ED with lower-case data

This note explains what I think needs to be adapted to train an ED model for REL with lower-case data. The steps are based on the tutorial for a new Wikipedia corpus and on the paper/my understanding of the paper, whose summary is in the `theory.md` document.

1. Overview

1. Creating a folder structure
    - No changes
    - But : There are duplicated rows in the database in the `lower` column. Currently the package just loads the first entry in this case. Is this the intended use? What to do about it? Issue a warning? Consolidate the scores from each duplicate into one (but two rows with the same entry in lowercase may differ in the column `word` , see details)? Issue warning?
2. Embeddings
    - run `Wikipedia2Vec` with the lower case option.
    - Then the embeddings are stored in the db (see "Storing Embeddings in DB").
    - Questions
        - Do the embeddings and p_e_m scores change as a result? *--I am not sure at the moment.*
        - What needs to be changed to make sure that REL finds the correct reference in the database? -- This depends on how casing impacts database keys and the queries that REL performs to the database.
        - Implementation: how should the database (and software) be set up for cased and uncased data? Is duplicating the database an option at all (ie, embeddings and p_e_m scores for cased and for uncased inputs)?
        - Or is the idea to leave the Wikipedia data as-is and just use the fallback query with lower when using REL with lower-case data?
3. Generating training, validation and test files
    - No direct change in the processing
    - See the file `generate_training_test`
        - it stores the data to `data/wiki_version/generated/test_train_data/` . The mentions are cased.
            - Implication: no need to re-generate the data here, but for the training the keys of the dictionary need to be put into lower case

4. Training your own Entity Disambiguation model
    - largely follow the existing instructions
    - uncase the keys in the method `TrainingEvaluationDatasets.load()` (from point 3 above).

Other questions: computations

- Can I run `Wikipedia2Vec` on my laptop? If not, where?
- Same for training the model