

题 目：_____基于关系抽取的答案自动提取系统的设计及实现_____

学院：_____专业：_____学生姓名：_____学号：_____

指导教师（签名）：_____提交日期：_____年____月____日

毕业设计（论文）基本内容和要求：

问答系统答案提取的准确性对问答系统本身的准确性起着关键作用，要求学生掌握机器学习基本算法，熟悉自然语言处理的方法和现有工具，本论文要求学生完成答案为简单实体类型的领域问题答案提取，并对客观答案（或者答案很长，需要整理信息）提取简要答案提供给用户。

1. 研究内容

- 问答系统答案自动提取方法
- 关系抽取技术
- 自动文本摘要技术

2. 技术指标和基本要求

- 学习并熟悉 Linux
- 学习并熟悉 SVN 版本控制
- 学习 Indri 检索
- 学习 Hadoop 和 Spark

毕业设计（论文）重点研究的问题：

- 问答系统答案提取的现有解决方案
- 关系抽取技术的现状
- 自动文本摘要的技术现状
- 网页正文提取
- 中文自然语言处理，包括切词、词性标注、句子依存分析、命名实体识别
- 文本特征提取
- 不同机器学习算法的调研与实现，计算其在本系统下的召回率和准确率，选择最优模型
- 自动文本摘要的实现，并对摘要文本可读性进行优化

毕业设计（论文）应完成的工作：

- 查阅文献数量需在 12 篇以上
- 对现有问答系统答案提取有详细的调研
- 答案提取需要进行实验，实验保证基本数据量足够大
- 对答案提取的准确率有数据计算和统计

参考资料推荐：

1. Mintz M, Bills S, Snow R, Jurafsky D. Distant supervision for relation extraction without labeled data. In: Proc. of the Joint Conf. of the 47th Annual Meeting of the ACL and the 4th Int'l Joint Conf. on Natural Language Processing of the AFNLP. Morristown: Association for Computational Linguistics, 2009. 1003–1011.
2. Etzioni O, Cafarella M, Downey D, Kok S, Popescu A, Shaked T, Soderland S, Weld D and Yates A. Unsupervised Named-entity Extraction from the Web: An Experimental Study [J]. Artificial Intelligence, 2005, 165(1): pp91-134.
3. Gunes E;Radev D R LexRank:Graph-based Centrality as Saliencin Text Summarization 2004
4. H. Zhang (2004). The optimality of Naive Bayes. Proc. FLAIRS.
5. Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision[J]. CS224N Project Report, Stanford, 2009, 1: 12.
6. Min B, Grishman R, Wan L, et al. Distant Supervision for Relation Extraction with an Incomplete Knowledge Base[C]//HLT-NAACL. 2013: 777-782.
7. Xu W, Hoffmann R, Zhao L, et al. Filling Knowledge Base Gaps for Distant Supervision of Relation Extraction[C]//ACL (2). 2013: 665-670.
8. Suttles J, Ide N. Distant supervision for emotion classification with discrete binary values[M]//Computational Linguistics and Intelligent Text Processing. Springer Berlin Heidelberg, 2013: 121-136.
9. Collins M. Ranking algorithms for named-entity extraction: Boosting and the voted perceptron[C]//Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2002: 489-496.
10. Nadeau D, Sekine S. A survey of named entity recognition and classification[J]. Lingvisticae Investigationes, 2007, 30(1): 3-26.

其他要说明的问题：