提供一套针对新闻评论的倾向性分析组件，包括如下功能：

1.评论倾向性分类，判别评论的正面、负面或中立观点倾向性

2.评论分类，给定类别体系，按照评论主题内容，将评论归入相应的类别。

3.代表性评论推荐，给定专题，从该专题的评论集中抽取代表性观点评论，从而反映该专题的主要评论信息。为此要考虑观点相关性、重要性及多样性等几方面要求。同时对所推荐的评论给出专题关键词和情感关键词，以概括该评论的主要信息。

在组件基础上，提供应用，包括按专题和时间的评论倾向性统计和走势分析和代表性评论观点推荐与统计两大功能。提供相应的远程过程调用、cgi或者独立后台程序，方便系统集成。

半自动的评论专用情感词典构建是离线模块，基于大规模评论语料和通用情感词典知识，利用自枚举方法或者统计话题模型，来抽取评论专用的候选情感词。通过人工检查确定，得到评论专用的情感词典。

评论倾向性分类、评论分类及代表性评论推荐等模块封装成独立组件，并提供一系列API接口，方便其他模块调用。该组件分为两层：基础的评论文本分析处理功能以及具体的评论倾向性分析功能。

“评论倾向性统计和走势分析”及“代表性评论观点推荐与统计”则是评论数据读取、评论倾向性分析及结果入库等流程的整合，并以远程过程调用、CGI或者独立后台程序形式进行系统集成。

本数据平台关注的是查全率，即系统在进行某一检索时，检索出的相关资料量与系统资料库中相关资料总量的比率。查准率则是保证我们找到最有用资料的一个关键，是系统在进行某一检索时，检索出的有用资料数量与检索出资料总量的比率。检索速度或者说响应时间是提高工作效率的保障，指的是从提交检索课题到查出资料结果所需的时间。最基本的检索速度是应该达“千万汉字，秒级响应"。还有诸如收录范围（所查找的范围）、用户负担（用户在检索过程中付出精力的总和）、输出形式 （输出信息表现形式）等指标也是衡量全文检索系统优劣的要素。全文检索对于站点检索是必不可少的，在通常情况下能够帮助人们很快地找到所要的网页；

能提供相关排序功能，因为当检索结果太多时，用户不可能一一浏览，大多数用户只浏览前面几条，没有相关排序，可能准确的检索结果排在后面，用户不能浏览到，而排在前面的检索结果却相关性很少，造成用户的错觉。

此外，还要考虑HTML/XML的特殊性、支持大量并发用户突发访问、Web站点的动态特性、要求索引维护效率很高等方面。