Probabilistic models for sentence level similarity.

Sentence similarity measures become increasingly important in text-related research and application areas such as text mining, information extraction, automatic question-answering, text summarization, text classification and machine translation.
The main goal of this seminar work is the swallow research on probabilistic models for sentence level similarity measures and implementation of the one such model, which is in literature considered as the baseline.

Task description:
Probabilistic models are based on idea of estimating the probability that one sentence is a translation of another. This translation probability will then serve as the basis of the similarity score for pairs of sentences. Your goal is to prepare the short review of existing methods and based on the research to implement the simplest one. Seminar is suitable for team work of two students.

Addition explanations, literature and questions will be clarified in consultation before the start of the seminar work.

References:
P. F. Brown, V. J. D. Pietra, S. A. D. Pietra, and R. L. Mercer.The mathematics of statistical machine translation: Parameter estimation. Computational Linguistics, 19(2):263–311, 1993.

Mentor:
Ercan Canhas
ercan_canhas@hotmail.com
Office hours: LKM, send e-mail.