Projects involved
- Archive of General Reference Corpora of Contemporary Written German
- Methods of Corpus Analysis and Corpus Mining

Researchers involved
- Dipl.-Ing. Cyril Belica
- Dr. Marc Kupietz
- Dipl.-Inf. Rainer Perkuhn
- Dr. Andreas Witt

Contact
Dipl.-Ing. Cyril Belica
Corpus Linguistics Programme Area
IDS – Institute for the German Language
Postfach 10 16 21
D-68016 Mannheim
Germany
Phone: +49-621-1581-0
Fax: +49-621-1581-200
Email: corpuslinguistics@ids-mannheim.de

The Institute for the German Language (IDS) is the central institution for the study and documentation of the contemporary usage and recent history of the German language. Together with 85 other non-university research institutes and service facilities, it belongs to the Leibniz Association, one of the four major research organisations in Germany.
Within a strictly corpus-driven paradigm, an in-depth profiling of many linguistic phenomena requires fast access to massive amounts of data derived from very large corpora. The collocation data base CCDB is a corpus-linguistic research and development workbench, which serves as a transparent lab for the study, development, and evaluation of a methodology of data-driven exploration and modeling of language use.

The current CCDB version

- serves internally as an empirical foundation and driving force for formulating novel usage-based linguistic hypotheses
- integrates early beta implementations of data-driven methods currently developed: e.g., the modeling of semantic similarities, the visualization of usage aspects, and contrasting near-synonyms
- is based on a 2.2 billion words subset of the German Reference Corpus (DeREKo)
- comprises rich collocation profiles, including syntagmatic patterns and other related information, for approx. 220,000 node words
- contains for each node word up to five different collocation profiles, with up to 100,000 concordance lines each
- provides a pre-computed network of similarity relations between collocation profiles to be exploited by the data-driven methods
- can be accessed at http://corpora.ids-mannheim.de/ccdb/