HttpUnitExample

Description

SpideringDemo.java is an example of how to use httpunit http://httpunit.sourceforge.net/ to index web pages. There is no guarante on how well this works. It's document parsing is limited to html pages and is inspired by the demo (see lucene-demos.jar).

Required Libraries

- httpunit-1.5.X.jar
- lucene-1.X.jar
- lucene-demos-1.X.jar
- Tidy.jar (should be with httpunit)

Source

SpiderDemo.java.ksh