一个简单的java网络爬虫(spider) PDF转换可能丢失图片或格式，建议阅读原文 https://www.100test.com/kao_ti2020/645/2021_2022__E4_B8_80_E4_B8_AA_E7_AE_80_E5_c104_645107.htm 一个简单的java网络爬虫,由于时间原因,没有进一步解释. 需要的htmlparser.jar包到官方网上去下. ----------Spider.java---------- import java.io.BufferedReader. import java.io.InputStreamReader. import java.net.URL. import java.net.URLConnection. import java.util.ArrayList. import java.util.HashMap. import java.util.Iterator. import java.util.List. import org.htmlparser.RemarkNode. import org.htmlparser.StringNode. import org.htmlparser.Node. import org.htmlparser.tags.*. import org.htmlparser.Parser. import org.htmlparser.filters.StringFilter. import org.htmlparser.util.NodeIterator. import org.htmlparser.util.NodeList. import org.htmlparser.util.ParserException. import java.util.Queue. import java.util.LinkedList. public class Spider implements Runnable { boolean search_key_words = false. int count = 0. int limitsite = 10. int countsite = 1. String keyword = "中国".//搜索关键字 Parser parser = new Parser(). // List linklist = new ArrayList(). String startsite = "".//搜索的其实站点 SearchResultBean srb.//保存搜索结果 List resultlist = new ArrayList().//搜索到关键字链接列表 List searchedsite = new ArrayList().//已经被搜索站点列表 Queue linklist = new LinkedList().//需解析的链接列表 HashMap 100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com