|  |
| --- |
| **NLP实验一：汉语分词系统** |
|  |
| **哈尔滨工业大学**  **1180300128-肖远** |
|  |
|  |
|  |
|  |

摘要

|  |
| --- |
| 000  001  002  003  004  005  006  007  008  009  010  011  012  013  014  015  016  017  018  019  020  021  022  023  024  025  026  027  028  029  030  031  032  033  034  035  036  037  038  039  040  041  042  043  044  045  046  047  048  049 |

绪论

|  |
| --- |
| 050  051  052  053  054  055  056  057  058  059  060  061  062  063  064  065  066  067  068  069  070  071  072  073  074  075  076  077  078  079  080  081  082  083  084  085  086  087  088  089  090  091  092  093  094  095  096  097  098  099 |

相关信息

**2.1 实验目标**

本次实验目的是对汉语自动分词技术有一个全面的了解，包括从词典的建立、分词算法的实现、性能评价和优化等环节。本次实验所要用到的知识如下：

* 基本编程能力（文件处理、数据统计等）
* 相关的查找算法及数据结构实现能力
* 语料库相关知识
* 正反向最大匹配分词算法
* N 元语言模型相关知识
* 分词性能评价常用指标

**2.2编程语言与环境**

实验所用编程语言为python，python版本为python2.7，笔者在win10家庭版上完成

词典的构建

正反向最大匹配分词实现

1. 正反向最大匹配分词效果分析
2. 基于机械分词系统的速度优化
3. 基于统计语言模型的分词系统实现

感谢

参考文献

|  |
| --- |
| 600  601  602  603  604  605  606  607  608  609  610  611  612  613  614  615  616  617  618  619  620  621  622  623  624  625  626  627  628  629  630  631  632  633  634  635  636  637  638  639  640  641  642  643  644  645  646  647  648  649 |

1. 附录
2. 补充材料