

南京农业大学
2009 年攻读硕士学位研究生入学考试试题

试题编号: 833 试题名称: 信息检索

注意: 答题一律答在答题纸上, 答在草稿纸或试卷上一律无效

一. 名词解释 (每小题 2 分, 共 14 分)

1. 文摘
2. 科技查新
3. 索引膨胀系数
4. 卢恩
5. 截词检索
6. 关键帧
7. 元搜索引擎

二. 名词辨析 (每题 6 分, 共 24 分)

1. 一次文献、二次文献、三次文献
2. 逻辑与、逻辑或、逻辑非
3. 书目数据库、全文数据库、多媒体数据库
4. 自动分类、自动聚类、自动归类

三. 计算题 (每小题 6 分, 共 18 分)

1. 文档库 D 中共有文献 100 篇, 以 f_{ki} 表示词 k 在第 i 篇文献中的词频 ($1 \leq i \leq 100$)。除 $i=10、18、84$ 时 f_{ki} 分别等于 137、657、339 外, 其余的 f_{ki} 均等于 0。求词 k 在 D 中的文献频率。
2. 设有文档 $d_1 = (2, 1, 1)$ 、 $d_2 = (1, 3, 2)$, 分别以余弦函数法求 $\text{Sim}(d_1, d_2)$ 和欧氏距离法求 $D(d_1, d_2)$ 。
3. 查全率与查准率分别为 0.8、0.5, 求调和数 F 及 b 参数等 1.2 时的 E 测度指标。

四. 简答题 (每小题 6 分, 共 54 分)

1. 什么是类? 给出类的任意一种形式化定义
2. 写出查全率、查准率、误检率、漏检率的计算公式。
3. google、百度等图片搜索的原理是什么?
4. 评价检索系统, 除了查全率、查准率这对金指标外, 还可以从哪些角度考虑?
5. kNN 分类法 k 值的取法。
6. 汉语机械式自动分词算法中, 最短匹配法有无实用价值? 为什么?

南京农业大学
2009 年攻读硕士学位研究生入学考试试题

7. 信息、知识、情报、文献的关系。

8. 信息检索的基本步骤。

9. 某篇文献在文末给出了下列三篇文献作为引文, 请按正确的文末参考文献格式按顺序对其进行著录。

(1) 题名 植物生理学——农业发展的基础 作者 沈允钢; 汤章城;

单位 中国科学院上海植物生理研究所, 中国科学院上海植物生理研究所 上海 200032,
上海 200032

刊名 植物生理学通讯

年 1993

期 02

(2) Alterations in cytosine methylation and species-specific transcription induced by interspecific hybridization between *Oryza sativa* and *O. officinalis* [electronic resource].

Jin, Huajun; Hu, Wei; Wei, Zhe; Wan, Linglin; Li, Gang; Tan, Guangxuan; Zhu, Lili; He, Guangcun; Theoretical and applied genetics TAG.; 2008 Nov.; (8); p. 1271

(3) 中文题名 梨火疫病病原菌核糖体基因结构序列的特征研究及病菌的分子检测

作者 刘华威

学位授予单位 新疆农业大学

中文关键词 梨火疫病病菌;; 分子检测;; 核糖体基因;; 序列分析;; 抗性鉴定

导师 王晓鸣; 郭庆元

论文级别 硕士

学科专业名称 植物病理学

三. 论述题 (每小题 10 分, 共 40 分)

1. 描述十大信息源。

2. 齐普夫定律的内容如何? 如何将其运用于信息检索。

3. TREC 的优缺点。

4. 文档间距离与文档间相似系数两者是什么关系?