

Justin Leitgeb

**Feedtosis – Efficient RSS and  
Atom feed retrieval and new entry  
detection for Ruby**

Tuesday July 13, 2009

# Features

- Feedtosis helps with efficient feed retrieval and new entry detection, meaning:
  - It keeps track of last\_modified and etag for URL, and uses HTTP conditional GET on subsequent requests
  - It keeps track of which feed entries you've seen, and presents you with ones you haven't for processing by your application

# Feedtosis is minimal

- It delegates RSS and Atom processing to FeedNormalizer
- It delegates retrieval of resources over HTTP to taf2-curb
- It isn't threaded, but easily able to work with your application's concurrency model

# Feedtosis internals

- Uses moneta, or other Hash-like key-value store for keeping track of feed details
- Stores a summary Hash for each feed, key is MD5 of url

- Summary structure:

```
{  :etag => "4c8f-46ac09fbbe940",
  :last_modified => "Mon, 25 May 2009 18:17:33 GMT",
  :digests => [ ["f2993783ded928637ce5f2dc2d837f10",
                 da64efa6dd9ce34e5699b9efe73a37a7" ] ]
}
```

# Usage

```
require 'feedtosis'  
require 'moneta'  
require 'moneta/memcache'  
  
url = "http://newsrss.bbc.co.uk/rss/newsonline_world" +  
      "_edition/south_asia/rss.xml"  
f = Feedtosis::Client.new(url, :backend =>  
  Moneta::Memcache.new(:server => 'localhost:1978'))  
res = f.fetch  
  
res.entries  
res.new_entries
```

# Conclusion

- Project “feedtosis” on github (user: jsl)
- Comments or questions? justin at phq.org