|  |
| --- |
| 基于知识图谱的限定领域问答系统 |



1. 项目背景

机构部门的干部任免需要考察干部的综合信息，而这些信息的来源、存储和检索需要一个集成部署的平台。以往的信息管理系统采用结构化的数据存储和基于规则的精确查询。为了优化存储、精简系统以及使搜索结果更丰富、可推理，本文基于知识图谱构建了一个基于知识图谱的组织干部领域问答系统。

1. 要解决的问题
2. 知识库的构造
3. 自然语言处理
4. 用到的软件和技术

软件：kettle、kafka、neo4j、elasticsearch

StandFord CoreNLP、Tinkerpop

技术：

1. 从关系数据库到图数据库

数据源为关系型数据库，将结构化、半结构化、文本数据抽取

处理后导入消息队列，再自动化的创建图到图数据库中。

1. 模型定义

实体定义、属性决策、关系建立

1. 语义网络搜索

对搜索关键词进行语义分析，提取相关概念，查询知识库，对返 回结果进行推理、统计、排序。

1. 系统整体架构

自然语言文本

|

分词、命名实体识别、词性标注 —— 分词模型

|

实体对齐（同义词替换。。） —— 词典

|

主题词（实体、关系）提取 —— 词典

|

实体链接 —— 知识库

关系抽取 —— 模型

|

1. 自然语言搜索

本文构建的是基于知识库的领域问答系统，即KB-QA。首先是建造一个领域的知识库。在现实世界中，知识多是非结构化展现的，如一句话、一段自然语言文本、一个网页。这些知识易于人的阅读和理解，