# Appendix

- **SCIERC** (Luan et al., 2018): SCIERC includes annotation for three different information extraction tasks, which are entity recognition, relation extraction and coreference resolution. The corpus are taken from the 500 abstracts of AI conference papers. The SCIERC corpus is available from `http://nlp.cs.washington.edu/sciIE/`.

- **BioNLP09** (Kim et al., 2009): The BioNLP09 Shared Task concerns the recognition of bio-molecular events (bio-events) that appear in biomedical literature. In our work, we focus on its upstream NER task and use the "Protein" annotations of the corpus. It has been published at `http://www.nactem.ac.uk/tsujii/GENIA/SharedTask/` with a pre-defined split into the train, development and test dataset.

- **BioNLP11** (Pyysalo et al., 2012): Similar to BioNLP09, BioNLP11 corpus is focused on Infectious Disease (*BioNLP11ID*) and Epigenetics and Post-translational Modifications (*BioNLP11EPI*). Both of them are available from `http://2011.bionlp-st.org`.

- **BioNLP13** (Nédellec et al., 2013): It is prepared for knowledge-based construction. In this work, we select three related parts: GENIA Event Extraction (*BiONLP13GE*), Cancer Genetics (*BiONLP13CG*) and Pathway Curation (*BiONLP13PC*). All these three dataset are available from `http://2013.bionlp-st.org`.