Technical Documentation (TD): high-level descriptions (in red) of lower-level code (in black, e.g., function signatures, code templates).

TD as a parallel corpus (Allamanis et al. 2015, Iyer et al. 2016) for studying translation, text → code (a synthetic semantic parsing (SP) task).

New Resources: Standard Library Documentation

- Pairs of text and function signature representations (a kind of KR), extracted automatically.

Task: Mapping text to function signatures

- Given a dataset of pairs, \( D = \{(x_i, y_i)\}_i \), learn model \( sp : x \rightarrow y \).
- Focus on developing baseline models, building on Deng and Chrupała 2014. Main model has two components:
  - Word SMT Model: Generates candidate signatures from text, uses a simple decoding strategy, specific to signature language.
  - Discriminative Model: Ranks translation candidates using additional word, (hiero)phrase and document-level features.
- Questions: What type of translation model to use? Do the additional phrase and document-level features help the translation?

Example Re ranker features

- Given a dataset of pairs, \( D = \{(x_i, y_i)\}_i \), learn model \( sp : x \rightarrow y \).
- Focus on developing baseline models, building on Deng and Chrupała 2014. Main model has two components:
  - Word SMT Model: Generates candidate signatures from text, uses a simple decoding strategy, specific to signature language.
  - Discriminative Model: Ranks translation candidates using additional word, (hiero)phrase and document-level features.
- Questions: What type of translation model to use? Do the additional phrase and document-level features help the translation?

References and Info

- Code retrieval prototype (see for data and code): zubr.ims.uni-stuttgart.de
- Supported by the German Research Foundation (DFG), project D2 of SFB 732.