Multithreaded Web Crawler

Hi all!

I wanted to spice/speed up the task of crawling by doing it in a multithreaded way.

This short README will cover the following aspects:

1. How to use
2. Implementation details

1. How to use

Assuming Java (tested with JRE 1.6) you can start crawling using the following command:

```
// The first argument to the JAR is the URL
// The second argument is the maximum number of threads that should be created
java -jar /path/to/Scraper.jar http://gocardless.com 40
```

2. Implementation details

Let's take a quick look at the project structure:

```
Scraper
| src
   |  Threading: Encapsulates Thread and Inter-Thread communication logic
   |  Util:  Contains Helper classes
   |  ScrapeClient: API Access point
   |  Scraper: Interface to the Command Line
   |  ScrapeTask: Thread subclass that carries out the scraping (using jsoup)
   |  URL Repository: Holds the data
```

The actual scraping is performed on top of jsoup.

Walking through the source should provide you with a clear picture. I recommend you to start by looking at the `Scraper.java` class as this acts as the interface to the command line.