

LSM: Language Sense Model for Information Retrieval

Shenghua Bao, Lei Zhang, Erdong Chen, Min Long, Rui Li, and Yong Yu

APEX Data and Knowledge Management Lab,
Department of Computer Science & Engineering,
Shanghai Jiao Tong University, Shanghai, P.R.China, 200230
{shhbao, zhanglei, edchen, mlong, rli, yyu}@apex.sjtu.edu.cn

Abstract. A lot of work has been done on drawing word senses into retrieval to deal with the word sense ambiguity problem, but most of them achieved negative results. In this paper, we first implement a WSD system for nouns and verbs, then the language sense model (LSM) for information retrieval is proposed. The LSM combines the terms and senses of a document seamlessly through an EM algorithm. Retrieval on TREC collections shows that the LSM outperforms both the vector space model and the traditional language model significantly for both medium and long queries (7.53%-16.90%). Based on the experiments, we can also empirically draw the conclusion that the fine-grained senses will improve the retrieval performance when they are properly used.