

NLPIR语义智能分类为企业解决深层信息挖掘

产品名称	NLPIR语义智能分类为企业解决深层信息挖掘
公司名称	灵玖中科软件（北京）有限公司
价格	.00/个
规格参数	
公司地址	北京市海淀区苏州街49-3号5层11号
联系电话	010-62648216

产品详情

随着Internet上文档信息的迅猛发展，文本分类成为处理和组织大量文档数据的关键技术。数据挖掘技术本身就是当前数据技术发展的新领域,文本挖掘则发展历史更短。传统的信息检索技术对于海量数据的处理并不尽如人意,文本挖掘便日益重要起来,可见文本挖掘技术是从信息抽取以及相关技术领域慢慢演化而成的。

利用数据挖掘进行数据分析常用的方法主要有分类、回归分析、聚类、关联规则、特征、变化和偏差分析、Web页挖掘等，它们分别从不同的角度对数据进行挖掘。

分类是找出数据库中一组数据对象的共同特点并按照分类模式将其划分为不同的类，其目的是通过分类模型，将数据库中的数据项映射到某个给定的类别。它可以应用到客户的分类、客户的属性和特征分析、客户满意度分析、客户的购买趋势预测等，如一个汽车零售商将客户按照对汽车的喜好划分成不同的类，这样营销人员就可以将新型汽车的广告手册直接邮寄到有这种喜好的客户手中，从而大大增加了商业机会。

文本分类是按照预先定义的主题类别，为文档集中的每个文档确定一个类别。这样用户不但能够方便地浏览文档，而且可以通过限制搜索范围来使文档的查找更容易、快捷。文本分类可以在较大程度上解决目前文本以及网络上信息杂乱的现象，方便用户准确地定位所需的信息和分流信息。

NLPIR文本语义智能平台中文分类系统是针对互联网内容处理的需要，融合了自然语言理解、网络搜索和文本挖掘的技术，在给定的分类体系下,根据文语义元是统计语义方法中的原子,是不可分本的内容自动确定文本类别的过程.当前的文本割的最小单位,在文本分类中语义元是词;

文本分类是按照预先定义的主题类别,为文档集中的每个文档确定一个类别.这样用户不但能够方便地浏览文档,而且可以通过限制搜索范围来使文档的查找更容易、快捷.

文本分类系统包括了文本的表达、分类器的选择与训练、分类结果的评价与反馈等过程，其中文本的表达又可细分为文本预处理、索引和统计、特征抽取等步骤。文本分类系统的总体功能模块为：

(1) 预处理：将原始语料格式化为同一格式，便于后续的统一处理；

(2) 索引：将文档分解为基本处理单元，同时降低后续处理的开销;

(3) 统计：词频统计，项(单词、概念)与分类的相关概率;

(4) 特征抽取：从文档中抽取反映文档主题的特征;

(5) 分类器：分类器的训练;

(6) 评价：分类器的测试结果分析

NLPIR采用深度神经网络对分类体系进行了综合训练，内置的算法支持类别自定义训练，该算法对常规文本的分类准确率较高，综合开放测试的F值接近88。NLPIR深度文本分类，可以用于新闻分类、简历分类、邮件分类、办公文档分类、区域分类等诸多方面。此外还可以实现文本过滤，能够从大量文本中快速识别和过滤出符合特殊要求的信息，可应用于品牌报道监测、垃圾信息过滤、敏感信息审查等领域。