Annotation guideline
for Entity Linking with Wikidata

To search for terms in the knowledge base, we allowed the following modifications of entities:

1. All terms are lemmatized before the search, ex. "*Linear equations*" -> "*Linear equation*".
2. If an entity was extracted from the text that matches the pattern "general concept + name" (ex. "*programming language Python*", "*operational system Windows*"), while the knowledge base contains only an entity with a name (for example, "*Python*" (Q28865)), then these two entities should be linked.
3. If the entity is written with a typo in the text, then in the knowledge base we are looking for an entity without a typo.
4. It is allowed to search for a synonym of an entity in the knowledge base (checked by a query to a search engine or Wikipedia), for example: "*statistical dependence*" -> "*correlation*", "*genetic sequence*" -> "*nucleotide sequence*", it is also possible to search for a translation of the entity, for example, in the English language.
5. Transformations as "*Architecture of the system*" -> "*System's architecture*" are admitted.
6. Abbreviations expansion, for example "*wps*" -> "*Wi-Fi Protected Setup*".
7. If two or more entities are represented as a set of homogeneous members with one common element, then each homogeneous member with a common element is considered as an entity, for example: "*satellite and mobile communications*" -> "*satellite communications*", "*mobile communications*".
8. Different kinds of coreferences are also associated with one entity, for example: if the "*k-means method*" is mentioned at the beginning of the text, and then "*the proposed [method]*" in the text, then these two entities should be linked by one identifier.
9. We also consider the terms "*approach*" and "*method*" to be synonyms.