

## 一、基本概念

- 搜索广告中的价格机制
  - 次高价拍卖
- 停用词的影响
  - 检索速度、正确率、保留语义信息、影响评分模型如 BM25
- 倒排索引
  - 倒排索引压缩的影响
  - 压缩方法：VB 编码、Gamma 编码

## 二、索引压缩与编码

- VB 编码和 Gamma 编码
  - 应用场景和具体实例计算
  - 总体空间消耗评估

## 三、信息检索评价

- TREC 缓冲池方法
  - Precision/Recall 的可靠性
- 评价指标
  - P/R/F 等基于集合的指标
  - 未插值 Average Precision, nDCG, bPref 等指标的设计思路和计算方法

## 四、检索模型

- 布尔检索的查询优化：合并查询次序的策略
- TF-IDF
- BM25
  - 2-Poisson 假设
- 语言模型
  - 平滑处理目的
- 网络信息采集
  - 采集策略和法规遵守
- 了解各类检索模型的基本原理、设计思路和优缺点

## 五、Web Search

- Pagerank 和 HITS 算法
  - Authority 和 Hub 值
  - 算法的动静态特性
  - 应用场景和效率比较