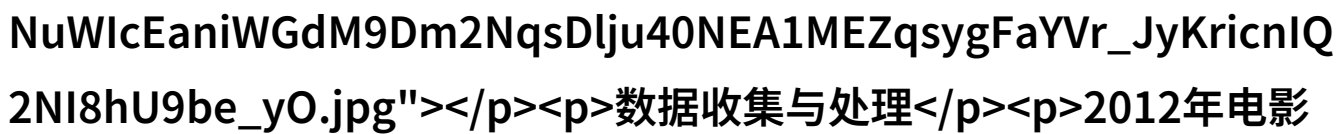
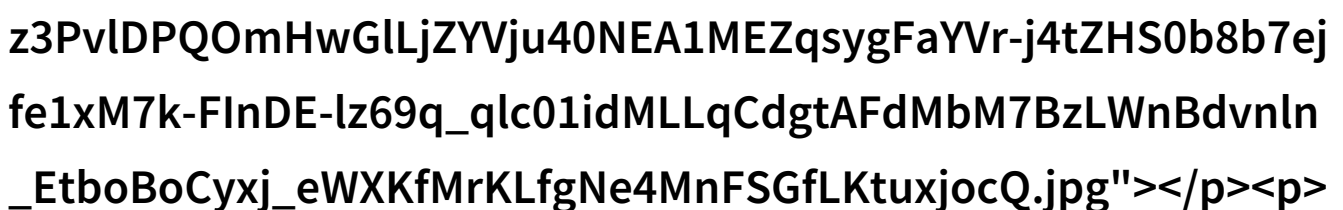


2012年电影字幕大数据分析报告

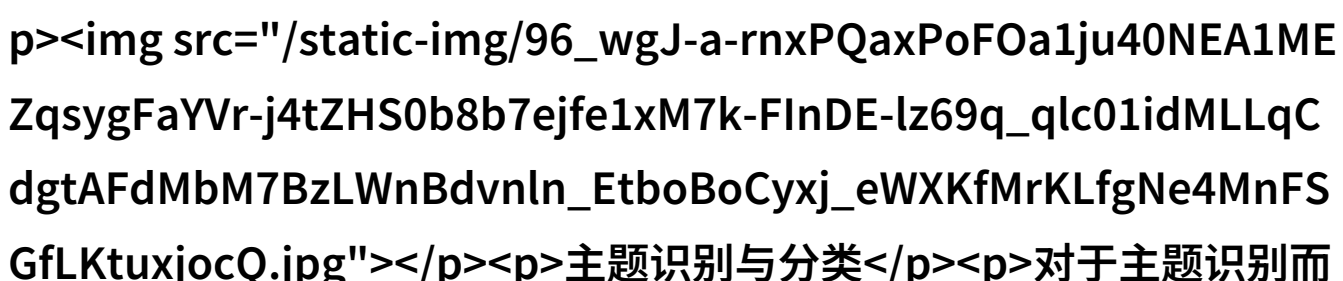
2012年电影字幕大数据分析报告

数据收集与处理

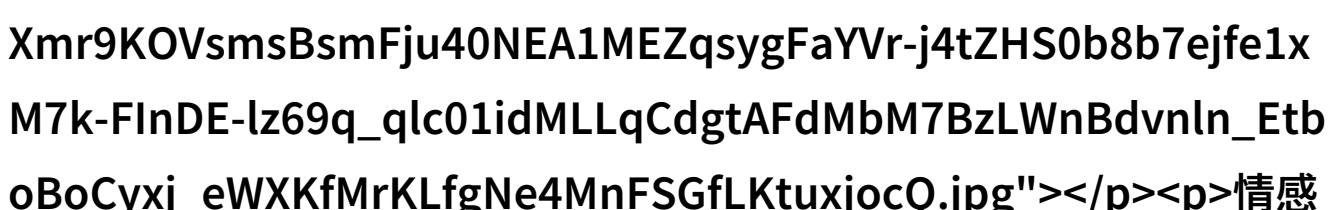
2012年电影字幕的数据收集是本次分析的起点。通过爬虫技术，我们成功地从互联网上抓取了大量的字幕文件。接下来，我们对这些文件进行了清洗和预处理，去除了不必要的信息，如无关文本、标记符号等，以确保分析时所依据的是纯净且有用的数据。

分析工具与方法

在数据处理完成后，我们选择了一系列适用于大规模文本数据分析的工具和方法。这些包括自然语言处理库、机器学习算法以及统计软件等。通过对比不同工具和方法在特定任务上的表现，我们找到了最合适应用于2012年电影字幕的大型数据库模型。

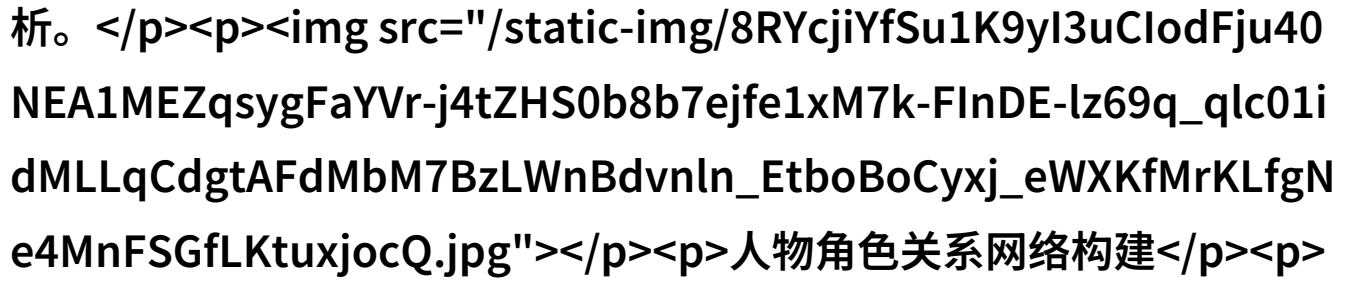
主题识别与分类

对于主题识别而言，关键词提取是非常重要的步骤。在这个过程中，我们运用了TF-IDF (Term Frequency-Inverse Document Frequency) 算法来确定每个句子的重要性，并将其转化为可视化图表，从而直观地展示出各个主题间的相似度和差异性。

情感倾向分析

随着情感倾向研究日益成熟，我们尝试将情感分析技术应用到2012年的电影字幕中。这一过程涉及到对正面负面情绪表达

量化，以及的情绪变化趋势探究。此外，还考虑到了文化背景因素对情感倾向影响的情况下，对不同国家或地区用户群体的情绪偏好进行了解析。



人物角色关系网络构建

为了揭示人物之间复杂多变的人际关系，本次研究还采用社交网络理论建立了基于角色之间互动频率、高度紧密度以及其他指标的人物角色网络。这一工作不仅对于理解剧情发展也为深入挖掘故事背后的隐喻提供了新的视角。

结果总结与未来展望

经过严谨科学的调查研究，总结性的结果显示某些明显趋势：如某些类型影片在一定程度上吸引同类观众，而另一些则超越此种界限；同时，一些演员或导演以其独特风格赢得市场。而未来可能会进一步利用人工智能技术来提升这一领域的实用性，为观众带来更加精准推荐服务。

[下载本文pdf文件](/pdf/652356-2012年电影字幕大数据分析报告.pdf)