

# 中央民族大学硕士研究生入学考试初试科目考试大纲

科目代码：**853** 科目名称：**语言信息安全基础综合**

## I. 考察目标

语言信息安全基础综合考试是为高等院校和科研院所招收语言信息安全学科的硕士研究生而设置的具有选拔性质的联考科目。考试内容涵盖数据结构和自然语言处理的主要内容。要求考生比较系统地掌握上述课程的基本概念、基本原理和基本方法，能够综合运用所学的基本原理和基本方法分析、判断和解决有关理论问题和实际问题。

## II. 考试形式和试卷结构

### 一、试卷满分及考试时间

本试卷满分为 150 分，考试时间为 180 分钟。

### 二、答题方式

答题方式为闭卷、笔试。

### 三、试卷内容结构

数据结构 90 分，自然语言处理 60 分。

### 四、试卷题型结构

单项选择题

简答题

综合应用题

## III. 考查范围

### 一、数据结构

#### 【考查目标】

1、掌握数据结构的基本概念、基本原理和基本方法。

2、掌握数据的逻辑结构、存储结构及基本操作的实现，能够对算法进行基本的时间复杂度与空间复杂度的分析。

3、能够运用数据结构基本原理和方法进行问题的分析与求解，具备采用 C 或 C++ 语言设计与实现算法的能力。

#### 一、线性表

(一) 线性表的定义和基本操作

(二) 线性表的实现

1、顺序存储

2、链式存储

3、线性表的应用

#### 二、栈、队列和数组

(一) 栈和队列的基本概念

(二) 栈和队列的顺序存储结构

(三) 栈和队列的链式存储结构

(四) 栈和队列的应用

(五) 特殊矩阵的压缩存储

#### 三、树与二叉树

(一) 树的基本概念

(二) 二叉树

1、二叉树的定义及其主要特征

2、二叉树的顺序存储结构和链式存储结构

3、二叉树的遍历

4、线索二叉树的基本概念和构造

(三) 树、森林

1、树的存储结构

2、森林与二叉树的转换

3、树和森林的遍历

(四) 树与二叉树的应用

1、二叉排序树

2、平衡二叉树

3、哈夫曼（Huffman）树和哈夫曼编码

#### 四、图

（一）图的基本概念

（二）图的存储及基本操作

1、邻接矩阵法

2、邻接表法

3、邻接多重表、十字链表

（三）图的遍历

1、深度优先搜索

2、广度优先搜索

（四）图的基本应用

1、最小(代价)生成树

2、最短路径

3、拓扑排序

4、关键路径

#### 五、查找

（一）查找的基本概念

（二）顺序查找法

（三）分块查找法

（四）折半查找法

（五）B 树及其基本操作、B+树的基本概念

（六）散列（Hash）表

（七）字符串模式匹配

（八）查找算法的分析及应用

#### 六、排序

（一）排序的基本概念

（二）插入排序

1、直接插入排序

## 2、折半插入排序

- (三) 气泡排序 (bubble sort)
- (四) 简单选择排序
- (五) 希尔排序 (shell sort)
- (六) 快速排序
- (七) 堆排序
- (八) 二路归并排序 (merge sort)
- (九) 基数排序
- (十) 各种内部排序算法的比较
- (十一) 排序算法的应用

## 二、自然语言处理

### 【考查目标】

1、理解自然语言处理的基本概念，掌握自然语言处理的步骤，掌握统计自然语言处理的相关算法，能够用某种程序语言实现算法。

2、掌握如何将自然语言处理的相关知识应用于实际应用问题中，如自动分词、命名实体识别、文本分类等。

### 一、基本概念

- (一) 自然语言处理的基本概念
- (二) 自然语言处理的主要研究内容
- (三) 自然语言处理存在的困难
- (四) 自然语言处理的基本方法及发展历程
- (五) 自然语言处理的研究现状

### 二、语言学基础

- (一) 语素的概念
- (二) 词与词性类别
- (三) 短语类型
- (四) 句子及简单句法树分析

### 三、数学基础及信息论基础

- (一) 概率论基础
- (二) 全概率公式
- (三) 期望与方差
- (四) 自信息量
- (五) 信息熵
- (六) 互信息

#### 四、语料库与语言知识库

- (一) 语料库类型
- (二) 语料库建设中的问题
- (三) 典型语料库介绍
- (四) 词汇知识库

#### 五、语言模型

- (一) N 元语法
- (二) 语言模型性能评价
- (三) 参数估计
- (四) 数据平滑

##### 1、加法平滑方法

##### 2、古德-图灵估计法

##### 3、绝对减值法

#### 六、文本表示及特征提取

- (一) 文本向量表示基本概念
- (二) 文本相似度计算
- (三) TF\*IDF 算法
- (四) 互信息方法
- (五) 信息增益方法

#### 七、自然语言处理应用

- (一) 词法分析

##### 1、基于词典的分词方法

##### 2、基于统计的分词方法

### 3、命名实体识别的简单方法

#### （二）文本分类

##### 1、贝叶斯的方法

##### 2、决策树的方法

#### （三）垃圾邮件分类