# 3. Theoretical Concepts

This section explains the most relevant theoretical concepts that were applied during the course of this project to provide a fundamental understanding.

## 3.1 Chatbot

**A chatbot is a computer program that interacts with its user via a chat interface1.** About 60 years ago, the first steps in the development of chatbots were taken by the computer scientists Alan Turing and Joseph Weizenbaum, proposing the concept of computers communicating like humans do*.* One of the first natural language processing programs was ELIZA, developed by Joseph Weizenbaum in 1966. Although some users were tricked into thinking that ELIZA was an actual human conversation partner ², this basic example was stretched to its limit quickly because of its simple rule-based structure. However, the fascination of computers being a conversation partner remained one of the big objectives of modern artificial intelligence.

The topic’s big revival occurred with the introduction of mobile devices in the early 2000s. All of the sudden, developers were faced with the task of transforming their well-known desktop applications and websites into apps to make them suitable for a mobile market. Still, in the last couple of years it turned out that users actually do not like to use a variety of apps, but rather concentrate on only a few, mainly messaging apps. That is how in 2016, the idea of the conversational interface resurged when a lot global big-players like Google, Facebook, Microsoft, IBM or Amazon decided to take part in the development of chatbots and conversational interfaces 3. Many of them provide free to use natural language processing platforms that make the development of conversational interfaces possible. These interfaces can be then integrated into instant messaging applications. Two of the key concepts that are widely used in most natural language processing systems are 4

* Intent: The mapping of the user’s message to an action that should be taken
* Entities: Objects the application or device takes action on, can be often considered as the action’s parameters. Entities can be either defined by the user or are commonly supported in natural language system, containing types such as Location or Date.

*! In this section, the following parts are going to be examined in a more detailed way: difference and relation between natural language processing, conversational interface and chatbot, key concepts and way of functioning*

## 3.2 OpenStreetMap

*OpenStreetMap5* (or short OSM)is a collaborative project with the aim to collect and update free to use geographic data. Its main purpose is to be a central data source which can be e.g. used for rendering maps. The stored data contains infrastructural information such as roads or buildings as well as variety of additional informational tags. In this project, the OpenStreetMap data is used to extract necessary tourist information in order to create user recommendations.

### 3.2. Data Structure

Regarding the data organization, OpenStreetMap’s structure consists of four principal elements6:

* **Nodes**: Points with a geographic position that are used to represent map features without a size, such as points of interest or mountain peaks.
* **Ways**: Ordered lists of nodes that are used both for representing linear features such as streets and rivers, and areas, like forests, parks, parking areas and lakes.
* **Relations**: Ordered lists of nodes, ways and relations (together called "members"), where each member can optionally have a "role" (a string). Relations are used for representing the relationship of existing nodes and ways. Examples include turn restrictions on roads, routes that span several existing ways (for instance, a long-distance motorway), and areas with holes.
* **Tags**: Key-Value pairs storing metadata of the geographic objects they are attached to (namely node, way or relation). Tags can include type, name and a broad variety of map features.

There are a big number of data dumps available that store the above mentioned data for either the whole planet or smaller regions or cities. These dumps can be downloaded in the file formats XML and PBF and imported into a PostgreSQL database to get access to the OSM data structure.

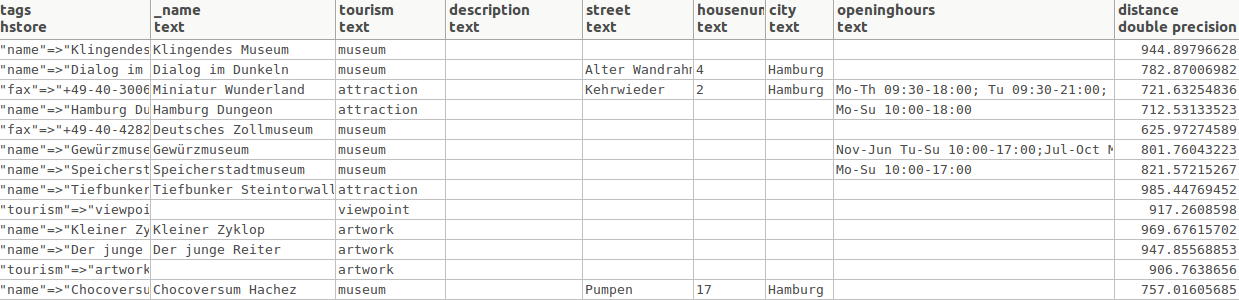
### 3.2.1 Retrieval of Point of Interests

In the context of geographic information, the expression *Point of Interest 7* is often used to describe a feature on a map that can have a certain significance. It is a broad term that reaches from practicable facts like post offices or car parks to tourist attractions. One of this project’s main elements is to retrieve essential tourist information from the vast amount of geographical data, meaning finding relevant points of interests and outputting them in a comprehensible way.

In order to achieve that aim, the data structures Nodes and Tags are mainly investigated. Like already stated, the Tags structure contains key-value pairs that give conclusions about the attached objects, in this case the nodes. Because of this fact, the first step was to look for keys that would indicate if the nodes are of touristic interest. The self-explanatory key *tourism* 8 was quickly found as well as the key *amenity 9* in combination with according values such as *restaurant* or *bar*. With these or similar tags, it is possible to filter the set of nodes, so only nodes are output that match the above mentioned criteria. The second step is to filter the nodes according to their location. Given the fact that the current user location is stated in coordinates (longitude, latitude), we only want to show nodes that are in a certain walking distance from that user.

These requirements lead to the following query that was additionally adjusted in a way that each relevant tag is output in an own column. This way, essential information can be seen at once:

SELECT id, tags,tags-> 'name' as poiname, tags-> 'tourism' as tourism, tags-> 'description' as description,tags-> 'addr:street' as street, tags->'addr:housenumber' as housenumber, tags-> 'addr:city' as city,tags-> 'opening\_hours' as openingHours,ST\_Distance(geography(geom), ST\_SetSRID(geography(ST\_Point(9.991636, 53.550090)), 4326)) as distanceFROM nodesWHERE ST\_DWithin(geography(geom), ST\_SetSRID(geography(ST\_Point(9.991636, 53.550090)), 4326), 1000)and tags ? 'tourism' and not (tags @> hstore('tourism','information') or tags @> hstore('tourism','hotel'));



# Bibliography:

1: https://chatbotsmagazine.com/the-complete-beginner-s-guide-to-chatbots-8280b7b906ca#.lcnq9f3ro

2: ELIZA – a computer program for the study of natural language communication between man and machine

3: <http://venturebeat.com/2016/08/15/a-short-history-of-chatbots-and-artificial-intelligence/>

4: <https://stanfy.com/blog/advanced-natural-language-processing-tools-for-bot-makers/>

5: <https://www.openstreetmap.org/about>

6: https://en.wikipedia.org/wiki/OpenStreetMap#Data\_storage

7: <http://wiki.openstreetmap.org/wiki/Points_of_interest>

8: http://wiki.openstreetmap.org/wiki/Key:tourism

9: http://wiki.openstreetmap.org/wiki/Key:amenity