Spider.conf

template.sync.website.url=http://127.0.0.1:7000/

定义模板管理系统的位置。

sourcedown.lastmodified.limit.days=140

新闻或博客的过期时间，超过这个时间会直接丢弃，会出现在iget管理端下载日志中

sourcedown.lastmodified.limit.bbs.days=40

使用bbs模板分析出信息的过期时间。在thread（主题列表）页面如果解析出最后更新时间，这个时间按早于设定，该主题的所有信息不下载。不出现在iget管理端下载日志。

component.processor=BBSSupport,toIM4

#bbs下载必选处理过程

component.process.BBSSupport.class=com.zhongsou.ispider.run.proc.BBSSupport

#测试用，数据保存到硬盘

component.process.debugdisk.class=com.zhongsou.ispider.run.proc.DebugDisk

#测试用，数据保存到数据库，在/debug里面可以查看

component.process.debug.class=com.zhongsou.ispider.run.proc.DebugProc

#数据输出到IM4中

component.process.toIM4.class=com.iget6.ispider.run.proc.ToIm4Proc

component.process.toIM4.Message.minimumChar.limit=20

component.process.toIM4.datelimit.delaydays=10

component.process.toIM4.savedisk=true

component.process.toIM4.output.path=E:\downdata\

component.process.toIM4.output.charset=UTF-8

component.process.toIM4.output.maxsize.oneZip.M=30

component.process.toIM4.output.msgnum.oneFile=1000

component.process.toIM4.data.location.info=conf\location\_keywords.txt

component.process.toIM4.data.emotion\_kewords.info=conf\emotion\_keywords.2.txt

信息源设定

1. 名称

名称对应检索库的CH（频道字段）。内容格式为“[频道名]@[网站域名]”

1. 信息源高级定义

domain\_level DL 网站级别

中央媒体/地方媒体/商业全国/商业行业/商业区域

举例：domain\_level=中央媒体

channel\_category CC 频道分类

国内/国际/财经/时政/军事/娱乐/体育/房产/汽车/互联网/科技/教育/社会/医疗/微博/博客

举例：channel\_category=国内

Spider的退出

请执行shutdown.bat退出spider，不能采用强行关闭，强行关闭会造成数据没有完全发送到检索端或者没有完全保存。

与分布式检索系统对接

Spider目录下Conf\searchsystem.conf文件中system.work.mode.distributed = true

检索系统icontroler.properties 文件中index.socket.enabled=true