# 《基于用户投诉文本的知识图谱构建研究与实现》文献综述

**1 引言**

近年来，大规模知识图谱的研究和应用在学术界和工业界引起了足够的注意力。知识图谱旨在描述现实世界中存在的实体以及实体之间的关系。知识图谱于2012年5月17日由Google正式提出，其初衷是为了提高搜索引擎的能力，改善用户的搜索质量以及搜索体验。在国内，百度和搜狗等搜索公司也相继推出了自己的知识图谱产品。随着人工智能技术的发展和应用，知识图谱逐渐成为关键技术之一，现已被广泛应用于智能搜索、智能问答、个性化推荐、内容分发等领域。

**2知识图谱基本概念**

网络是一个巨大的知识库，其中的知识是并不是孤立存在的，而是相互关联的。从互联网上海量数据中提取相互关联的知识可以构成一个巨大的图。这个图即为知识图谱。知识图谱本质上是一种语义网络。其结点代表实体（entity）或者概念（concept），边代表实体或概念之间的各种语义关系。

**3知识图谱构建技术研究**

知识图谱的构建可分为三个部分。第一个部分是实体或概念与关系的提取。第二个部分是对提取出来的数据进行融合，除去伪信息和互斥信息，并建立信息之间的关联。第三个部分是以知识图谱为工具进行相关推理。

**3.1实体或概念及关系的提取**

一般而言，实体或概念及关系的提取数据源可分为结构化、半结构化和非结构化三种数据类型。常见的结构化数据有：MusicBrainz，YAGO等。