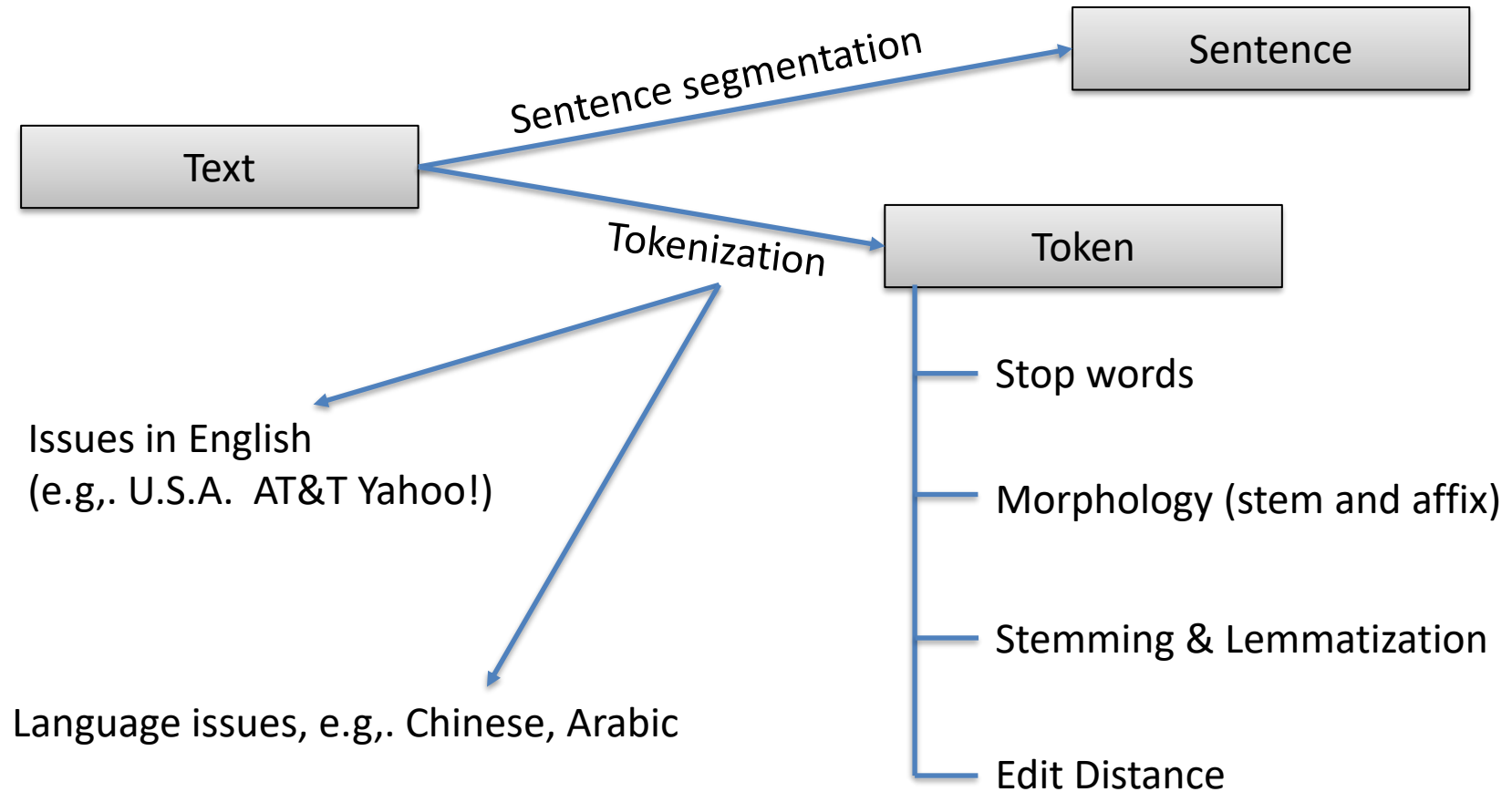


AI6122 Text Data Management & Analysis

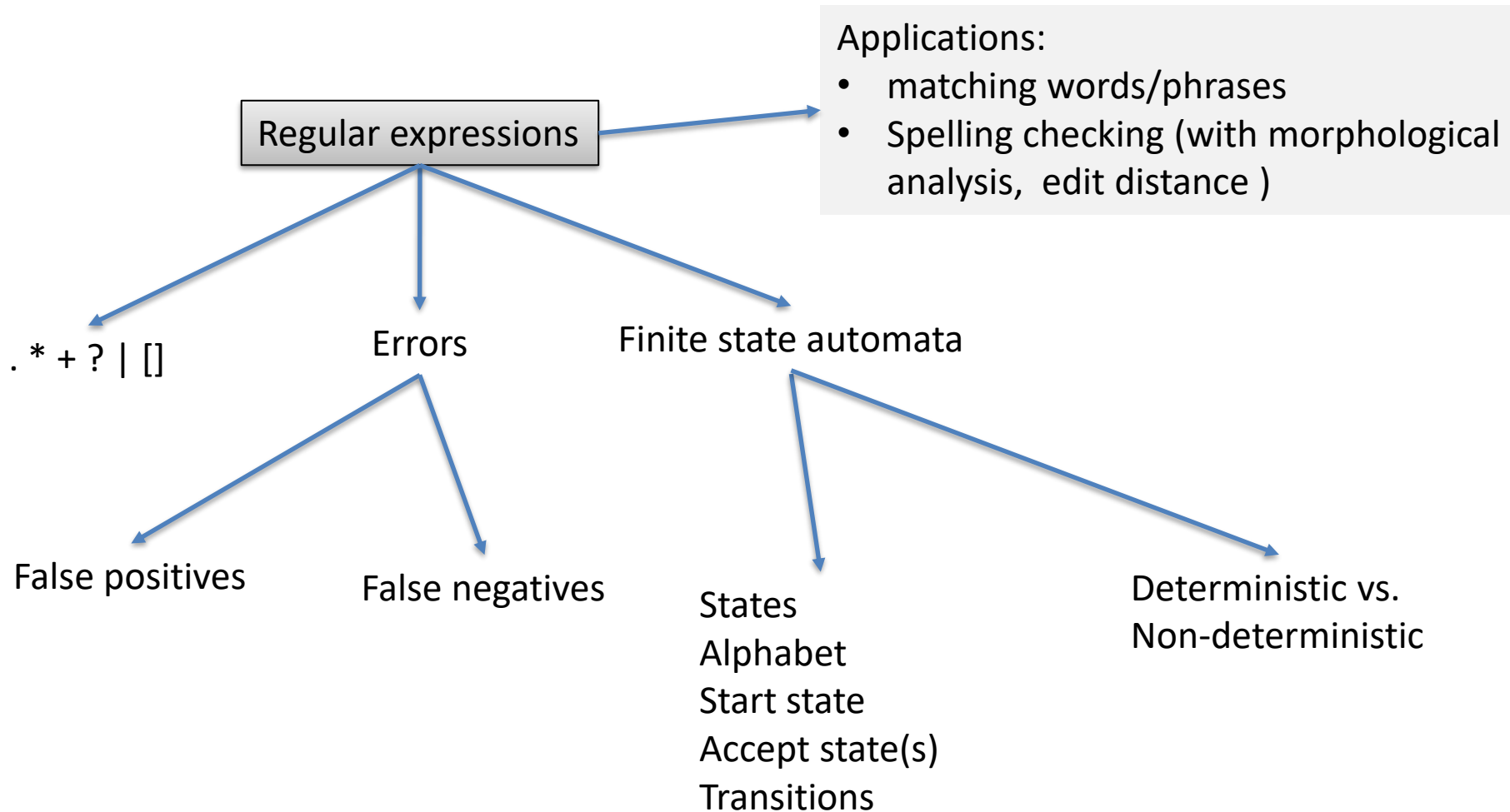
Lecture Review 1



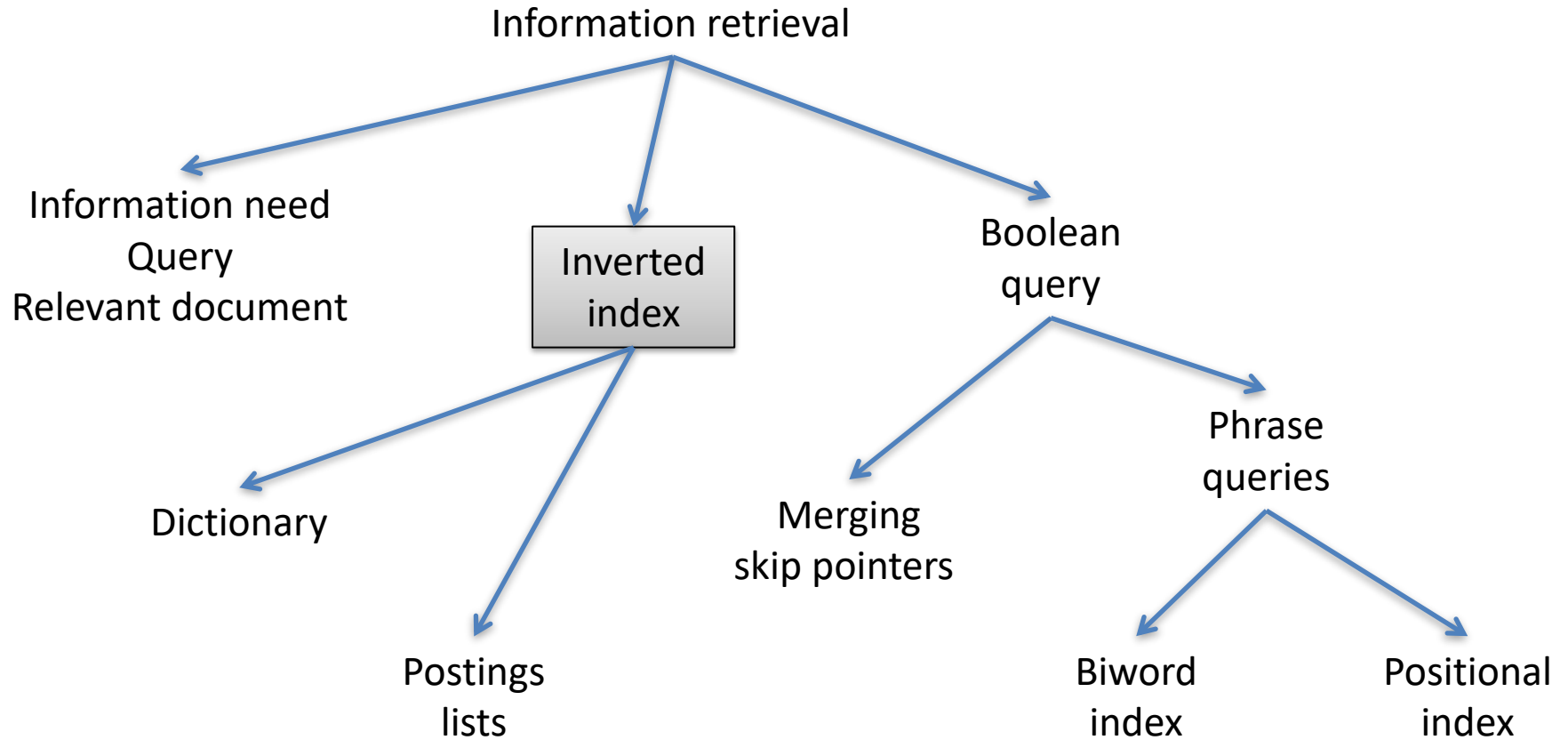
Tokenization, Morphology



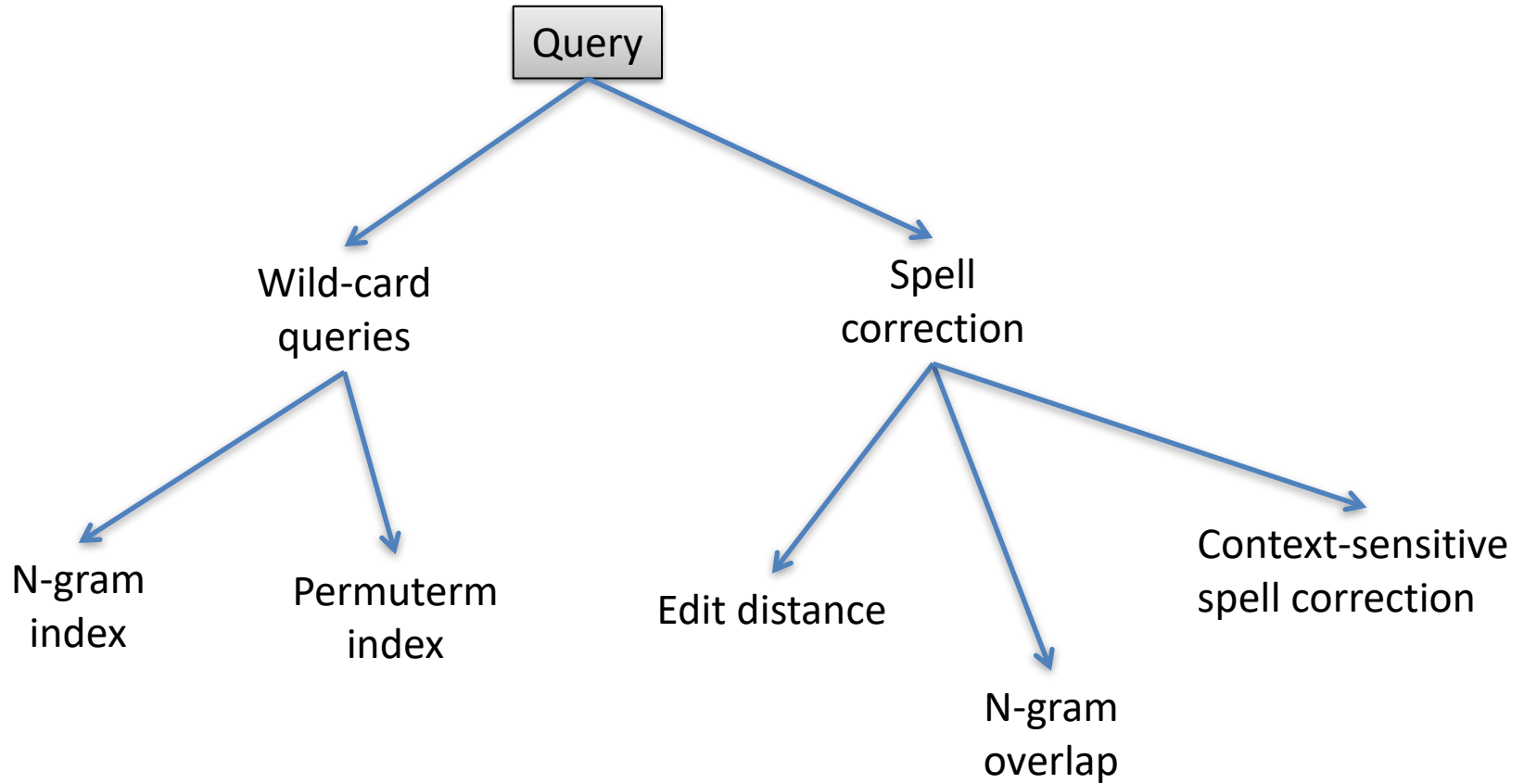
Regular expressions



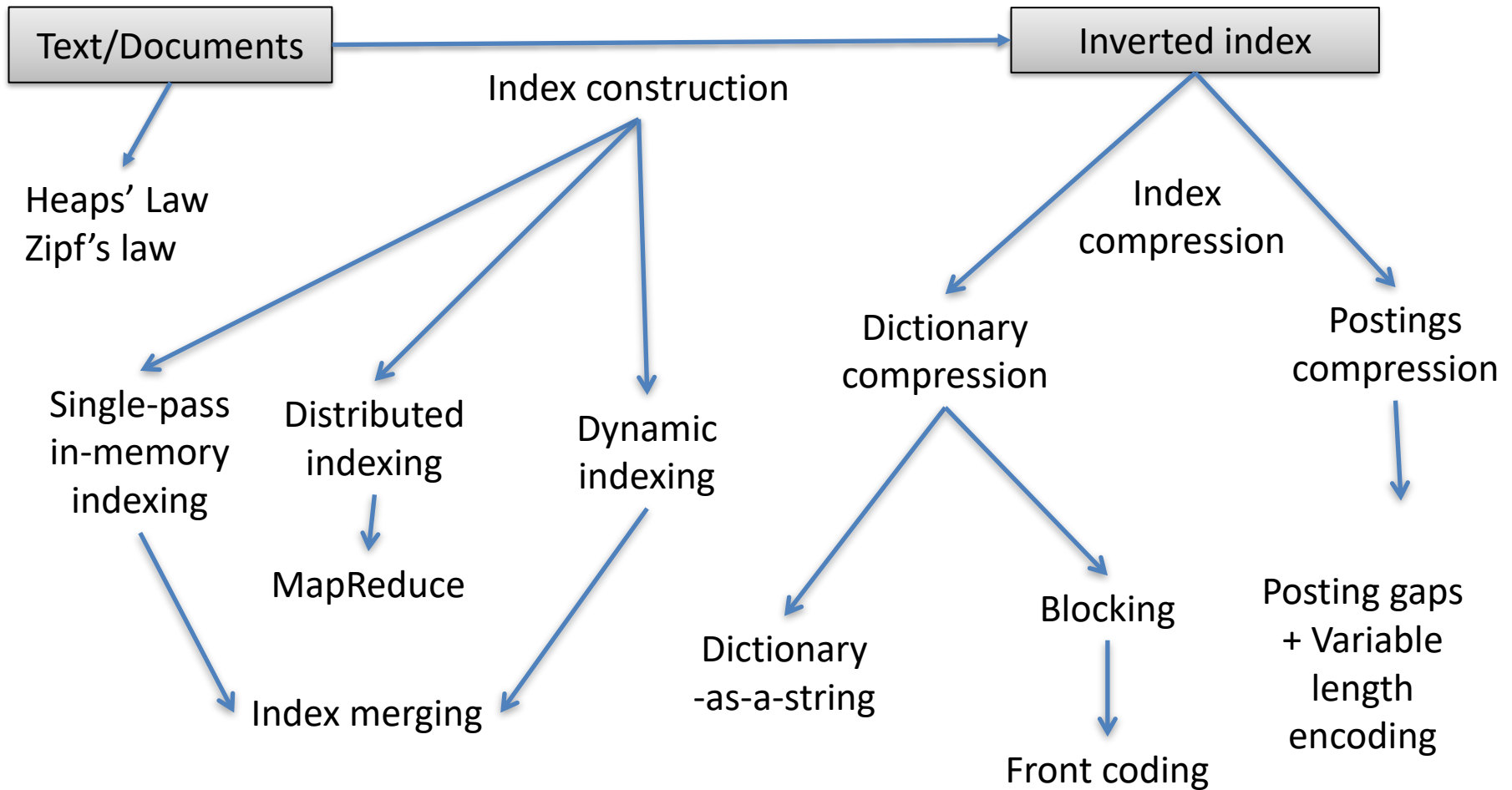
Boolean retrieval



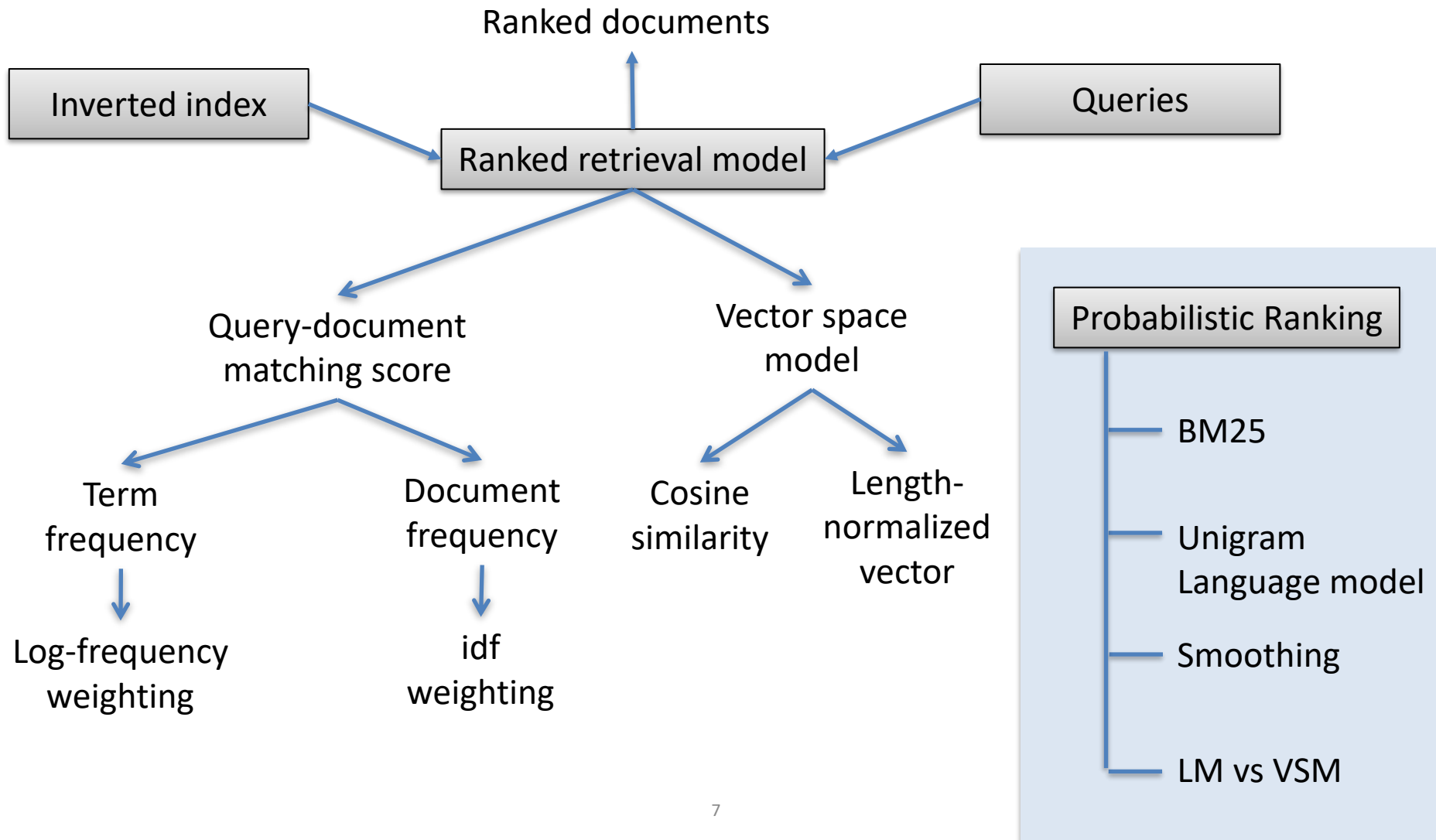
Tolerant retrieval



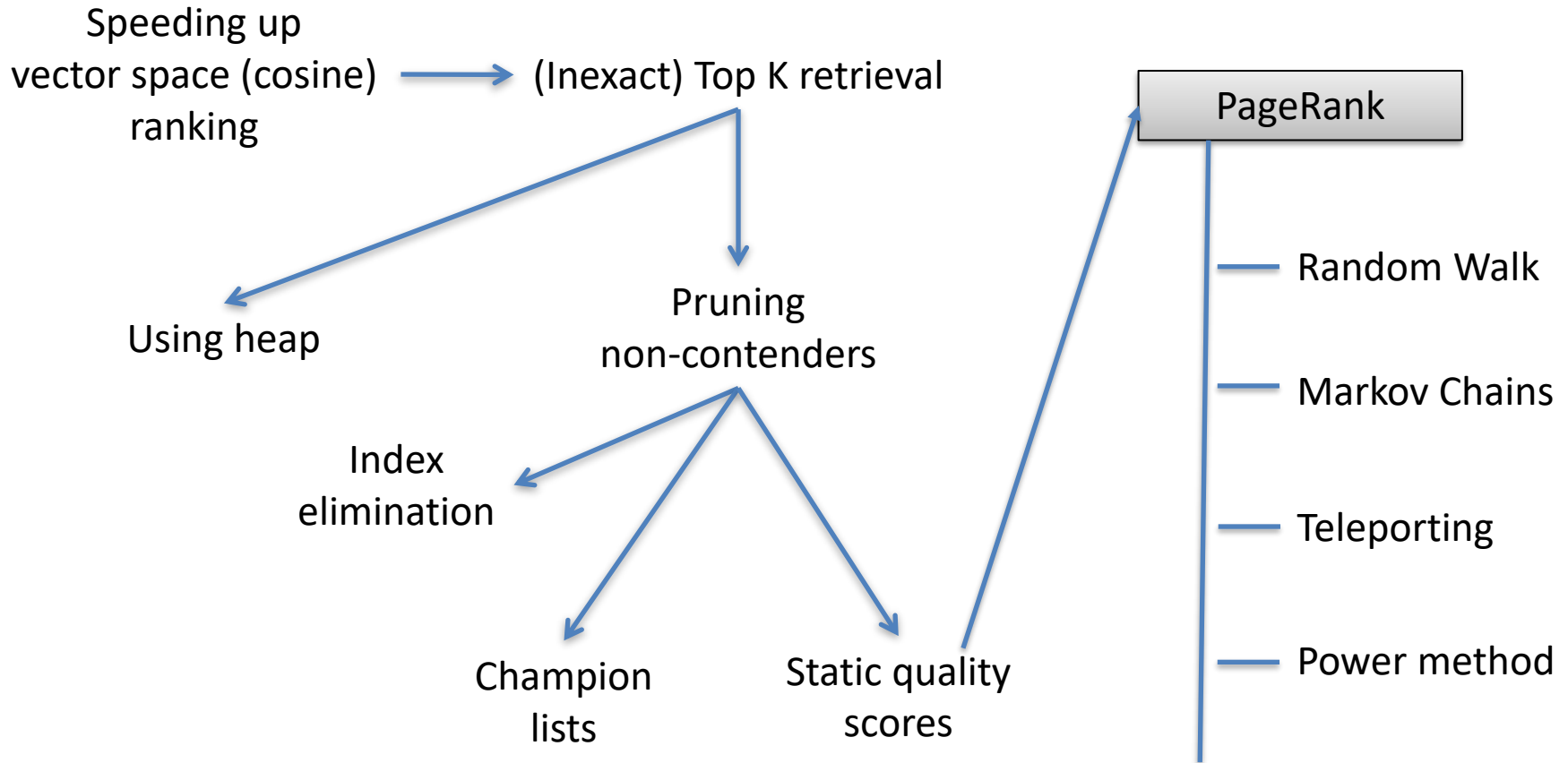
Index construction and Compression



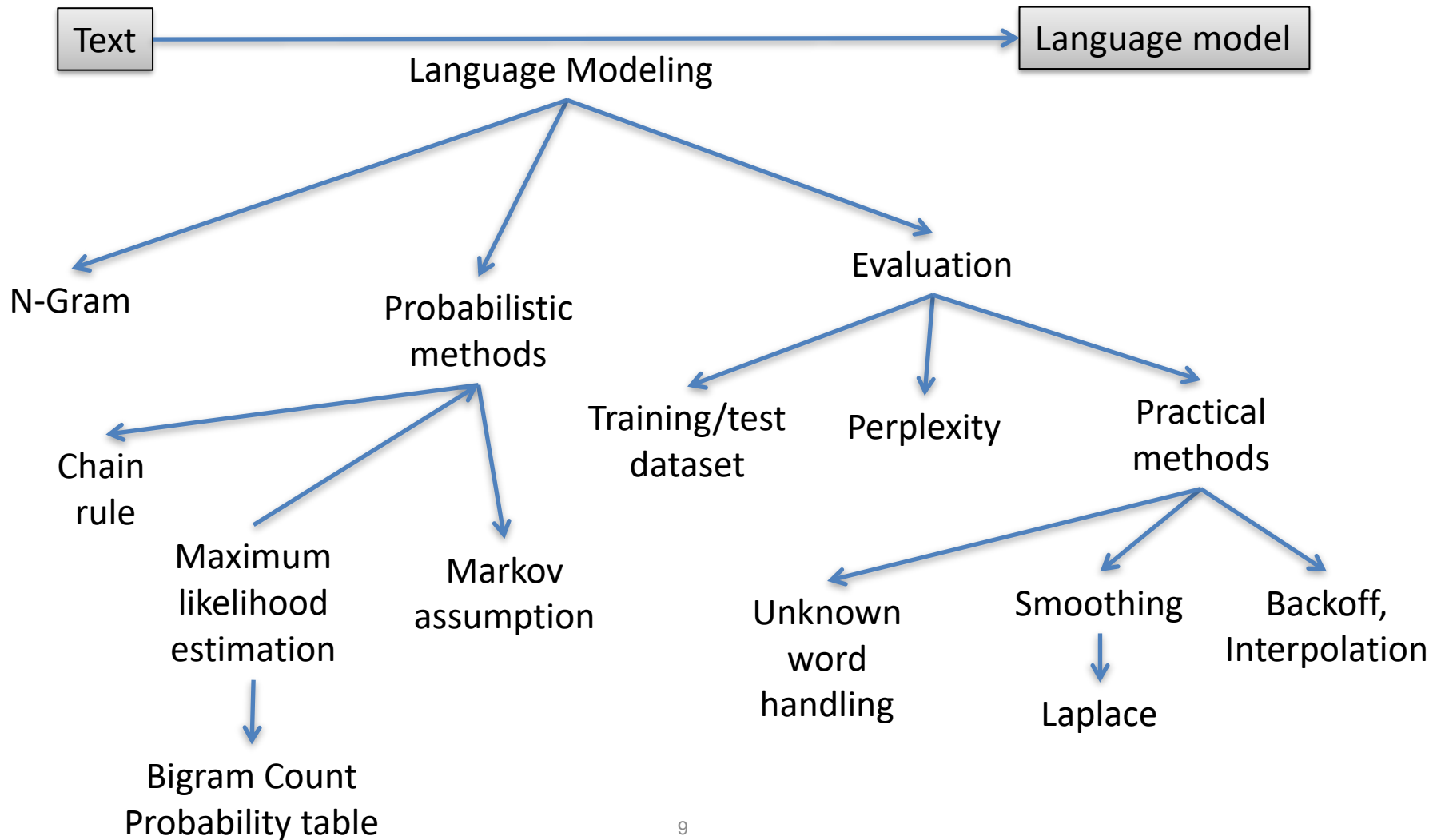
TFIDF and Vector Space Model



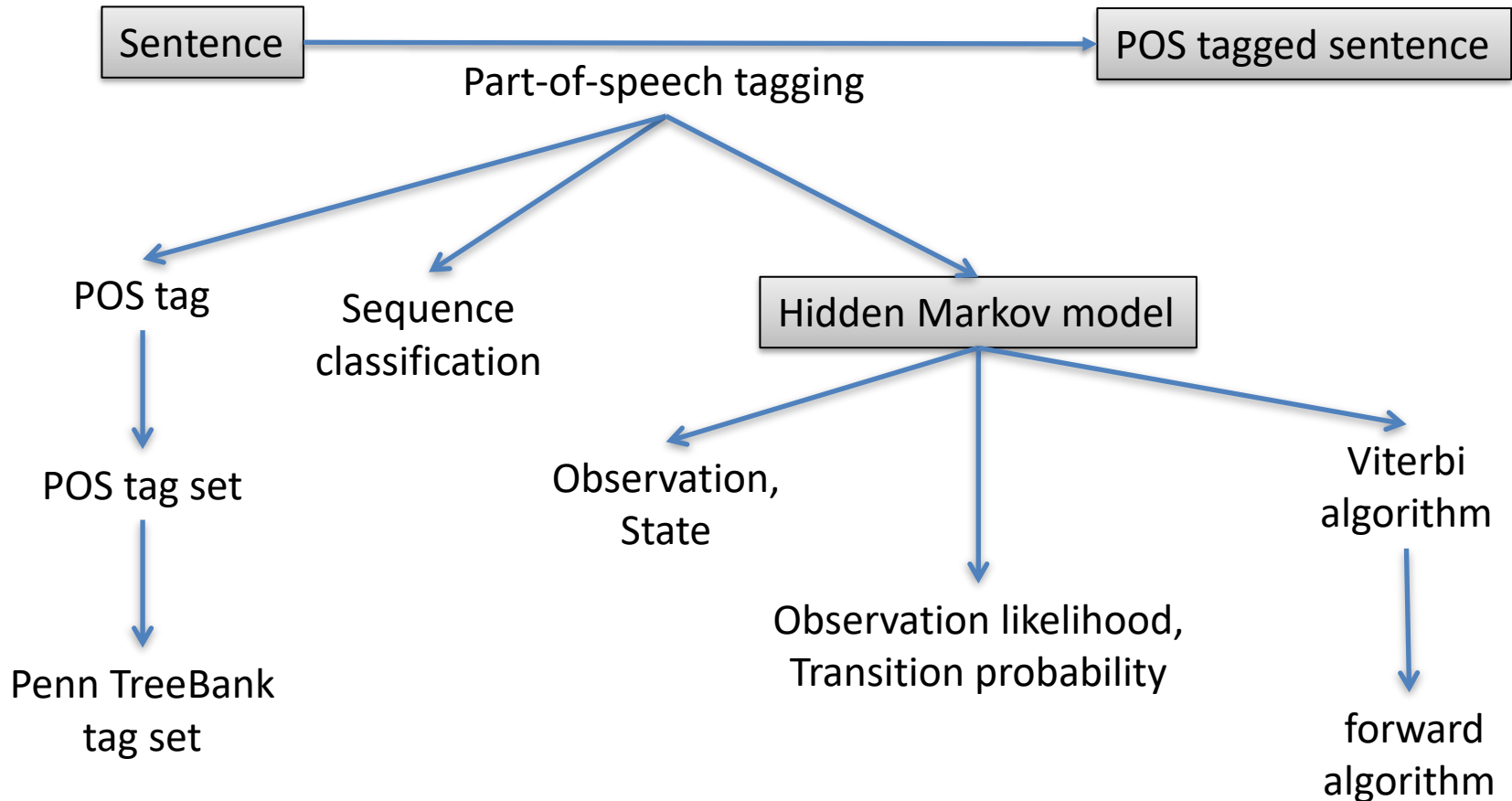
Efficient scoring



Language modeling and N-Gram Model



Part-of-speech tagging and HMM



Parsing

