Background
The engineering of ontologies, especially with a view to a text-mining use, is still a new research field. There does not yet exist a well-defined theory and technology for ontology construction. Many of the ontology design steps remain manual and are based on personal experience and intuition. However, there exist a few efforts on automatic construction of ontologies in the form of extracted lists of terms and relations between them.

Results
We share experience acquired during the manual development of a lipoprotein metabolism ontology (LMO) to be used for text-mining. We compare the manually created ontology terms with the automatically derived terminology from four different automatic term recognition methods. The top 50 predicted terms contain up to 89% relevant terms. For the top 1000 terms the best method still generates 51% relevant terms. In a corpus of 3066 documents 53% of LMO terms are contained and 38% can be generated with one of the methods.
Secondly we present a use case for ontology-based search for toxicological methods.

Conclusions
Given high precision, automatic methods can help decrease development time and provide significant support for the identification of domain-specific vocabulary. The coverage of the domain vocabulary depends strongly on the underlying documents. Ontology development for text mining should be performed in a semi-automatic way; taking automatic term recognition results as input.

Availability
The automatic term recognition method is available as web service, described at http://gopubmed4.biotec.tu-dresden.de/IdavollWebService/services/CandidateTermGeneratorService?wsdl

Dagstuhl Seminar Proceedings 08131
Ontologies and Text Mining for Life Sciences : Current Status and Future Perspectives
http://drops.dagstuhl.de/opus/volltexte/2008/1506