

南京农业大学  
2008 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 12 分）

1. 信息可视化
2. 文献语词矩阵
3. 机械分词法
4. 萨尔顿
5. 多媒体数据压缩比
6. 自动抽词标引

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. 文献检索、事实检索、数据检索
3. 目录、题录、文摘
4. 标引、标目、参照

三. 计算题（前三题每小题 6 分，第四小题 8 分，共 26 分）

1. 将测试文档分到十个类目，程序运行结束后，得到的测评数据如下：

$$\#TP_j=6 \quad j=1, 2, \dots, 10$$

$$\#FP_j=2 \quad j=1, 2, \dots, 10$$

$$\#FN_j=4 \quad j=1, 2, \dots, 10$$

求宏观分准率、宏观分全率、微观分准率、微观分全率及各类的分准率、分全率。

2. 对某一检索系统的测评发现，当查全率为 0.25、0.5、0.75 时，对应的查准率分别为 0.8、0.5、0.2，求此时的 AP 值。可否求得此时的 AR？该测评方法的名称分别是什么？
3. 推导并计算等级聚类法的时间复杂性。
4. 设有文档  $d_1=(2, 1)$ 、 $d_2=(1, 3)$ 、 $d_3=(6, 7)$ 、 $d_4=(4, 7)$  和类目中心  $c_1=(4, 3)$ 、 $c_2=(5, 5)$ ，以余弦函数法为相关性测度计算公式，将  $d_1$ 、 $d_2$ 、 $d_3$  全部就近归类（归类过程中类目中心不移动）。

四. 简答题（每小题 6 分，共 48 分）

1. 何谓信息检索经典模型，都有哪几种？
2. 基于篇章结构的网页自要摘要算法中，可用什么方法合并段落？
3. 从信息检索的角度如何看待邮件过滤技术？

南京农业大学  
2008 年攻读硕士学位研究生入学考试试题

---

4. TREC 有哪些局限性?
5. 信息检索中的同义词与语言学中的同义词有何区别?
6. 试比较印刷型文献和电子型文献。
7. 引文索引的用途是什么?
8. 简介世界三大农业文献数据库。

五. 论述题 (每小题 10 分, 共 40 分)

1. 请归纳综述目前多媒体信息检索技术的发展现状。
2. 请评价向量空间模型在信息检索发展史上的意义。
3. 请分析查全率、查准率的用户独立性, 并给出理由。
4. 请看下列索引片段或检索式, 然后回答问题:
  - (1) Grape Sugar **see** Glucose (索引片段)
  - (2) Seed **see also** Grain (索引片段)
  - (3) Mercury (**Planet**) (索引片段)
  - (4) (rice and breed\$).ti. (检索式)
  - (5) **not** English in la (检索式)

请分别写出每个表达式的含义。