**可视化文本挖掘软件采购技术需求**

为在跨境电商风险评估项目中利用用户评论信息挖掘产品风险，以及支持课题组在智能决策领域的后续研究，现需采购一款文本挖掘软件，能够以非编程的方式对文本和数字信息进行知识挖掘，具体功能需求如下：

• 异常检测算法(outlier detection)

• Apriori关联发现算法

• Bayesian Networks图形概率模型

• C5.0等决策树算法，包括交互式树构建

• PCA等数据压缩算法

• 回归以及逻辑回归(regression and logistic regression)

• K均方、K近邻聚类(K-means and KNN)

• 神经网络(Neural Network)

• 支持向量机(Support Vector Machine)

• 从文件、数据库和RSS订阅源(如博客、网页订阅)中提取文本数据

• 自然语言处理功能处理英语、法语、德语、日语

• 提取特定领域的概念，如单元词、表达、缩写、首字母缩写等

• 识别同义词

• 提取个人、组织、术语、产品、地点和其他用户定义的类型命名概念

• 提取非语言实体，如地址、货币、时间、电话号码、社会保险(Social Security)号码

• 使用和自定义预建的模板和程序库进行情感分析、客户关系管理(CRM)和智能、市场智能

• 使用文本分类算法根据内容智能分组文本文档和记录

• 使用基于文本和可视化的报告查询概念的关系、出现、频率和类型