

实训5：多线程MapReduce词频统计

· ***目标***：掌握多线程和分布式计算。

· ***任务***：使用MapReduce思想统计大文本词频，分片处理，合并结果。

· ***详细要求***：

- o 读取大文本文件（large_text.txt，约1MB）。
- o 分成4份，使用4个线程计算每部分词频。
- o 使用Counter统计词频，合并结果，输出前10高频词到word_counts.txt。
- o 过滤停用词（从stopwords.txt加载）。
- o 验证文件大小和分片准确性，处理文件读取错误。
- o 添加日志记录，记录每个线程的处理时间和词数。
- o 提供配置选项（如线程数可调）。

· ***技能***：多线程、Counter、文件分片、日志记录。