

bin/nutch parsechecker

Parsechecker is an alias for org.apache.nutch.parse.[ParserChecker](#)

This class takes an URL, fetches it (without storing) and returns the URL, the parse_data and all of the parse_text of that URL. It is extremely useful for checking parser implementations from the command line.

Usage:

```
bin/nutch parsechecker [-dumpText] [-forceAs mimeType] url
```

[-dumpText]: TEnables us to dump the parse_text into a text file

[-forceAs mimeType]: Forces mimeType for the given URL argument.

url: The URL you wish to check the parser on.

e.g. bin/nutch parsechecker -dumpText <http://nutch.apache.org> > check.log

[CommandLineOptions](#)