

The logo for TXODDS, featuring the word "TXODDS" in white, bold, uppercase letters on a red rectangular background.

WINNING DATA

Here is the task we would like you to consider. You should be able to make reasonable assumptions if some of the requirements are not precise enough and we will be able to discuss during the discussion. If you do not have any parts of the requirements implemented, please, provide your design thoughts, if you do not have them yet, please, let us know as well. It's not all or nothing, try to do what you can picking the items you feel more comfortable with.

Allowed time: 24 hours

Delivery: Push to a git repository

A Simple Producer/Consumer Web Link Extractor

The Producer

1. The producer receives a list of URLs - it can be from file, command line etc; doesn't matter.
2. It extracts the markup from each of the URLs and places this output onto some form of queue.

The Consumer

1. The consumer reads the queue until it is empty and the producer is no longer extracting markup.
2. It parses the HTML and extracts and hyperlinks into a list. This list is output (file or command line) against each parsed URL.

Requirements

1. The producer and consumer must run concurrently.
2. Error handling should ensure isolation - one bad fetch or parse should not affect processing of others.
3. Some unit tests.
4. Create a GitHub account and put the project there, before sending us a link.

Bonus Points

1. URLs fetched concurrently.
2. Trimming oldest queue entries if queue size balloons.
3. Comprehensive test coverage.
4. Other considerations/enhancements that we have neglected here.