

信息检索课程设计 (修订)



目录

- 课程介绍与背景
- 信息检索基础知识
- 文本信息检索技术
- 图像与视频信息检索技术
- 语音信息检索技术
- 信息检索应用与实践
- 课程总结与展望



01

课程介绍与背景





信息检索课程的目的



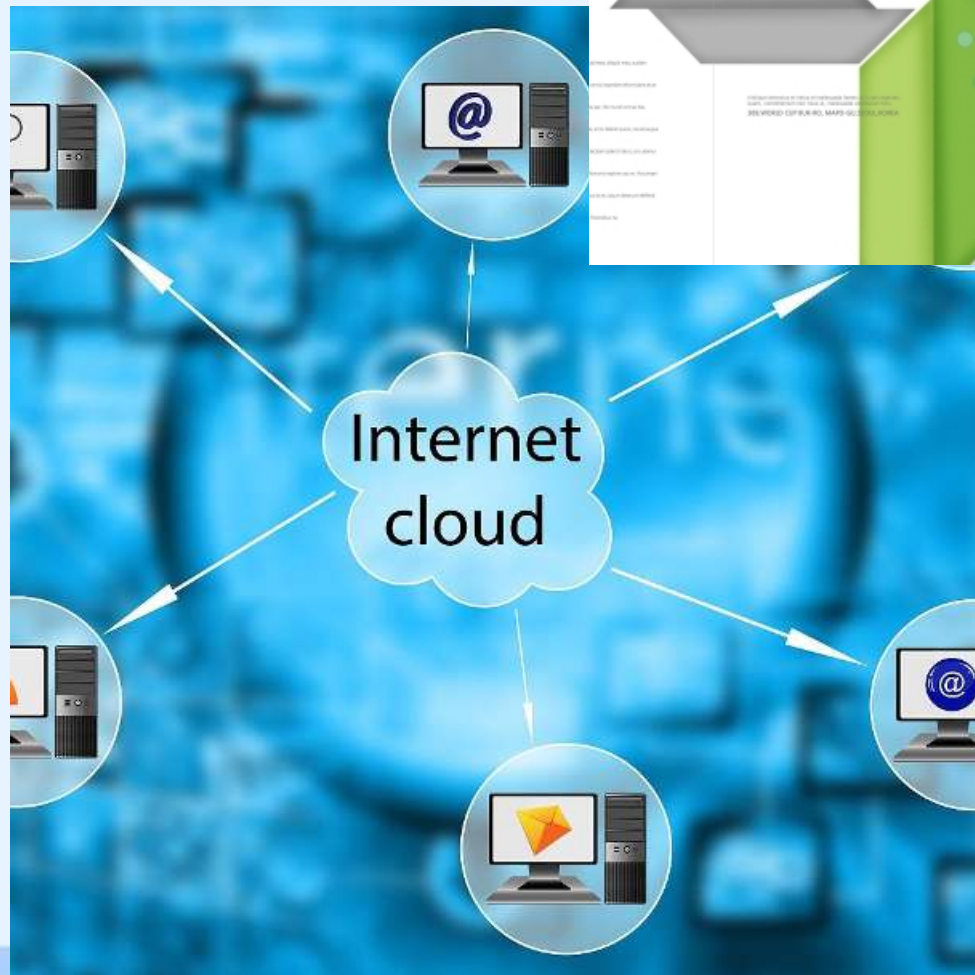
培养学生掌握信息检索的基本理论和方法，提高信息获取和利用的能力。



让学生了解信息检索系统的基本原理、结构、算法和设计方法。



培养学生具备独立分析和解决信息检索问题的能力，以适应信息化社会的需求。





信息检索的历史与发展

1

信息检索的起源

从图书馆的书目检索到计算机化的信息检索系统。

2

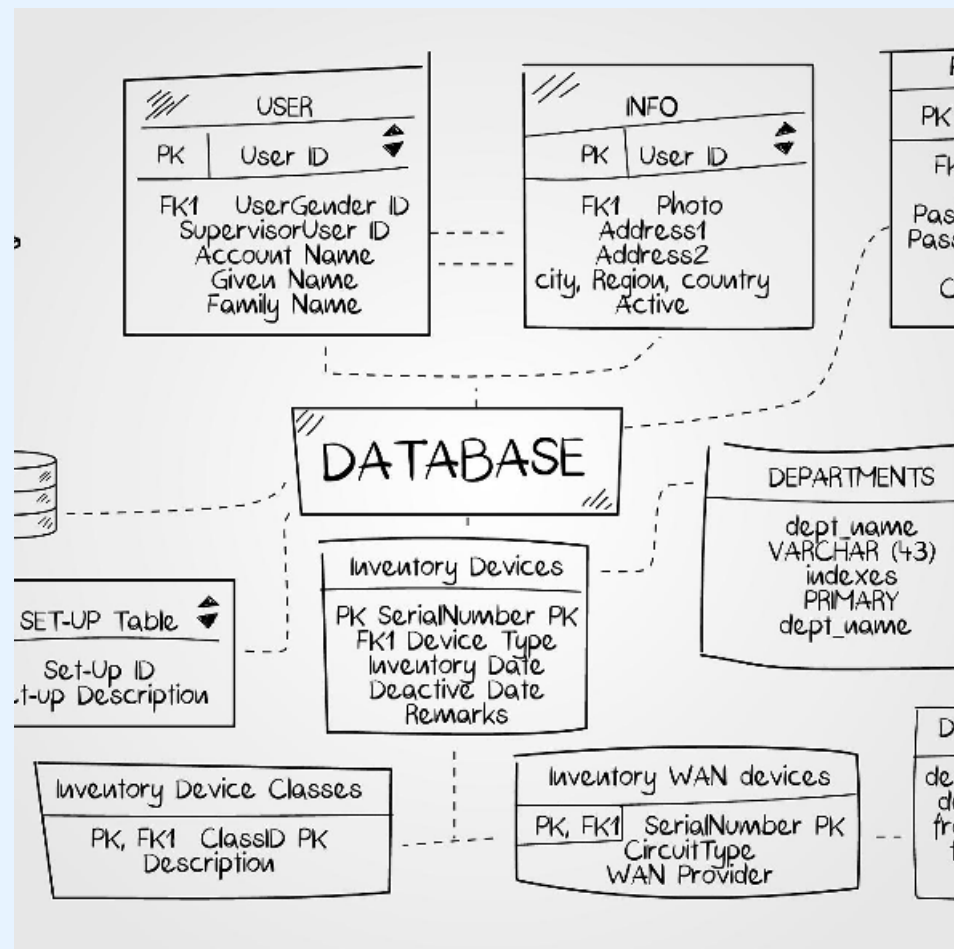
信息检索技术的发展

布尔逻辑检索、向量空间模型、概率检索模型等。

3

网络信息检索的崛起

搜索引擎的出现和发展，以及社交媒体、学术数据库等信息资源的涌现。





修订后的课程特色

强化实践环节

增加实验课程和项目实践，让学生在实践中掌握信息检索技能。



引入最新技术

介绍深度学习、自然语言处理等技术在信息检索中的应用。



拓展应用领域

探讨信息检索在数据挖掘、推荐系统、问答系统等领域的应用。

加强跨学科融合

结合计算机科学、图书馆学、情报学等多学科知识，培养学生综合素质。

02

信息检索基础知识



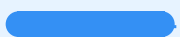


信息检索的定义与原理



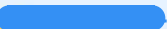
信息检索的定义

信息检索是指从大量的、无序的、不确定的信息集合中，根据用户的信息需求，按照一定的方法和策略，找出满足用户需求的信息，并对这些信息进行加工、整理、组织，以使用户有效利用的过程。



信息检索的原理

信息检索的原理主要包括信息存储和信息检索两个过程。在信息存储过程中，系统对收集到的信息进行标引、分类、组织等处理，形成可供检索的信息资源集合。在信息检索过程中，用户通过输入查询请求，系统根据查询请求与信息资源集合的匹配程度，返回相关的信息资源。



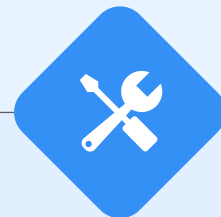


信息检索系统的构成



信息资源集合

包括各种类型的信息资源，如文本、图像、音频、视频等。



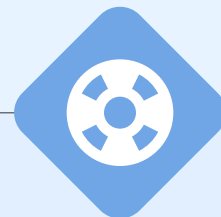
信息处理模块

对信息资源进行标引、分类、组织等处理，形成可供检索的信息资源集合。



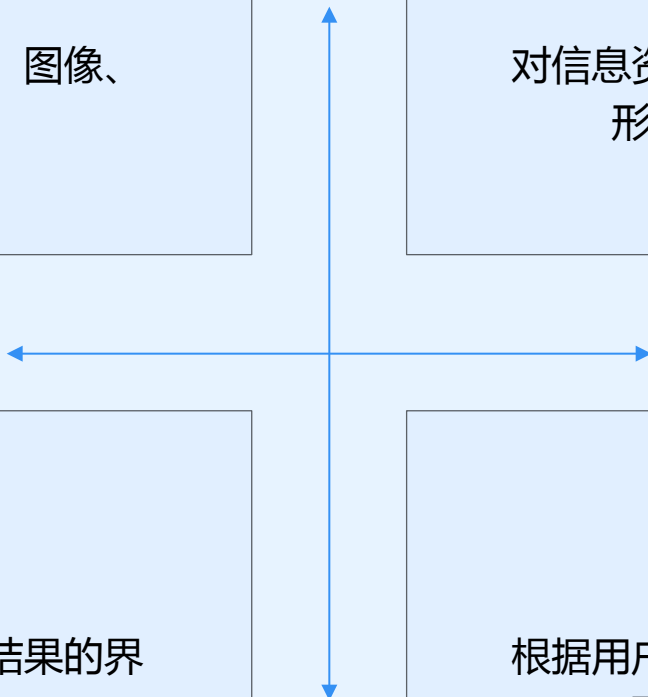
用户接口模块

提供用户输入查询请求和接收检索结果的界面。



检索算法模块

根据用户输入的查询请求和信息资源集合的匹配程度，返回相关的信息资源。





信息检索的评价指标



查全率

指系统检出的相关文献量与检索系统中相关文献总量的比率，是衡量信息检索系统检出相关文献能力的尺度。



查准率

指系统检出的相关文献量与检出文献总量的比率，是衡量信息检索系统精确度的尺度。



响应时间

指从用户输入查询到系统返回结果所需的时间，是衡量信息检索系统性能的重要指标。



用户满意度

指用户对检索结果的主观感受和评价，是衡量信息检索系统实用性和易用性的重要指标。

03

文本信息检索技术





文本预处理技术



● 分词技术

基于规则、统计或混合方法进行中文文本分词，处理歧义和未登录词问题。

● 停用词处理

去除对文本主题贡献较小的停用词，如助词、介词等。

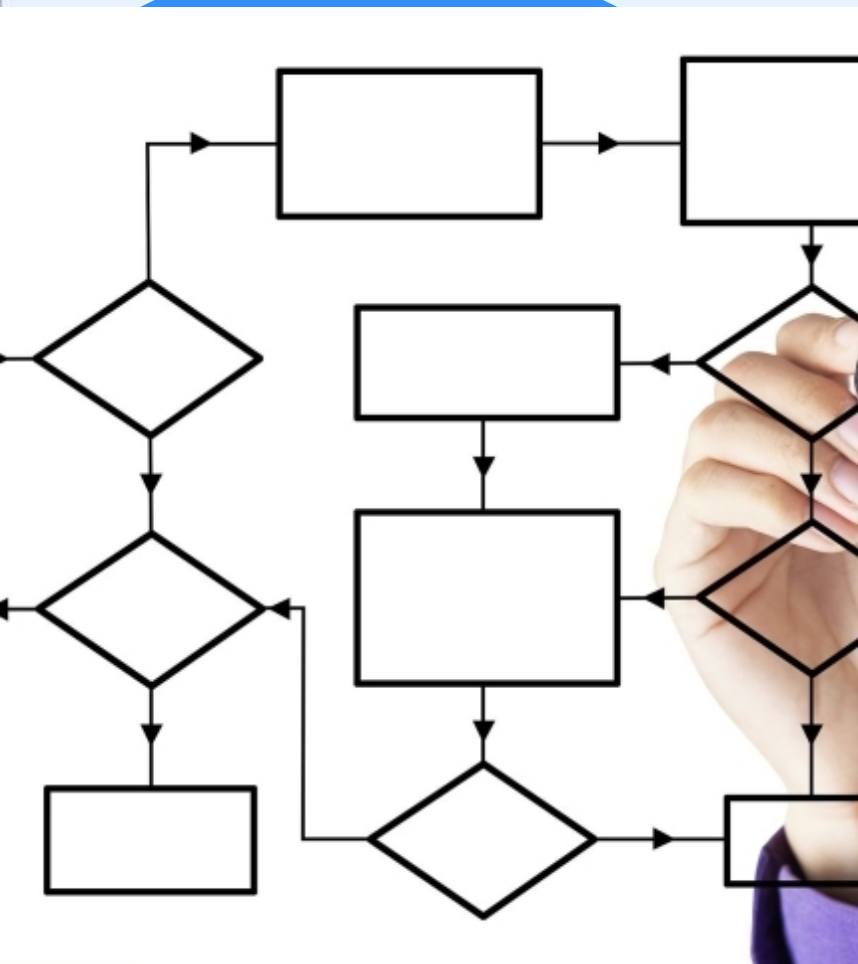
● 文本清洗

去除文本中的噪音和无关信息，如HTML标签、特殊符号等。





文本表示与建模方法



01

词袋模型

将文本表示为词频的向量形式，简单易实现但忽略词序信息。

02

N-gram模型

考虑相邻词之间的关联信息，捕捉局部上下文。

03

主题模型

如LDA等，挖掘文本中隐藏的主题信息，实现文本的降维表示。



文本相似度计算与匹配

基于词袋模型的相似度计算

如余弦相似度、Jaccard相似度等。

基于语义的相似度计算

利用词向量（如Word2Vec、BERT等）计算词语或文本的语义相似度。



文本匹配算法

如BM25、TF-IDF等，实现文本之间的快速匹配与检索。

04

图像与视频信息检索技术





图像特征提取与表示



颜色特征

提取图像的颜色直方图、颜色矩等统计特征，用于描述图像的全局颜色分布。

形状特征

基于轮廓、区域等方法提取图像的形状特征，用于描述图像中目标的形状属性。



纹理特征

利用灰度共生矩阵、Gabor滤波器等提取图像的纹理特征，刻画图像的局部细节信息。

深度特征

利用深度学习技术提取图像的高层语义特征，如卷积神经网络（CNN）的特征图等。



以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：
<https://d.book118.com/916142221003010105>