

## 信息检索技术

一种基于本体论和潜在语义索引的文本语义处理方法\*

秦春秀<sup>1</sup>; 刘怀亮<sup>1</sup>; 赵捧未<sup>1,2</sup>

西安电子科技大学经济管理学院<sup>1</sup>

收稿日期 2006-6-20 修回日期 2006-7-3 网络版发布日期 2006-9-15 接受日期

**摘要** 提出一种基于潜在语义索引和本体论的文本语义处理方法。首先构建一个基于本体论的虚拟标准文本特征向量, 然后采用潜在语义索引方法以虚拟标准文本特征向量为参照对文本集进行语义聚类, 最后在虚拟标准文本特征向量的导引下利用本体库中的知识对聚类获得的文本集合的类别和语义进行显性标注。实验表明, 该方法能较好地语义层面对文本进行有效的聚类, 而且聚类结果能显性地显示类聚所属的类别。

**Abstract** This paper presents a method for text semantic information processing based on Ontology and latent semantic indexing. Firstly, virtual standard text characteristic vectors are constructed; then, the texts are semantically classified into document sets according to virtual standard text characteristic vectors by using latent semantic indexing method; finally, semantically explicit annotations to the document sets are obtained from Ontology-base by guidance of virtual standard text characteristic vectors. Experiments show that method can achieve good text clustering of semantic level, and the clustering can explicitly indicate categories of the clustered documents.

**关键词** [潜在语义索引](#) [本体论](#) [文本聚类](#) [语义](#) [标注](#)

**Key words** Latent semantic indexing; Ontology; Clustering; Semantics; Annotation

**分类号** [G354.2](#)

**DOI:**

通讯作者:

秦春秀 [qinchx@126.com](mailto:qinchx@126.com)

作者个人主页: 秦春秀 刘怀亮 赵捧未

## 扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF](#) (OKB)

▶ [\[HTML全文\]](#) (OKB)

▶ [参考文献\[PDF\]](#)

▶ [参考文献](#)

服务与反馈

▶ [把本文推荐给朋友](#)

▶ [加入我的书架](#)

▶ [加入引用管理器](#)

▶ [引用本文](#)

▶ [Email Alert](#)

▶ [文章反馈](#)

▶ [浏览反馈信息](#)

相关信息

▶ [本刊中 包含“潜在语义索引”的  
相关文章](#)

▶ 本文作者相关文章

• [秦春秀](#)

• [刘怀亮](#)

• [赵捧未](#)

•