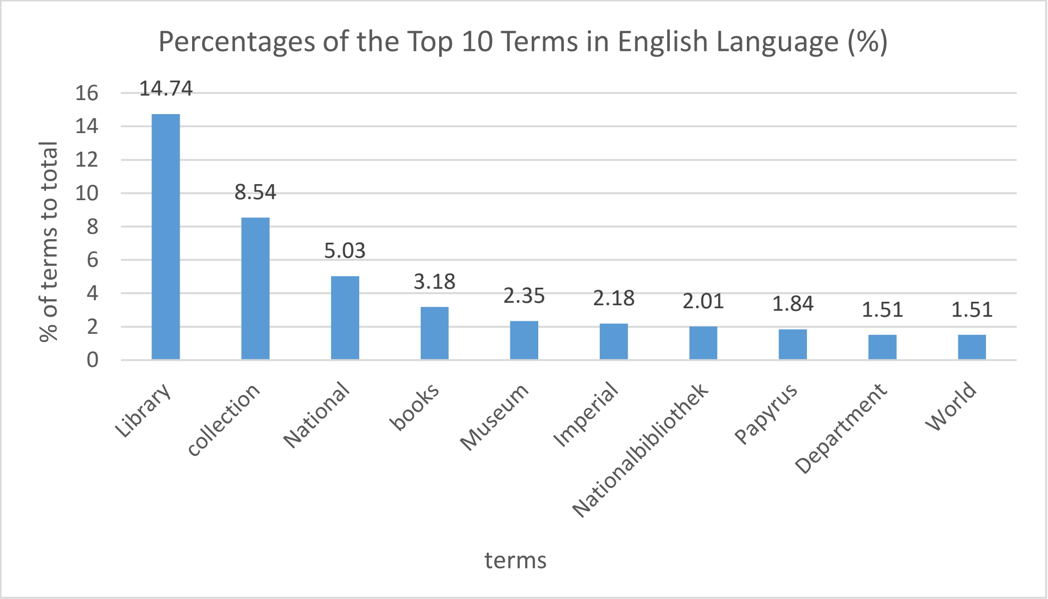
Republik
historischen
Jahrhundert
Bücher
Sammlungen
weltweit
Johannes
Jahr
Josefsplatz
Prunksaal
Hofburg
Bestände
Teil
wertvollen
Maria
Hofbibliothek
Habsburger
Besitz
Bibliothek
umfasst
Palais
Blotius
Gebäude
Burg
Bände
Franz
Fischer
wertvolle
Sammlung
Tod
Millionen
Johann
Handschriften
Bestand
Juni
Kaiser
kaiserliche
Werke
Katalog
Geschichte
Esperantomuseum



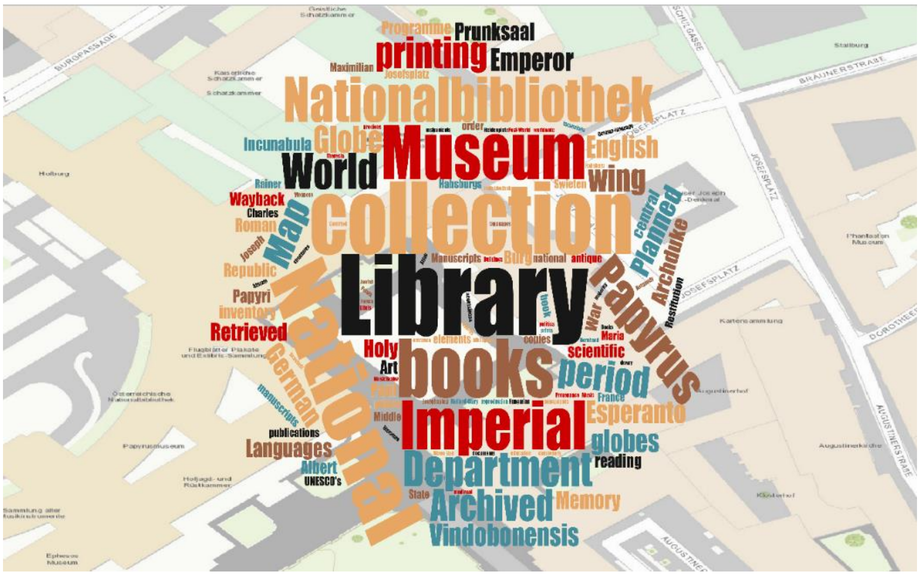
National Library Top Terms in English Language vs REST of Terms (%)



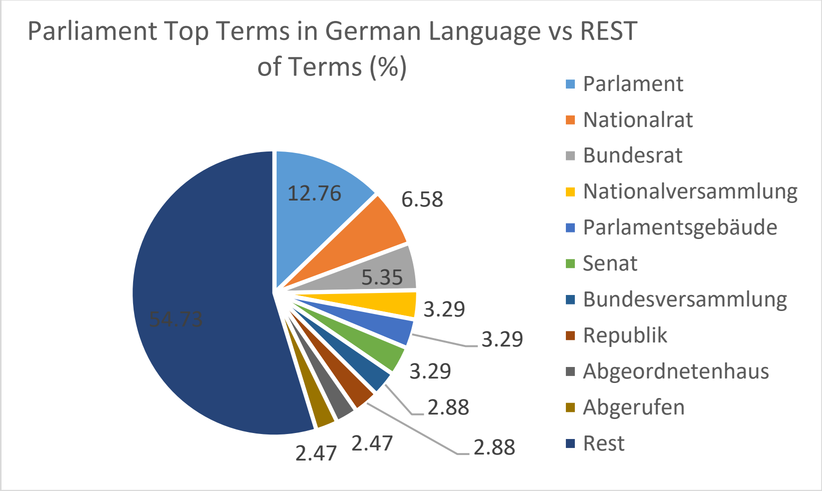
National Library 10 Top Terms in English Language (%)

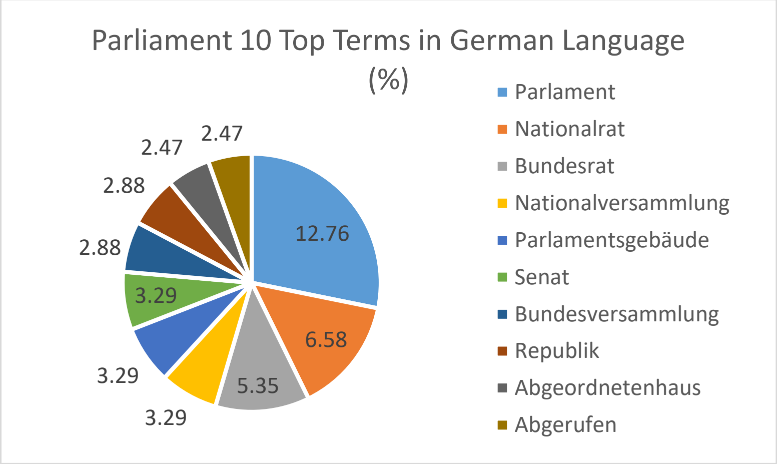


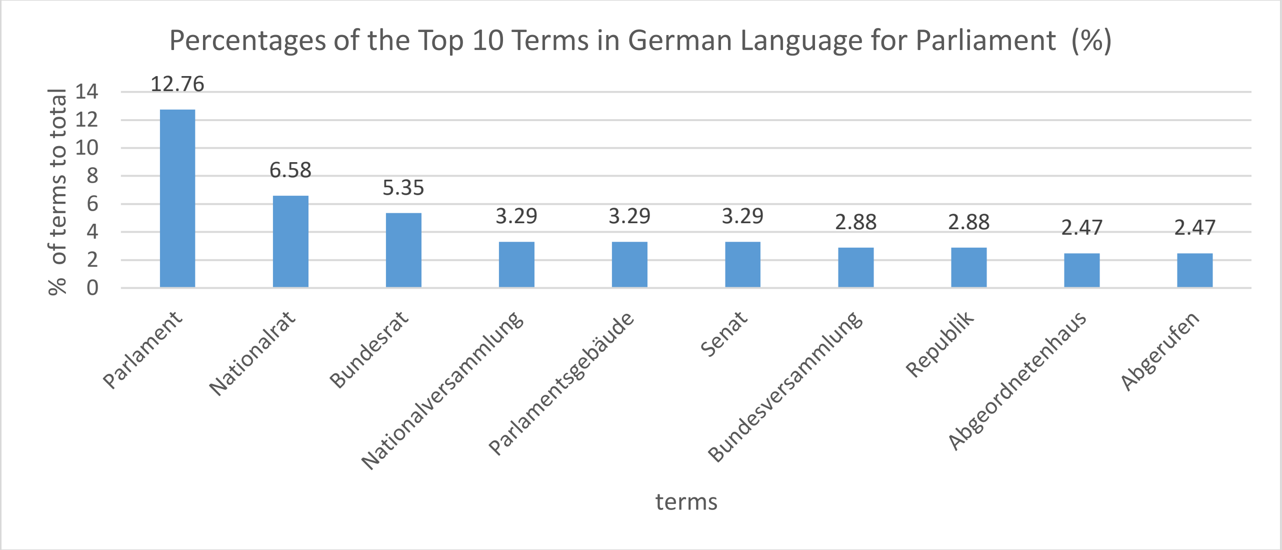
Percentages of the Top 10 Terms in English Language (%)



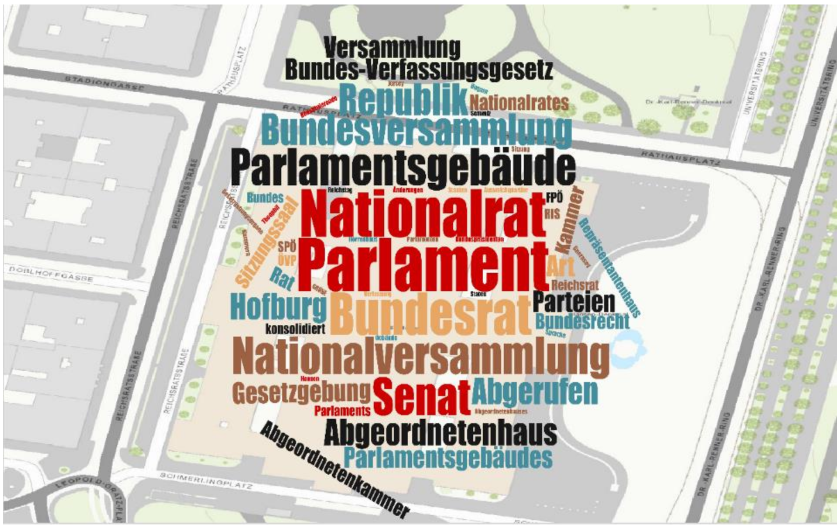
Parliament:

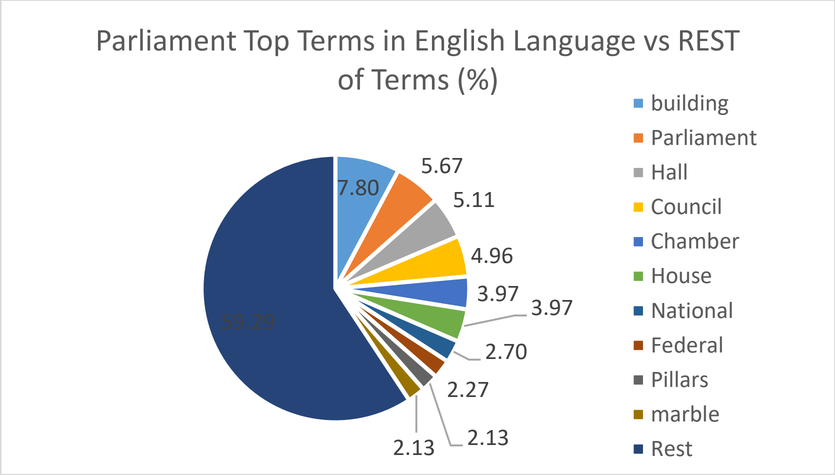

Parliament Top Terms in German Language vs REST of Terms (%)


Parliament 10 Top Terms in German Language (%)

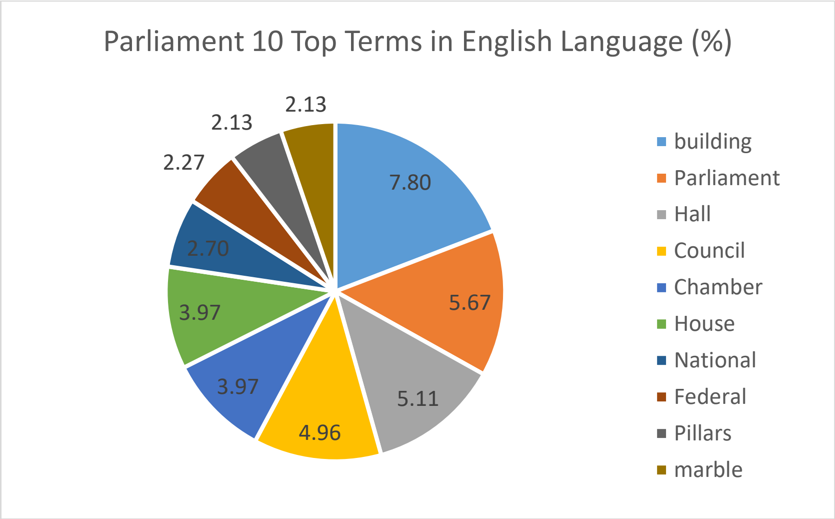


Percentages of the Top 10 Terms in German Language for Parliament (%)

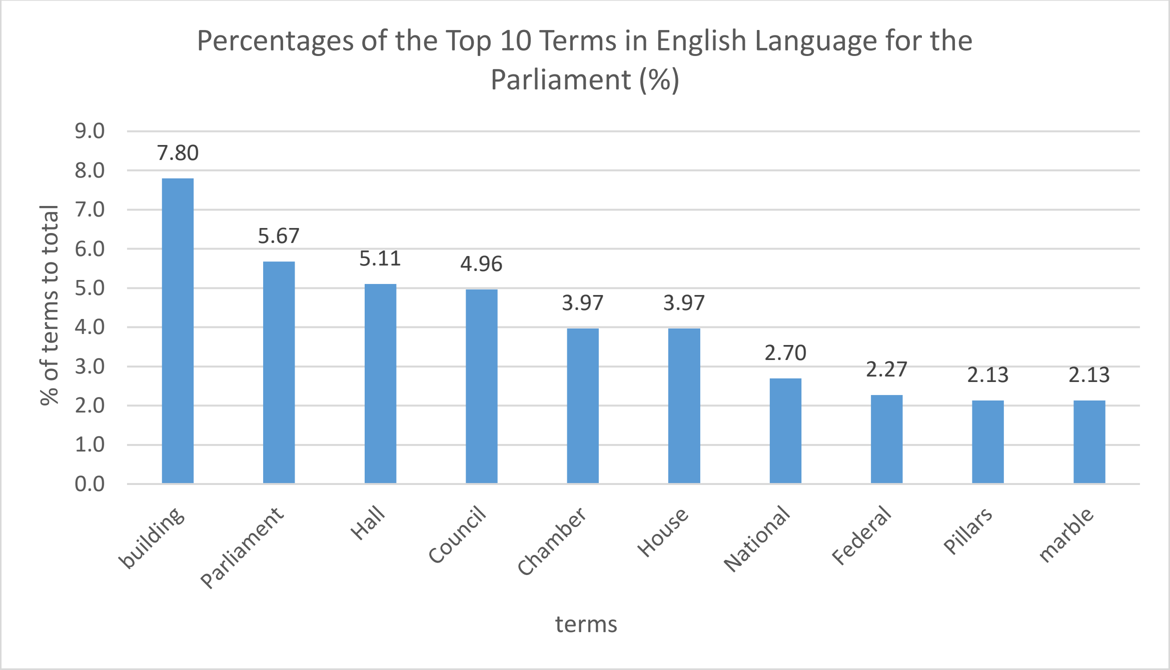




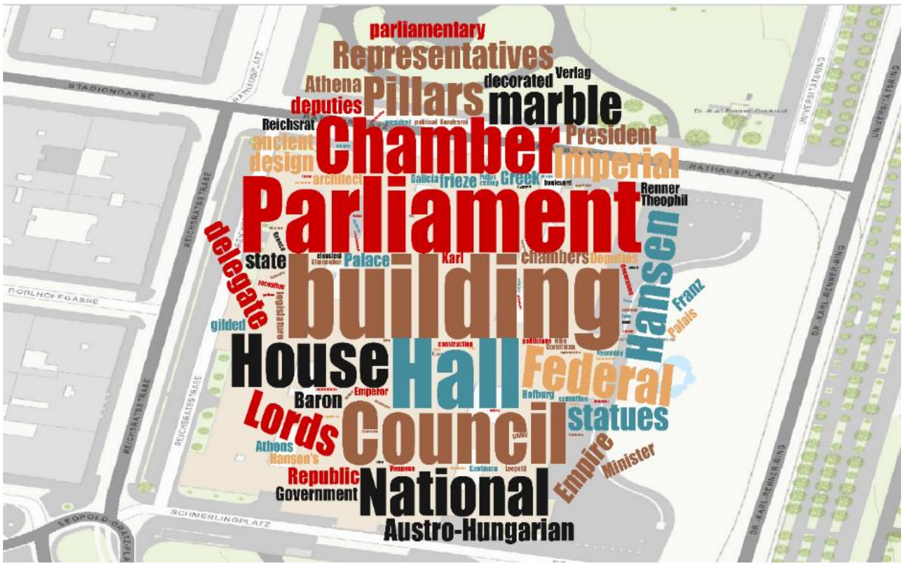
Parliament Top Terms in English Language vs REST of Terms (%)



Parliament 10 Top Terms in English Language (%)
7.80
5.67
5.11
4.96
3.97
3.97
2.70
2.27
2.13
2.13
building
Parliament
Hall
Council
Chamber
House
National
Federal
Pillars
marble

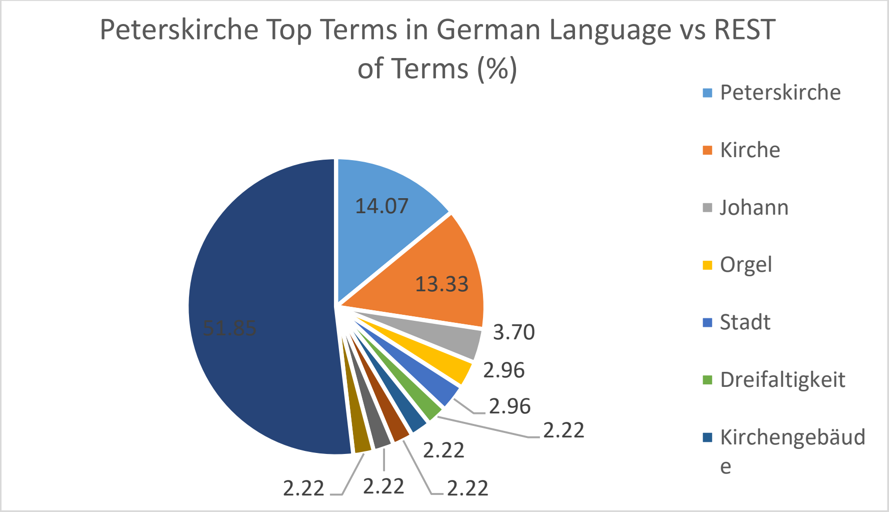


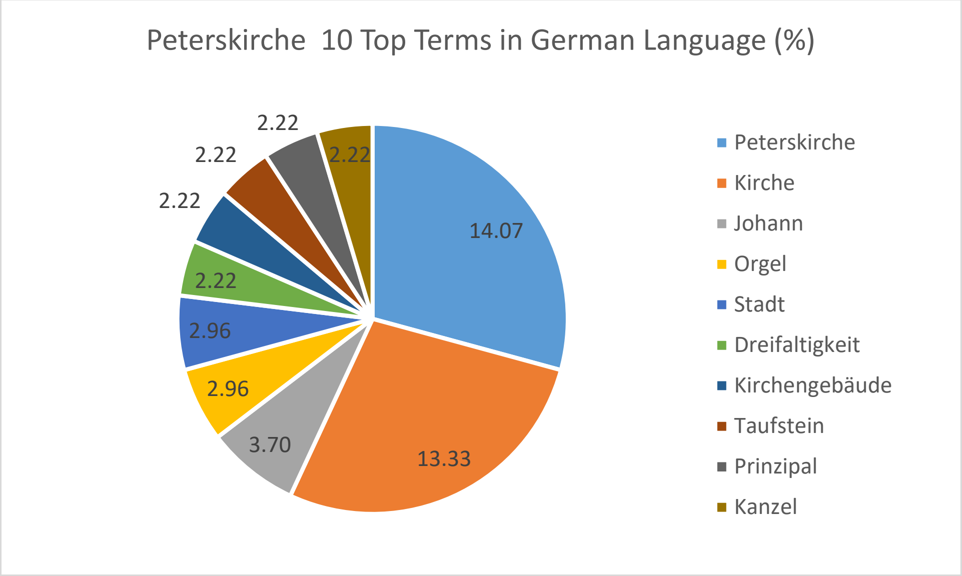
Percentages of the Top 10 Terms in English Language for the Parliament (%)
% of terms to total
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
7.80
5.67
5.11
4.96
3.97
3.97
2.70
2.27
2.13
2.13
building
Parliament
Hall
Council
Chamber
House
National
Federal
Pillars
marble
terms

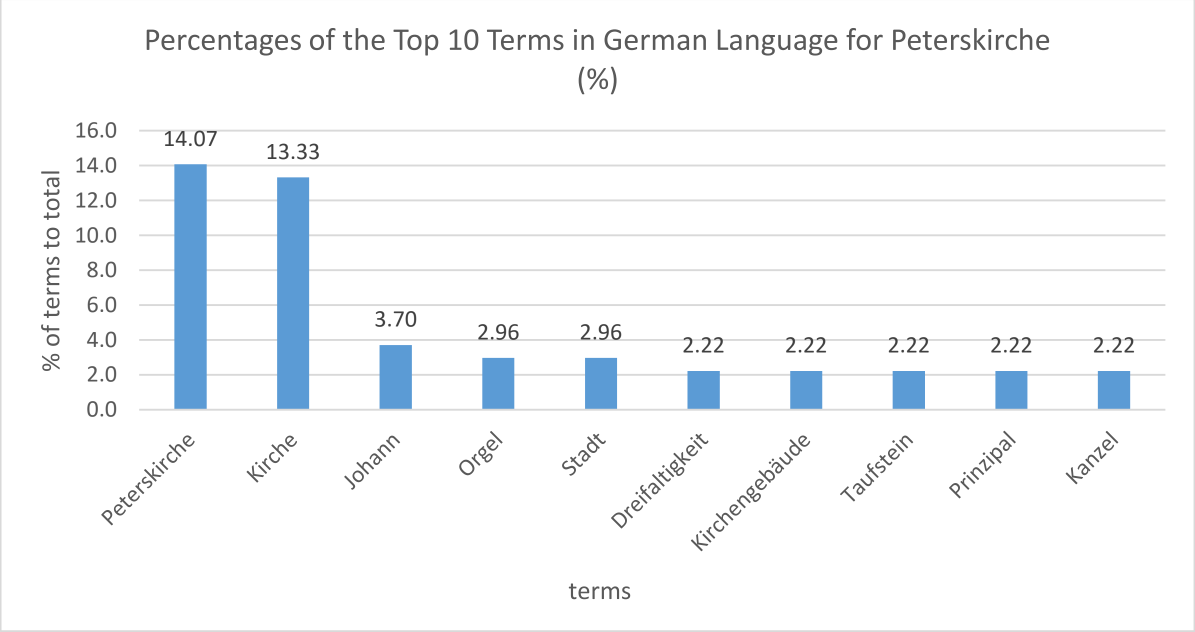


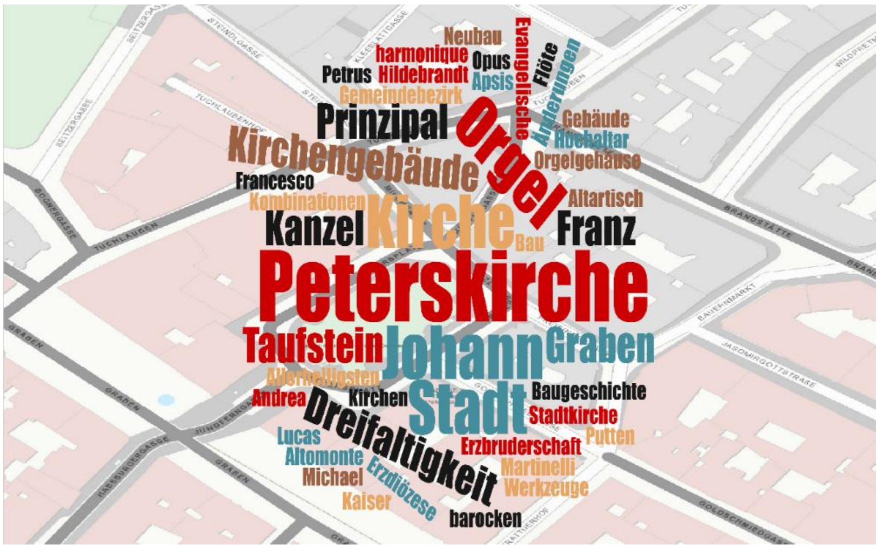
parliamentary
Representatives
Athena
Pillars
decorated
Verlag
deputies
marble
Reichsrat
ancient
Chamber
President
design
architect
Galicia
frieze
Greek
Imperial
Renner
Theophil
Parliament
delegate
state
Palace
Karl
chambers
Hansen
building
Franz
Palais
gilded
House
Hall
Federal
Baron
Emperor
Lords
Council
statues
Athens
Hansen's
Empire
Minister
Republic
Government
National
Austro-Hungarian

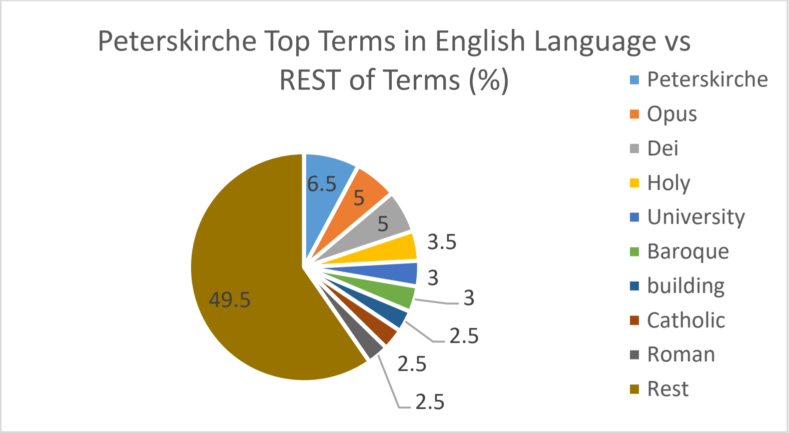
Peterskirche:

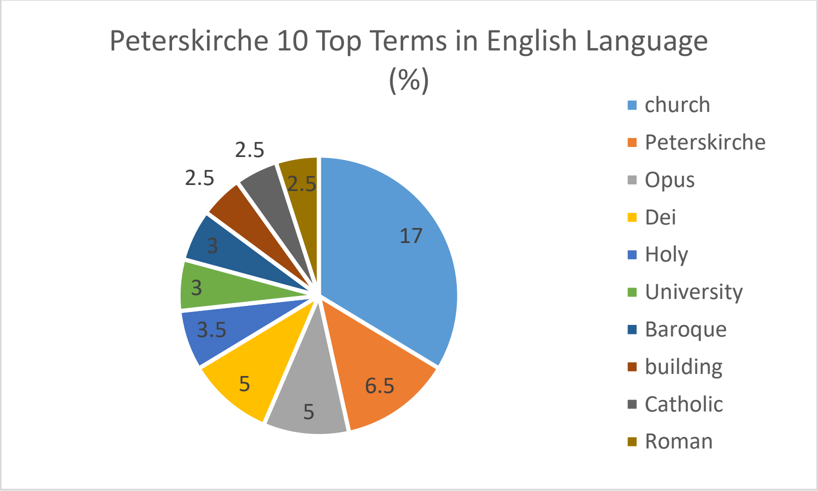


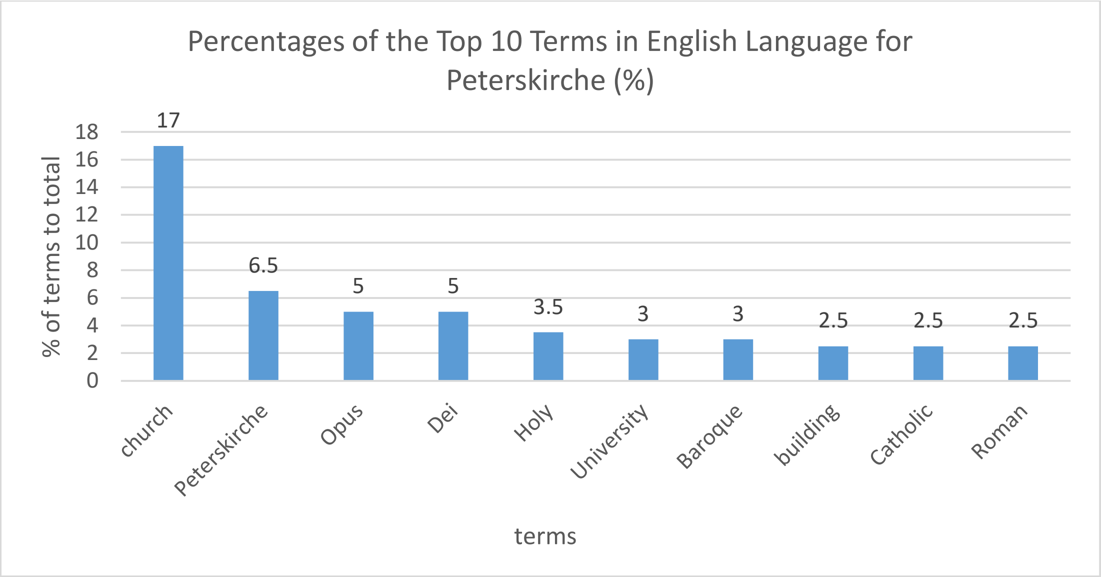





Neubau
harmonique
Opus
Evangelische
Flöte
Petrus
Hildebrandt
Apsis
Änderungen
Gemeindebezirk
Prinzipal
Orgel
Gebäude
Hochaltar
Kirchengebäude
Orgelgehäuse
Francesco
Kombinationen
Altartisch
Kanzel
Kirche
Bau
Franz
Peterskirche
Taufstein
Johann
Graben
Allerheiligsten
Andrea
Kirchen
Stadt
Baugeschichte
Stadtkirche
Dreifaltigkeit
Lucas
Putten
Erzbruderschaft
Altomonte
Erzdiözese
Martinelli
Michael
Werkzeuge
Kaiser
barocken

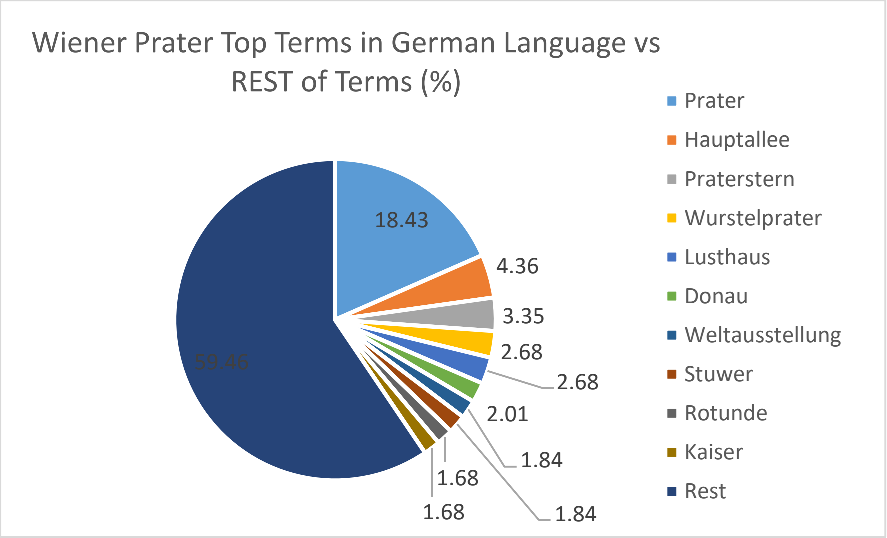

Peterskirche Top Terms in English Language vs REST of Terms (%)
6.5
5
5
3.5
3
3
2.5
2.5
2.5
49.5
Peterskirche
Opus
Dei
Holy
University
Baroque
building
Catholic
Roman
Rest

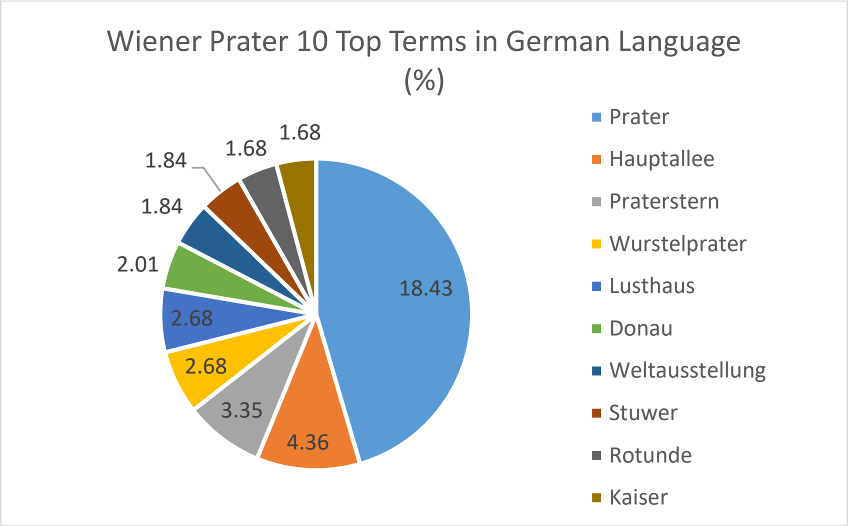

Peterskirche 10 Top Terms in English Language (%)
17
6.5
5
5
3.5
3
3
2.5
2.5
2.5
church
Peterskirche
Opus
Dei
Holy
University
Baroque
building
Catholic
Roman

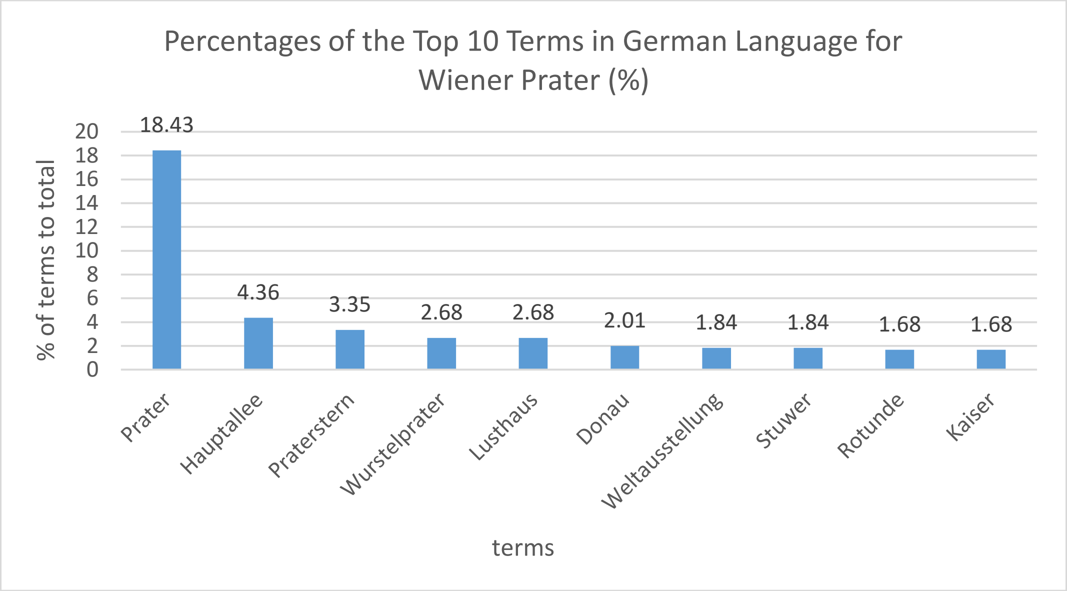


University
chapel
Austria
Peter
Dei
Baroque
pulpit
Peterskirche
church
altarpiece
paintings
Michael
Peter's
Trinity
building
Holy
Opus
Roman
consecrated
interior
Church
artist
dome
Graben
painted
Retrieved
German
Catholic

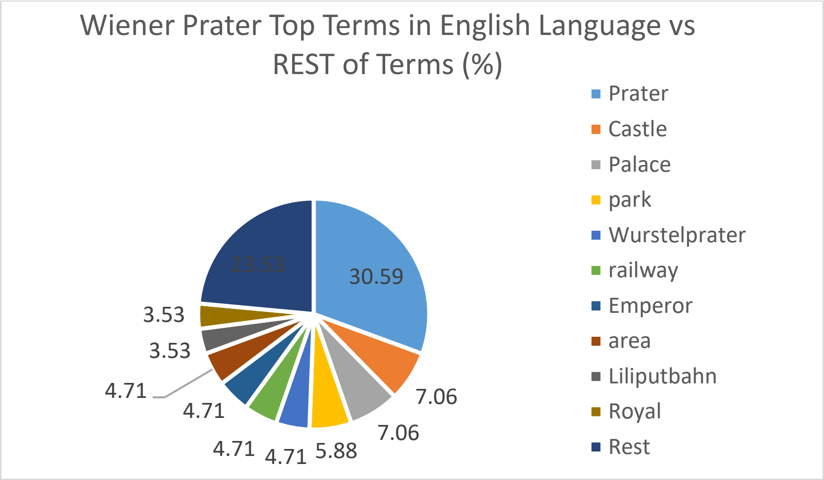
Wiener Prater:

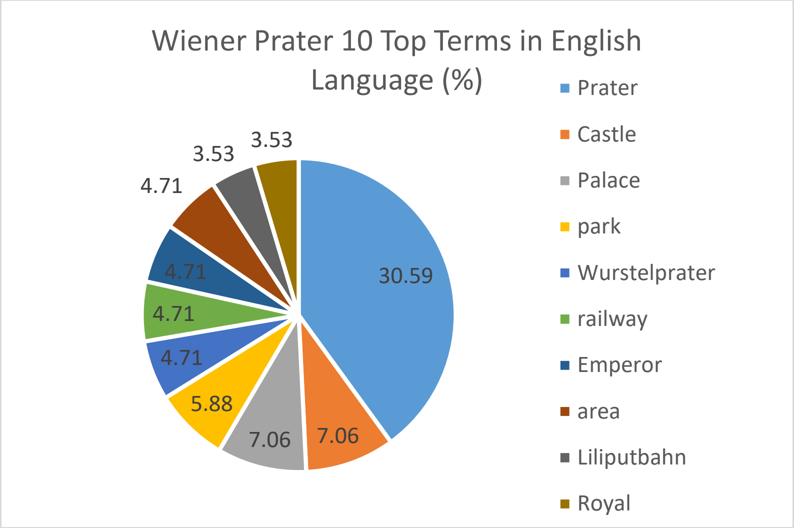


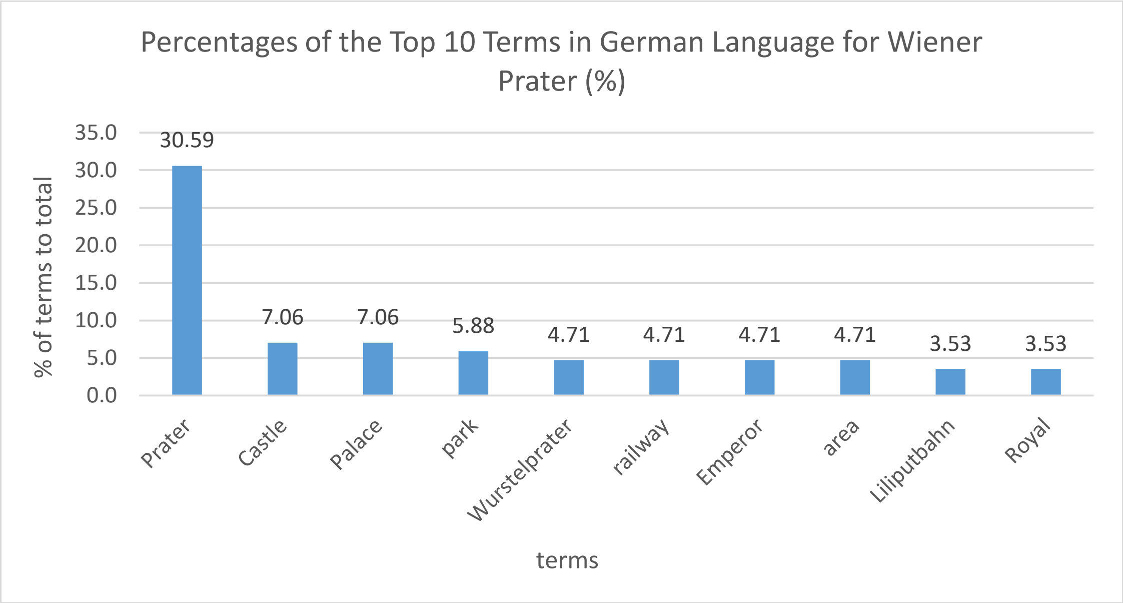

Wiener Prater 10 Top Terms in German Language (%)
18.43
4.36
3.35
2.68
2.68
2.01
1.84
1.84
1.68
1.68
Prater
Hauptallee
Praterstern
Wurstelprater
Lusthaus
Donau
Weltausstellung
Stuwer
Rotunde
Kaiser


Percentages of the Top 10 Terms in German Language for Wiener Prater (%)
% of terms to total
20
18
16
14
12
10
8
6
4
2
0
18.43
4.36
3.35
2.68
2.68
2.01
1.84
1.84
1.68
1.68
Prater
Hauptallee
Praterstern
Wurstelprater
Lusthaus
Donau
Weltausstellung
Stuwer
Rotunde
Kaiser
terms

Messegelände
Mauthnerwasser
Naturdenkmal
Wirtschaftsuniversität
Trabrennbahn
Meiereistraße
Geschichte
Peter
Hans
Weltausstellung
Praterstadion
Richtung
Wurstelprater
Jesuitenwiese
Planetarium
Praterstern
Bezirk
Schloss
Feuerwerk
Liliputbahn
Prater
Donau
Stuwer
Kaiser
Kaffeehaus
Lusthauswasser
Zirkus
Architekt
Hauptallee
Donaukanal
Brücke
Rotunde
Grün
Wien
Freudenau
Wiesen
Gewässer
Lusthaus
Messe
Stadion
Krebsenwasser
Fischer
Heustadelwasser
Galopprennbahn
Konstantinhügel
Ausstellungsstraße

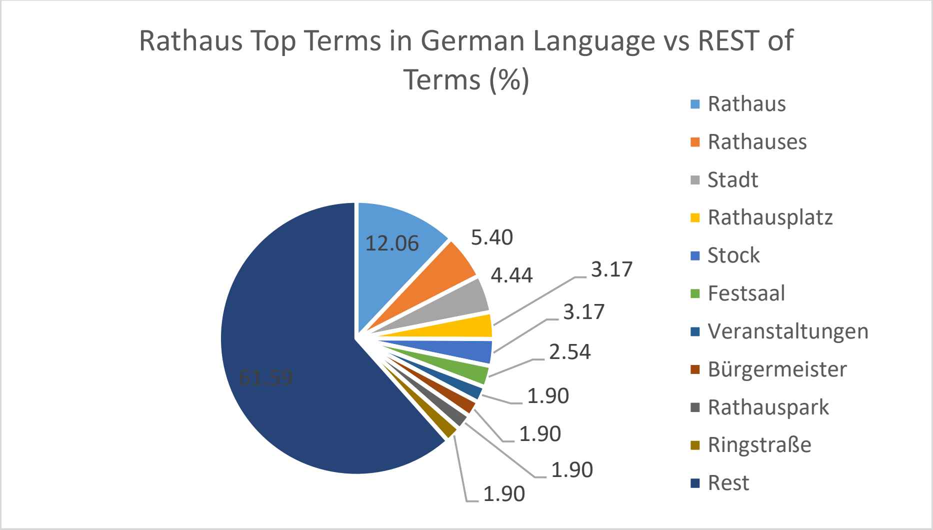

Wiener Prater Top Terms in English Language vs REST of Terms (%)
Prater
Castle
Palace
park
Wurstelprater
railway
Emperor
area
Liliputbahn
Royal
Rest
30.59
7.06
7.06
5.88
4.71
4.71
4.71
4.71
3.53
3.53
23.53

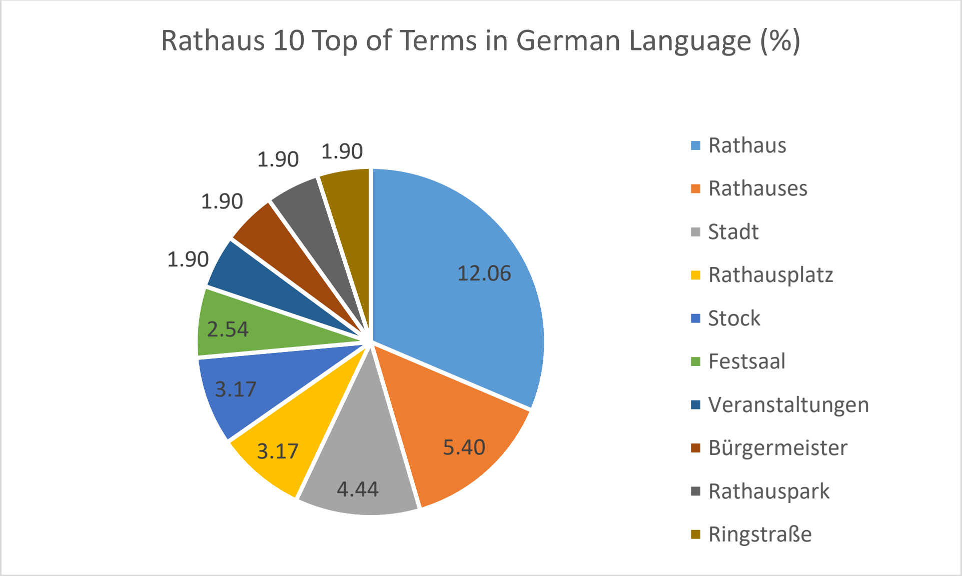

Wiener Prater 10 Top Terms in English Language (%)
Prater
Castle
Palace
park
Wurstelprater
railway
Emperor
area
Liliputbahn
Royal
30.59
7.06
7.06
5.88
4.71
4.71
4.71
4.71
3.53
3.53

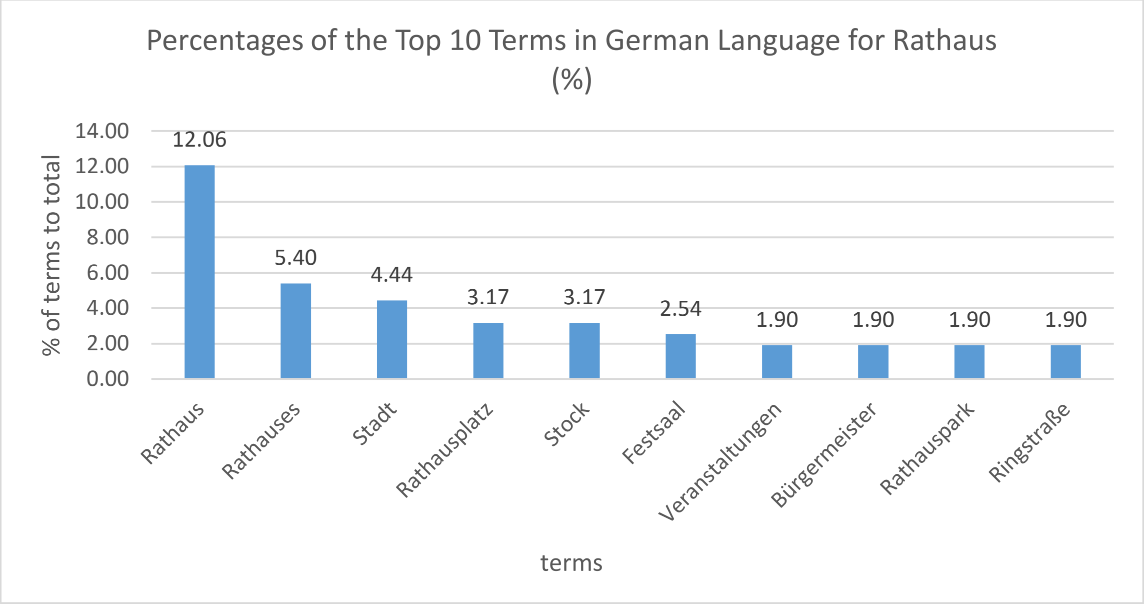

Percentages of the Top 10 Terms in German Language for Wiener Prater (%)
% of terms to total
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
30.59
7.06
7.06
5.88
4.71
4.71
4.71
4.71
3.53
3.53
Prater
Castle
Palace
park
Wurstelprater
railway
Emperor
area
Liliputbahn
Royal
terms

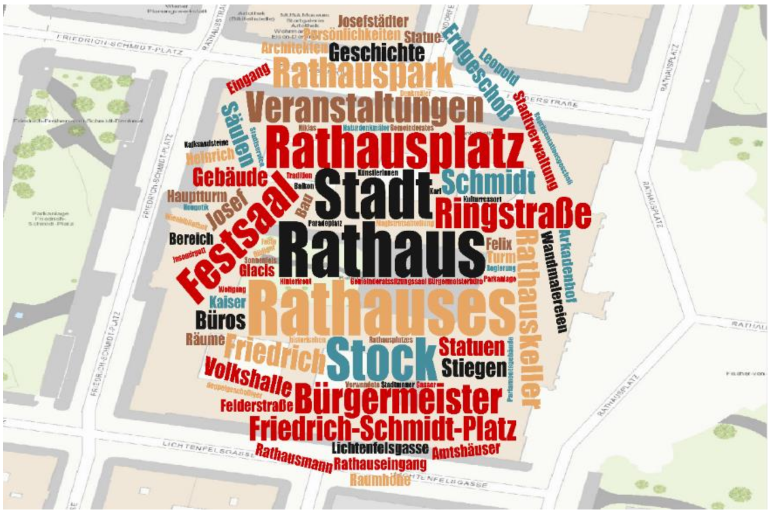
Riesenrad
Kugelmugel
Emperor
German
Leopoldstadt
park
railway
Palais
Palace
Hofburg
attractions
area
Prater
amusement
Royal
Castle
Schloss
Villa
Wurstelprater
Liliputbahn

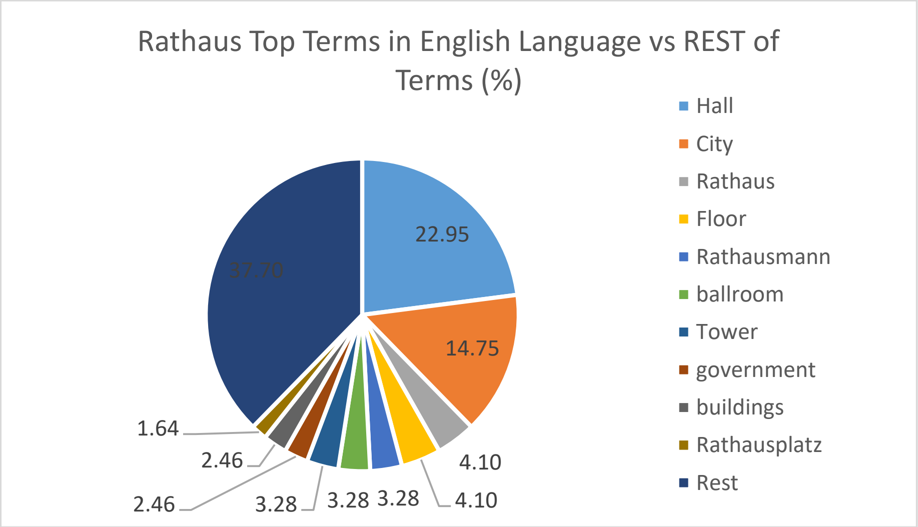
Rathaus:

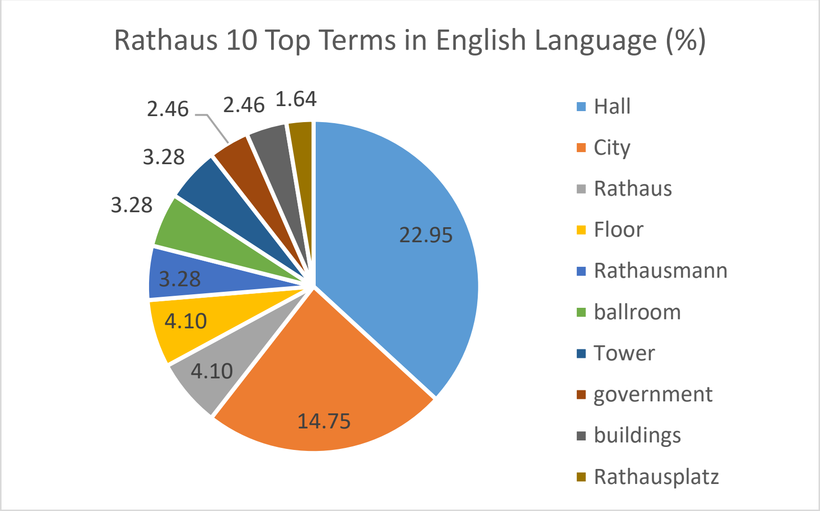

Rathaus Top Terms in German Language vs REST of Terms (%)

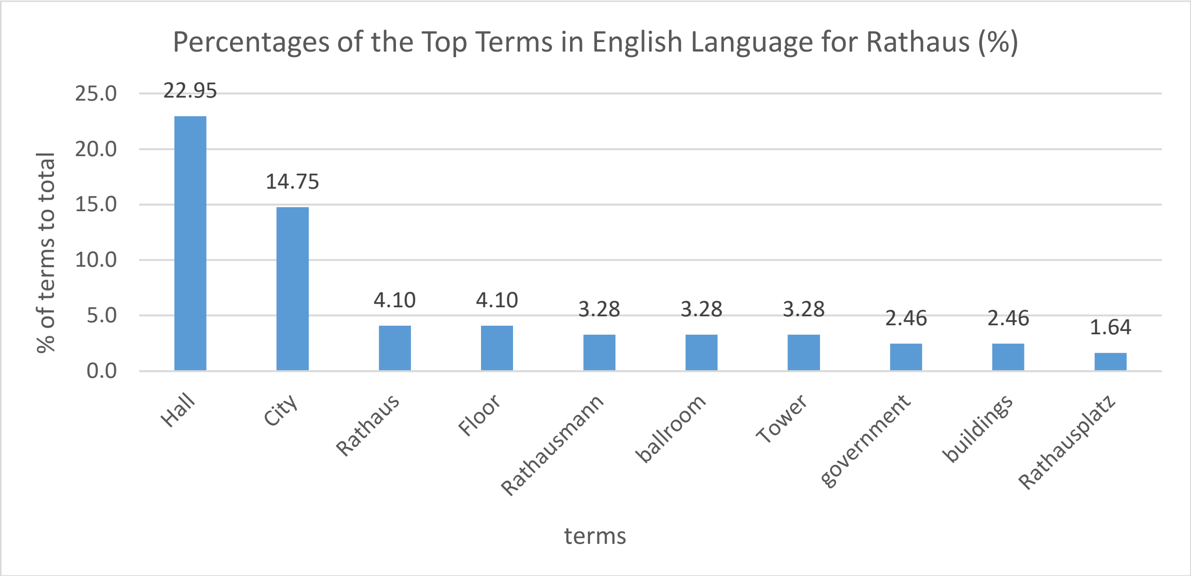

Rathaus 10 Top of Terms in German Language (%)

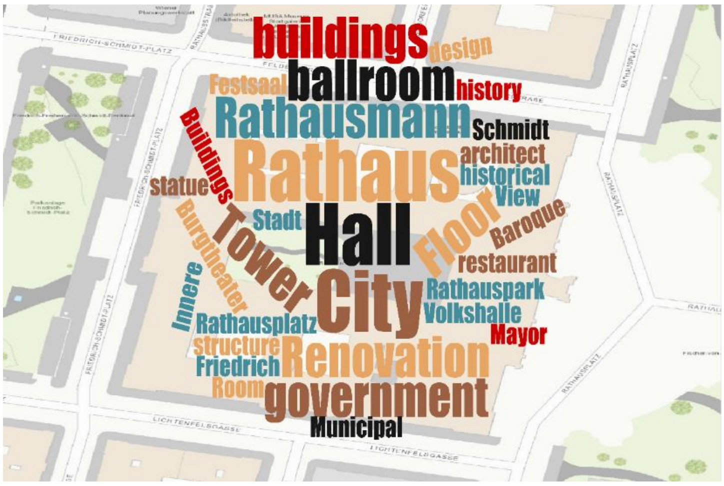

Percentages of the Top 10 Terms in German Language for Rathaus (%)
% of terms to total
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
12.06
5.40
4.44
3.17
3.17
2.54
1.90
1.90
1.90
1.90
Rathaus
Rathauses
Stadt
Rathausplatz
Stock
Festsaal
Veranstaltungen
Bürgermeister
Rathauspark
Ringstraße
terms


Josefstädter
Persönlichkeiten
Statue
Architekten
Geschichte
Erdgeschoß
Leopold
Eingang
Rathauspark
Veranstaltungen
Stadtverwaltung
Säulen
Heinrich
Rathausplatz
Gebäude
Stadt
Schmidt
Hauptturm
Josef
Ringstraße
Festsaal
Bereich
Rathaus
Felix
Turm
Arkadenhof
Wandmalereien
Glacis
Kaiser
Rathauses
Rathauskeller
Büros
Räume
Friedrich
Stock
Statuen
Stiegen
Volkshalle
Felderstraße
Bürgermeister
Friedrich-Schmidt-Platz
Lichtenfelsgasse
Rathausmann
Rathauseingang
Amtshäuser
Baumhöhe

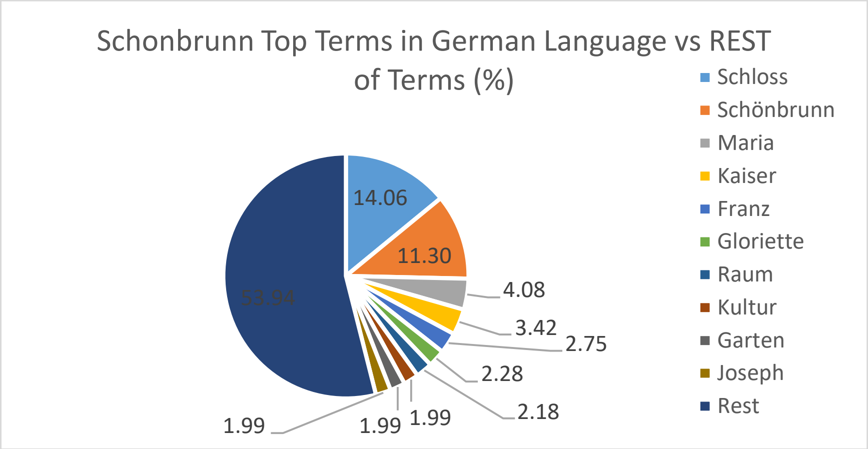

Rathaus Top Terms in English Language vs REST of Terms (%)
22.95
14.75
37.70
4.10
4.10
3.28
3.28
3.28
2.46
2.46
1.64
Hall
City
Rathaus
Floor
Rathausmann
ballroom
Tower
government
buildings
Rathausplatz
Rest

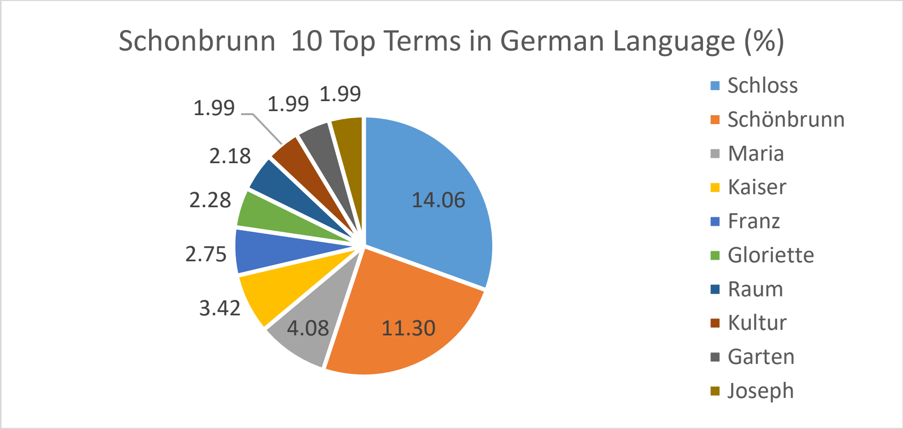

Rathaus 10 Top Terms in English Language (%)
22.95
14.75
4.10
4.10
3.28
3.28
3.28
2.46
2.46
1.64
■ Hall
■ City
■ Rathaus
■ Floor
■ Rathausmann
■ ballroom
■ Tower
■ government
■ buildings
■ Rathausplatz

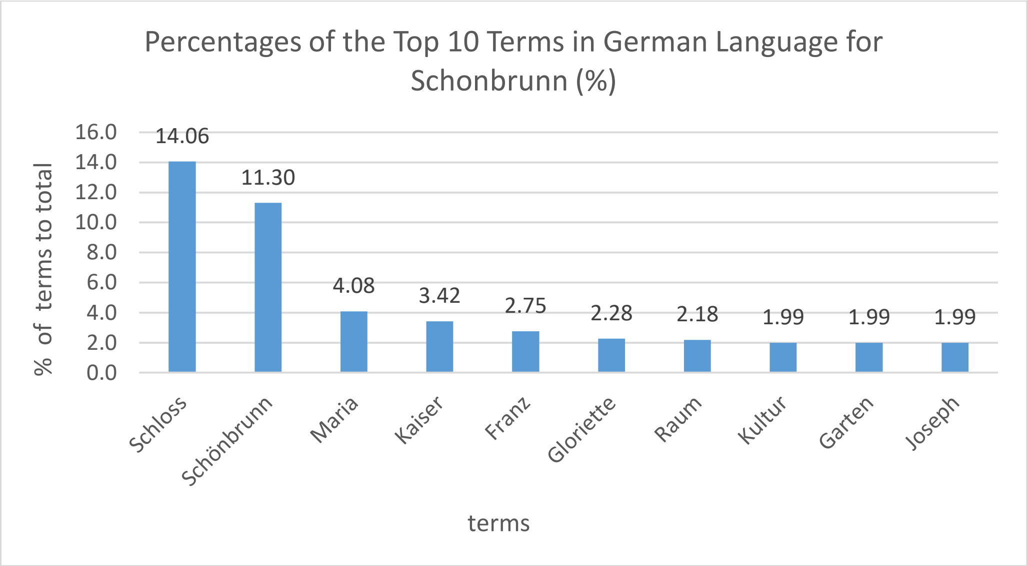

Percentages of the Top Terms in English Language for Rathaus (%)
% of terms to total
25.0
20.0
15.0
10.0
5.0
0.0
22.95
14.75
4.10
4.10
3.28
3.28
3.28
2.46
2.46
1.64
Hall
City
Rathaus
Floor
Rathausmann
ballroom
Tower
government
buildings
Rathausplatz
terms

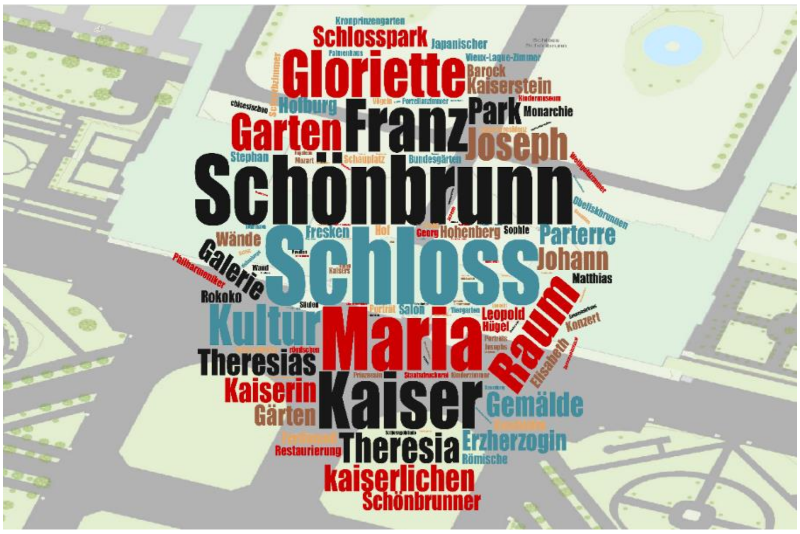

buildings
design
Festsaal
ballroom
history
Rathausmann
Schmidt
Buildings
architect
Rathaus
historical
statue
View
Tower
Stadt
Hall
Floor
Baroque
Burgtheater
restaurant
Innere
City
Rathauspark
Volkshalle
Rathausplatz
Mayor
structure
Renovation
Friedrich
Room
government
Municipal

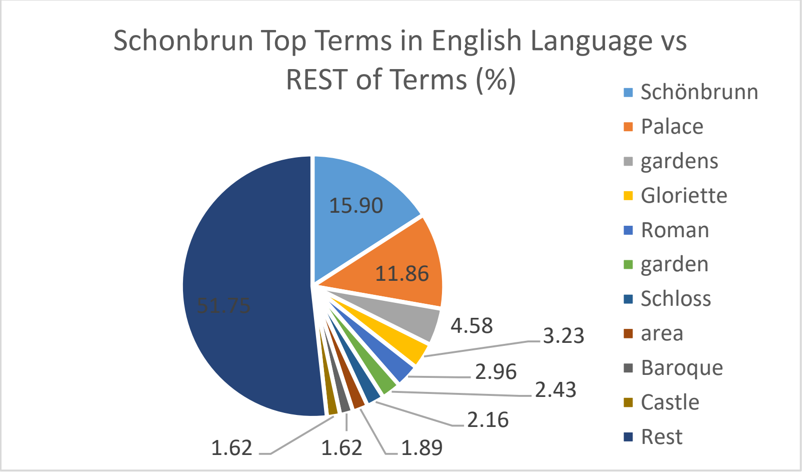
Schonbrunn Palace:



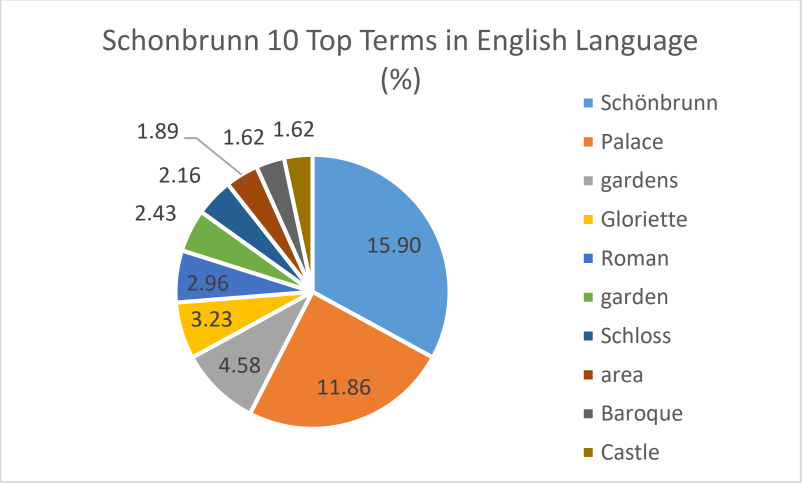




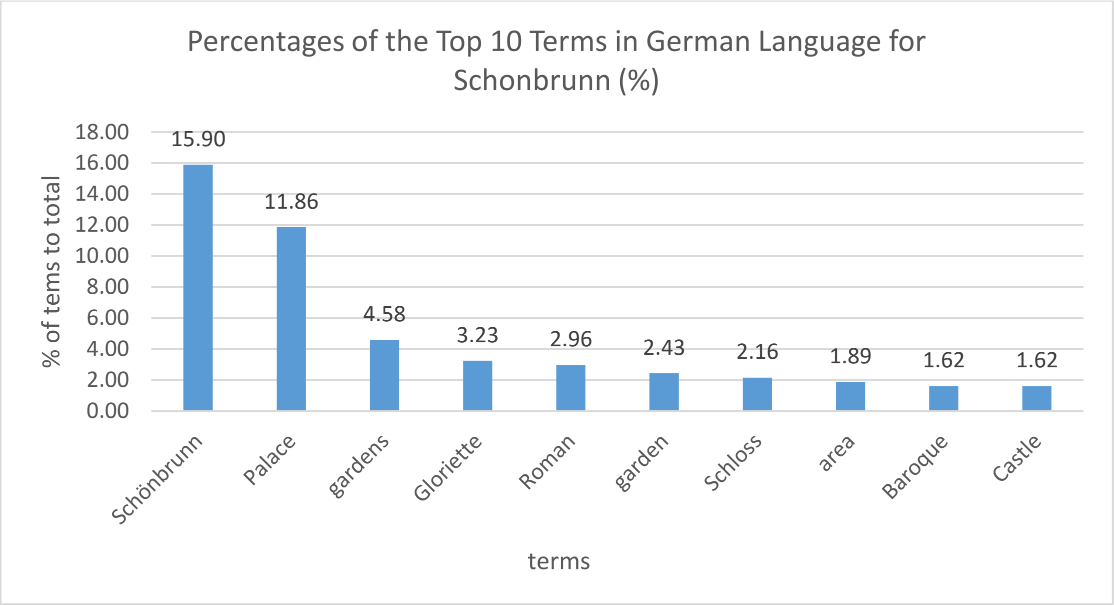


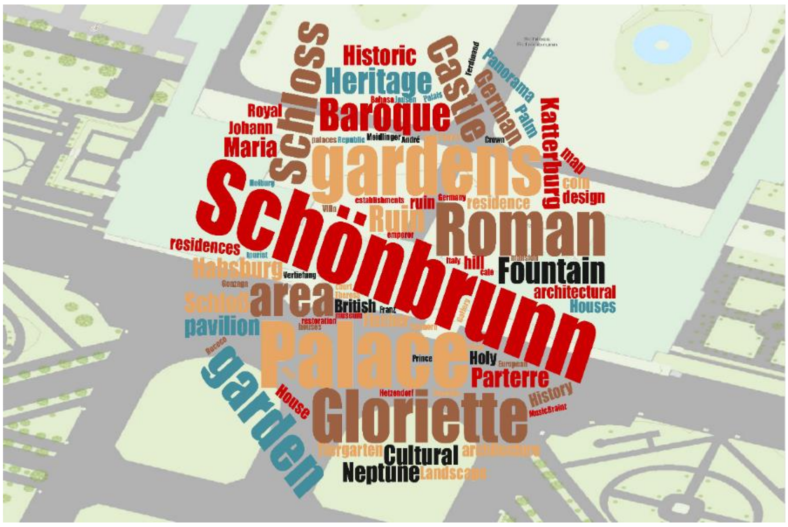


Schonbrun Top Terms in English Language vs REST of Terms (%)

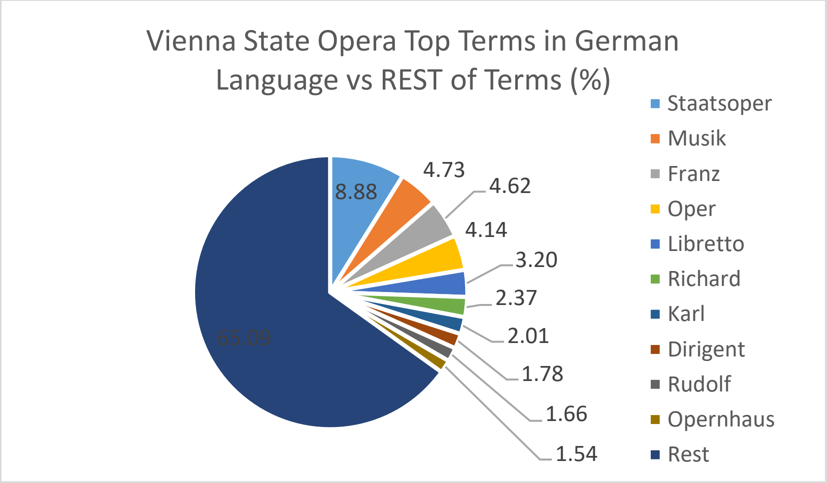


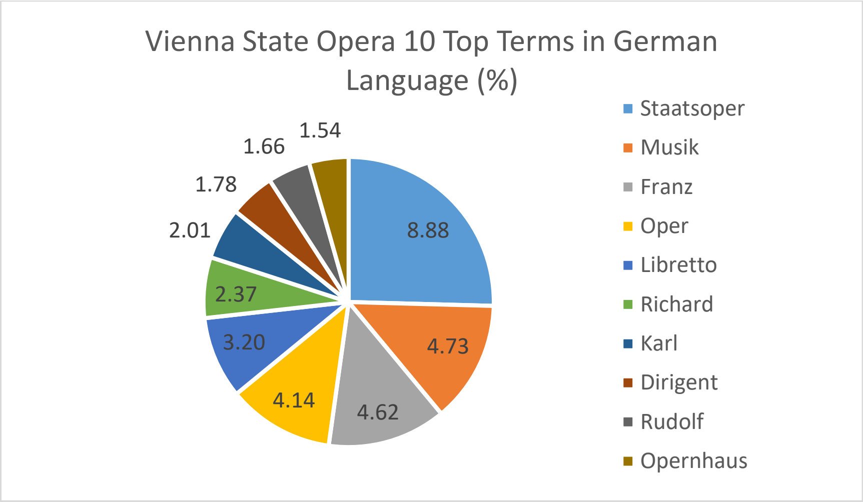
Schonbrunn 10 Top Terms in English Language (%)

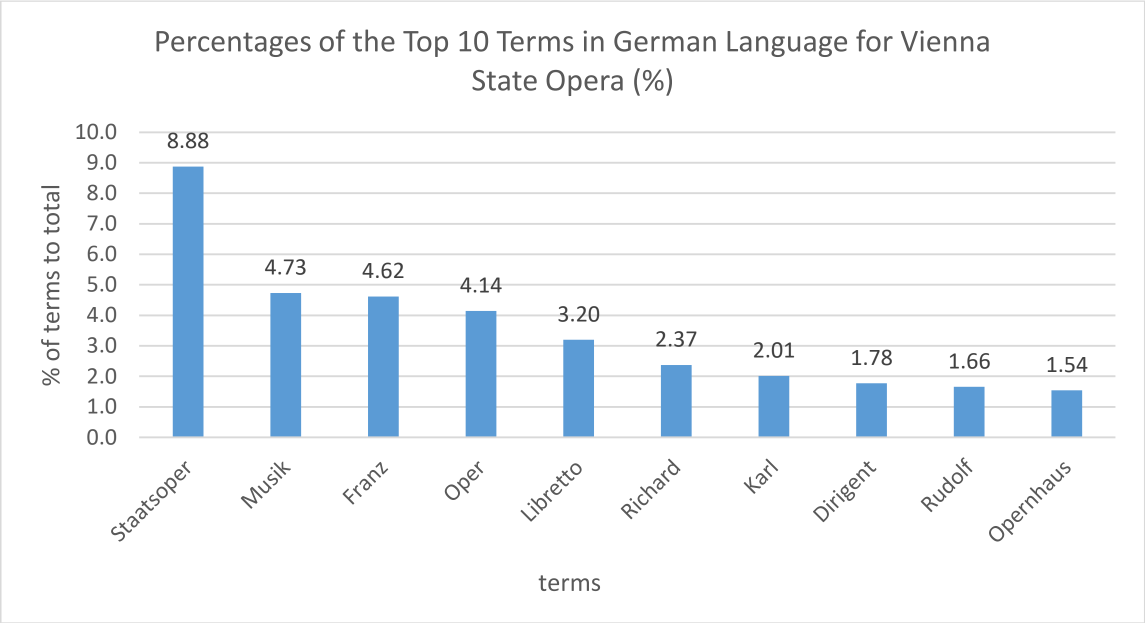




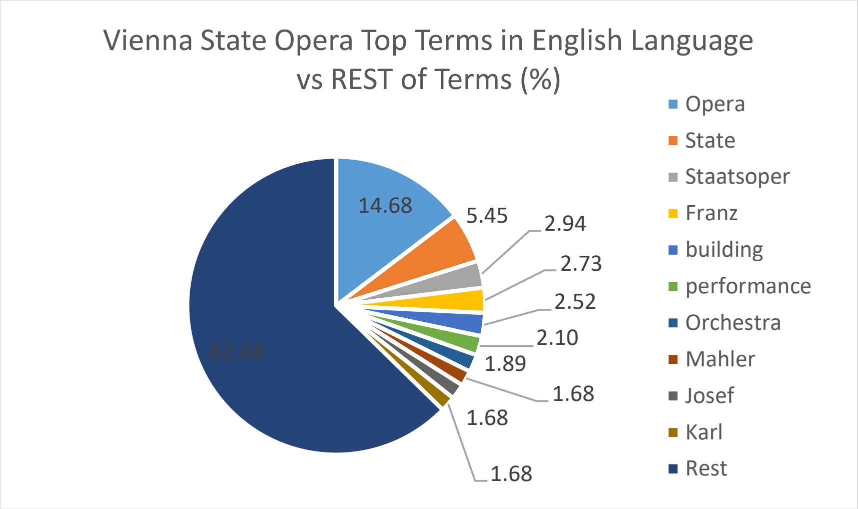
Vienna State Opera:

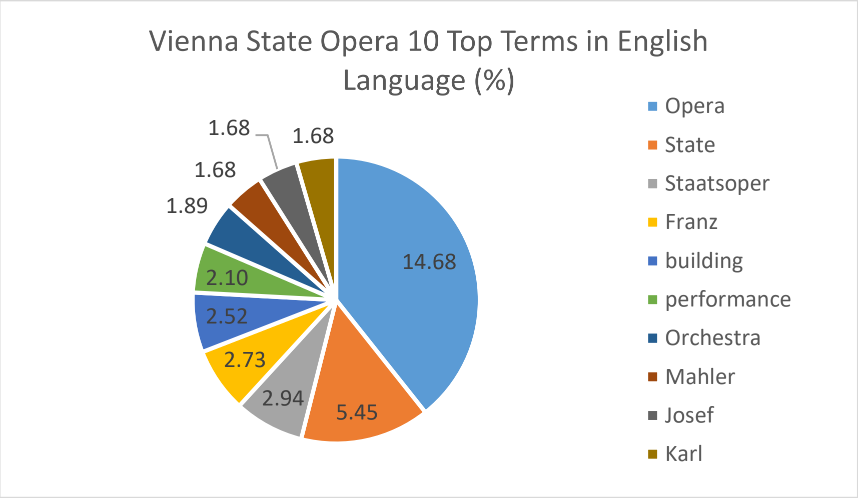


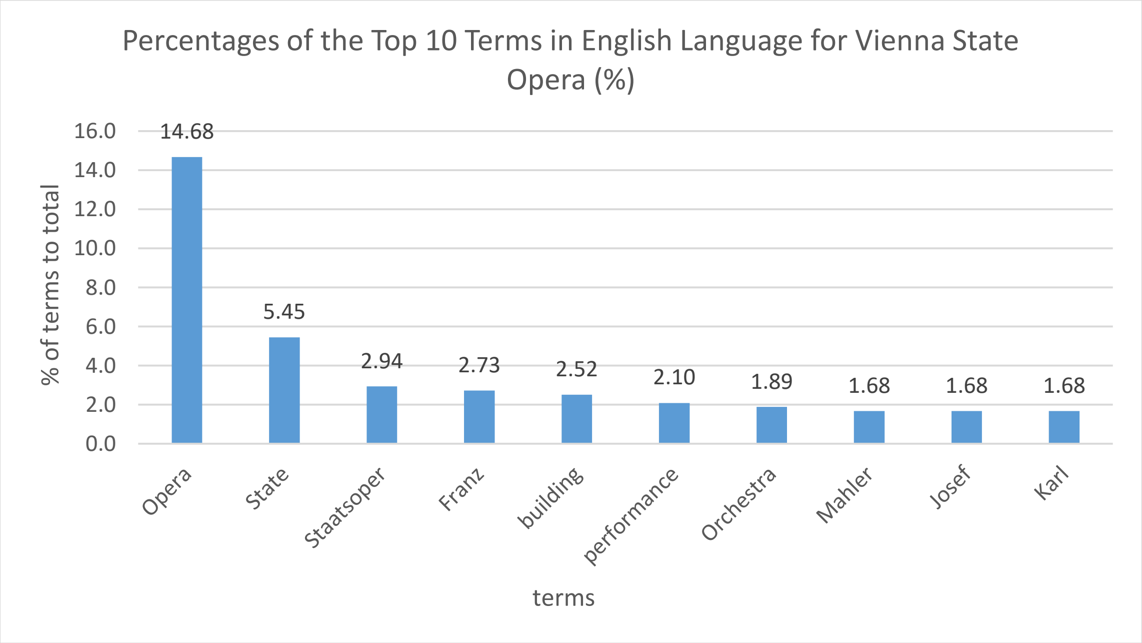

Vienna State Opera 10 Top Terms in German Language (%)
8.88
4.73
4.62
4.14
3.20
2.37
2.01
1.78
1.66
1.54
Staatsoper
Musik
Franz
Oper
Libretto
Richard
Karl
Dirigent
Rudolf
Opernhaus


Percentages of the Top 10 Terms in German Language for Vienna State Opera (%)
% of terms to total
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
8.88
4.73
4.62
4.14
3.20
2.37
2.01
1.78
1.66
1.54
Staatsoper
Musik
Franz
Oper
Libretto
Richard
Karl
Dirigent
Rudolf
Opernhaus
terms

Wiederaufbau
Hof-Operntheater
Opernhauses
Künstler
Wagners
Regie
Friedrich
Hermann
Dirigent
Rudolf
Opernball
Josef
Jahn
Haus
Thomas
Ballett
Gustav
Musik
Karl
Mozart
Hofoper
Hans
Alfred
Staatsoper
Heinrich
Erich
Franz
Strauss
Volksoper
Direktor
Richard
Egon
Wilhelm
Stein
Opern
Libretto
Chor
Platz
Operntheater
Theater
Sänger
Opernhaus
Joseph
Eisenmenger
Konzertvereinigung

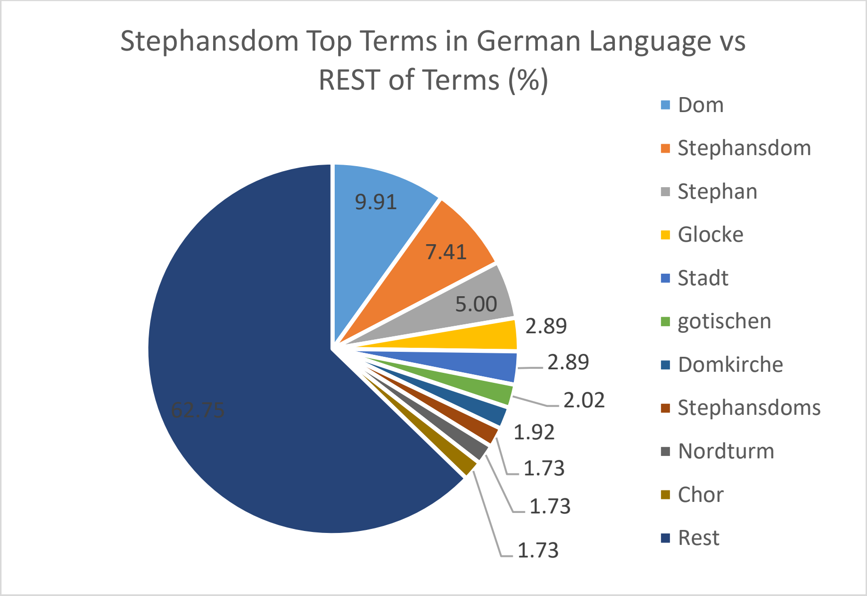

Vienna State Opera Top Terms in English Language vs REST of Terms (%)

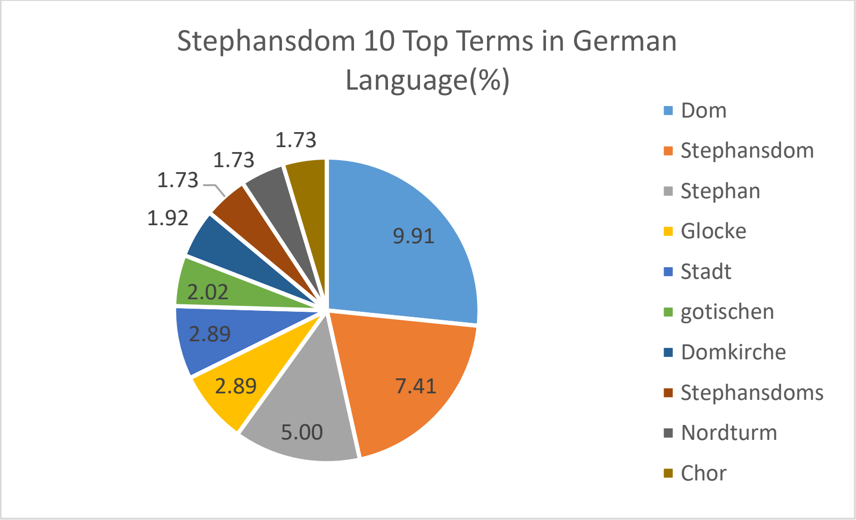

Vienna State Opera 10 Top Terms in English Language (%)

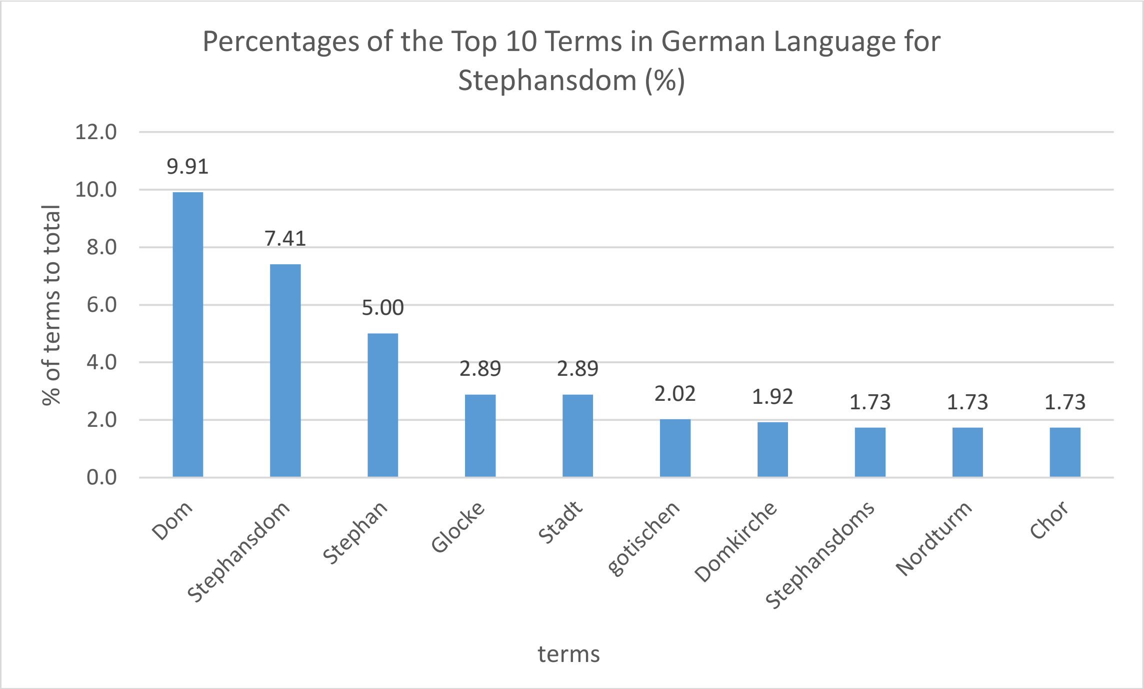

Percentages of the Top 10 Terms in English Language for Vienna State Opera (%)

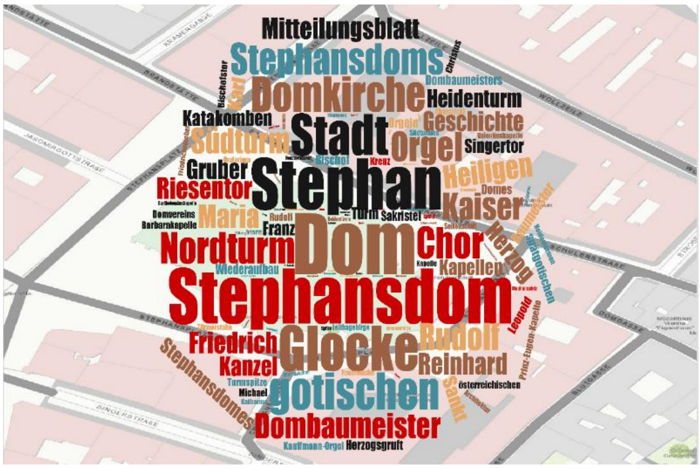
German-language
Weingartner
Giuseppe
museum
Felix
Richard
Egon
Eduard
Volksoper
performance
singers
Helmut
Peter
Johann
Heinrich
Herbert
Franz
Hlávka
Orchestra
Otto
Salzburg
State
opening
Clemens
Theater
Strauss
live
ball
Josef
operas
Idomeneo
Jordan
Opera
Mahler
German
Daniel
director
Norma
artists
Mozart
Directors
Staatsoper
Walter
Giovanni
broadcast
Krips
Hans
building
music
Ballet
Böhm
Eberhard
Philippe
Krauss
History
Karl
ballet
theatre
ensemble
Michael
Philharmonic
Wilhelm
auditorium
stage
Erich
Elisabeth
Dominique
Mozart-Ensemble

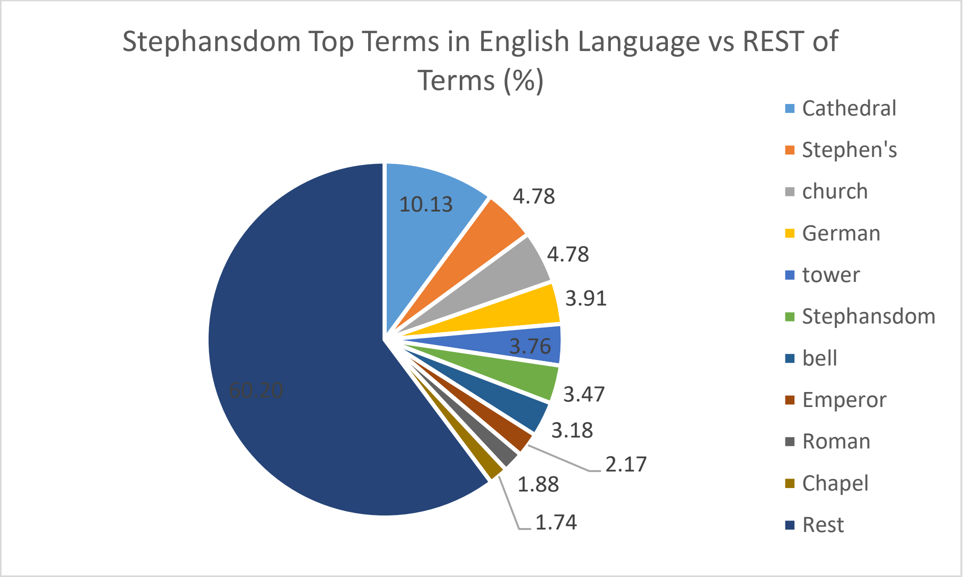
Stephansdom:

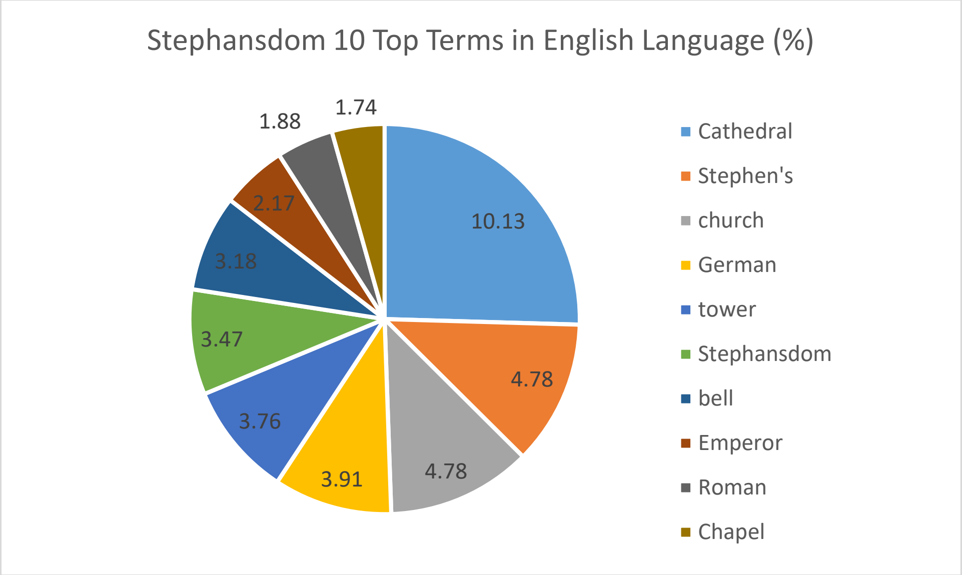


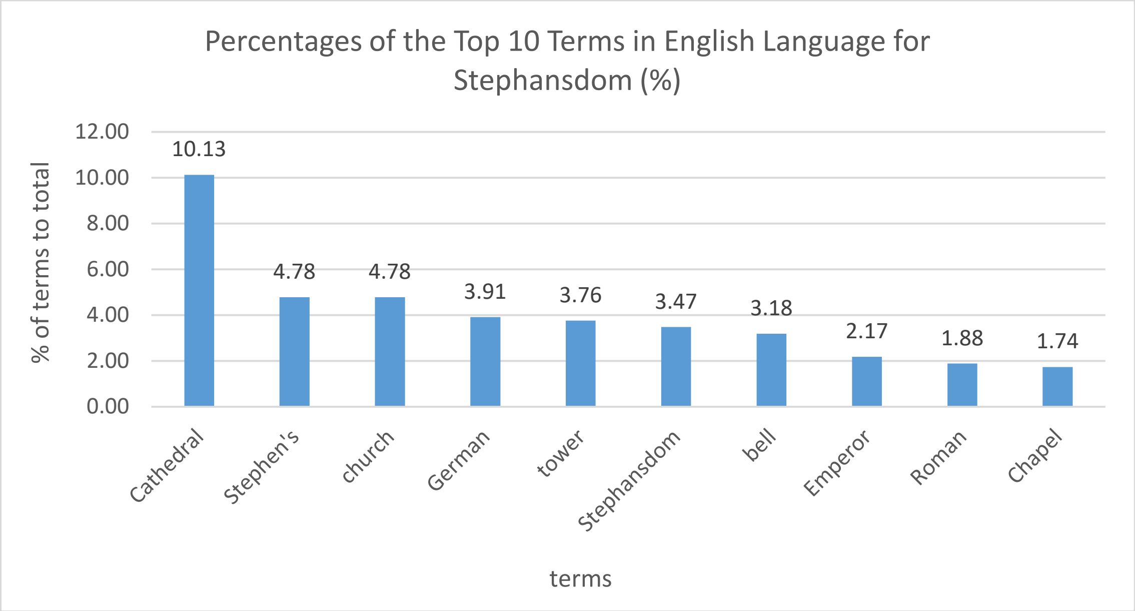


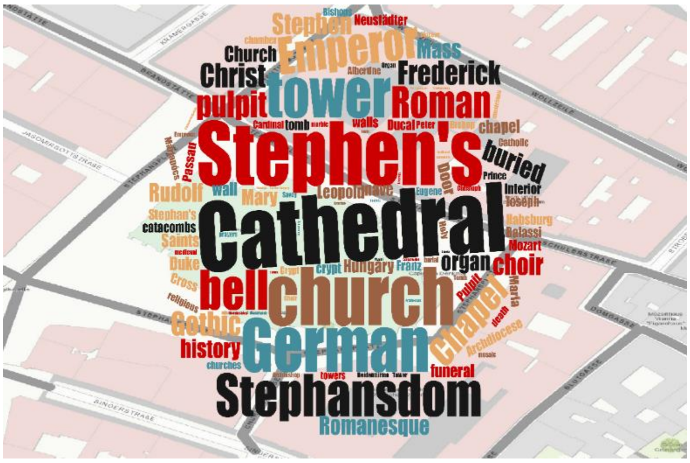

Percentages of the Top 10 Terms in German Language for Stephansdom (%)
% of terms to total
12.0
10.0
8.0
6.0
4.0
2.0
0.0
9.91
7.41
5.00
2.89
2.89
2.02
1.92
1.73
1.73
1.73
Dom
Stephansdom
Stephan
Glocke
Stadt
gotischen
Domkirche
Stephansdoms
Nordturm
Chor
terms


Mitteilungsblatt
Stephansdoms
Domkirche
Heidenturm
Katakomben
Südturm
Stadt
Geschichte
Orgel
Singertor
Gruber
Riesentor
Stephan
Heiligen
Kaiser
Maria
Turm
Sakristei
Franz
Nordturm
Dom
Chor
Herzog
Kapellen
Wiederaufbau
Stephansdom
Friedrich
Glocke
Rudolf
Kanzel
Reinhard
Stephansdomes
gotischen
Dombaumeister
Herzogsgruft

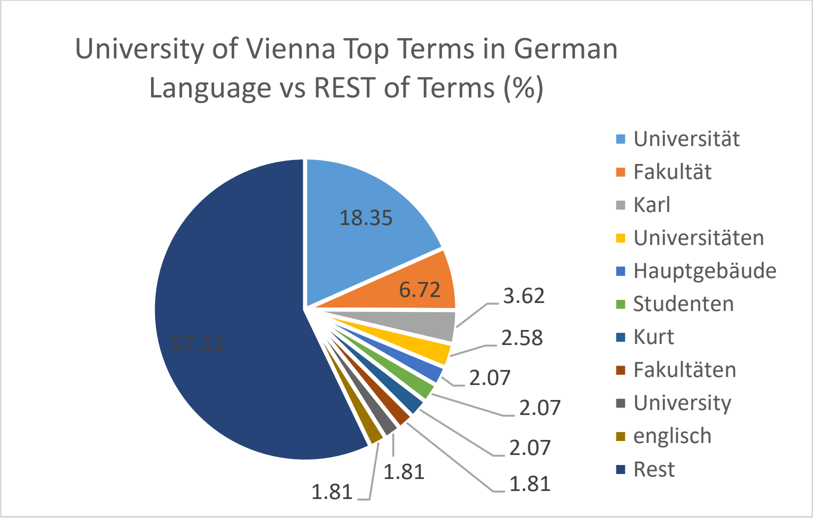

Stephansdom Top Terms in English Language vs REST of Terms (%)
10.13
4.78
4.78
3.91
3.76
3.47
3.18
2.17
1.88
1.74
60.22
Cathedral
Stephen's
church
German
tower
Stephansdom
bell
Emperor
Roman
Chapel
Rest

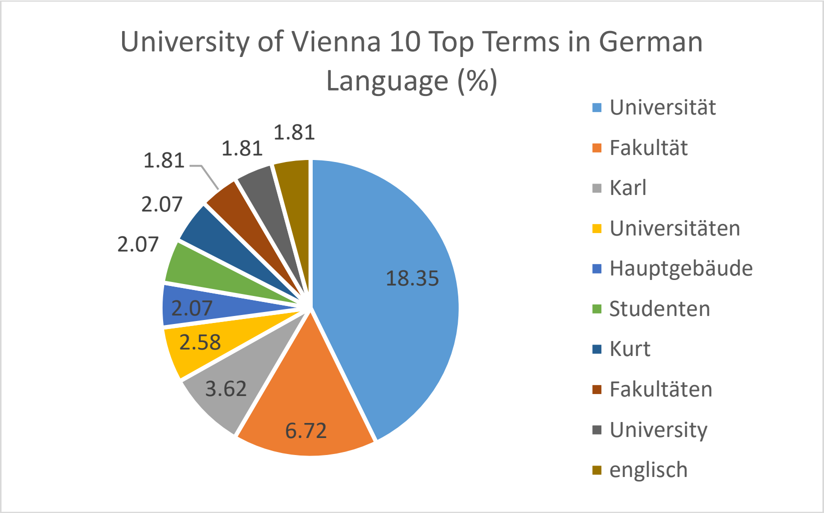

Stephansdom 10 Top Terms in English Language (%)
10.13
4.78
4.78
3.91
3.76
3.47
3.18
2.17
1.88
1.74
Cathedral
Stephen's
church
German
tower
Stephansdom
bell
Emperor
Roman
Chapel

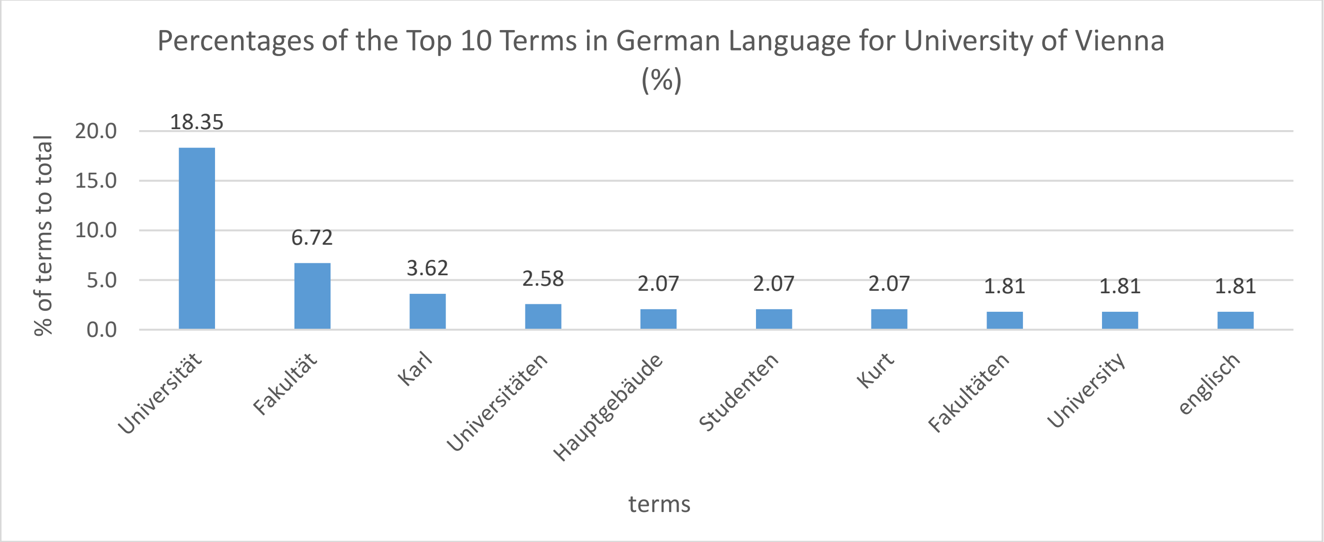

Percentages of the Top 10 Terms in English Language for Stephansdom (%)
% of terms to total
12.00
10.00
8.00
6.00
4.00
2.00
0.00
10.13
4.78
4.78
3.91
3.76
3.47
3.18
2.17
1.88
1.74
Cathedral
Stephen's
church
German
tower
Stephansdom
bell
Emperor
Roman
Chapel
terms

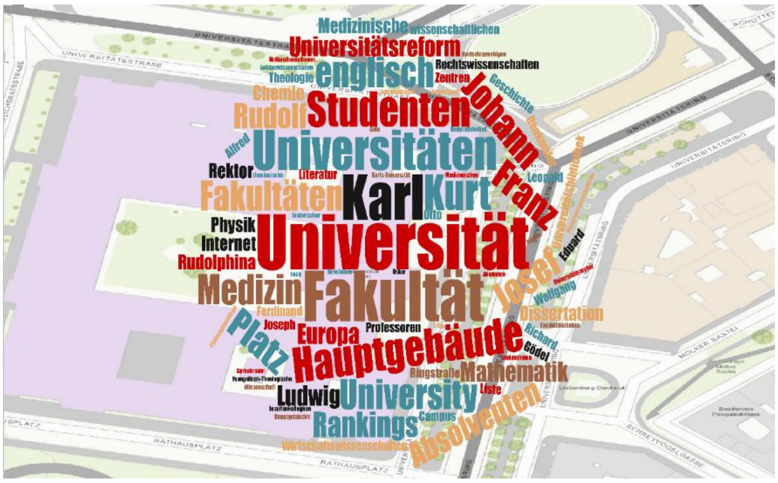

Stephen
Emperor
Mass
Church
Christ
Frederick
pulpit
tower
Roman
chapel
Stephen's
buried
Rudolf
Mary
Cathedral
catacombs
choir
organ
bell
church
Hungary
Gothic
history
German
Chapel
funeral
Stephansdom
Romanesque

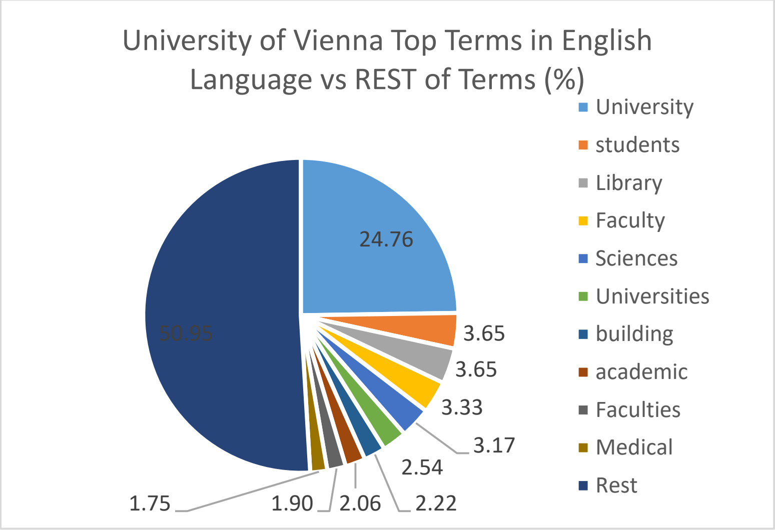
University of Vienna:



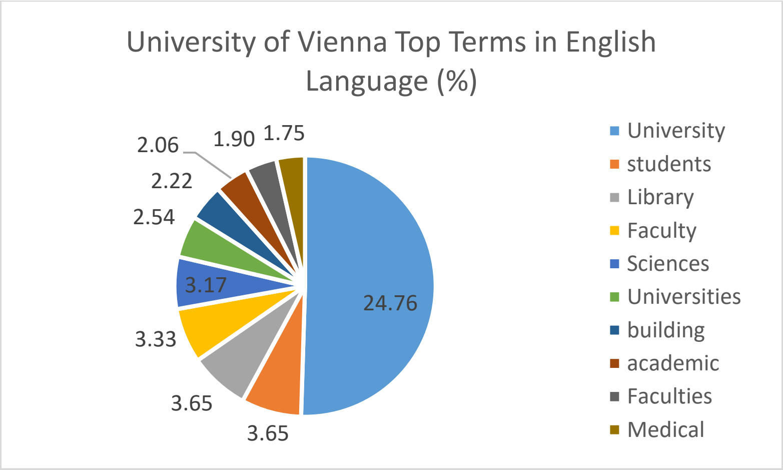




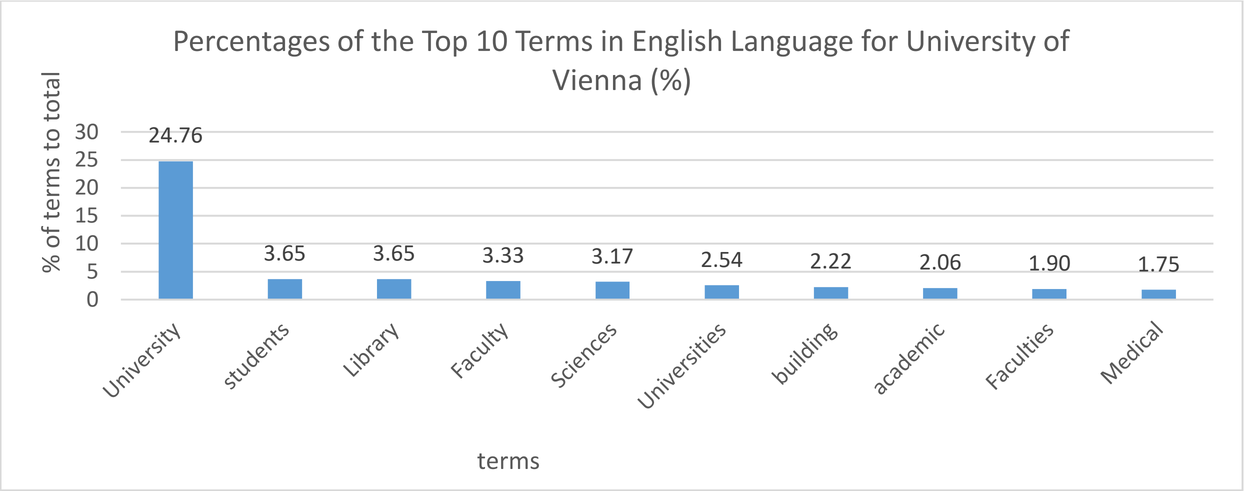


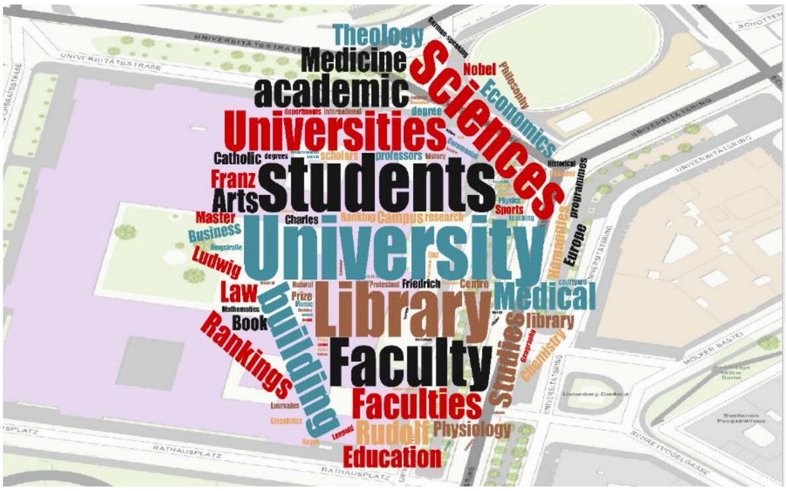


University of Vienna Top Terms in English Language vs REST of Terms (%)

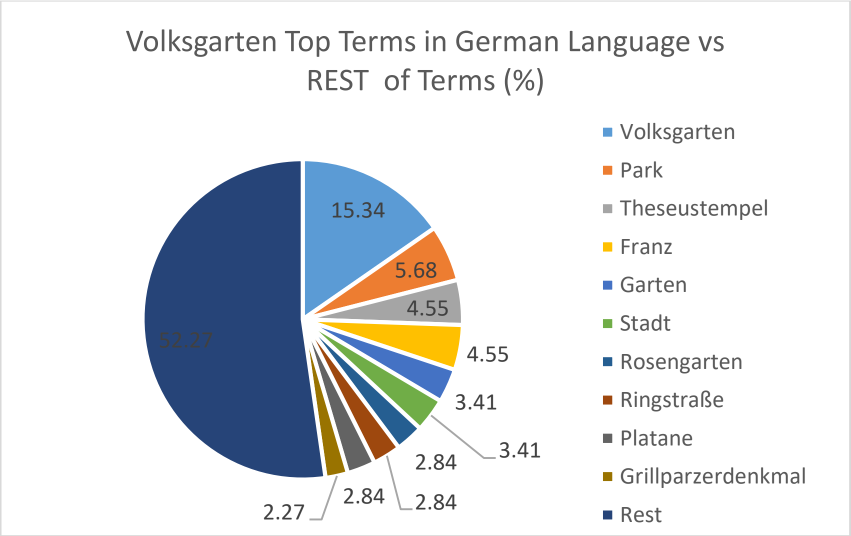


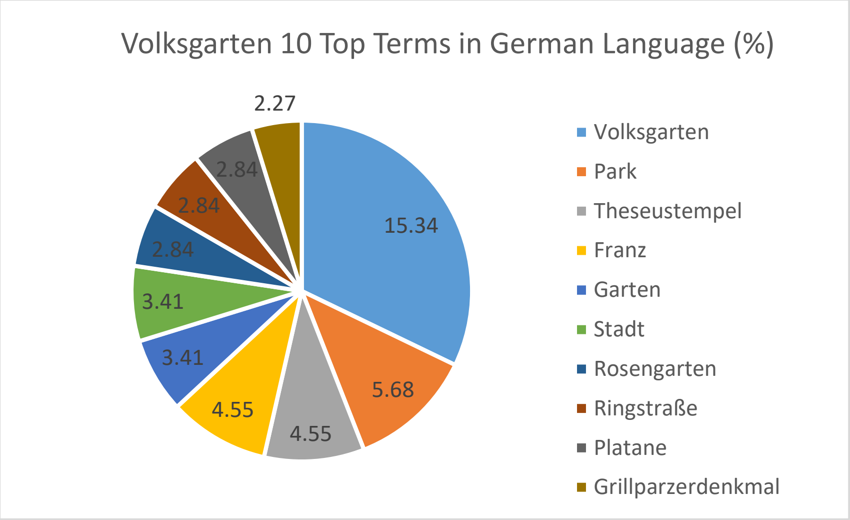
University of Vienna Top Terms in English Language (%)

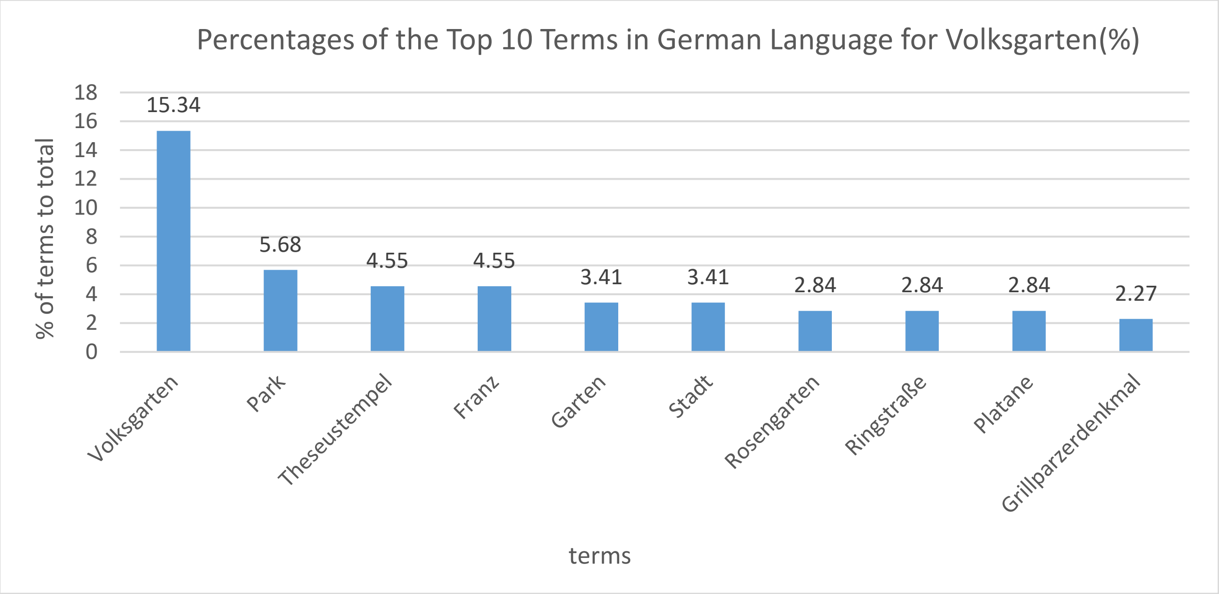


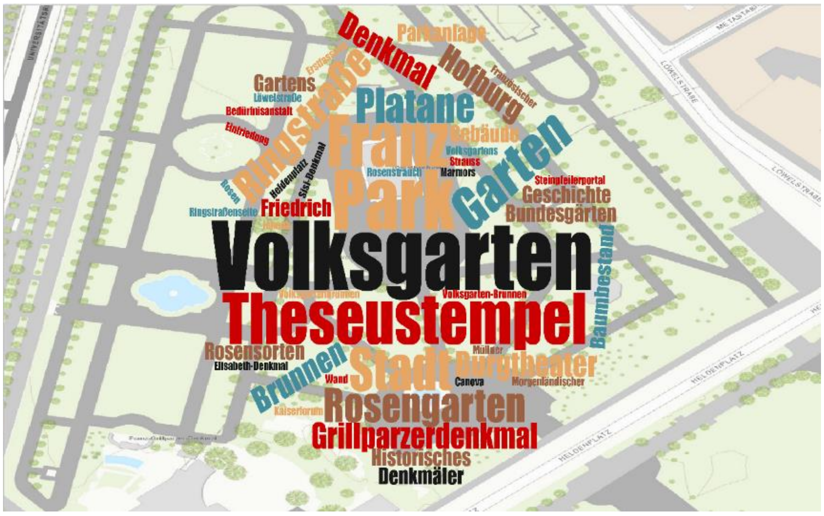


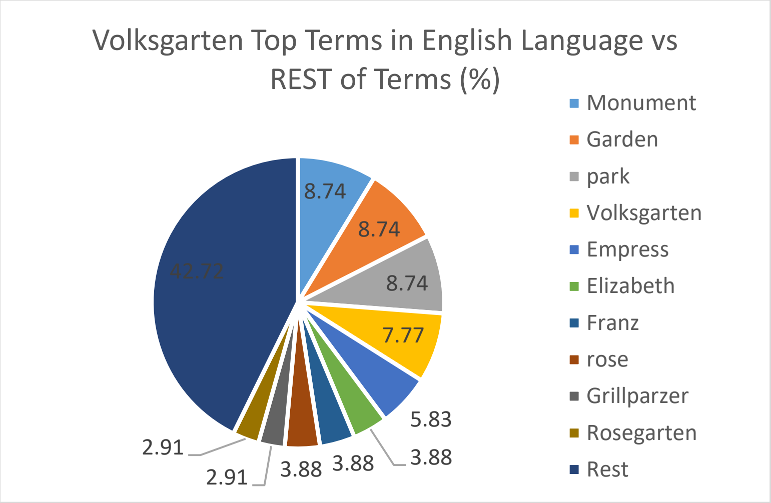
Volksgarten:



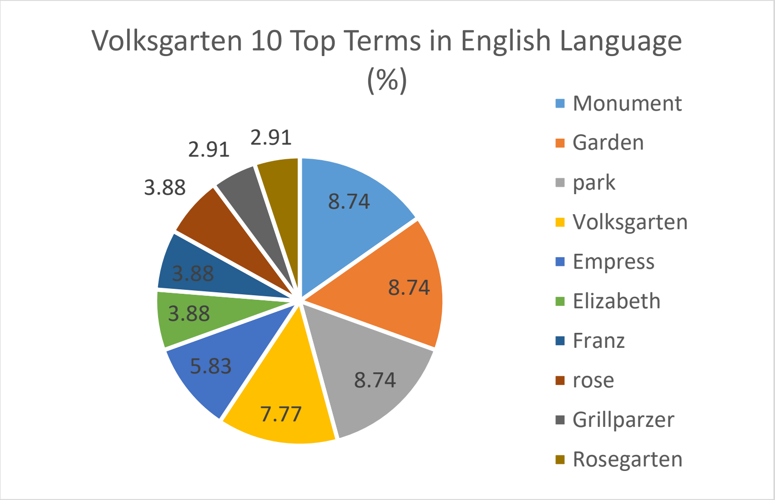

Volksgarten 10 Top Terms in German Language (%)
15.34
5.68
4.55
4.55
3.41
3.41
2.84
2.84
2.84
2.27
Volksgarten
Park
Theseustempel
Franz
Garten
Stadt
Rosengarten
Ringstraße
Platane
Grillparzerdenkmal


Percentages of the Top 10 Terms in German Language for Volksgarten(%)
% of terms to total
18
16
14
12
10
8
6
4
2
0
15.34
5.68
4.55
4.55
3.41
3.41
2.84
2.84
2.84
2.27
Volksgarten
Park
Theseustempel
Franz
Garten
Stadt
Rosengarten
Ringstraße
Platane
Grillparzerdenkmal
terms

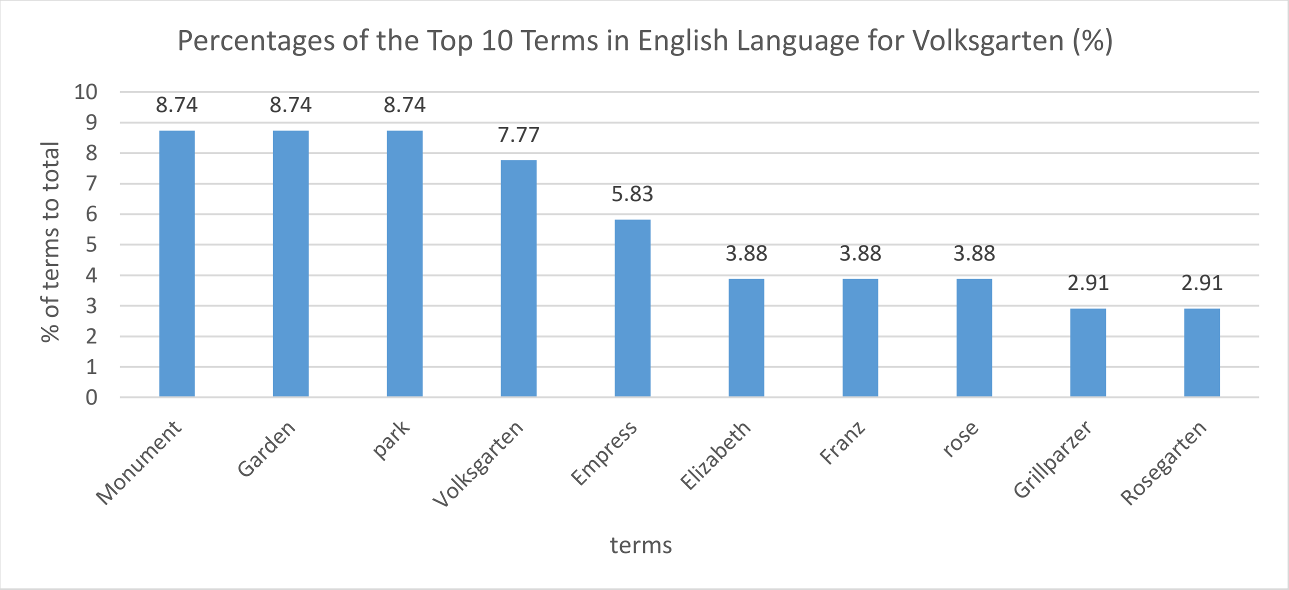

Denkmal
Parkanlage
Hofburg
Gartens
Löwelstraße
Bedürfnisanstalt
Ringstraße
Platane
Franz
Gebäude
Volksgartens
Strauss
Marmors
Rosenstrauch
Heldenplatz
Sisi-Denkmal
Rosen
Ringstraßenseite
Friedrich
Park
Garten
Steinpfeilerportal
Geschichte
Bundesgärten
Volksgarten
Baumbestand
Volksgarten-Brunnen
Theseustempel
Rosensorten
Elisabeth-Denkmal
Brunnen
Stadt
Burgtheater
Wand
Canova
Morgenländischer
Kaiserforum
Rosengarten
Grillparzerdenkmal
Historisches
Denkmäler



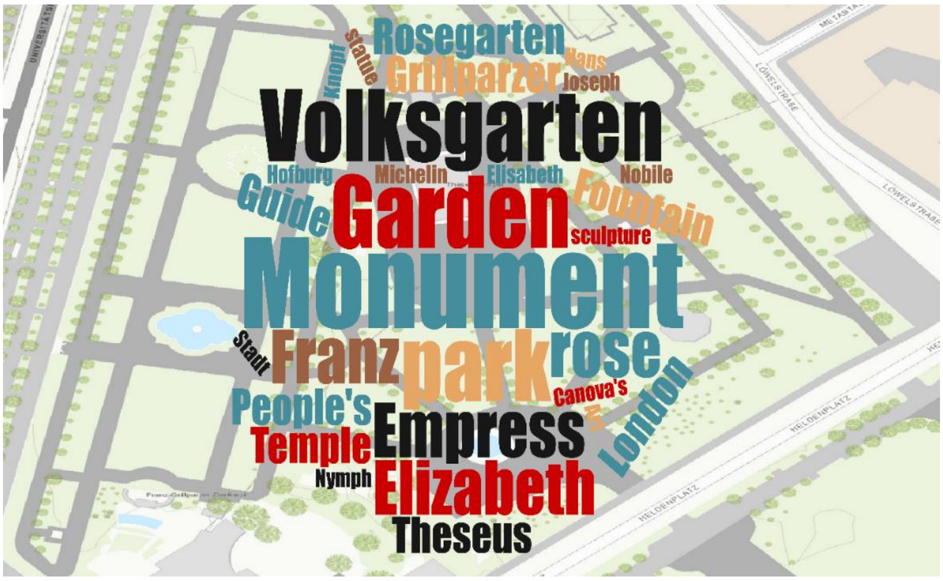
Volksgarten Top Terms in English Language vs REST of Terms (%)



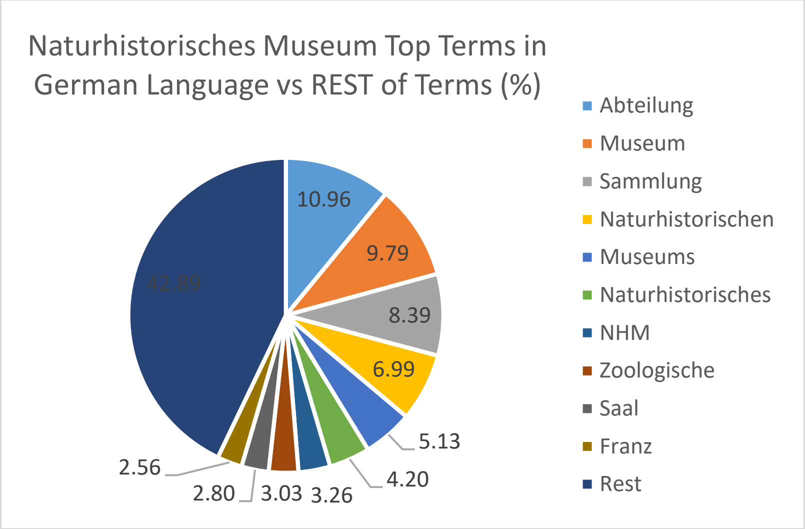
Volksgarten 10 Top Terms in English Language (%)

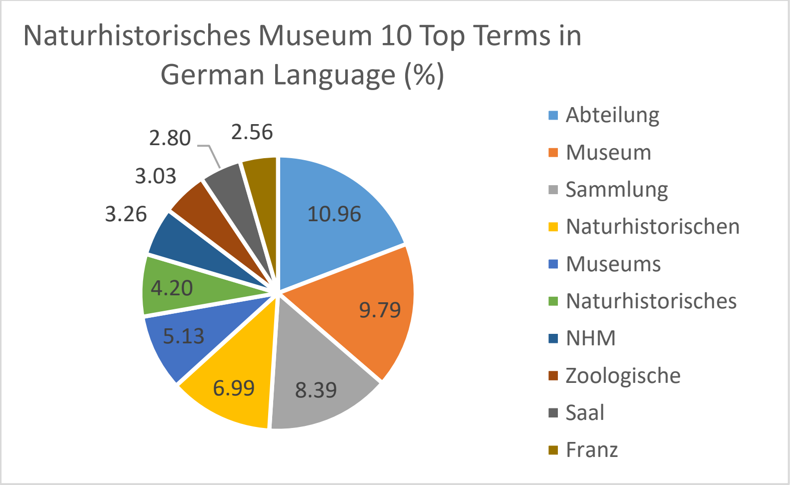


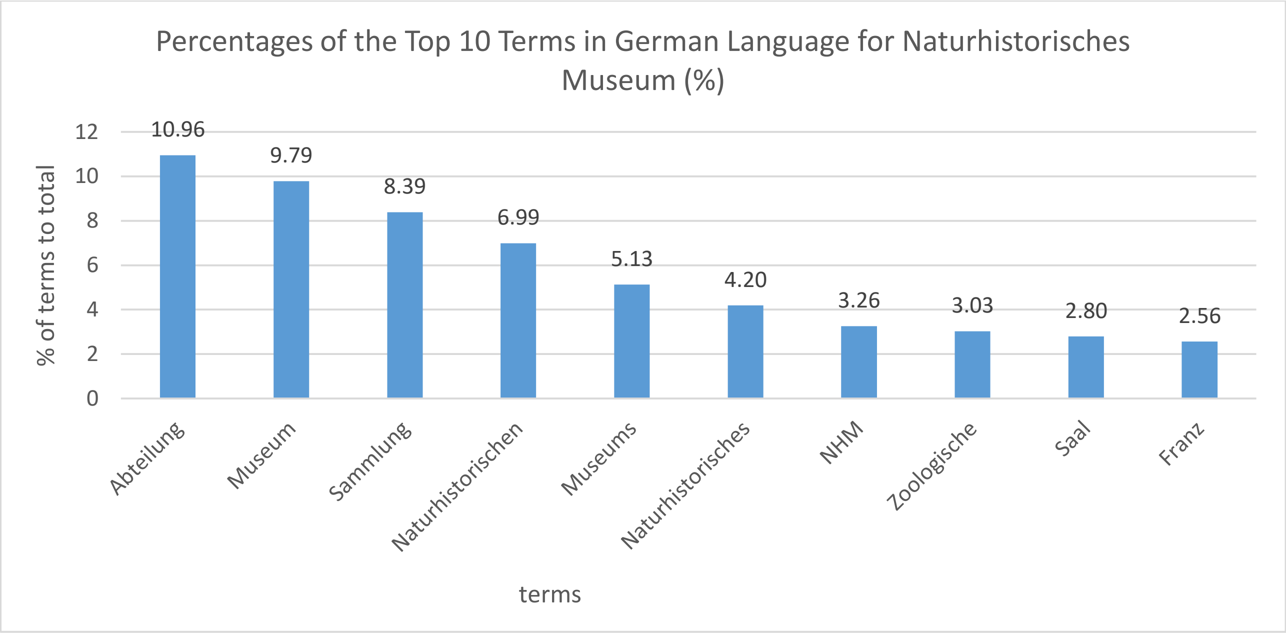
Percentages of the Top 10 Terms in English Language for Volksgarten (%)

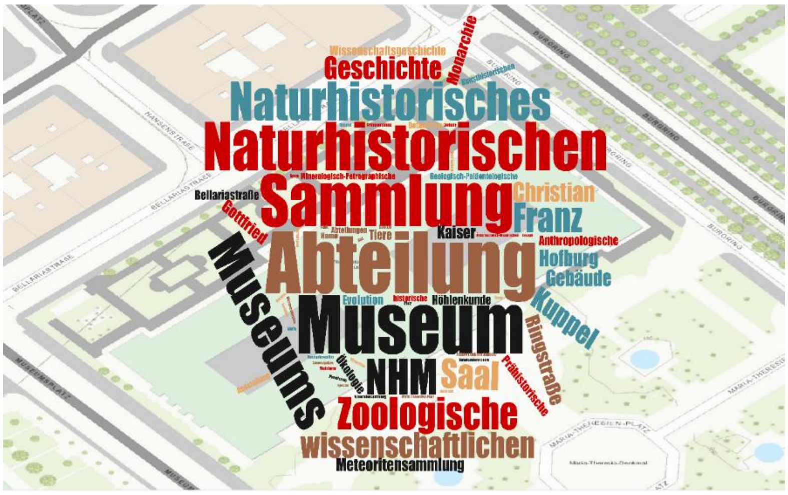


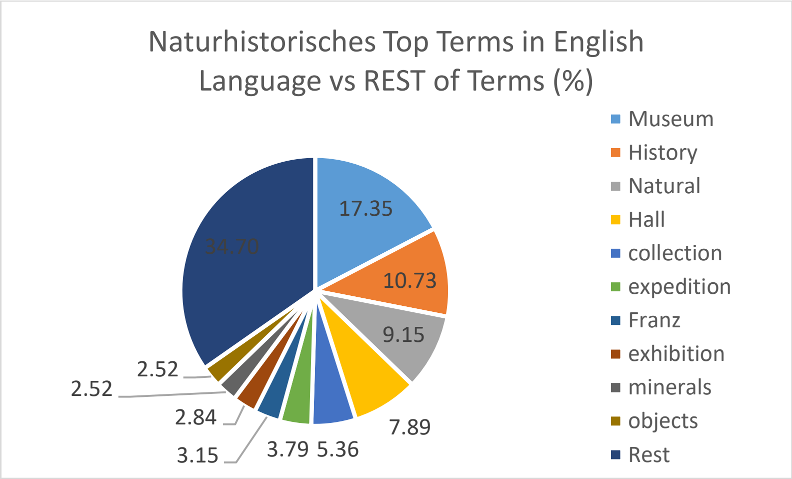
Naturhistorisches Museum:

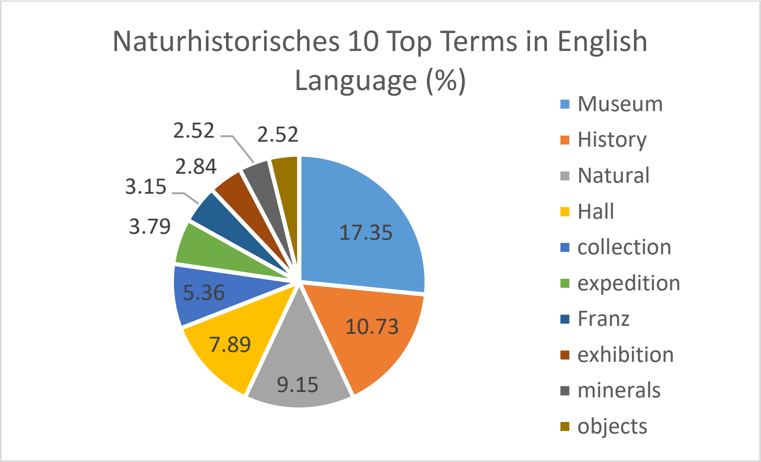


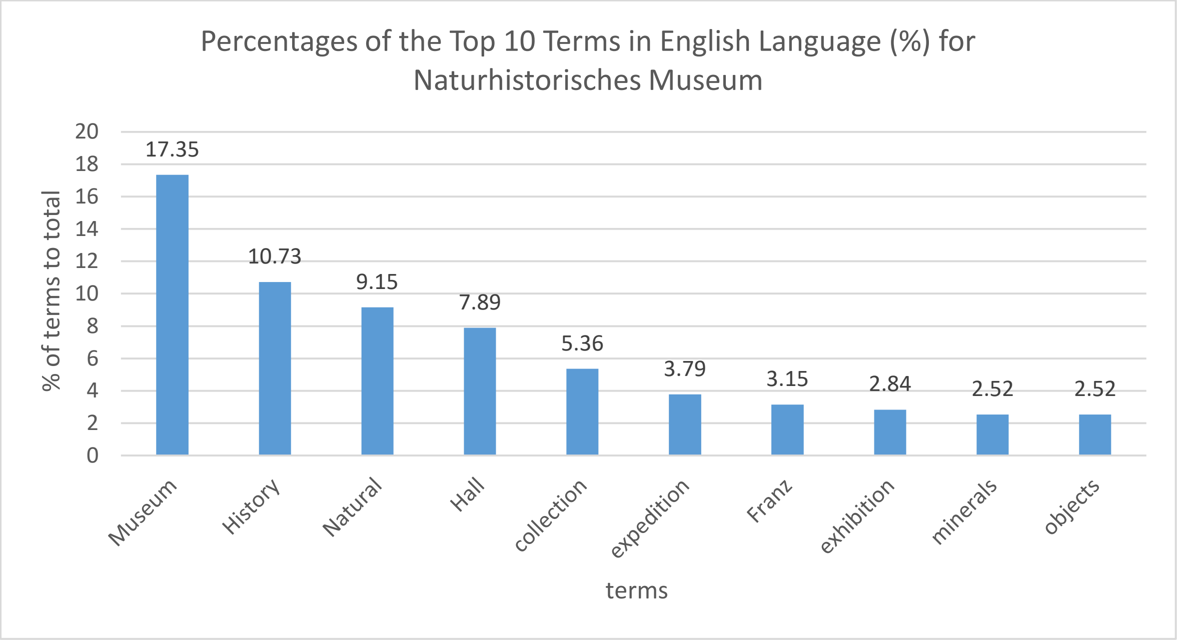



Percentages of the Top 10 Terms in German Language for Naturhistorisches Museum (%)
% of terms to total
12
10
8
6
4
2
0
10.96
9.79
8.39
6.99
5.13
4.20
3.26
3.03
2.80
2.56
Abteilung
Museum
Sammlung
Naturhistorischen
Museums
Naturhistorisches
NHM
Zoologische
Saal
Franz
terms


Wissenschaftsgeschichte
Geschichte
Monarchie
Naturhistorisches
Naturhistorischen
Mineralogisch-Petrographische
Geologisch-Paläontologische
Bellariastraße
Sammlung
Christian
Franz
Gottfried
Tiere
Kaiser
Anthropologische
Abteilung
Hofburg
Gebäude
Museums
Evolution
Höhlenkunde
Museum
Kuppel
Ringstraße
NHM
Saal
Ökologie
Prähistorische
Zoologische
wissenschaftlichen
Meteoritensammlung


Naturhistorisches Top Terms in English Language vs REST of Terms (%)
17.35
10.73
9.15
7.89
5.36
3.79
3.15
2.84
2.52
2.52
34.70
Museum
History
Natural
Hall
collection
expedition
Franz
exhibition
minerals
objects
Rest


Naturhistorisches 10 Top Terms in English Language (%)
17.35
10.73
9.15
7.89
5.36
3.79
3.15
2.84
2.52
2.52
Museum
History
Natural
Hall
collection
expedition
Franz
exhibition
minerals
objects

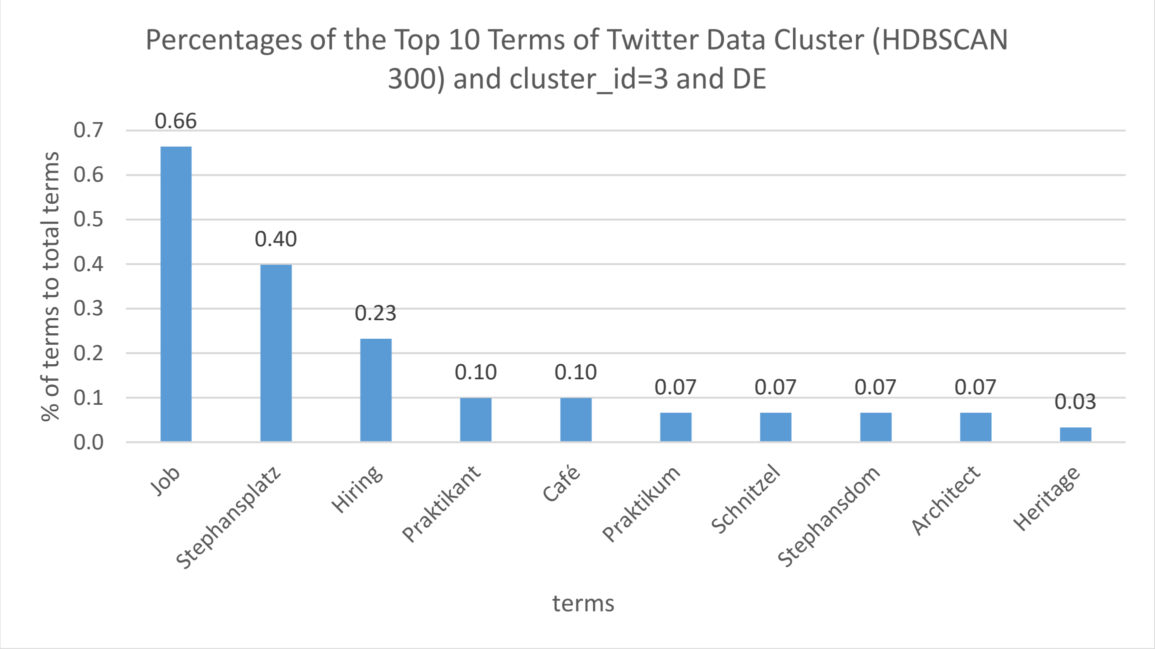

Percentages of the Top 10 Terms in English Language (%) for Naturhistorisches Museum
% of terms to total
20
18
16
14
12
10
8
6
4
2
0
17.35
10.73
9.15
7.89
5.36
3.79
3.15
2.84
2.52
2.52
Museum
History
Natural
Hall
collection
expedition
Franz
exhibition
minerals
objects
terms

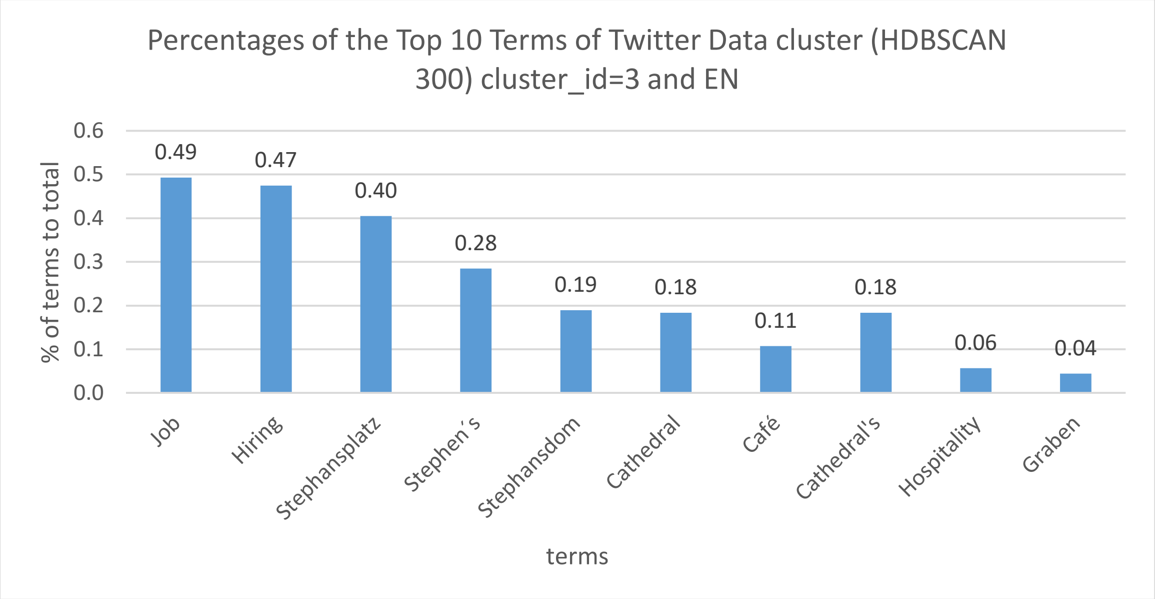
Jovanovic-Kruspel
expedition
Cabinet
Mineralogy
Hall
Joseph
scientific
Franz
Ceiling
Ferdinand
NHM
History
Tegetthoff
human
Hans
Museum
minerals
objects
Emperor
Earth
Natural
Scientists
collection
exhibition
Naturhistorisches

58

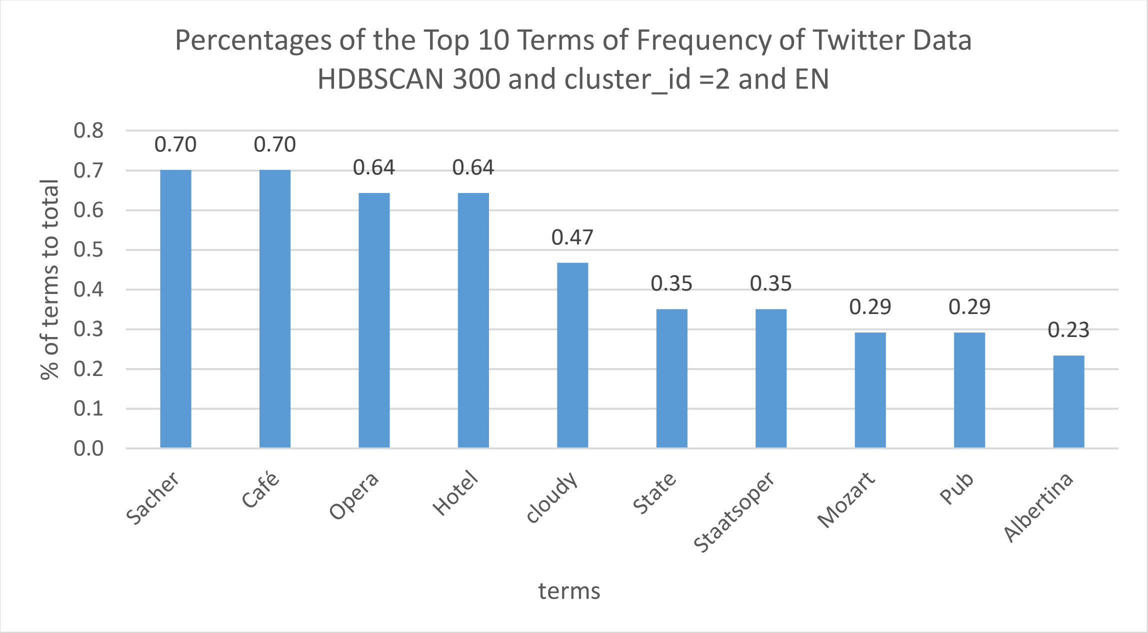
Twitter Data Charts

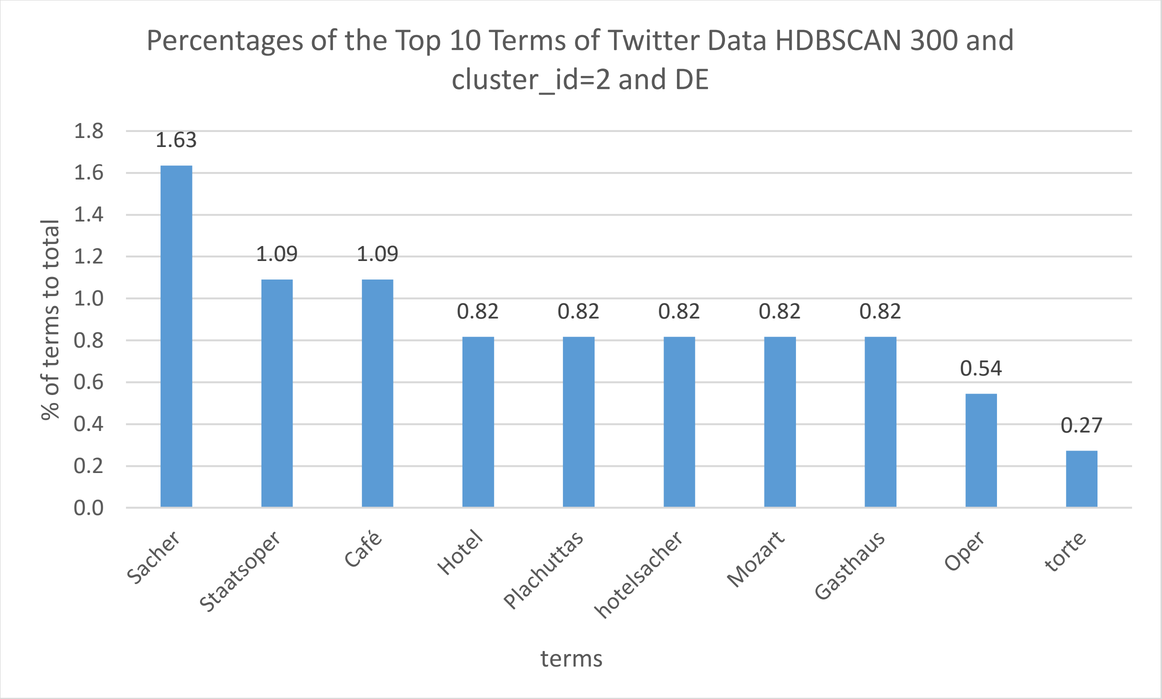
HDBSCAN 300 Twitter Data: Cluster_id=3 and DE and EN



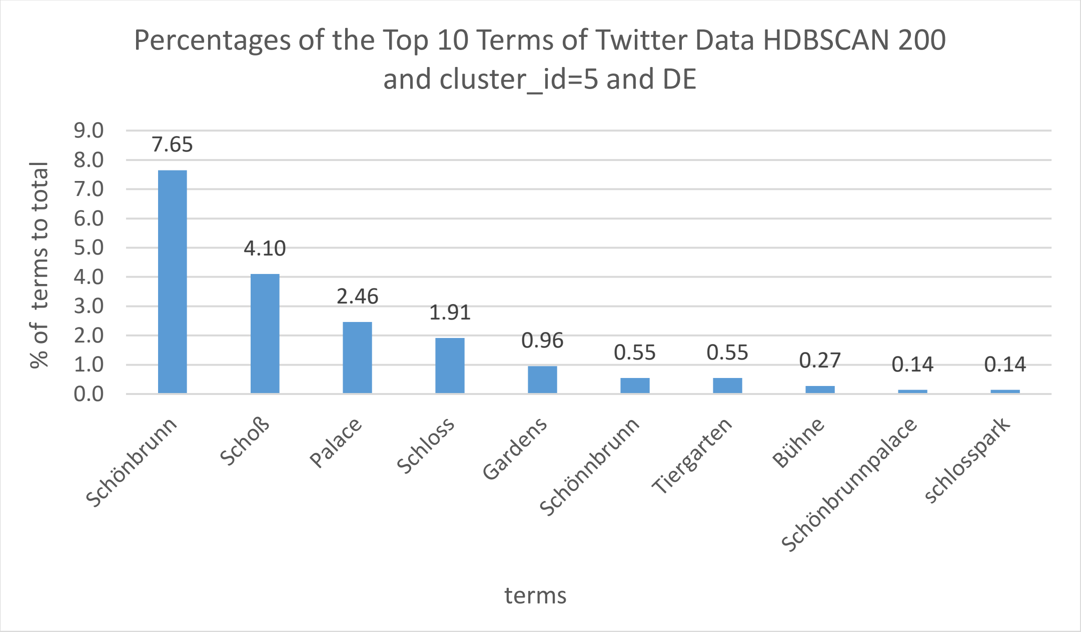


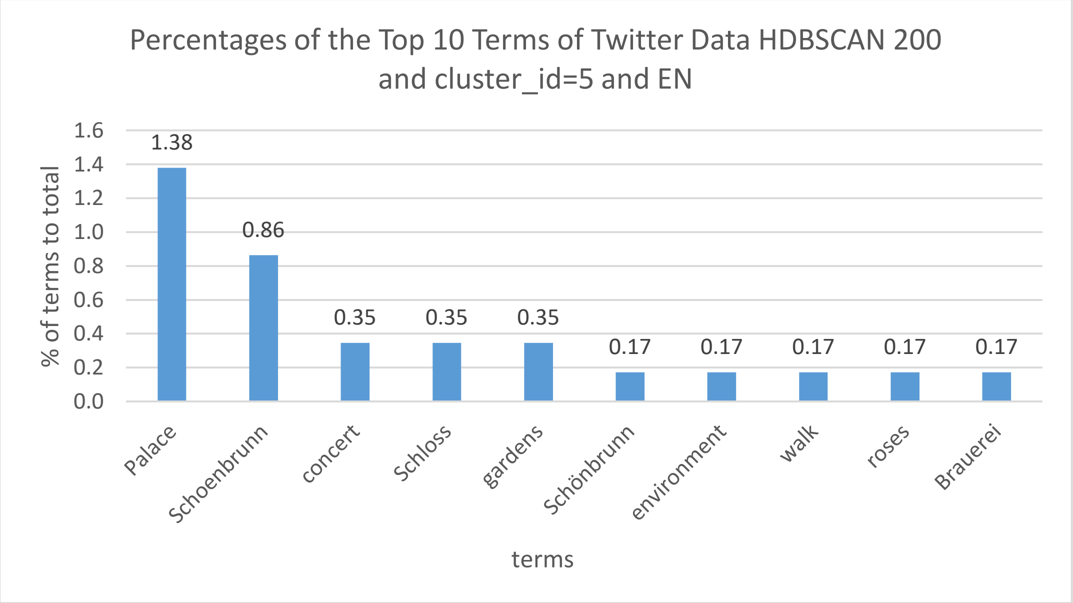
HDBSCAN 300 Twitter Data: Cluster_id=2 and DE and EN



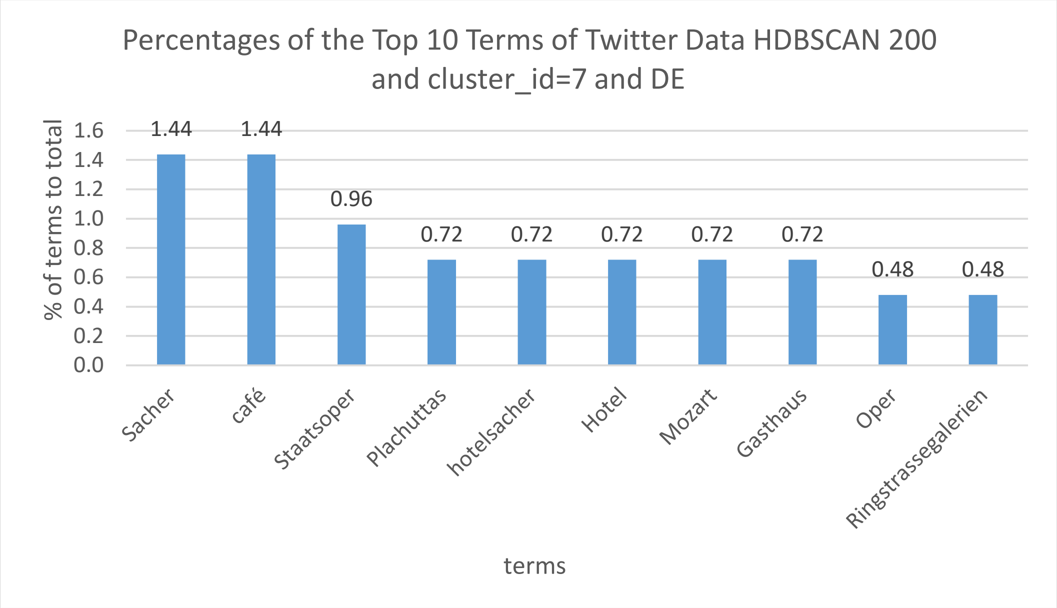


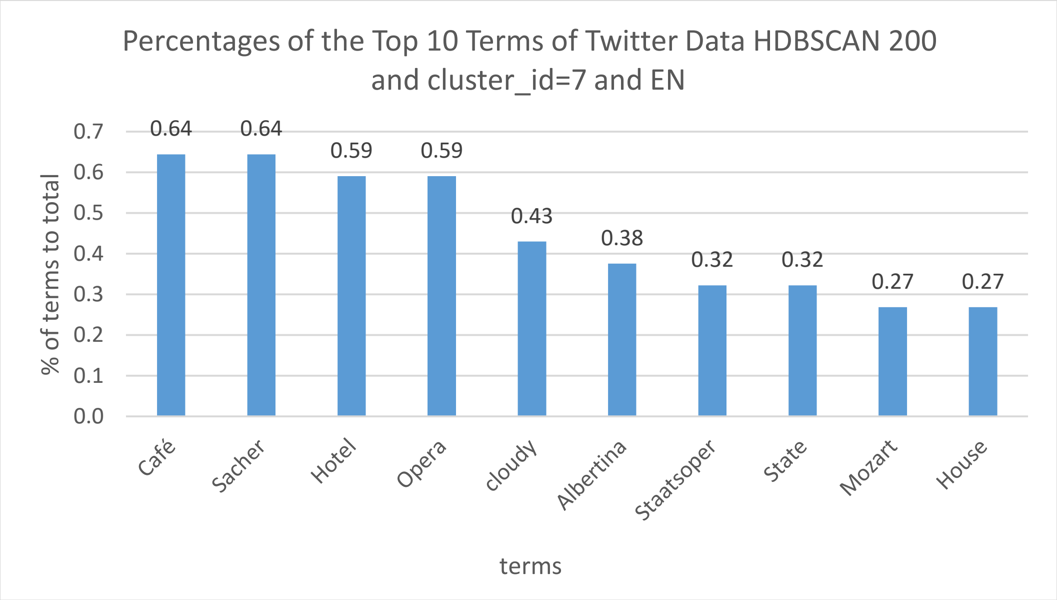
HDBSCAN 200 Twitter Data: Cluster_id=5 and DE and EN



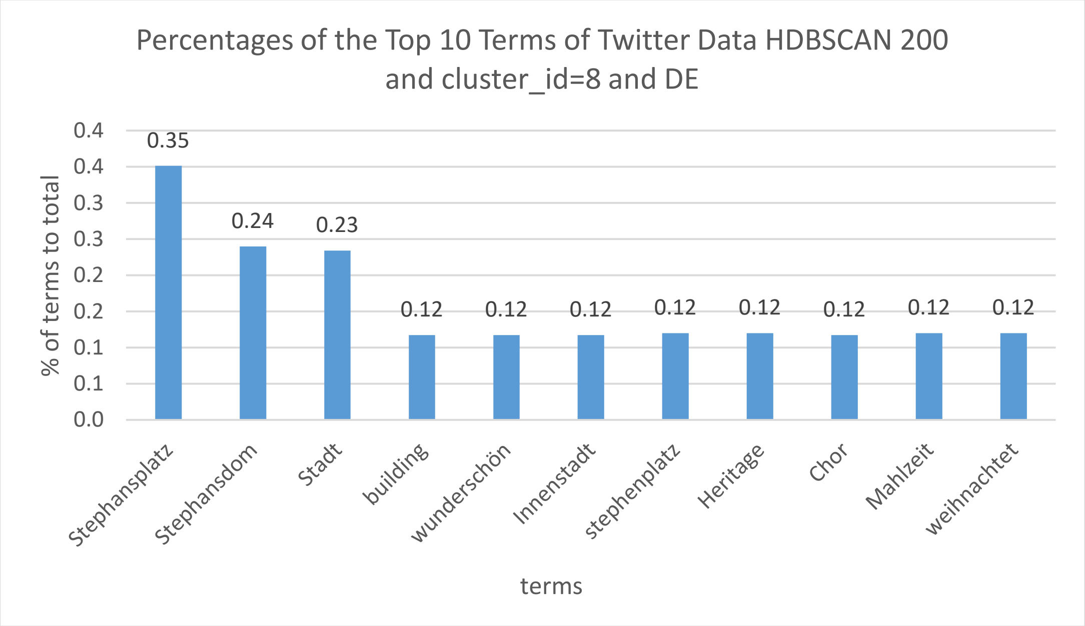


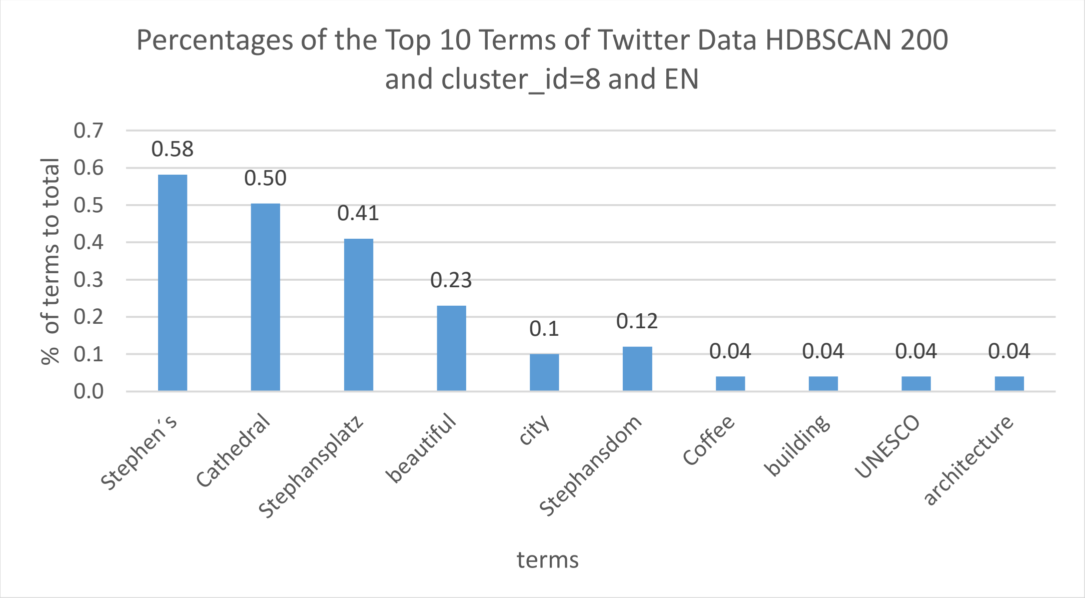
HDBSCAN 200 Twitter Data: Cluster_id=7 and DE and EN



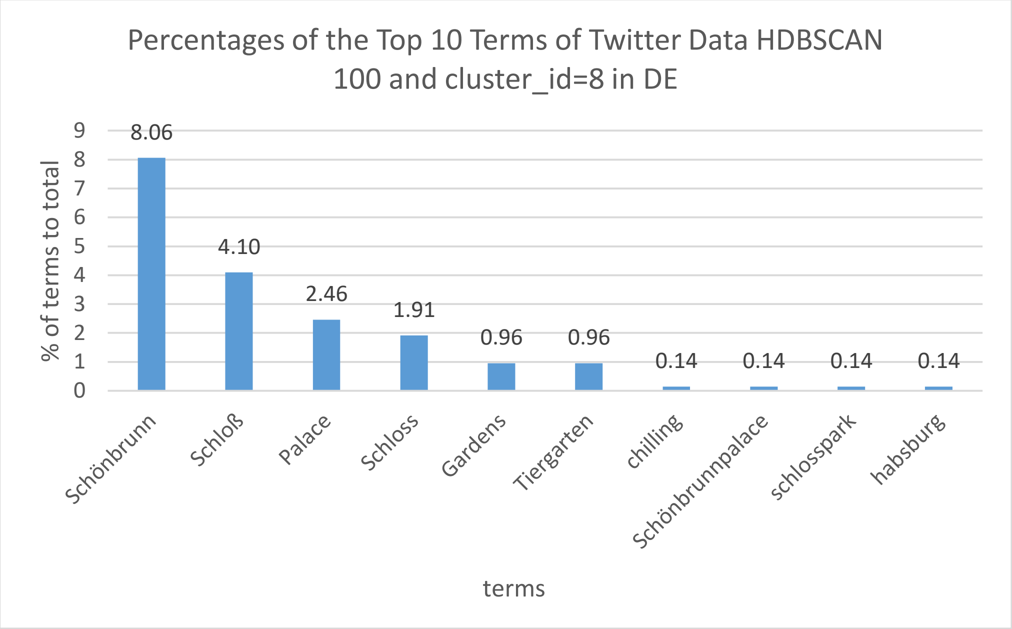


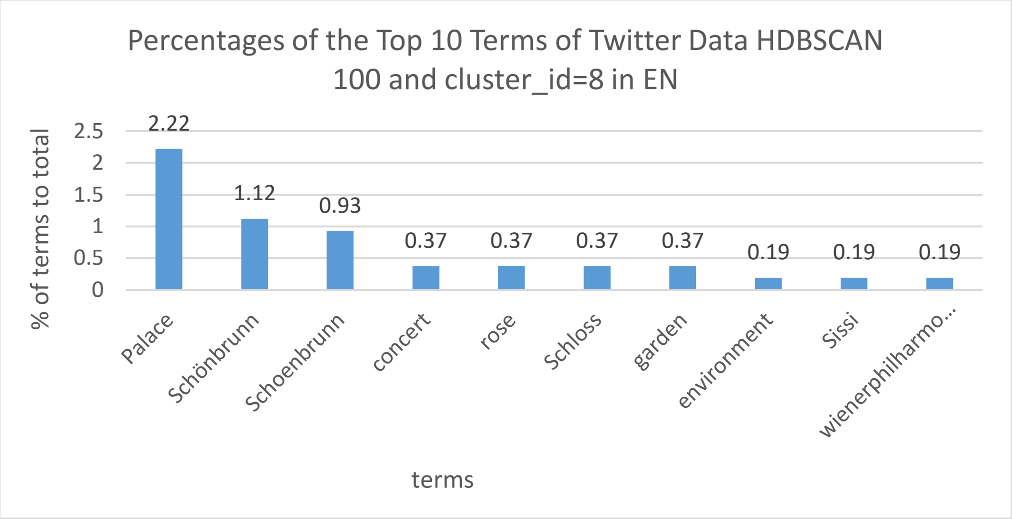
HDBSCAN 200 Twitter Data: Cluster_id=8 and DE and EN



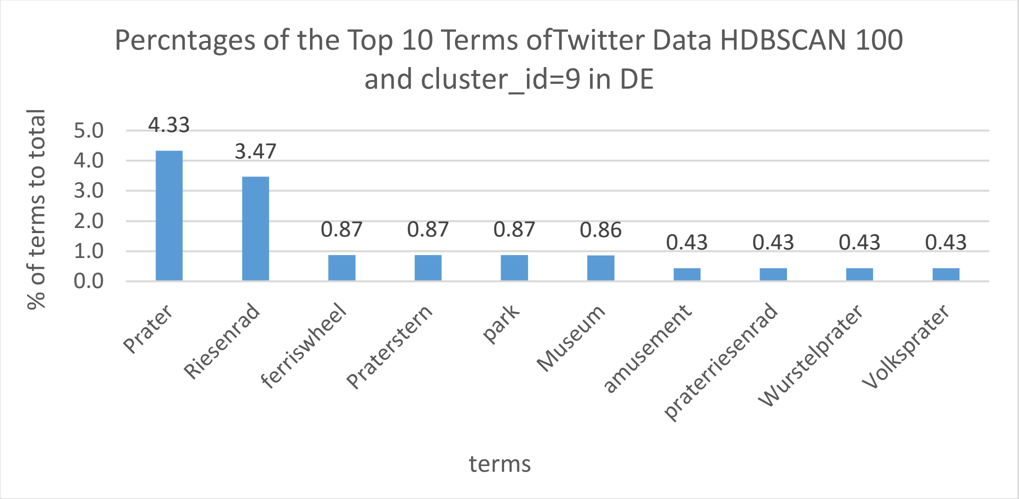


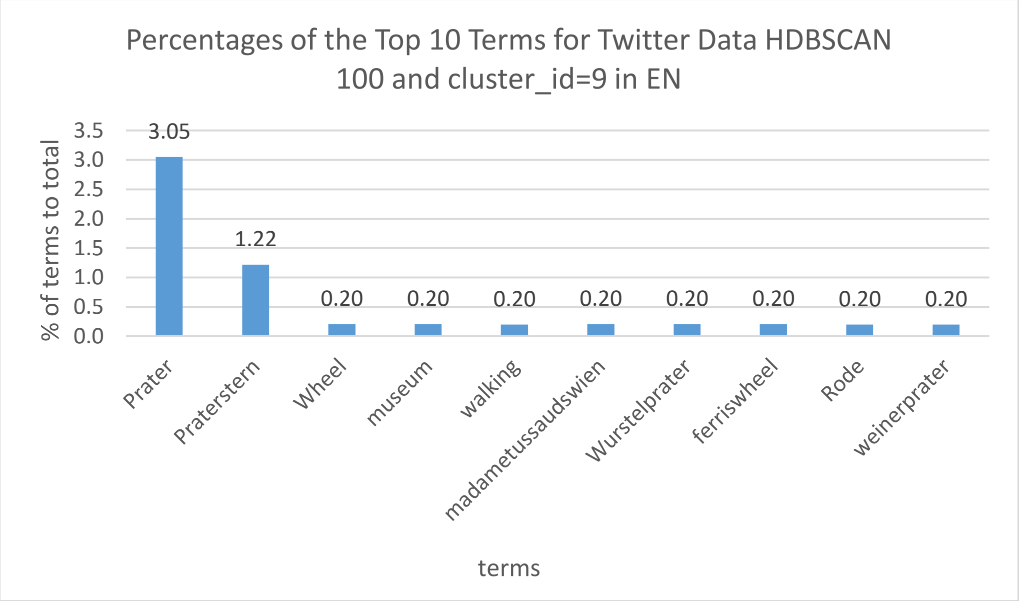
HDBSCAN 100 Twitter Data: Cluster_id=8 and DE and EN



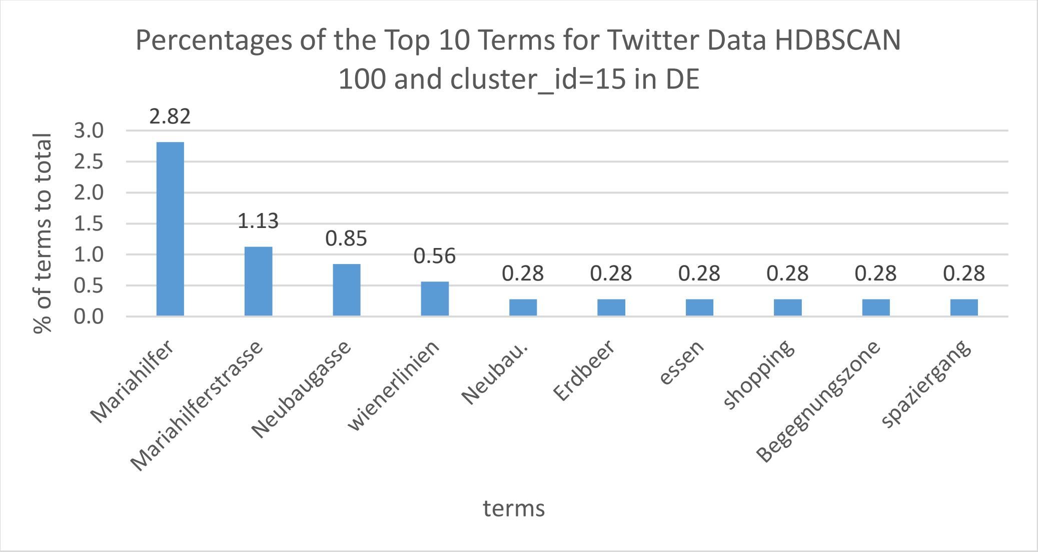


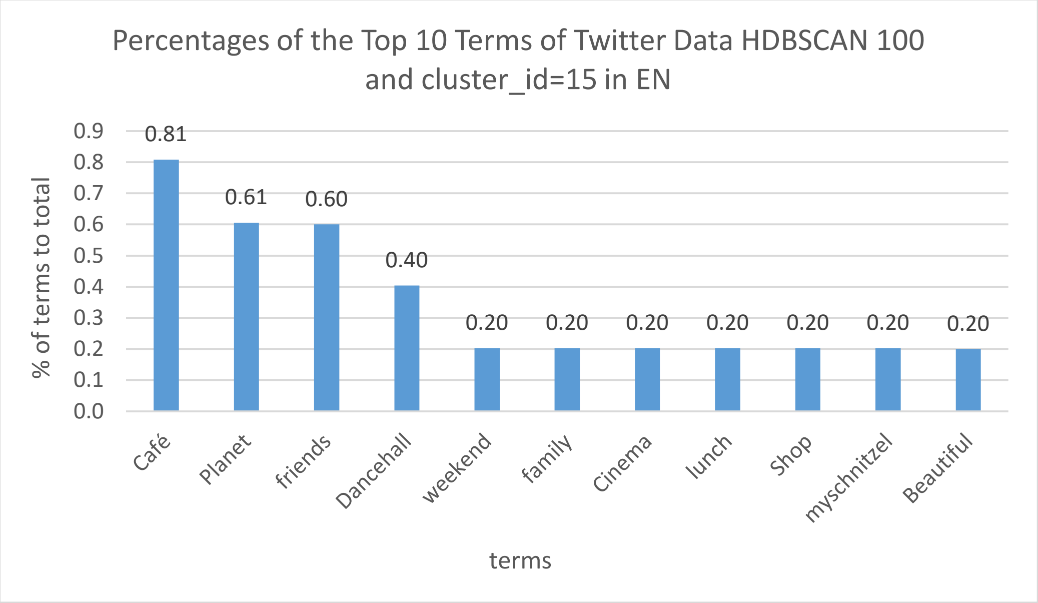
HDBSCAN 100 Twitter Data: Cluster_id=9 and DE and EN



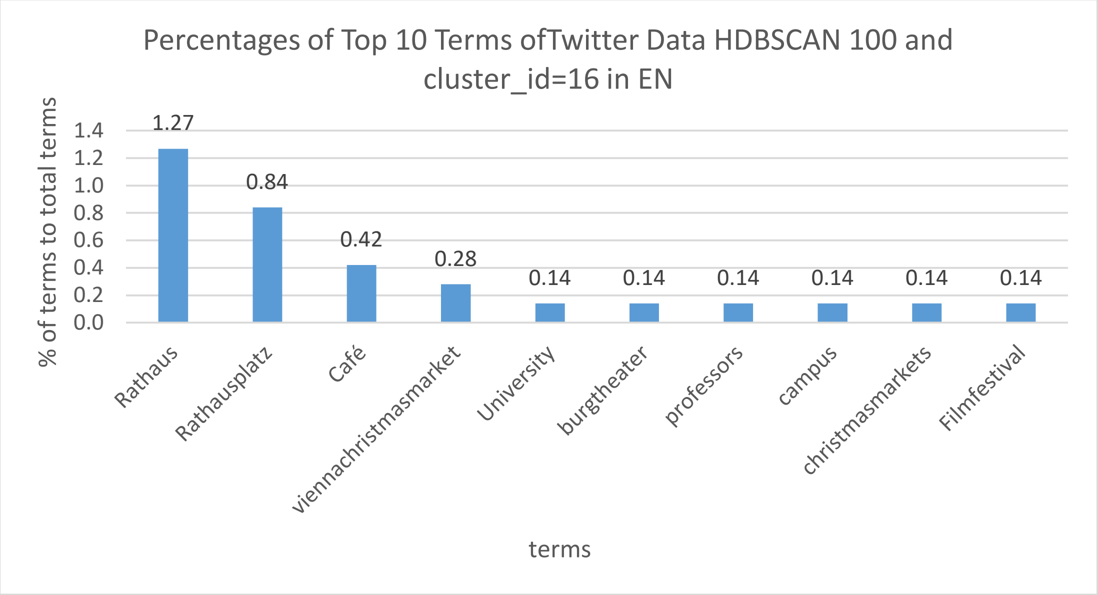


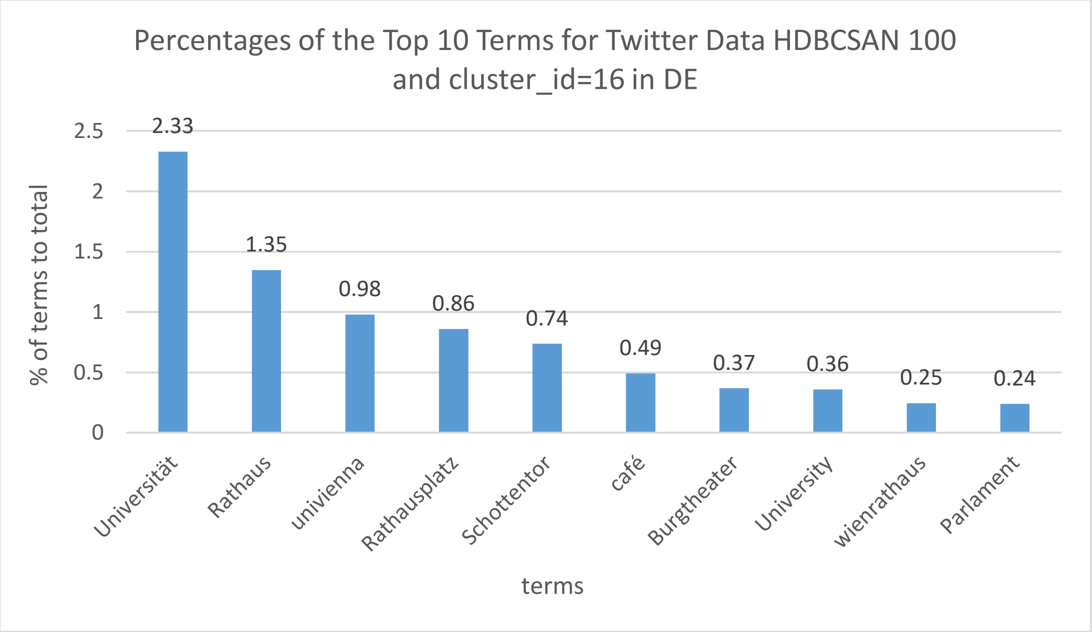
HDBSCAN 100 Twitter Data: Cluster_id=15 and DE and EN



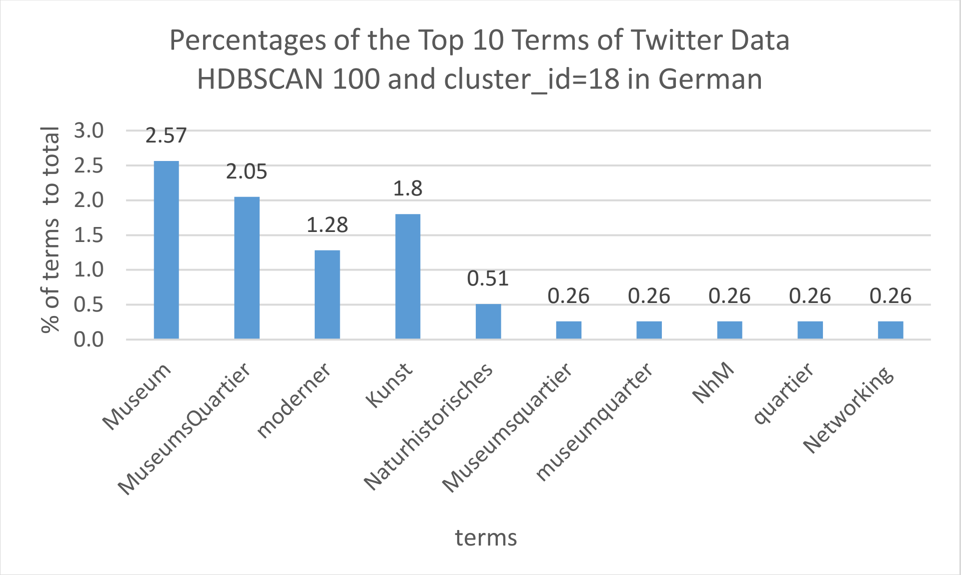


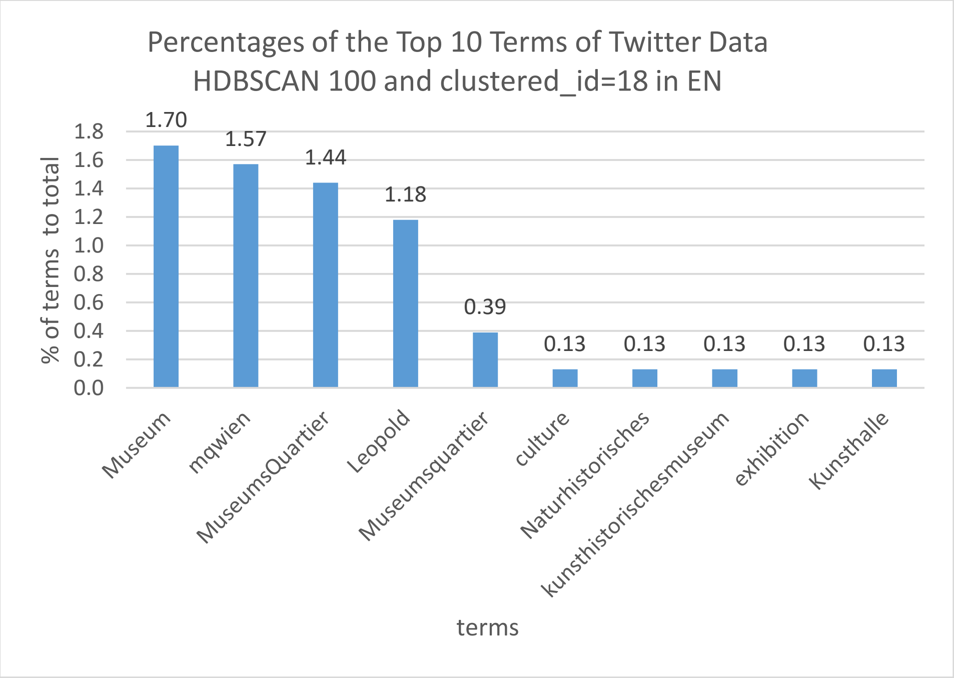
HDBSCAN 100 Twitter Data: Cluster_id=16 and DE and EN



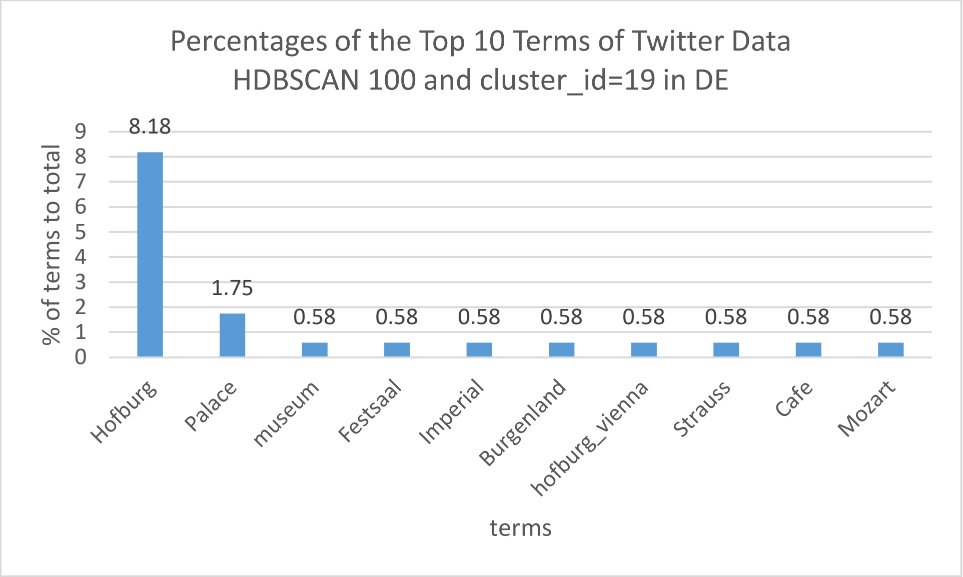


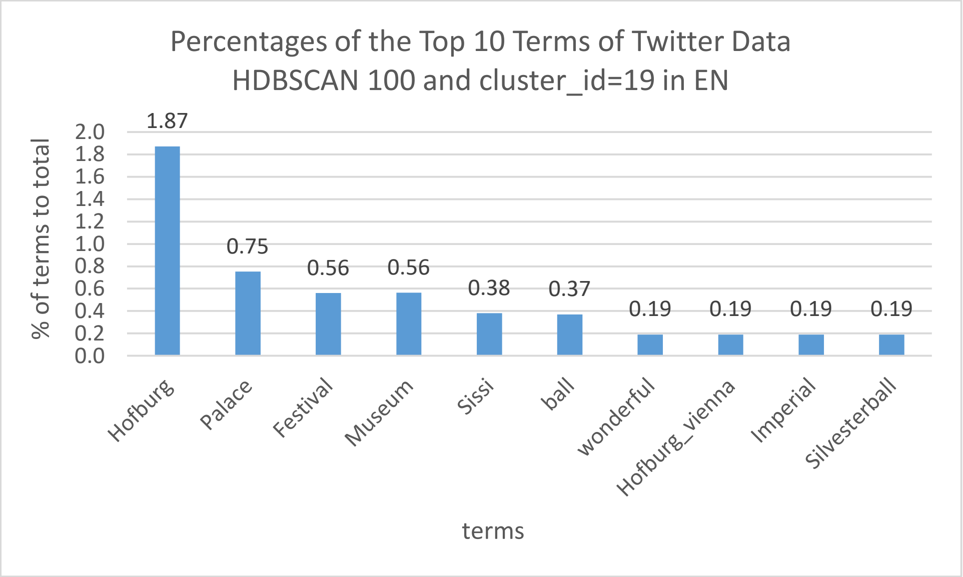
HDBSCAN 100 Twitter Data: Cluster_id=18 and DE and EN



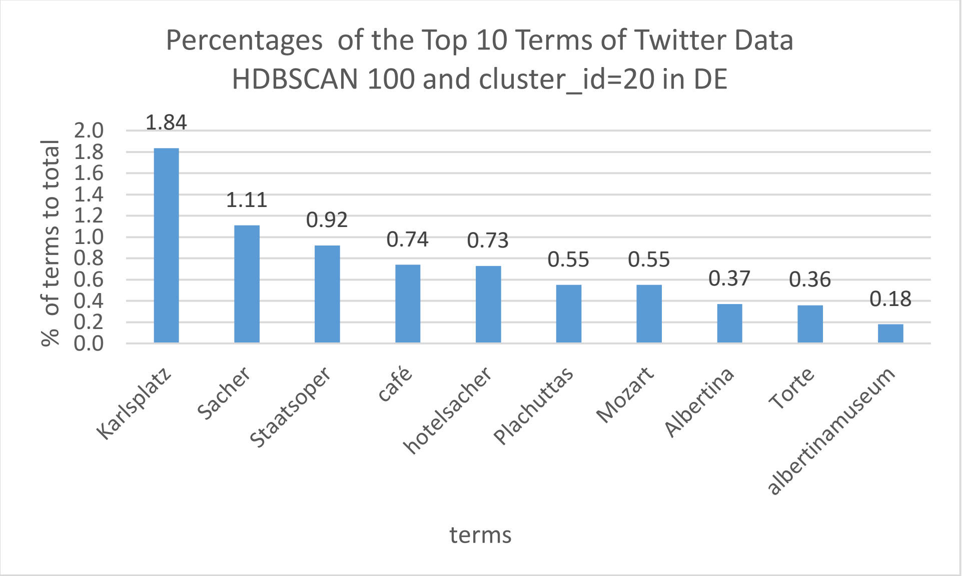


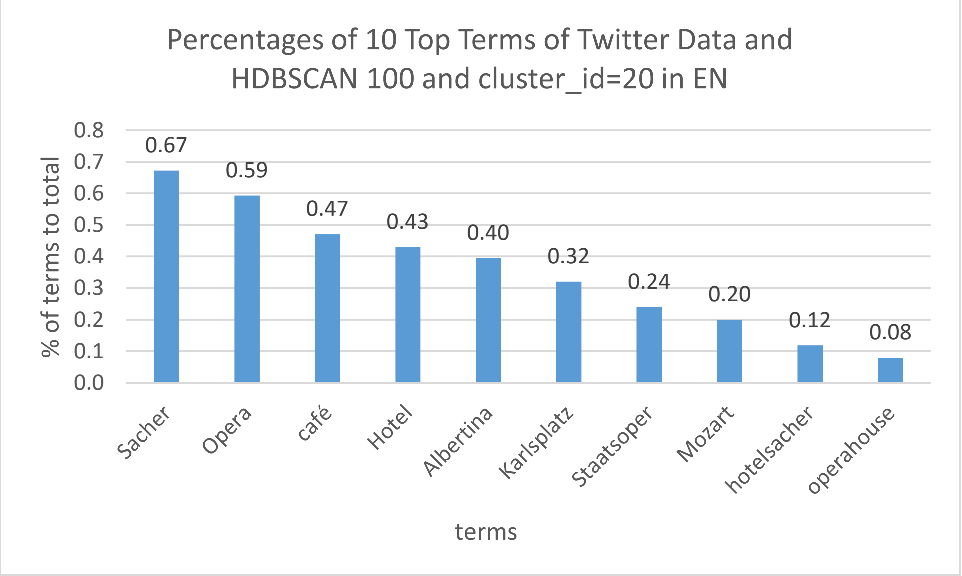
HDBSCAN 100 Twitter Data: Cluster_id=19 and DE and EN



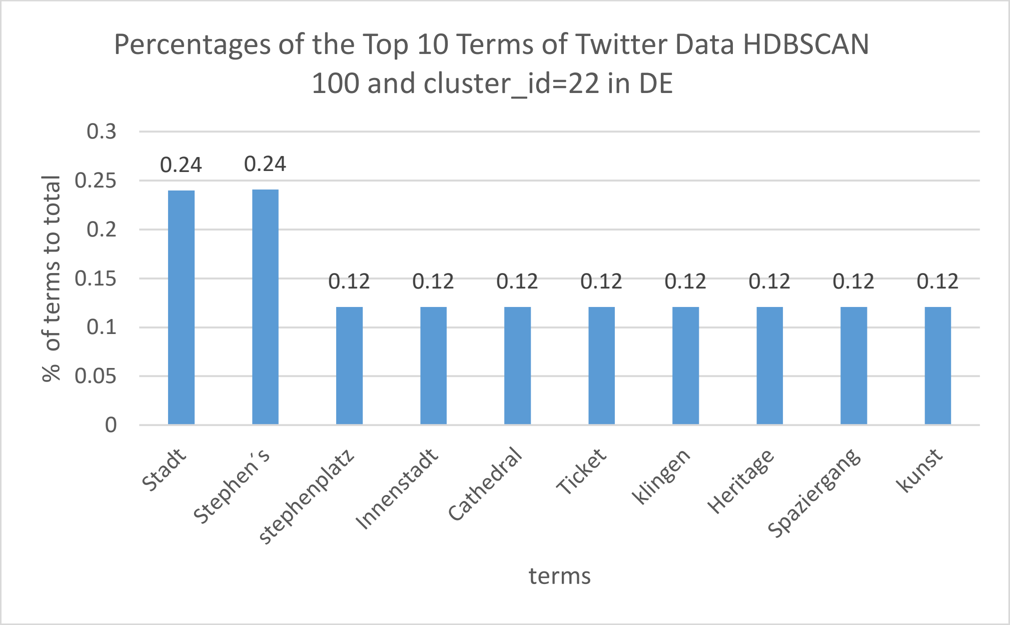


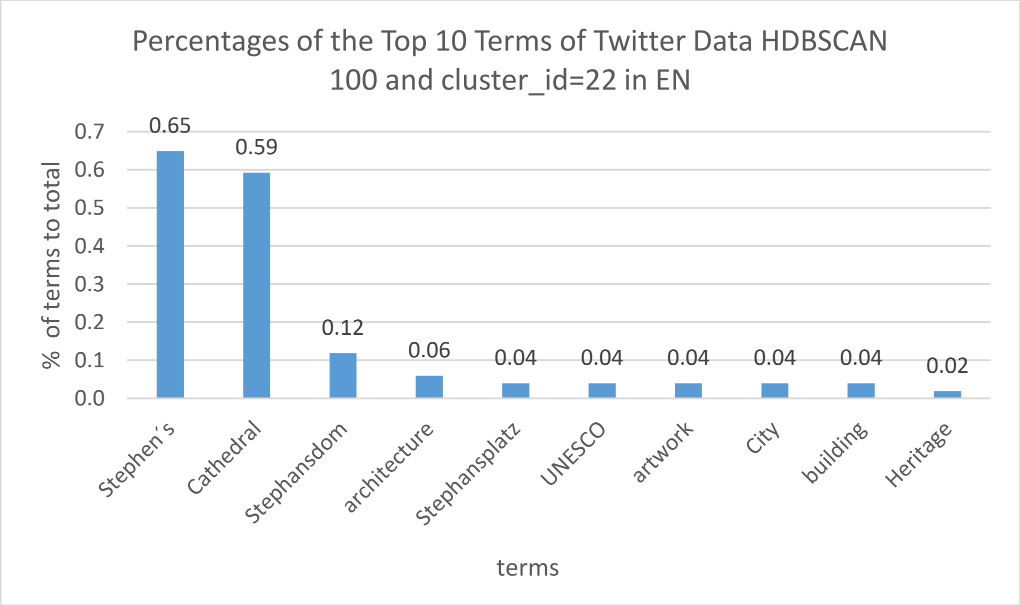
HDBSCAN 100 Twitter Data: Cluster_id=20 and DE and EN





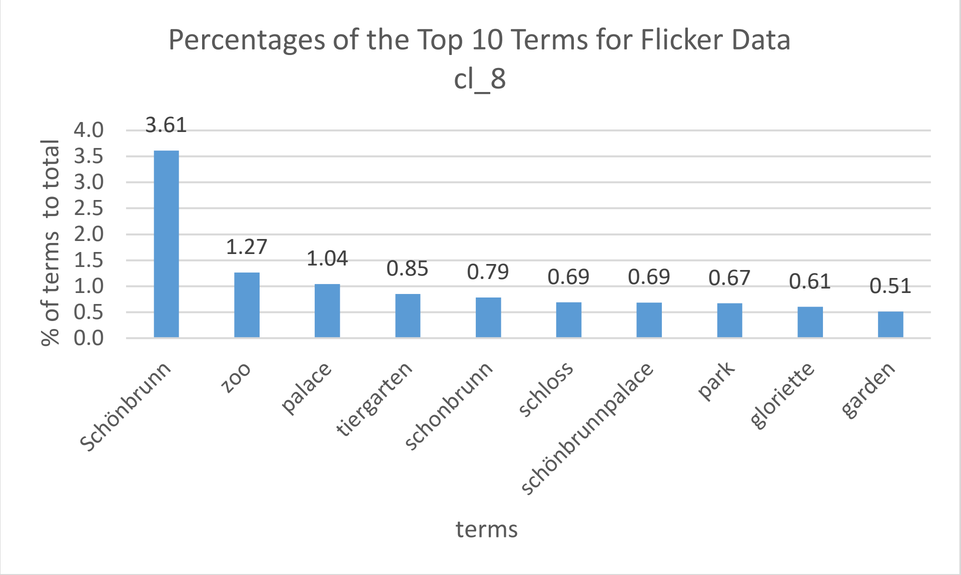
HDBSCAN 100 Twitter Data: Cluster_id=22 and DE and EN



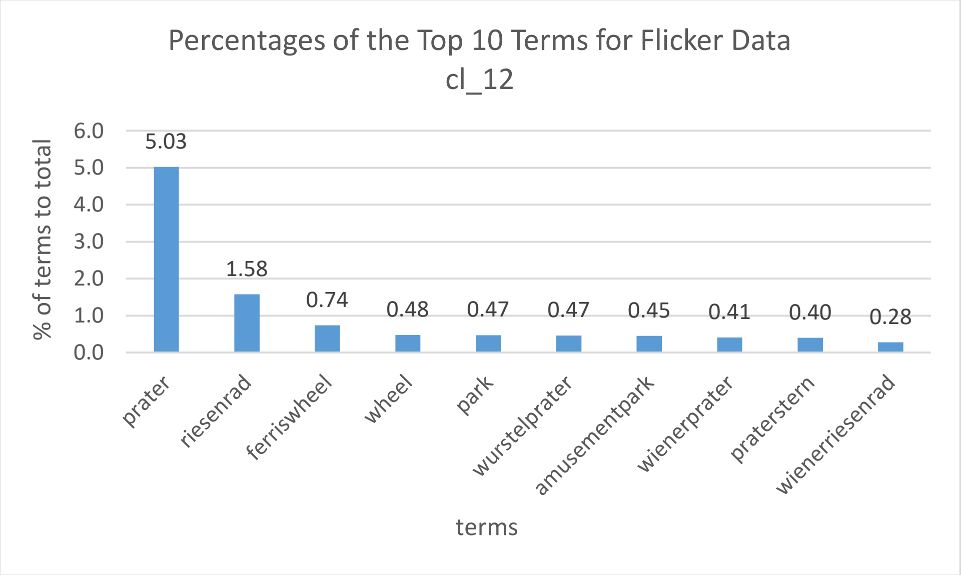


Flickr Data Charts

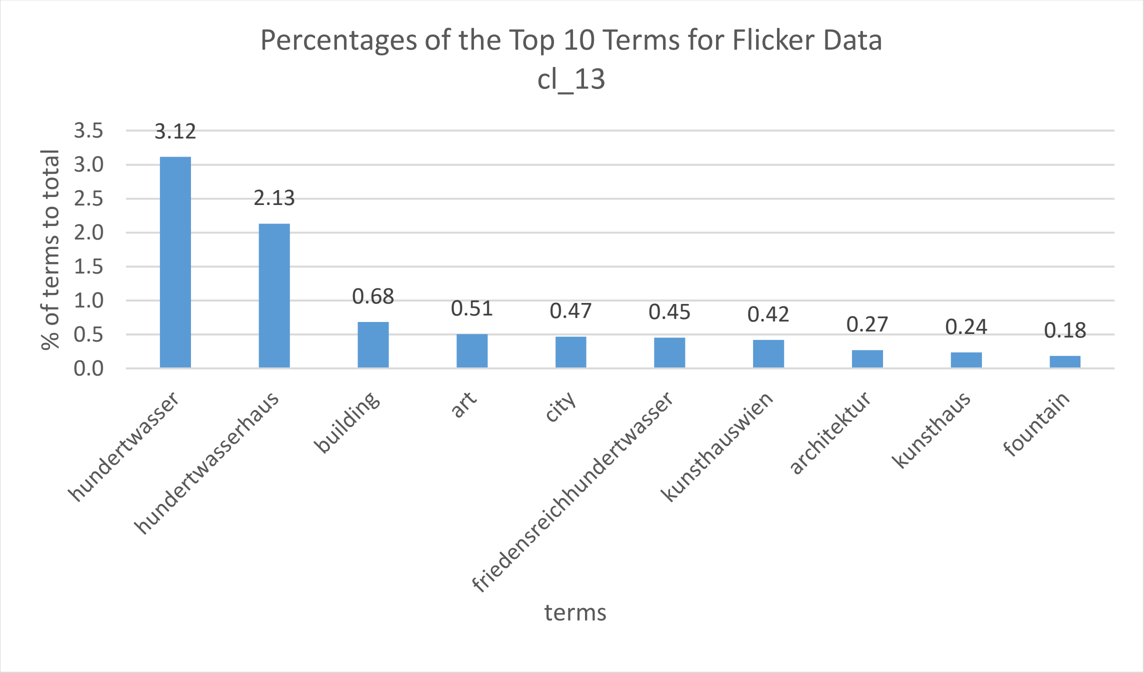
Flicker: cluster_id=8



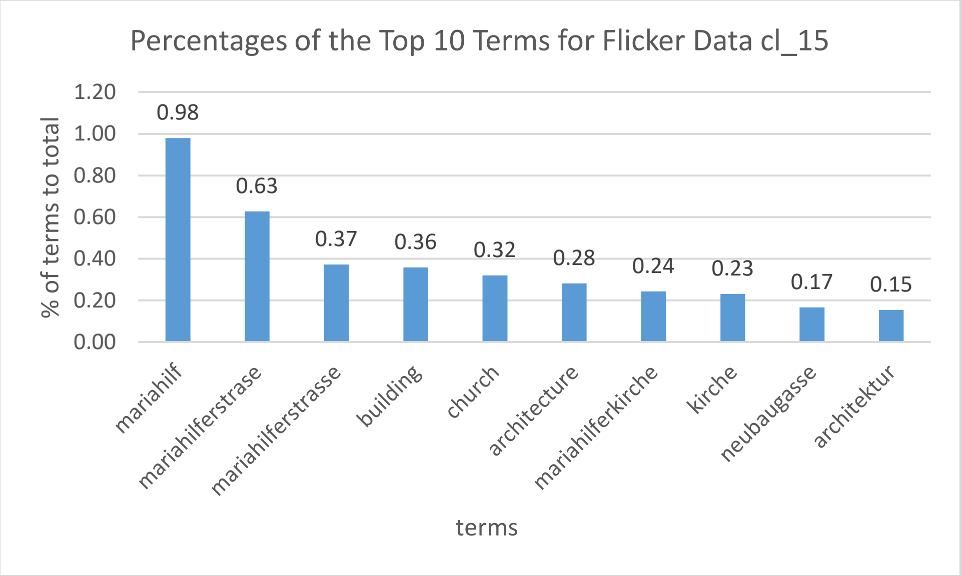
Flicker: cluster_id=12



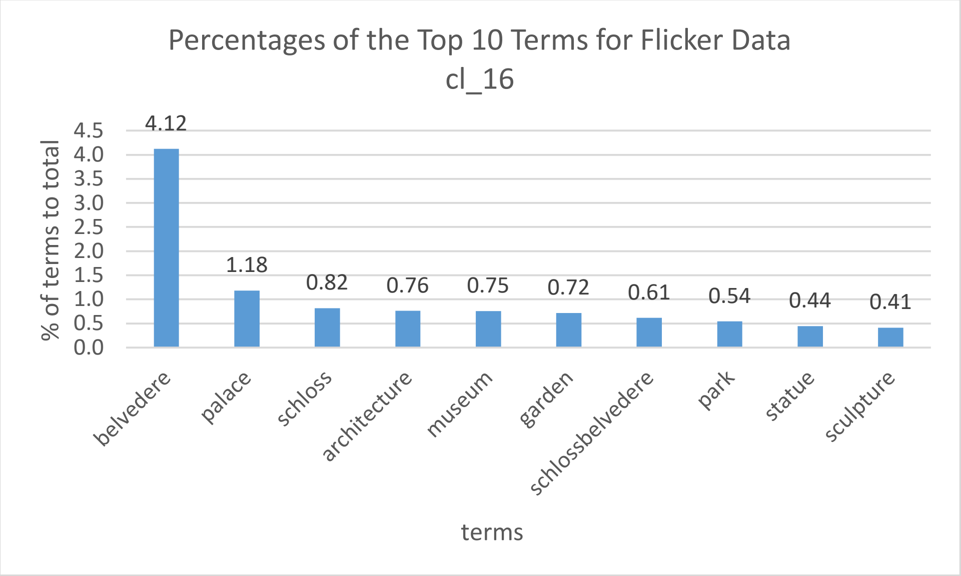
Flicker: cluster_id=13



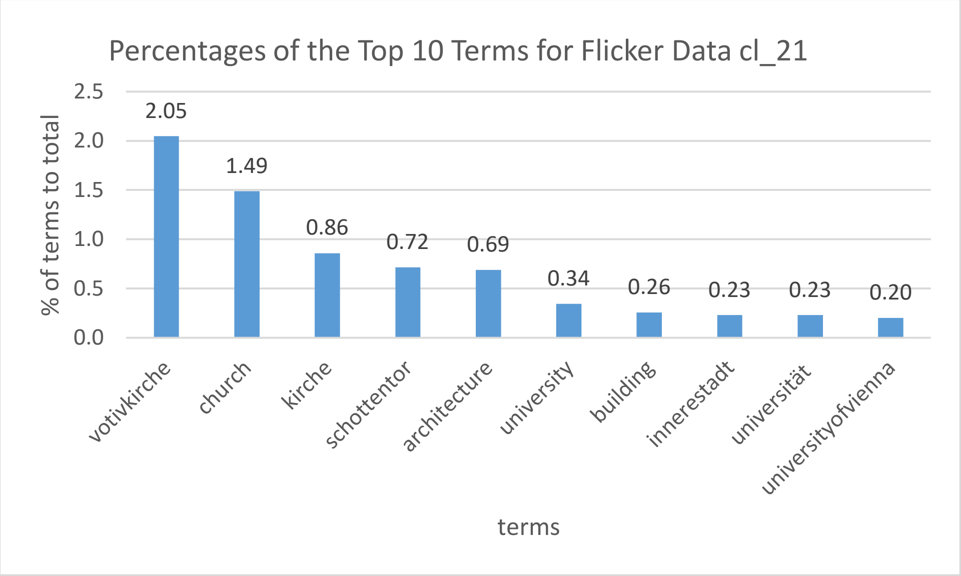
Flicker: cluster_id=15



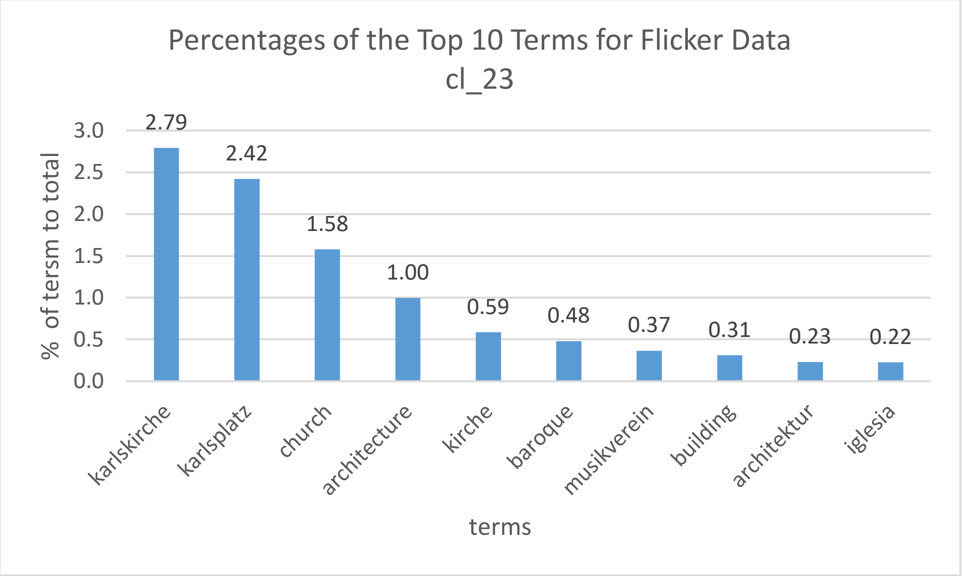
Flicker: cluster_id=16



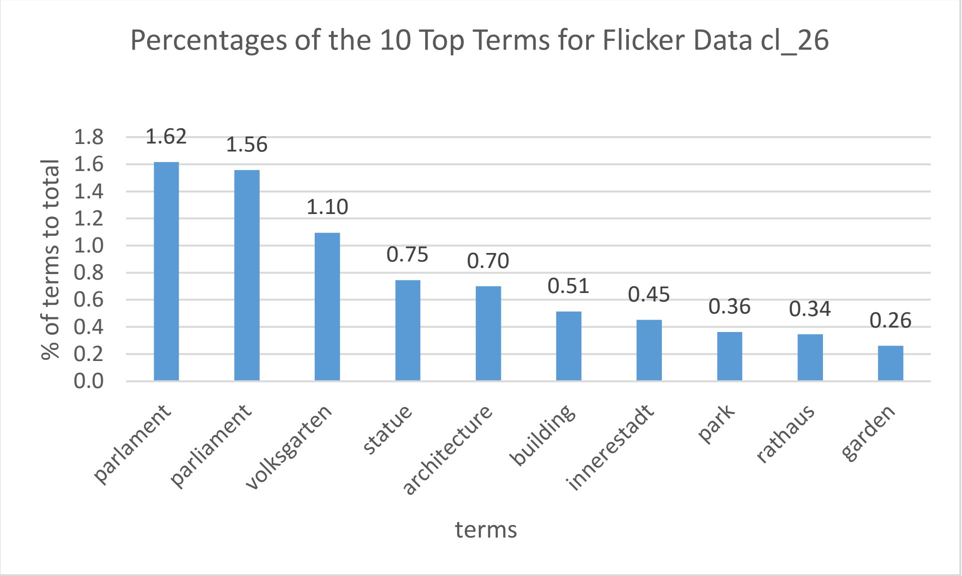
Flicker: cluster_id=21



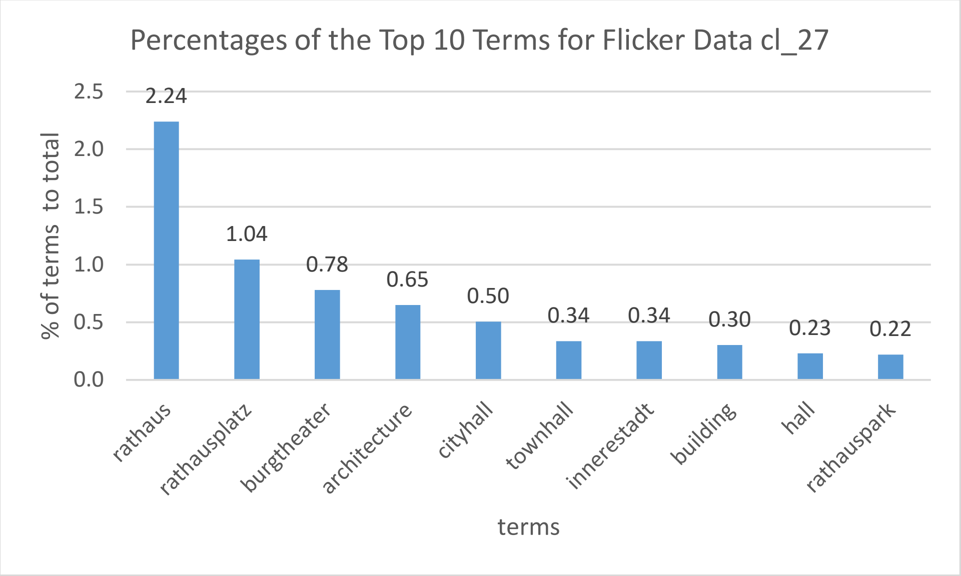
Flicker: cluster_id=23



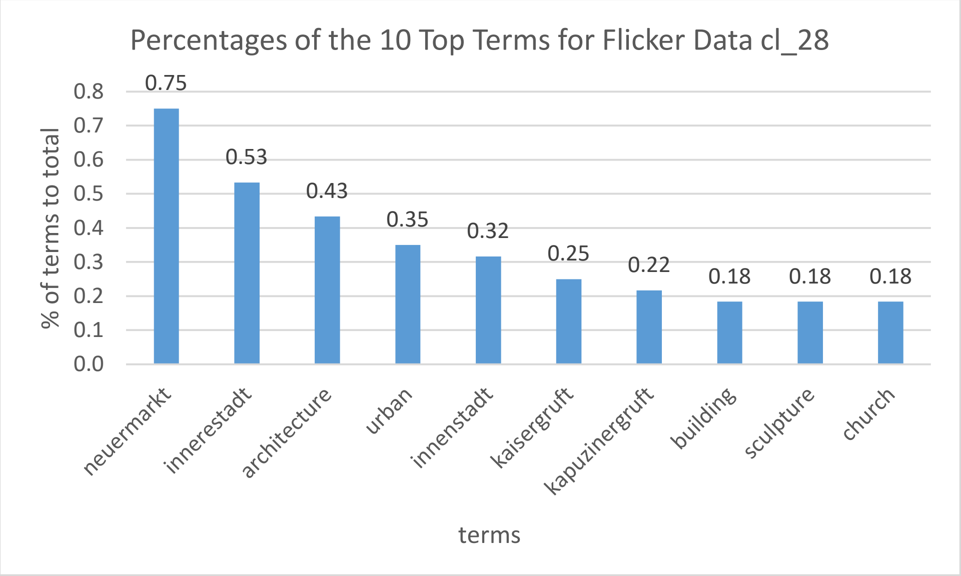
Flicker: cluster_id=26



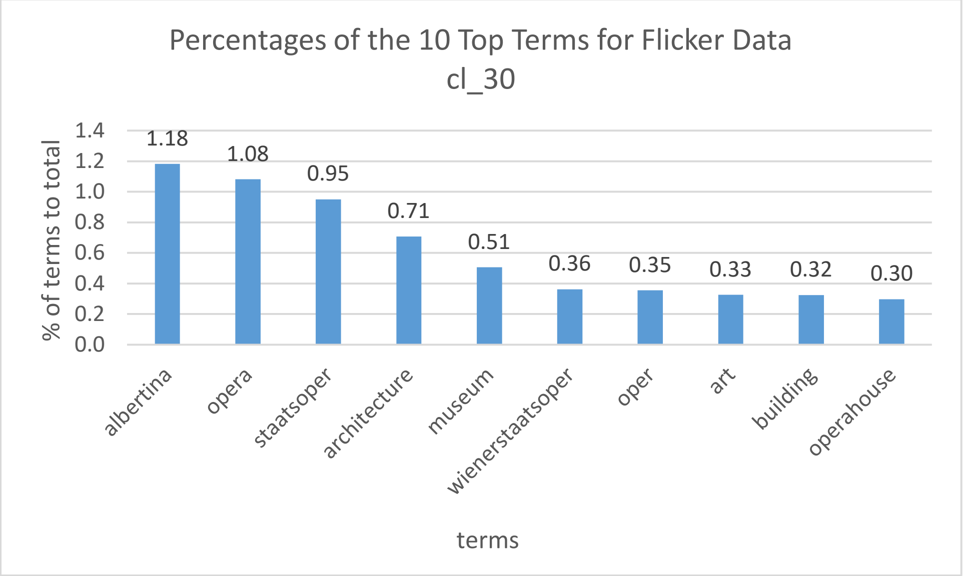
Flicker: cluster_id=27



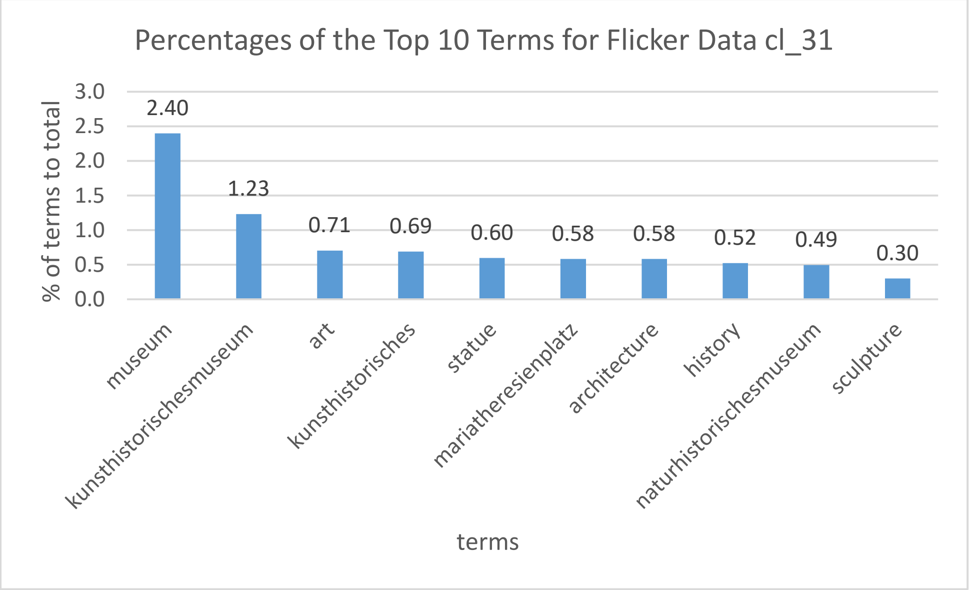
Flicker: cluster_id=28



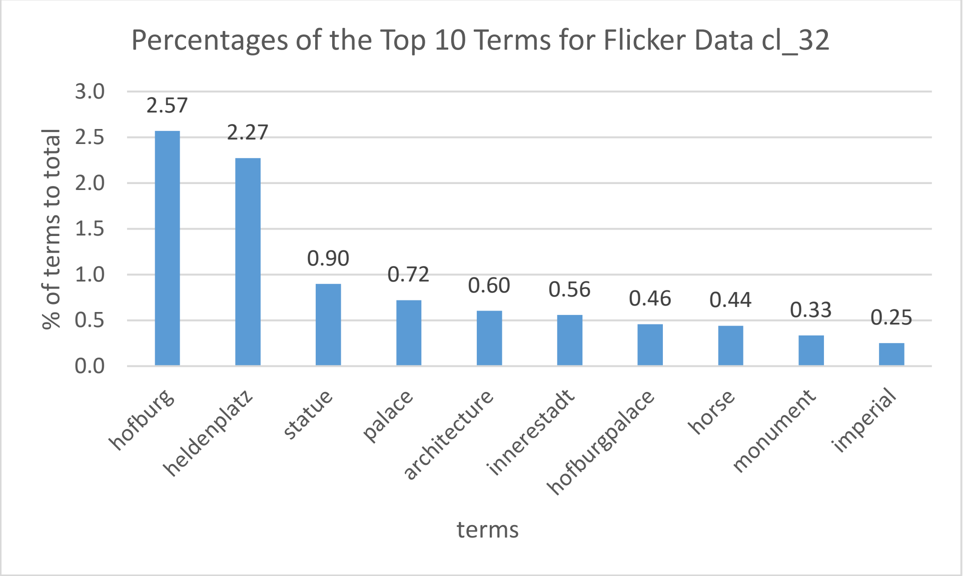
Flicker: cluster_id=30



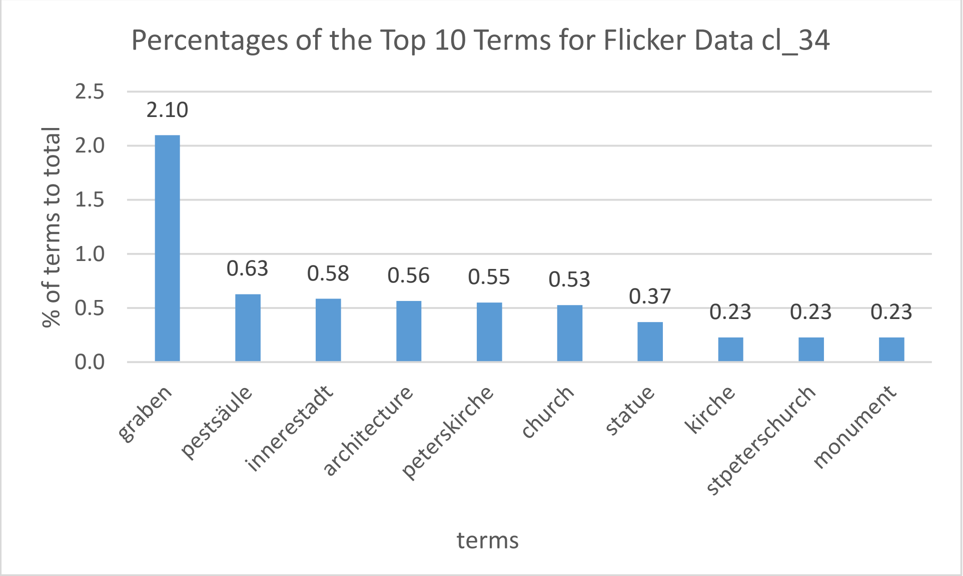
Flicker: cluster_id=31



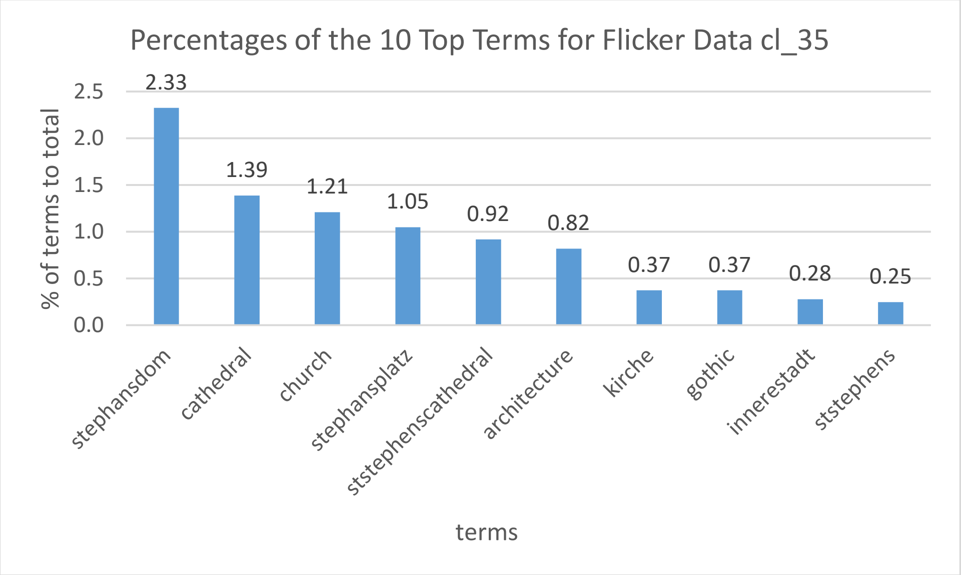
Flicker: cluster_id=32



Flicker: cluster_id=34



Flicker: cluster_id=35



Comparing Tables of Flickr and Twitter Data

Top 10 Terms twitter	fr(i)%	Sum(%)	Top 10 Terms flickr	fr(i)%	Sum(%)
Schönbrunn	4.71		Schönbrunn	3.61	
Palace	2.38		zoo	1.27	
Schloß	2.07		palace	1.04	
Schloss	1.40		tiergarten	0.85	
Tiergarten	0.26	11.74	schonbrunn	0.79	10.72
Schlosspark	0.10		schloss	0.69	
concert	0.10		schönbrunnpalace	0.69	
garden	0.10		park	0.67	
Schoenbrunn	0.31		gloriette	0.61	
Schonbrunn	0.31		garden	0.51	
Rest	88.26		Rest	89.28	

Top 10 Terms twitter	fr(i)%	Sum(%)	Top 10 Terms flickr	fr(i)%	Sum(%)
Prater	2.55		prater	5.03	
Riesenrad	0.72		riesenrad	1.58	
Praterstern	0.59		ferriswheel	0.74	
Wurstelprater	0.40		wheel	0.48	
park	0.13	4.84	park	0.47	10.30
ferriswheel	0.20		wurstelprater	0.47	
Rode	0.07		amusementpark	0.45	
Volksprater	0.07		wienerprater	0.41	
tlandmark	0.07		Praterstern	0.40	
Lunapark	0.07		wienerriesenrad	0.28	
Rest	95.16		Rest	89.70	

Top 10 Terms twitter	fr(i)%	Sum(%)	Top 10 Terms flick	fr(i)%	Sum(%)
Mariahilfer	0.82		mariahilf	0.98	
Straße	0.82		mariahilferstrase	0.63	
Café	0.44		mariahilferstrasse	0.37	
Mariahilferstraße	0.44		building	0.36	
Neubaugasse	0.25	3.34	church	0.32	3.73
Planet	0.19		architecture	0.28	
Cinema	0.13		mariahilferkirche	0.24	
Dancehall	0.13		kirche	0.23	
Mariahilf	0.06		neubaugasse	0.17	
coffee	0.06		architektur	0.15	
Rest	96.51		Rest	96.27	

Top 10 Terms twitter	fr(i)%	Sum	Top 10 Terms flick	fr(i)%	Sum(%)
Universität	0.96		rathaus	2.24	
Rathaus	1.12		rathausplatz	1.04	
Rathausplatz	0.46		burgtheater	0.78	
Schottentor	0.42	4.08	architecture	0.65	
univienna	0.31		cityhall	0.50	6.64
café	0.27		townhall	0.34	
Hall	0.20		innerestadt	0.34	
Burgtheater	0.15		building	0.30	
Rathauspark	0.12		hall	0.23	
University	0.08		rathauspark	0.22	
Rest	95.92		Rest	93.36	

Top 10 Terms twitter	fr(i)%	Sum(%)	Top 10 Terms fickr	fr(i)%	Sum(%)
MuseumsQuartier	2.15		museum	2.40	
Museum	1.11		kunsthistorischesmuseum	1.23	
Leopold	0.44		art	0.71	
Kunst	0.33		kunsthistorisches	0.69	
moderner	0.28	5.35	statue	0.60	8.11
art	0.28		mariatheresienplatz	0.58	
Naturhistorisches	0.22		architecture	0.58	
Museumsquartier	0.22		history	0.52	
Maria-Theresien-Platz	0.22		naturhistorischesmuseum	0.49	
exhibition	0.11		sculpture	0.30	
Rest	94.65		Rest	91.89	

Top 10 Terms twitter	fr(i)%	Sum(%)	Top 10 Terms flickr	fr(i)%	Sum(%)
Hofburg	3.18		hofburg	2.57	
Palace	1.03		heldenplatz	2.27	
Imperial	0.52		statue	0.90	
Museum	0.34		palace	0.72	
ball	0.26	6.10	architecture	0.60	9.11
Festival	0.26		innerestadt	0.56	
Sisi	0.17		hofburgpalace	0.46	
hofburg_vienna	0.17		horse	0.44	
Horses	0.09		monument	0.33	
hofburgsilvesterball	0.09		imperial	0.25	
Rest	93.90		Rest	90.89	

Top 10 Terms twitter	fr(i)%	Sum	Top 10 Terms flickr	fr(i)%	Sum(%)
Staatsoper	0.57		albertina	1.18	
Sacher	0.55		opera	1.08	
Café	0.47		staatsoper	0.95	
Oper	0.42	3.67	architecture	0.71	
Karlsplatz	0.39		museum	0.51	6.09
Albertina	0.39		wienerstaatsoper	0.36	
Mozart	0.26		oper	0.35	
Opera	0.24		art	0.33	
State	0.24		building	0.32	
Museum	0.14		operahouse	0.30	
Rest	96.33		Rest	93.91	

Total 10 Terms twitter	fr(i)%	Sum	Top 10 Terms flickr	fr(i)%	Sum(%)
Stephens	0.45		stephansdom	2.33	
Cathedral	0.41		cathedral	1.39	
Stephansdom	0.10		church	1.21	
Heritage	0.07		stephansplatz	1.05	
architecture	0.06	1.22	ststephenscathedral	0.92	8.99
UNESCO	0.04		architecture	0.82	
Art	0.02		kirche	0.37	
Albertina	0.02		gothic	0.37	
Stephansplatz	0.02		innerestadt	0.28	
Stadt	0.02		ststephens	0.25	
Rest	98.78		Rest	91.01	