

随着近几年文本信息的爆发式增长,从大量文本信息中提取重要内容已成为我们的一个迫切需求。自动文本摘要技术提供了一个高效的解决方案。目前该技术具有广泛的应用场景,可以在新闻标题生成、科技文献摘要生成、搜索结果排序、商品评论摘要、信息个性化推荐等多方面发挥重要作用。

自动文本摘要具有较强的可行性:从 90 年代开始,随着机器学习技术在自然语言处理中的应用,自动文本摘要技术中开始融入人工智能的元素,主要针对新闻、学术论文等主题明确、结构清晰的文档。自动摘要技术在今后主要有两个个发展趋势:第一,基于深度学习的摘要技术研究将成为热点。当前深度学习的各项技术飞速发展,标记语料不断增加,给其广泛应用提供了基础条件。基于深度学习的方法可以更加全面地把握全文信息,更好地表示语义信息,在自动摘要领域的应用将不断完善,效果也会逐渐提升。第二、在应用层面,与现实场景相结合的技术改进方法,也将成为研究的趋势。

实施本项目应该做好以下工作,研究经典的文本摘要方法,以现有的自动文本摘要技术为基础,复现并掌握文本摘要提取技术,在不同的数据集上训练测试与实验;提出新的模型和方法,改进现有问题,更高效、精准地完成文本摘要提取。本项工程目标是:建立完善的文本数据集,设计并实现新型自动文本摘要模型。完成论文的撰写与专利发表,并将其市场化。