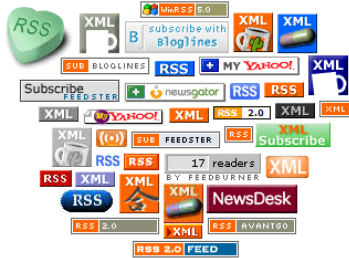


## Short introduction to RSS feeds (flux ou fil RSS)



AIDN-SD&FW  
pfm 2020-2021

pfm-rss, 2020

1/18

## What is a RSS feed?

- RSS is an acronym that stands for **Rich Site Summary**, though it's often alternatively defined as **Really Simple Syndication**. The simplest definition of RSS is an eXtensible Markup Language(XML) format that uses the Resource Description Framework (RDF) for representing information about resources on the World Wide Web.
- Similarly, **Atom** (Open Data framework) is another format based on XML technology.
- A RSS feed is a mean to access new information content published on a web site, without needing to manually access it (using a classical web browser).
- A RSS feed is a file that contain a title, a short description of the data content and a URI toward a web page describing more in details the published information.
- RSS feeds are identified via numerous icons.



pfm-rss, 2020

2/18

## What is a RSS feed?

RSS format is XML

– Simple example :

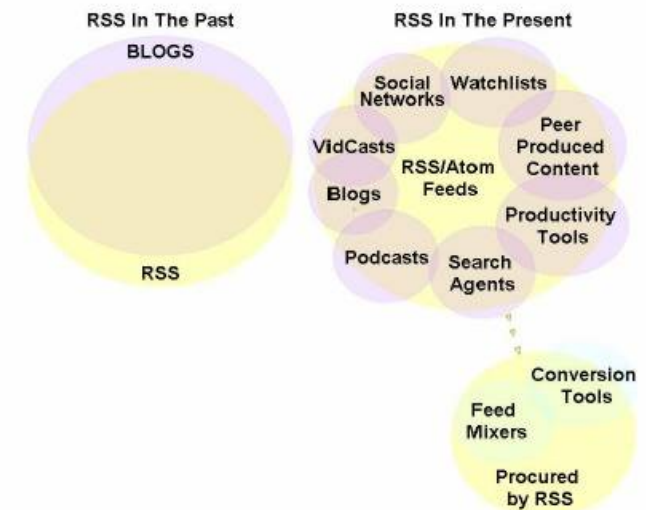
```
<?xml version="1.0" encoding="UTF-8" ?>
<rss version="2.0">
<channel>
  <title>RSS Title</title>
  <description>This is an example of an RSS feed</description>
  <link>http://www.someexamplessdomain.com/main.html</link>
  <lastBuildDate>Mon, 06 Sep 2010 00:01:00 +0000 </lastBuildDate>
  <pubDate>Mon, 06 Sep 2009 16:45:00 +0000 </pubDate>
  <ttl>1800</ttl>

  <item>
    <title>Example entry</title>
    <description>Here is some text containing an interesting description.</description>
    <link>http://www.wikipedia.org/</link>
    <guid>unique string per item</guid>
    <pubDate>Mon, 06 Sep 2009 16:45:00 +0000 </pubDate>
  </item>
  ....
</channel>
</rss>
```

pfm-rss, 2020

3/18

## RSS from 2003 to 2012



pfm-rss, 2020

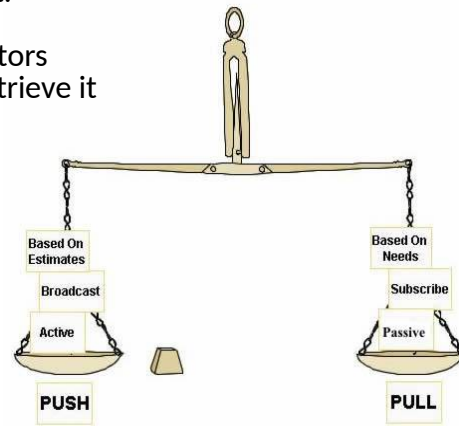
4/18

# RSS feeds

Feeds are most commonly referred to as **pulling**, rather than a pushing technology.

Pushing technology is broadcasting a message and hoping that someone will come by your website or your television station to listen to it.

**RSS is pulling** technology because visitors subscribe to your message and can retrieve it whenever they want.



pfm-rss, 2020

5/18

# Accessing RSS feeds ?

- The information contained in the RSS feed is accessed by a piece of software called an **aggregator**. The aggregator holds a list of web sites and resources that offer RSS feeds. The aggregator may be a desktop application, a browser or mail application extension, or web-based.

## • Web-Based Aggregators

- Bloglines - <http://www.bloglines.com/>
- NewsIsFree - <http://www.newsisfree.com/>
- Newsgator - <http://www.newsgator.com/>
- Google Reader - <http://www.google.com/reader/>

## • Desktop Aggregator

- Fuzzy Duck - <http://www.fuzzyd.co.uk/RSSreader/>
- FeedReader - <http://www.feedreader.com/>
- CITA RSS Aggregator - <http://www.seeita.com/RSSA/>

## • Extensions and Toolbars

- Wizz RSS (Firefox) - <http://www.wizzcomputers.com/WizzRSS.php>
- My RSS Toolbar – (Internet Explorer) - <http://www.myRSStoolbar.com/>

pfm-rss, 2020

6/18

## How to subscribe to RSS feeds?

- So several ways exist to access RSS feeds:
  - Using a dedicated software that allows to aggregate several feeds to provide a single view of the RSS news.
  - Using a web browser
  - Using a mailing application service
  - Developping your own code

pfm-rss, 2020

7/18

Example of a site publishing free RSS feeds

<http://www.freep.com/> **Detroit Free Press**

### RSS feeds for Freep.com headlines

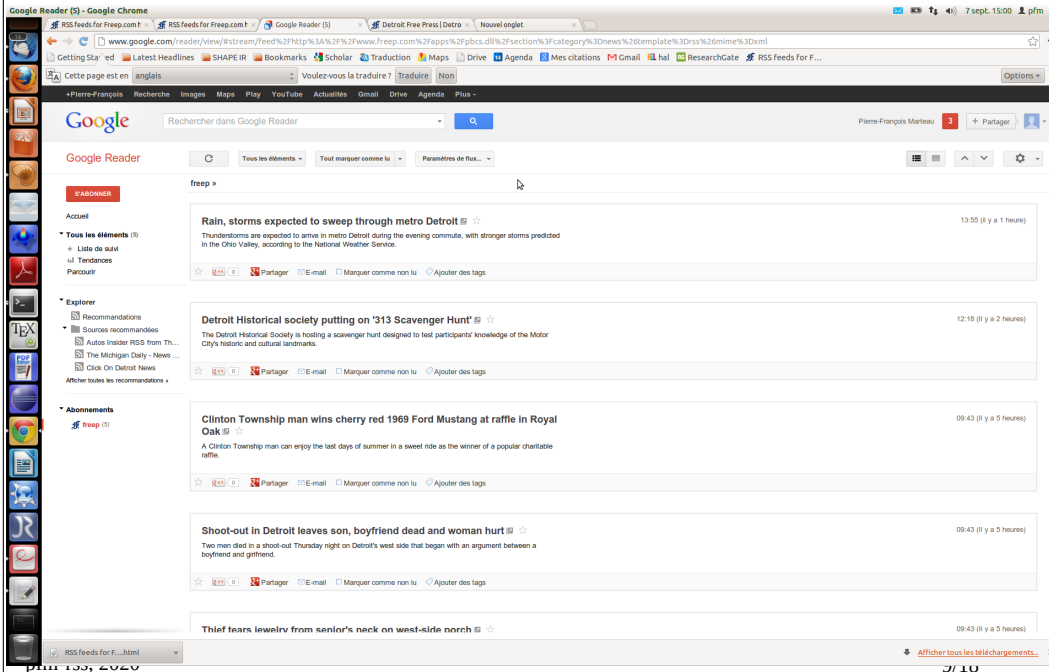
The screenshot shows the Detroit Free Press website. At the top, there's a date 'September 7, 2012' and a 'Recommend' button. Below that, there's a list of categories with RSS icons: Business, Auto News, Sports, Sports - Lions, Sports - Tigers, Sports - Pistons, Sports - Red Wings, Sports - Wolverines, Sports - Spartans, Entertainment, Features, Opinion, Columns, and Blogs. To the right, there's a large red octagon with 'ABP' inside. Below that, there's a 'Most Popular' section with a list of articles, including 'Fact check: A look at Bill Clinton's speech' and 'Secret Service investigates alleged theft of Mitt Romney's tax records'.

pfm-rss, 2020

8/18

## Google Reader on freep Headlines RSS feed

<http://www.freep.com/apps/pbcs.dll/section?category=news&template=rss&mime=xml>



## Example of a **python** code for accessing a RSS feed

```
#!/usr/bin/python3
```

```
import feedparser
import time
from subprocess import check_output
```

```
# -----
# CNN Collector (feedparser)
# -----
```

```
d = feedparser.parse("http://rss.cnn.com/rss/edition.rss")
```

```
# print all posts
count = 1
blockcount = 1
for post in d.entries:
    if count % 5 == 1:
        print("\n" + time.strftime("%a, %b %d %I:%M %p") + ' ((('
        TRIBUNE - ' + str(blockcount) + ' )))')
        print("-----\n")
        blockcount += 1
    print(post.title + "\n")
    count += 1
```

pfm-rss, 2020

10/18

## Executing the code

...

Tue, Sep 26 08:23 AM ((( CNN - 3 )))

Merkel's bruising election makes coalition building tricky

Showdown over NFL protests moves to Monday night

Murder of Indian journalists sparks press freedom fears

Kurdish referendum: Iraq parliament OKs use of force to maintain 'unity'

US: 'No indication' of Iran missile launch, despite Trump tweet

...

pfm-rss, 2020

11/18

## What's up in java?

- My small experience would advice for the ROME initiative
  - ROME is a set of RSS and Atom Utilities for Java that is open source under the Apache 2.0 license.
  - ROME makes it easy to work in Java with most syndication formats:
  - Compliant to RSS 0.90, RSS 0.91 Netscape, RSS 0.91 Userland, RSS 0.92, RSS 0.93, RSS 0.94, RSS 1.0, RSS 2.0, Atom 0.3, Atom 1.0
- <https://rometools.jira.com/wiki/display/ROME/Home>

pfm-rss, 2020

12/18

# Getting started with ROME

```
//ROME small example
import java.net.HttpURLConnection;
import java.net.URL;
import java.util.Iterator;
import java.util.List;

import com.sun.syndication.feed.synd.SyndEntry;
import com.sun.syndication.feed.synd.SyndFeed;
import com.sun.syndication.io.SyndFeedInput;
import com.sun.syndication.io.XmlReader;

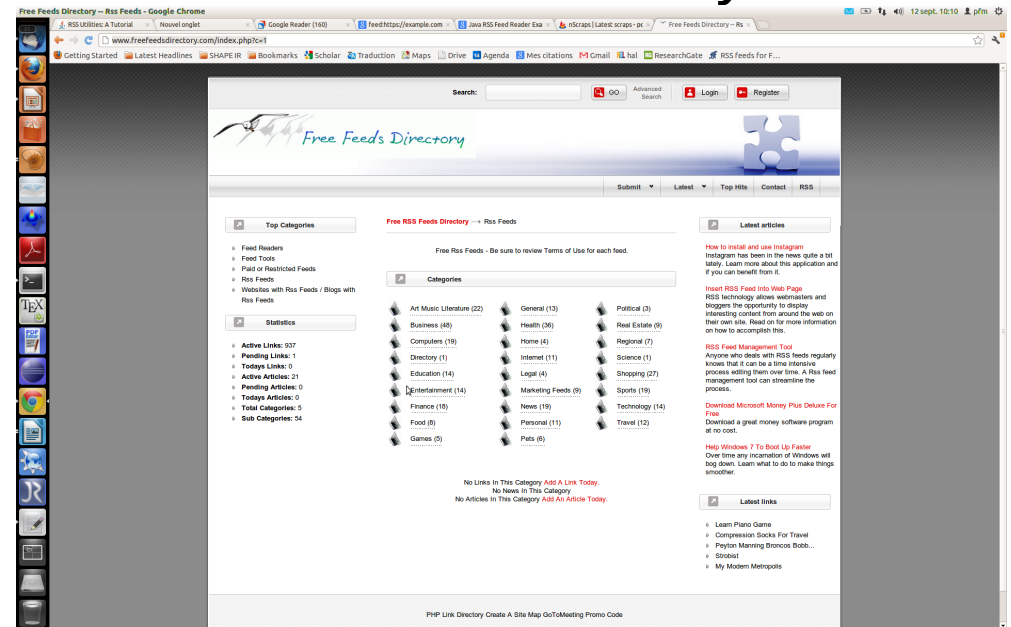
public class HelloFeedParser {
    public static void main(String[] args) throws Exception {
        URL url = new URL("http://movie-revs.blogspot.com/atom.xml");
        if (args.length > 0) {
            try {
                url = new URL(args[0]);
            } catch (Exception ex) {
                ex.printStackTrace();
            }
        } // end if (args...
        HttpURLConnection httpcon = (HttpURLConnection)url.openConnection();
        // Reading the feed
        SyndFeedInput input = new SyndFeedInput();
        SyndFeed feed = input.build(new XmlReader(httpcon));
        List entries = feed.getEntries();
        Iterator itEntries = entries.iterator();

        while (itEntries.hasNext()) {
            SyndEntry entry = (SyndEntry) itEntries.next();
            System.out.println("Title: " + entry.getTitle());
            System.out.println("Link: " + entry.getLink());
            System.out.println("Author: " + entry.getAuthor());
            System.out.println("Publish Date: " + entry.getPublishedDate());
            if (entry.getDescription() != null)
                System.out.println("Description: " + entry.getDescription().getValue());
            System.out.println();
        }
    }
}
```

pfm-rss, 2020

13/18

# Free feed directory



pfm-rss, 2020

14/18

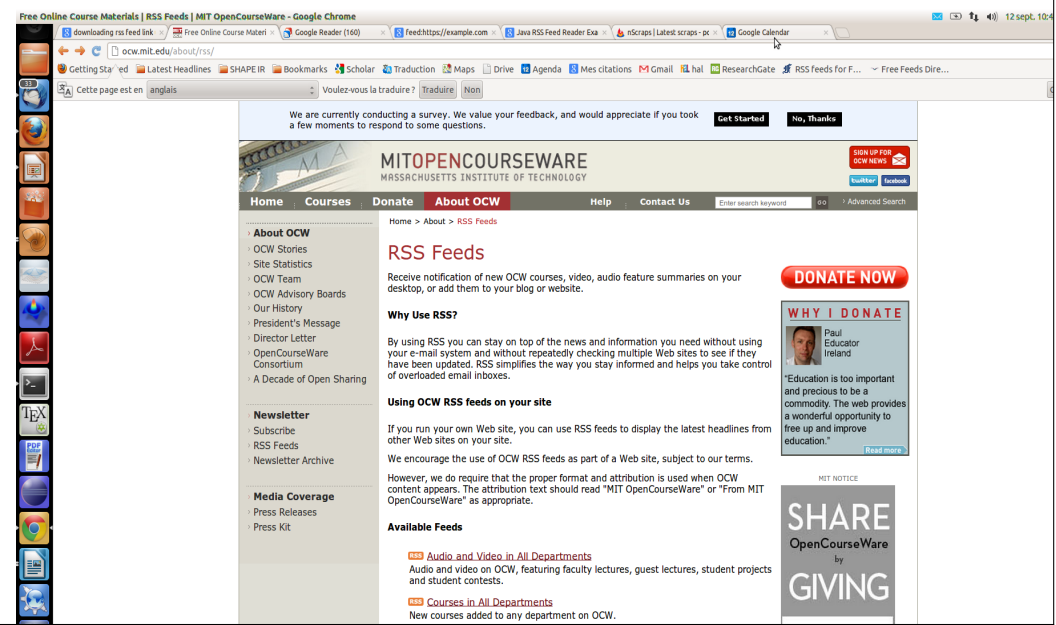
# Free feed directory

- Art-Music\_Litterature
  - Movie Revs:  
<http://movie-revs.blogspot.com/atom.xml>
- Technology
  - Techblology:  
<http://feeds.feedburner.com/techblology>
- Health
  - Healthcare-naturalway :  
<http://feeds.feedburner.com/robusthealth>
  -

pfm-rss, 2020

15/18

# MIT Open Course Ware



# MIT Open Course Ware

- Brain and Cognitive Sciences:  
<http://feeds.phedo.com/OcwWeb/rss/new/mit-newcourses-9>
- Computer Sciences
  - Element of Software Construction:  
<http://www.phedcontent.com/click.phdo?i=00b72f141b1a2b21e6bb0dfe7c30d54e>
  - Automata and Computability:  
<http://www.phedcontent.com/click.phdo?i=afa451f2d8d44031fa282ee6d65fb62d>
- Media Arts and Sciences
  - Pattern Recognition and Analysis:  
<http://www.phedcontent.com/click.phdo?i=05c155611aa327e562cb92720599c6c3>

# For more ...

- RSS on WIKIPEDIA :  
<http://en.wikipedia.org/wiki/Rss>
- Open Data Protocol (Odata) :  
<http://www.odata.org/>
- Developping Feeds with RSS and Atom,  
Ben Hammersley, O'Reilly ed. 2005

