Selected references for Text Mining

Named Entity Extraction

- a recent survey

- early paper in a long series on using CRFs for named entity recognition

Indrajit Bhattacharya and Lise Getoor, Collective Entity Resolution in Relational Data ACM Transactions on Knowledge Discovery from Data (TKDD) 1(1), 2007

Machine Learning methods

- CRFs are the most widely used method in NER

- application to schema matching and data integration

Information Extraction

Andrew McCallum, Information Extraction, ACM Queue 2005.
- good introduction with references

- entity resolution in databases


- old but classic

Open IE and Relation extraction


Oren Etzioni, Michele Banko, Michael J. Cafarella, Machine Reading, extended version; AAAI, 2006

Michele Banko and Oren Etzioni
The Tradeoffs Between Open and Traditional Relation Extraction Proc. 46th Annual Meeting of the Association for Computational Linguistics (ACL 2008)
SRES,

Sentiment analysis and Market Structure analysis

- seminal paper


Lise Getoor Link mining: a new data mining challenge
ACM SIGKDD Explorations Newsletter 84 - 89, 2003

Visualization


Scalable algorithms

Jeffrey Dean and Sanjay Ghemawat,
MapReduce: Simplified Data Processing on Large Clusters,
OSDI’04: Sixth Symposium on Operating System Design and Implementation, 137-150, 2004,

Koby Crammer, Ofer Dekel, Joseph Keshet, Shai Shalev-Shwartz and Yoram Singer

Abhinandan Das, Mayur Datar, Ashutosh Garg, Shyam Rajaram
Google News Personalization: Scalable Online Collaborative Filtering
WWW, 2007

S. Guha, N. Mishra, R. Motwani, L. O'Callaghan, Clustering Data Streams,