

功能测试报告

本报告详细记录了对中文分词系统（N-最短路径算法）的功能测试过程。测试采用由浅入深的策略，从基础的单条最短路径分词开始，逐步过渡到复杂的 N-最短路径分词及歧义处理。

1. 相关信息

- 操作系统: Windows
- 编译器: g++ (MinGW64)
- 测试数据:
 - 基础短句 (验证基本切分逻辑)
 - 歧义句子 (验证词性权重调整效果)
 - 长难句 (验证 N-最短路径的多样性)

2. 1-最短路径测试 (N=1)

该部分测试旨在验证系统在默认模式下（N=1），能否正确输出概率最大（路径最短）的唯一分词结果。

2.1 基础切分

T-01 我是人大学生

- 测试目的: 验证基础词库匹配与切分。
- 预期结果: 应切分为 我/是/人大/学生/，避免切分为 我/是/人/大学生/。
- 测试结果:

```
----- 分词结果 -----  
1 : 我 / 是 / 人大 / 学生 /  
-----
```

2.2 歧义处理 (词性加权验证)

T-02 这的确是一个好主意

- 测试目的: 验证副词降权与虚词惩罚机制。
- 预期结果: 应切分为 这/的确/是/一个/好/主意/。若未优化权重，可能会错误切分为 这/的/确/是/...。
- 测试结果:

```
----- 分词结果 -----  
1 : 这 / 的确 / 是 / 一个 / 好 / 主意 /  
-----
```

2.3 专有名词识别

T-03 邓小平与撒切尔夫人在北京会谈

- **测试目的:** 验证人名、地名等专有名词的高优先级权重。
- **预期结果:** 应正确识别 邓小平、撒切尔、北京。
- **测试结果:**

```
----- 分词结果 -----
1 : 邓小平 / 与 / 撒切尔 / 夫人 / 在 / 北京 / 会谈 /
-----
```

3. N-最短路径测试 (N>1)

该部分测试旨在验证系统在 N>1 时，能否正确输出前 N 条最优路径，并观察不同路径之间的细微差异。

3.1 短语的多样性切分 (N=3)

T-04 结合成分子

- **测试目的:** 验证交集型歧义的多种切分可能。
- **预期结果:** 结合/成/分子/，语义最通顺，应放在第一位。
- **测试结果:**

```
----- 分词结果 -----
1 : 结合 / 成 / 分子 /
2 : 结合 / 成分 / 子 /
3 : 结 / 合成 / 分子 /
-----
```

3.2 复杂长句切分 (N=3)

T-05 乒乓球拍卖完了

- **测试目的:** 验证组合型歧义。
- **预期结果:** 乒乓球/拍卖/完/了/ 与 乒乓球拍/卖完/了/ 应该出现在前两位，顺序不定。
- **测试结果:**

```
----- 分词结果 -----
1 : 乒乓球 / 拍卖 / 完 / 了 /
2 : 乒乓球拍 / 卖完 / 了 /
3 : 乒乓球拍 / 卖 / 完 / 了 /
-----
```

3.3 专名与普通词冲突 (N=3)

T-06 王小明在上海

- **测试目的:** 验证人名识别与单字切分的竞争。
- **预期结果:**
 1. 王小明/在/上海/ (人名优先)
 2. 王/小/明/在/上海/ (单字切分)
 3. 其他组合。
- **测试结果:**

```
----- 分词结果 -----
1 : 王小明 / 在 / 上海 /
2 : 王 / 小明 / 在 / 上海 /
3 : 王 / 小 / 明 / 在 / 上海 /
-----
```

3.4 虚词连用测试 (N=3)

T-07 的的确确

- **测试目的:** 验证叠词与单字切分的权重差异。
- **预期结果:**
 1. 的的确确/ (成语/副词整体优先)
 2. 的/的确/确/ (部分切分, 的确 作为副词被降权)
 3. 可能为纯单字切分 的/的/确/确/。
- **测试结果:**

```
----- 分词结果 -----
1 : 的的确确 /
2 : 的 / 的确 / 确 /
3 : 的 / 的 / 确 / 确 /
-----
```

3.5 较长句测试 (N=5)

T-08 有同学反馈部分作业硬件条件受限环境安装有冲突因此给大家一些小建议

- **测试目的:** 验证句子较长时, 系统是否能给出更合理的切分组合。
- **预期结果:** 输出 5 条路径, 能正确“解读”到句意。
- **测试结果:** 第一名确实切分合理, 其他路径则展示了不同的切分可能性。

----- 分词结果 -----

1：有 / 同学 / 反馈 / 部分 / 作业 / 硬件 / 条件 / 受限 / 环境 / 安装 / 有 / 冲突 / 因此 / 给 / 大家 / 一些 / 小 / 建议 /

2：有 / 同学 / 反馈 / 部分 / 作业 / 硬件 / 条件 / 受限 / 环境 / 安装 / 有 / 冲突 / 因此 / 给 / 大家 / 一些 / 小建 / 议 /

3：有 / 同学 / 反馈 / 部分 / 作业 / 硬件 / 条件 / 受限 / 环境 / 安装 / 有 / 冲突 / 因此 / 给 / 大家 / 一 / 些小 / 建议 /

4：有 / 同学 / 反馈 / 部分 / 作业 / 硬件 / 条件 / 受 / 限 / 环境 / 安装 / 有 / 冲突 / 因此 / 给 / 大家 / 一些 / 小 / 建议 /

5：有 / 同学 / 反馈 / 部分 / 作 / 业 / 硬件 / 条件 / 受限 / 环境 / 安装 / 有 / 冲突 / 因此 / 给 / 大家 / 一些 / 小 / 建议 /
