Abstract

As there is availability of large amount of data on the web, but due to constraints web is only used for browsing and searching. Traditional IE uses NLP techniques such as lexicons, grammars, whereas web applies machine learning and pattern mining techniques to exploit the syntactical patterns or layout structures of the template-based documents. Information Retrieval is the art of presentation, storage, organization of and access to information items. IR now—days mainly deals with retrieving information based on user queries. The paper deals with basic understanding of IR and IR models and shows Support Vector Machines is a good technique for classification of huge data sets.

References

Comparison of IR Models for Text Classification

corcepts and technology behind search. Addison Wesley.

- Hang Li, Oct 2011, A Short Introduction to Learning to Rank, Special Section on Information-Based Induction Sciences and Machine Learning, IEICE TRANS. INF. & SYST., VOL. E94-D, NO. 10
- Tristan Fletcher, March 1, 2009, "Support Vector Machines Explained," ucl
- Dustin Boswell, August 6, 2002, Introduction to Support Vector Machines,
Comparison of IR Models for Text Classification


Index Terms

Computer Science
Information Sciences

Keywords

Boolean retrieval  Vector Space model  BM25  Language models  Inference networks  Learning to Rank

SVM