<<智能检索技术>>

图书基本信息

书名：<<智能检索技术>>

13位ISBN编号：9787030253286

10位ISBN编号：7030253280

出版时间：2009-8

出版时间：科学出版社

作者：陆建江，张亚非，徐伟光，苗壮　编著

页数：240

前言

面对海量信息，信息的精确检索就像大海捞针一样困难。
智能检索技术吸取多个学科的研究成果，力图通过对文本、图像和视频信息的智能处理，实现信息的精确检索。
本书系统地阐述了文本、图像和视频检索的理论方法和实现技术，并重点突出了本领域的最新研究成果。
本书涵盖智能检索技术的主要内容，全书共分14章：第1～4章介绍文本的智能检索技术，包括文本检索技术、文本自动分词、概念语义空间、基于本体的文本检索技术等；第5～10章介绍图像的智能检索技术，包括MPEG一7标准中图像的视觉特征、图像的局部特征、基于视觉特征的图像检索技术、基于语义的图像检索技术、Web图像的检索技术等；第11～14章介绍视频的智能检索技术，包括视频的结构化技术、语音识别技术、视频的标注技术等。
本书的成果是集体智慧的结晶，由陆建江、张亚非、徐伟光和苗壮负责撰稿。
另外，感谢赵天忠、李阳、肖琪、谢正辉、周波、李冉、李言辉、康达周、王进鹏、王家宝、田豫龙等同学为本书撰写工作付出的辛勤工作，这些同学参与了全书的校对工作，在此深表感谢。
全书每一章的内容组织和细节都经过多次讨论和修改才定稿，力求深入浅出，让读者轻松掌握相关的知识。
尽管每一节、每一句、每篇参考文献，甚至每个标点我们都精心检查，但难免还存在一些缺点和遗漏，殷切希望广大读者批评指正。
希望本书的出版能够对智能检索技术相关领域的研究人员有所裨益，并希望通过阅读本书，读者能够很快进行相关领域的研究工作。

<<智能检索技术>>

内容概要

面对海量信息，信息的精确检索就像大海捞针一样困难。
智能检索技术吸取多个学科的研究成果，力图通过对文本、图像和视频信息的智能处理，实现信息的精确检索。
本书系统地阐述了文本、图像和视频检索的理论方法和实现技术，并重点突出了本领域的最新研究成果。

本书可作为高等院校计算机科学与技术、模式识别与智能系统等学科方向高年级本科生和研究生的教材，也可作为相关领域学生的参考书。

<<智能检索技术>>

书籍目录

章节摘录

第1章 文本检索技术1.3 文本预处理1.3.1 停用词删除我们知道如果一个词在某个文本中多次出现，那么这个词就很有可能与文本的主题密切相关。
然而如果一个词在多个文本中出现，而且频率过高，那么它对文本的区别能力就非常低。
一般地，在文档库的文本中出现频率超过80%的词对检索过程根本起不到作用。
这部分词被称为停用词(stopword)。
在选择构建索引的词时，停用词需要被过滤，以提高索引效率。
一般地，冠词、介词、连词等都是停用词，实际检索系统都会设置一个停用词表。
删除停用词可以大大缩小索引空间的大小，一般可以缩小40%左右。
删除停用词的缺点是可能会影响检索系统的查准率，有的文本检索系统为了克服这一缺点采用全文索引，并不剔除停用词，对所有的词都建立索引。
1.3.2 词干提取词干提取是为了解决英文检索中存在的问题而采取的操作。
词于是指将词的词缀(前缀和后缀)删除后剩下的部分。
例如单词“compete”是它的变形“competes”、“competitor”、“competition”、“competin9”和“competed”的词干。
在英文检索中，如果用户输入的词是信息库中某个相关文本中词的一种变形，词的变形可以是该词的复数、动名词或者过去分词形式等，那么这些相关文本将被视作与查询无关的文本，这将大大影响召回率。
为解决这个问题，在构建索引时，用词干来代替词干的所有变形。
词干提取不仅在很大程度上提高召回率，改善信息检索的性能，同时由于词干的众多变形都由词干代替，用于构建索引的词数量也大大减少，索引空间也进一步缩小。
目前，词干提取技术可以分为：词缀删除、表格查询、后续变形、N-连字。
词缀删除技术比较直观、简单、有效。
在词缀删除中，最重要的就是对词中后缀的删除，因为大多数词的变形是通过后缀来实现的。
目前已经有多种关于词缀删除的算法，其中，Porter算法以其简单性和有效性而得到广泛应用。
表格查询技术通过在表格中查找某个词的词干来实现，表格中的信息依赖于整个语言中词的词干，因此通常需要相当大的存储空间来存放表格，这就制约了表格查询技术的应用。
后续变形技术主要是通过结构化语言的知识来确定词素的边界，这种技术比词缀删除技术复杂。
N-连字技术判断单词中的字母是否连在一起，这一过程实际上是词条聚类的过程。

<<智能检索技术>>

编辑推荐

《智能检索技术》特点：智能检索技术是国内外学术界研究的热点，《智能检索技术》吸取计算机科学与技术、模式识别与智能系统等多个学科的研究成果，系统地阐述了文本、图像和视频检索的理论方法和实现技术，并重点突出语义检索技术的最新研究成果。

版权说明