

Summary

- Text representation determines what kind of mining algorithms can be applied
- **Multiple ways** of representing text are possible
 - string, words, syntactic structures, entity-relation graphs, predicates...
 - can/should be **combined** in real applications
- This course focuses on **word-based representation**
 - **General and robust**: applicable to any natural language
 - **No/little manual effort**
 - **“Surprisingly” powerful** for many applications (not all!)
 - **Can be combined** with more sophisticated representations