NLMaps Annotation Guidelines

Principles

In essence, we need a dataset that fulfills three criteria:

1. The natural language (NL) queries should be linguistically diverse, i.e. a mix of short search-engine-style queries and full questions like you would ask another person.

2. The queries should cover a large part of the commonly used OpenStreetmap (OSM) tags that are relevant for our system.

3. The queries should cover the most useful OSM tags in particular depth. E.g., asking for restaurants and shops is arguably one of the most useful areas.

Linguistic Diversity

When entering queries, please diversify your language use. Valid variations of the same question include:

- *closest swimming pool near Eiffel Tower in Paris*
- *In Paris, give me the swimming pool that is closest to the Eiffel Tower!*
- *Closest place where I can take a swim near Eiffel Tower in Paris*

However, keep it natural and don't make your queries artificially complex. If a particular query style comes more natural to you, it's alright to use it more often. Just make sure that not all of you queries look alike.

Similarly, avoid using the same place name more than a couple of times. Also use place names in other languages than German and English, e.g. *České Budějovice* or *València*.

It's **very important** that you don't only base your wording on the name of the OSM tags. E.g., for `highway=speed_camera` you can ask for a *speed camera*, but you can also ask for a *speed trap* or a *radar trap*.

Tag Diversity and Depth

Take a *quick* look at the most important OSM features to get a feel for what things you can ask for in OSM. You can use this as an inspiration if you run out of ideas about what to ask. Choose tags that you find relevant.

Beware of the `building=*` tag! It is used to tag what the building was built as, not to tag its current use. E.g., a place tagged `building=church` may not be a church anymore; churches are tagged as `amenity=place_of_worship + religion=christian`. If in doubt, don't use the `building=*` tag, at all.

In general, enter more than one query for a chosen tag or tag combination, especially if the system fails answering the query correctly.

Most relevant keys

Please enter at least the given amount of queries for each of the following. The linked wiki pages give you a feel for what the most common tags in each category are.

- shop=*: 30
- leisure=*: 30
- sport=*: 20
- craft=*: 20
- man_made=*: 20
- amenity=restaurant, amenity=fast_food: 30
- cuisine=*: 30.
- diet:*=*: 20.
- wheelchair=yes: 50. Just sprinkle phrases like *wheelchair-accessible* [*place*] or [*places*] *that are wheelchair-accessible* into your queries now and then.