

CodeConPresentation

This is the outline for the [CodeCon](#) presentations on [FeedParser](#) I'm giving in February.

Outline

- Introduction
 - Based on the [NewsMonster](#) parser infrastructure (XSLT)
 - Designed for use within Rojo (an online RSS aggregator)
 - Event based not DOM based
 - Jakarta Commons
 - Apache 2.0 Open Source License
- Challenges with building a feed parser
 - Too many standards
 - RSS (0.9, 0.91, 0.92, 1.0, 2.0)
 - Atom (0.3-0.5 and all draft specs (IETF work in progress))
 - OPML
 - FOAF
 - Changes.xml
 - RDF
 - XFN
 - HTML (link parsing, relations, nofollow, meta tags, generators, etc)
 - Modules (dc, aggregation, content, etc)
 - Semantic confusion:
 - rss:entry vs atom:item
 - title issues across specifications (dc, rss, atom, etc)
 - Encoding issues
 - Invalid entity references
 - XML prefix prior to <?xml?> (usually XML comments)
 - Date encoding issues:
 - RFC822 (RSS 2.0)
 - ISO8601 (RSS 1.0 and Atom)
- Feed Event Model
 - SAX model
 - DOM on top (in the future)
 - SAX is about 12x faster
 - [FeedParserListener](#):
 - init()
 - onChannel(state, title, link description): void
 - onItem(state, title, link description): void
 - onItemEnd(): void
 - General API not wire API
- HTTP issues (network API):
 - Timeouts
 - ETags (If-None-Modified)
 - If-Modified-Since
 - [UserAgent](#)
 - Correct string support via Content-Type
- Problems with DOM models:
 - Namespace matching doesn't line up correctly.
 - Doesn't (easily) support ad-hoc schema updates with extensions
 - Plugin API to pass events with vendor specific interfaces.
 - mod_bigcompany
 - [BigCompanyEventListener](#)
 - Totally isolated development.
 - Just register a SAX [DefaultHandler](#) to intercept your own events
- Autodiscovery
 - [FeedLocator](#) API
 - Atom + RSS autodiscovery support
 - Feed location via href
 - URL fishing (disabled by default)
- Blog Profiles
 - Flickr doesn't support HEAD
 - Invalid autodiscovery implementations
 - Avoid URL fishing
 - Profile discovery support
- Feed Creation
 - Same API can be used to create RSS feeds
- API
 - [FeedParsing](#):
 - [FeedParserListener](#) (rss, atom)
 - Directory Parsing
 - [FeedDirectoryParserListener](#) (opml, foaf, changes.xml)
- - Content Parsing
 - Tag Parsing
- Thanks

- Brad Neuberg
- Rojo Team!