专家知识抽取与规则化描述

1知识抽取方式

1.1人工抽取

人工抽取是指特定的人员对大量数据进行分析并手动写下抽取出的信息。这种方式受到时间和经济成本的约束，很难实现大规模知识库的构建，所以在当前状况下并不适用。

1.2半自动抽取

半自动知识抽取主要采用包装器( wrapper) 归纳或包装器学习方法，其具体流程是首先由用户手工标注出训练数据，然后利用机器学习方法从训练数据中学习得到抽取规则，用来从具有相同结构的其他内容中抽取出相应的信息。

采用半自动抽取时可以把文档当作字符串进行训练，如果文档是比较结构化的还可以提取出结构化文档中的树结构去训练。

使用半自动抽取在学习到抽取规则后可以有较好的效果，但具有两方面不足：a) 手工标注训练数据代价高昂，不仅要标注一定数量的训练数据，而且要选择适当的训练数据; b) 包装器的维护代价高昂。

由于这两个缺点，这种方式只适合在数据量较小或数据格式统一时使用，在数据量较大且数据格式不统一时较为麻烦。

1.3自动抽取

2存储规范

2.1知识抽取方式（包括人工、半自动、自动三种方式的流程、特点、适用条件、推荐方法）

2.2知识规则化（5种方法的录入模板形式，特点、适用性。5种方法的流程、特点、推荐方法及适用范围）

2.2.1产生式规则

2.2.2知识图谱

2.2.3框架法

2.2.4谓词逻辑

2.2.5人工智能

2.3存储规范（方法、根据规则化的输出结果如图、表，每种方法推荐一种工具）