《信息资源与信息检索》教学大纲

课程编号：2121552B

课程类型：□通识教育必修课 □通识教育选修课

□学科基础课 □专业核心课

☑专业提升课 □专业拓展课

总 学 时：32 讲课学时：32 实验（上机）学时：0

学　　分：2

考试类型：□考试 ☑考查

适用对象：信息管理与信息系统、数据科学与大数据（信息技术）、

计算机科学与技术专业

□是 ☑否 适合作为其他专业学生的个性化选修课

先修课程：数据结构、程序设计基础与应用

一、教学目标

说明本课程的性质以及在人才培养方案中的地位、作用和任务，明确学生在学完本课程后，在思想、知识和能力等方面应达到的目标以及对后续课程的影响。

《信息资源与信息检索》课程是信息管理与信息系统及其相关专业的一门专业拓展课程。该课程是一个跨学科的课程，其中包括计算机技术、数据结构和数据组织。教学主要内容有：信息检索的基本概念与用户界面、经典的信息检索模型及其发展、文档组织和查询、索引和搜索基础、Web爬取和检索基础、数字图书馆基本知识等。

通过本课程的学习，能够达到以下目标：

目标1：理解和掌握信息检索的基本概念和原理。

目标2：掌握信息检索的技术基础。

目标3：了解信息检索技术在网络搜索引擎和在数字图书馆等领域的应用。

目标4：（课程思政目标）在互联网+时代，对学生进行信息道德教育,信息法律与法规教育,信息检索技术的科学精神和创新精神教育等内容，最后通过设计检索案例,设置课程结构和设计课堂教学将思想政治教育元素融入《信息资源和信息检索课程》,为开展课程思政提供了理论基础，培养学生的科学创新精神和爱国精神。

二、教学内容及其与毕业要求的对应关系

（一）教学内容

1．知识体系

第一部分：信息检索的基本概念，包括检索、信息检索系统、查全率和查准率等；

第二部分：用户界面，包括用户界面的重要性、用户界面设计的基本原则和技术等；

第三部分：信息检索模型，包括布尔模型、一般线性模型等经典模型以及广义线性模型等扩展模型；

第四部分：文档组织与查询，包括文档组织的基本概念与基础技术方法、文档查询的基本原理与方法；

第五部分：索引与搜索基础，包括索引的作用和基本技术、基于索引的搜索引擎的基本原理；

第六部分：Web爬取和检索基础，包括Web爬取的基本概念与技术发展，简单的Web爬取实现等。

第七部分：数字图书馆概述，包括数字图书馆的基本概念，检索理论与技术在数字图书馆中的应用。

2．核心内容介绍

本课程的核心内容包括信息检索的基本概念与用户界面、经典的信息检索模型及其发展、文档组织和查询、索引和搜索基础、Web爬取和检索基础、数字图书馆基本知识等。通过本课程的学习，要求学生理解和掌握信息检索的基本概念和原理，掌握信息检索的技术基础，了解信息检索技术在网络搜索引擎和在数字图书馆等领域的应用。

（二）教学方法和手段

根据教学目标，拟采用的教学方法有：课堂讲解基本概念和核心知识，讲授和讨论相结合领会知识要点，案例教学训练解决问题的能力，最后对常用的数据库工具和检索系统进行讲授示范。

（三）学习要求

为有效学习本课程，要求学生首先具备计算机基础、数据结构和程序设计等方面的基本知识，较熟练掌握数据结构等方面的完整知识，具备管理学、计算机科学和数学等专业知识。

自学时应该熟读大纲，提纲挈领地掌握信息检索的基础知识，随后，按照大纲熟读教材，还要参考大量的参考文献，通过课后思考和练习题进行多角度和层次的反复学习。

（四）该课程与毕业要求的对应关系

该课程与毕业要求中的第二点“具备问题分析能力”中的“掌握信息技术和信息管理的基础理论和实践方法；掌握文献研究方法和手段，具有文献检索能力”相对应，通过该课程的学习，学生能够掌握信息资源组织和检索的原理和方法，具有文献检索能力。

三、各教学环节学时分配

以表格方式表现各章节的学时分配，表格如下：

**教学课时分配**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 章节内容 | 讲课 | 实验 | 其他 | 合计 |
| **1** | 第一章 信息检索概述 | **2** | **0** | **0** | **2** |
| **2** | 第二章 用户界面 | **2** | **0** | **0** | **2** |
| **3** | 第三章 信息检索建模 | **6** | **0** | **0** | **6** |
| **4** | 第四章 文档组织与查询 | **6** | **0** | **0** | **6** |
| **5** | 第五章 索引与搜索基础 | **6** | **0** | **0** | **6** |
| **6** | 第六章Web爬取和检索基础 | **6** | **0** | **0** | **6** |
| **7** | 第七章 数字图书馆概述 | **4** | **0** | **0** | **4** |
| **合计** |  | **32** | **0** | **0** | **32** |

四、教学内容

第一章 信息检索概述

第一节 信息检索的发展历史

第二节 信息检索问题

1. 用户的任务

2. 信息检索与数据检索

第三节 信息检索系统

1. 信息检索系统的软件架构

2. 检索和排序过程

教学重点、难点：信息检索的概念、发展历史、查全率与查准率的概念与算法。

课程的考核要求：通过本章的学习,使学生掌握信息检索的基本概念，了解信息检索的发展历史，掌握查全率和查准率的概念。

课程思政切入点：使同学意识到信息社会中信息资源的重要性，在浩若烟海的信息中，信息检索技术的重要性，培养同学的科学精神和创新精神。“信息是跨国界流动的，信息流引领技术流、资金流、人才流，信息资源日益成为重要生产要素和社会财富，信息掌握的多寡成为国家软实力和竞争力的重要标志。信息技术和产业发展程度决定着信息化发展水平，要加强核心技术自主创新和基础设施建设，提升信息采集、处理、传播、利用、安全能力，更好惠及民生。”习近平2014年2月27日在中央网络安全和信息化领导小组第一次会议上的讲话。

复习思考题：

1. 简述信息检索的发展历史经历的阶段。

2. 信息检索和数据检索有什么区别？

3. 检索和排序的过程具体是怎样的？

第二章 用户界面

第一节 人们如何搜索

1. 信息查找与探索式搜索

2. 信息搜寻的经典模型与动态模型

3. 导航与搜索

4. 对搜索过程的观察

第二节 现今的搜索界面

1. 启动搜索

2. 查询描述与查询描述界面

3. 检索结果显示

4. 查询重构

5. 组织搜索结果

第三节 搜索界面可视化

1. 可视化布尔语法

2. 可视化查询结果中的查询项

3. 可视化词语和文档间的关系

4. 文本挖掘的可视化

第四节 搜索界面的设计和评价

教学重点、难点：用户界面设计的基本原则与技术。

课程的考核要求：通过本章的学习，学生应掌握为什么检索系统如此重视用户界面，用户界面设计的原则和基本技术。

复习思考题：

1. 信息搜寻有哪些经典模型？

2. 文本挖掘的可视化技术有哪些？

3. 搜索界面的设计原则是什么？

4. 搜索界面评价有哪些方法？

第三章 信息检索建模

第一节 信息检索模型

1. 建模和排序

2. 信息检索模型描述

3. 信息检索模型的分类体系

第二节 经典信息模型

1. 基本概念

2. 布尔模型

3. 向量模型

4. 概率模型

5. 经典模型之间的简单比较

第三节 其他集合论模型

1. 基于集合的模型

2. 扩展布尔模型

3. 模糊集模型

第四节 其他代数模型

1. 广义向量空间模型

2. 潜在语义索引模型

3. 神经网络模型

第五节 其他模型

1. 超文本模型

2. 基于Web的模型

3. 结构化文本模型

4. 多媒体检索

5. 企业和垂直搜索

教学重点、难点：经典检索模型与扩展检索模型的基本算法。

课程的考核要求：通过本章的学习，使学生掌握布尔模型、一般线性模型等经典模型以及广义线性模型等扩展模型。

课程思政切入点：信息道德教育，具体包括很多内容，如保护知识产权、尊重个人隐私、抑制不良信息、维护信息安全等。高校大学生正处在人生观、世界观和价值观形成和确立的关键时期，大学生要在网络信息的海洋中白由地航行，就必须具备抵御风浪的能力，即良好的信息道德素质。促使遵学生遵循一定的信息伦理与道德准则来规范自身的信息行为活动，在信息检索课上，可重点引导学生了解学术写作中的引用规范，知道参考文献的著录规则，明确信息活动的权利和义务，认识信息安全的重要性，遵循相关的法律规范，抵制各种违法、淫秽、迷信信息和虚假信息，尊重个人隐私等，避免由于缺乏信息伦理道德，在信息活动中引起诸如信息泄密、信息犯罪等社会问题。

复习思考题：

1. 试比较几种经典模型之间的优缺点。

2. 简述广义向量空间模型的基本思想。

3. 说明结构化文本检索的基本架构。

4. 多媒体检索的主要技术有哪些？

第四章 文档组织与查询

第一节 元数据基础

第二节 文档格式

1. 文本

2. 多媒体

3. 图形和虚拟现实

第三节 文档预处理

1. 文本的词汇分析

2. 去除禁用词

3. 词干提取

4. 关键词选择

5. 同意词典

第四节 组织文档

1. 分类体系法

2. 分众分类法

教学重点、难点：文档组织的基本概念与关键技术、文档查询的原理与方法。

课程的考核要求：通过学习本章内容，使学生掌握文档组织的基本方法与关键技术、文档查询的基本原理与方法。

复习思考题：

1. 简述元数据的主要思想。

2. 多媒体文档格式的主要特点是什么？

3. 文档处理的主要过程和步骤是什么？

4. 分众分类法的主要思想是什么？

第五章 索引与搜索基础

第一节 介绍

第二节 带排索引

1. 基本概念

2. 完全倒排索引

3. 搜索

4. 排序

5. 建构

6. 结构化查询

第三节 后缀树和后缀数组

1. 结构：trie树和后缀树

2. 简单字符串搜索

3. 复杂模式的搜索

4. 建构

教学重点、难点：索引的概念与基本原理、搜索引擎的基本原理的主要技术及其实现。

课程的考核要求：通过本章的学习，使学生理解索引的概念与常用技术、搜索引擎的基本原理、索引在搜索引擎中的应用。

课程思政切入点：培养学生要弘扬科学精神，提升自身科学素养。《全民科学素质行动计划纲要实施方案(2016-2020年)》将“引导大学生树立科学思想，弘扬科学精神，激发大学生创新创造创业热情，提高大学生开展科学研究和就业创业的能力”列为重点任务。崇尚科学精神、树立创新意识既是时代的要求，也是大学生立身之本。布置作业让学生搜索我国搜索引擎技术的发展，搜索网站存在的问题以及与国外先进技术的差距，培养创新精神和使命感。

复习思考题：

1. 倒排索引的基本特征是什么？

2. 建立倒排索引的主要过程是什么？

3. 简单字符串搜索的基本思想是什么？

第六章 Web爬取和检索基础

第一节 介绍

第二节 网络爬虫的应用

1. 通用Web搜索

2. 聚焦爬取

3. Web刻画

4. 镜像

5. 网站分析

第三节 架构和实现

1. 爬虫架构

2. 实际问题

3. 并行爬取

第四节 评价

教学重点、难点：Web爬取的关键技术及其实现。

课程的考核要求：通过本章的学习，使学生理解Web爬取的基本原理与方法、掌握简单的Web爬取实现技术。

复习思考题：

1. 网络爬虫主要有哪些应用？

2. 爬虫的基本架构是怎样的？

3. 如何评价网络爬虫效率的高低？

第七章 数字图书馆概述

第一节 基本概念

1. 数字对象和馆藏

2. 元数据和目录

3. 资源库/档案库

4. 服务

第二节 社会经济问题

1. 社会问题

2. 经济问题

第三节 软件系统

第四节 数字图书馆案例

教学重点、难点：数字图书馆的概念和基本技术架构、信息检索技术在数字图书馆中的应用。

课程的考核要求：通过本章的学习，使学生理解网络环境下数字图书馆的重要意义、数字图书馆的基本概念以及信息检索技术在数字图书馆中的应用。

课程思政切入点：课堂中使用合适的具有思想政治教育意义的检索案例可以使学生在提高检索能力的同时接受良好的思想政治教育。通过检索案例将思想政治教育元素融入课堂，的教学设计来实现。例如，通过讲述“日本人获知大庆油田的秘密”事件，在阐述信息素养重要性的同时，可以进行爱国主义教育；通过介绍诺贝尔医学奖获得者、药学家屠坳“幼的研究历程向学生进行科学精神和创新精神教育；通过让学生撰写课程论文可以对学生进行信息道德、科学精神和创新精神教育；通过让学生以小组形式完成检索课题作业比如让学生搜索建党100周年相关内容，“中国共产党为什么能”等培养学生的搜索能力和团结协作精神。

复习思考题：

1. 什么叫数字图书馆？

2. 建立数字图书馆会遇到哪些主要的社会问题？

3. 选择1~2种常见的数字图书馆系统，开展检索体验实践。

五、考核方式、成绩评定

本课程的考核分为平时考核及期末考核两种形式。本课程平时成绩占40%，期末考试成绩占60%。

平时考核采用课堂思考与提问、课堂案例讨论、课下作业等方式。

平时成绩的分配比例为：考勤占10%，课堂思考与提问占5%，课堂案例讨论成绩占10%、课下作业占15%。

期末考核采用开卷形式，试题内容要注重知识应用能力和解决问题能力的考核上，而知识点的记忆和理解要服务于学生能力的提高和巩固。

六、主要参考书及其他内容

[1] 贝泽-耶茨(Ricardo Baeza-Yates)、 里贝罗-内特(Berthier Ribeiro-Neto). 现代信息检索（英文版）.机械工业出版社,2012.

[2] Ricardo Baeza-Yates，Berthier Ribeiro-N. 现代信息检索（原书第2版）.机械工业出版社，2012.

[3] 克罗夫特(W.Bruce Croft)，Donald Metzler，Trevor Strohman.刘挺译. 搜索引擎：信息检索实践. 机械工业出版社，2010.

[4] 祁延莉,赵丹群. 信息检索概论(第2版).北京大学出版社，2013.

[5] 黄如花,信息检索（第三版）, 武汉大学出版社,2019.

执笔人：杨艳红 教研室主任：胡磊　　 系教学主任审核签名：