Oleguer Ortega Fernandez
Francesc Garcia Lafuente

THE NEWSMASTERING ENGINE “NEWS HUNTER”
ŽINIŲ TVARKYMO VARYTUVĖ “NEWS HUNTER”

Baigiamasis magistro darbas

Informacinių technologijų
Duomenų gavybos technologijos
Informatikos inžinerija

Vilnius, 2008
THE NEWSMASTERING ENGINE “NEWS HUNTER”
ŽINIŲ TVARKYMO VARYTÜVĖ “NEWS HUNTER”

Baigiamasis magistro darbas

Informacinės technologijos
Duomenų gavybos technologijos
Informatikos inžinerija

Vadovas doc. dr. Mindaugas Rybokas

Vilnius, 2008
Informatics Engineering study programme master thesis
Title: The Newsmastering Engine "News Hunter"
Authors: Oleguer Ortega Fernandez, Francesc Garcia Lafuente. Supervisor dr. M. Rybokas

Annotation
This project is a newsmastering search engine based on RSS feeds technologies and using web technologies to create a user-friendly interface to use the search engine.

The search engine will be programmed using Microsoft .NET technologies and it will be a windows application running in the server.

By the other hand, the website user interface will be programmed using PHP programming language and MySQL database and it will be also running in the same server under Apache HTTP Server.

With this application the user will be able to generate custom newspapers deciding from which sources he wants to take the news and which can of news he wants in the newspaper. After that, the search engine will look for the news that suits more the users' preferences and will generate a newspaper that the user will be able to change before the final publication.

Structure: introduction, analysis, specification, design, conclusions and suggestions, references and appendices.


Keywords: Newsmastering, RSS, RSS feeds, Web application, windows application, search engine.
<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Name</th>
<th>Description/Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>18/10/2007</td>
<td>Oleguer Ortega Fernandez</td>
<td>First Version</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Francesc Garcia Lafuente</td>
<td></td>
</tr>
</tbody>
</table>
CONTENT

1 INTRODUCTION ............................................................................................................. 8

1.1 RESEARCH OBJECT ......................................................................................... 8
1.2 THE OBJECTIVE AND TASKS FOR THE WORK ............................................. 8
1.3 NOVELTY OF THE TOPIC .................................................................................. 9
1.4 IMPORTANCE OF THE TOPIC .......................................................................... 9
1.5 RESEARCH METHODS ....................................................................................... 9
1.6 SCIENTIFIC VALUE OF THE WORK ............................................................ 10
1.7 APPROBATION OF THE WORK ....................................................................... 10

2 ANALYSIS .............................................................................................................. 11

2.1 WHAT IS RSS .................................................................................................... 11
2.1.1 A brief history ............................................................................................... 11
2.1.1.1 So which one do I use? ........................................................................ 12
2.2 WHAT DOES RSS LOOK LIKE? ..................................................................... 12
2.3 HOW TO ACCESS RSS FEEDS ...................................................................... 15
2.3.1 Aggregator........................................................................................................ 15
2.3.1.1 Functions .................................................................................................. 15
2.3.1.2 Varieties ................................................................................................... 16
2.3.1.3 Web-based ................................................................................................ 16
2.3.1.4 Client software .......................................................................................... 17
2.3.1.5 Client Libraries ......................................................................................... 17
2.3.2 RSS Readers or "News Aggregators" .............................................................. 17

2.4 APPLICATIONS CURRENTLY WORKING USING RSS TECHNOLOGY .......... 18

- RSS 2 PDF (http://rss2pdf.com/) ........................................................................... 18

2.5 RSS: USES AND APPLICATIONS NO TRADICIONALS ................................... 19

2.5.1.1 Monitoring Newsgroups/Usenet ................................................................ 19
2.5.1.2 Shared Calendars: .................................................................................. 20
2.5.1.3 Weather: ................................................................................................ 20
2.5.1.4 Classifieds (Houses/Jobs) ...................................................................... 21
2.5.1.5 Ego Searches ......................................................................................... 21
2.5.1.6 Tracking Shipments: ............................................................................. 22
2.5.1.7 Offers and Special Stamps: ................................................................... 22
2.5.1.8 Monitoring online auction houses: ......................................................... 23
2.5.1.9 Monitoring prices: ................................................................................. 23
2.5.1.10 Podcasting and Podcatching: ................................................................ 24
2.5.1.11 Video Edition: ........................................................................................ 25
2.5.1.12 The Discover Discoveries from Its Friends ........................................... 25
2.5.1.13 Newsmastering: .................................................................................... 26
2.5.1.14 Social Networking for music lovers ......................................................... 26
2.5.1.15 Swarming ............................................................................................... 26
2.5.1.16 X-Events .............................................................................................. 27
2.5.1.17 Video Search ......................................................................................... 27
2.5.1.18 Bug reporting ......................................................................................... 27
2.5.1.19 Crime Monitoring ................................................................................ 27
2.5.1.20 RSS Mashups ...................................................................................... 28
2.5.1.21 BroadCatching .................................................................................... 28

(RSS + P2P in order to download fast a custom tv) .................................................. 28

2.5.2 RSS is still very young: .................................................................................. 28

2.6 OBJECTIVES .................................................................................................... 28

2.6.1 Main objectives of the project....................................................................... 28
2.6.2 Functional objectives of the project ............................................................. 29

3 SPECIFICATION .................................................................................................... 31

3.1 FUNCTIONAL REQUIREMENTS ...................................................................... 31

3.1.1 User registration: .......................................................................................... 32
3.1.2 Newspaper management: ........................................................................... 32
3.1.3 Modify Newspaper ................................................................. 33
3.1.4 Publish newspaper ............................................................ 34
3.1.5 Read newspaper library ..................................................... 34
3.2 NON-FUNCTIONAL REQUIREMENTS ........................................ 34
3.3 DEVELOPMENT ENVIRONMENT ............................................ 35
3.3.1 Software development environment .................................... 36
3.3.2 Technologies analysis ....................................................... 37

4 DESIGN ................................................................................. 45

4.1 SYSTEM ARCHITECTURE ...................................................... 45
4.1.1 The web interface ............................................................. 45
4.1.1.1 The design ................................................................. 45
4.1.1.2 Packages ................................................................. 48
4.1.2 The database ..................................................................... 69
4.1.2.1 Administration .......................................................... 69
4.1.2.2 Design ..................................................................... 69
4.1.3 News Hunter application .................................................. 74
4.1.3.1 How the Hunter Program works ................................... 74
4.1.3.2 Document Generation ................................................ 84
4.1.3.3 What is XSLT ............................................................ 87
4.1.3.4 Experience improve method ....................................... 93
4.1.3.5 How to use News Hunter Program ............................... 94

5 CONCLUSIONS .................................................................... 101

5.1 AIMS OF OUR PROJECT ....................................................... 101
5.1.1 General aims ................................................................. 101
5.1.2 Concrete aims ............................................................... 101
5.1.3 Others ........................................................................... 101
5.2 PERSONAL EXPERIENCE .................................................. 101

6 BIBLIOGRAPHY ................................................................. 103

6.1 USED .................................................................................... 103
6.2 ESSENTIAL READING ....................................................... 104
6.3 FORUM .............................................................................. 104
6.4 OTHER PAGES ................................................................. 104

7 APPENDICES ...................................................................... 105

7.1 THE RSS NEWSMASTER ....................................................... 105
7.1.1 The Issues Catalyzing The Change .................................... 107
• TOO MUCH INFORMATION COMING IN ................................ 107
  7.1.2 The Network Starts To Organize Itself ............................... 108
  7.1.3 RSS Dolby ................................................................. 108
  7.1.4 What So Great About This? ............................................ 109
  7.1.5 What's The Great Benefit? .............................................. 109
  7.1.6 The Network Is Already Doing It! ................................. 110
  7.1.7 So What Is A Newsmaster And What Does sHe Do? .......... 111
  7.1.8 The NewsMaster Toolkit - An Opportunity To Support Robin Good ............................................ 113
  7.1.9 Personal Thoughts ....................................................... 113

7.2 THE BIRTH OF THE NEWSMASTER: THE NETWORK STARTS TO ORGANIZE ITSELF ................................. 114
7.2.1 NewsMasters: RSS And The Opportunity For Sustainable Filtering And Aggregation Of Online Content Into Niche Websites And Dedicated Information Channels ................................................................. 114
  7.2.2 What Is The Discovery? .................................................. 115
  7.2.3 Where Is The Real Value? ................................................. 115
  7.2.4 Summary: Critical Issues/ Problems At Hand .................. 115
  7.2.5 The Filtering Mechanisms Available To Us ........................ 115
  7.2.6 A Learning Voyage Into RSS .......................................... 117
  7.2.7 Possible Solutions: RSS Aggregators? ............................. 118
  7.2.8 How Do I Want To Find Information Relevant To Me ....... 118
1 Introduction

1.1 Research Object

The object of research of this project is the newsmastering engines technologies based on RSS news feeds that everyday it passes are taking more prominence in the internet network. In the next sections will be explained deeply what are those technologies about.

Nowadays, the World Wide Web has changed his philosophy giving place to what is known as Web 2.0. This new generation doesn’t lie in creating new technologies to create new sites, but in web-based communities and hosted services — such as social-networking sites, wikis, and folksonomies — which aim to facilitate creativity, collaboration, and sharing between users. In order to achieve that aim those websites must be simple and easy to use in order that everybody can use them. Newsmastering websites and RSS news feeds are Web 2.0 technologies.

1.2 The Objective and Tasks for the Work

The objective of this project is to create a website based on RSS news feed that will provide a newsmastering RSS search engine to generate custom newspaper with the user preferences.

In order to achieve that it will be necessary to do a deeply research about the way of working of the RSS news feeds technologies in order to apply those new knowledge in a website that uses web technologies according to the days we are living. At the same time it will be necessary to research in other Web 2.0 technologies in order to generate the website that will allow the users to generate the newspapers.

After that research task it will be necessary to apply the knowledge acquired during the university years to generate the final project. At the same time, this project is addressed to be used for all kind of users so it’s necessary to generate an easy interface to use the application and easily accessible for everybody.
This website will help people to create newspapers depending on the interests that each one has in an easy way and without the need of checking all the sources the user need. It means that automatically the Newsmastering website will check for the news in the different sources the user decide.

1.3 Novelty of the Topic

Although nowadays there are a few different websites that give you the opportunity of getting information from blog sites or that kind of stuff, there’s no really a website that gives the same service that News Hunter gives. That kind of websites are starting to appear now so News Hunter can be an innovative site that can be very useful to professional investigators, information librarians, news editors or just people that are interested in read only some kind of news from different sources.

1.4 Importance of the Topic

As it was said before, the service that gives News Hunter is quite new and each day it passes it’s taking more importance in Internet. That kind of websites forms the second generation of websites and nowadays millions of people are using the new functionalities that they give.

At the same time, creating a website like News Hunter makes practice a lot of things that have been learned during the Computer Engineering studies and it’s a complete project to put in practice our knowledge.

1.5 Research methods

In order to create the project it will be used Internet as the main source for helping in the different problems that could be found in the creation of the News Hunter project. Besides, it will be useful to check different books and university material used during all those years of studies.

1.6 Scientific value of the work

As it was said before, this project collects a lot of different knowledge from our studies. For example, programming and algorism creation for the search engine, programming for the website interface or
designing an attractive interface for the user. Of course, is difficult to collect all the thinks learned during the years of studies but News Hunter it's an ambitious and complete project.

1.7 Approbation of the work

Due to the complexity of the project, the coordinator of the project approved it without doubt because it carries out the different requirements asked for it.
2 Analysis

2.1 What Is RSS

RSS is a format for syndicating news and the content of news-like sites, including major news sites like Wired, news-oriented community sites like Slashdot, and personal weblogs. But it's not just for news. Pretty much anything that can be broken down into discrete items can be syndicated via RSS: the "recent changes" page of a wiki, a changelog of CVS checkins, even the revision history of a book. Once information about each item is in RSS format, an RSS-aware program can check the feed for changes and react to the changes in an appropriate way.

RSS-aware programs called news aggregators are popular in the blogging community. Many weblogs make content available in RSS. A news aggregator can help you keep up with all your favorite weblogs by checking their RSS feeds and displaying new items from each of them.

2.1.1 A brief history

But coders beware. The name "RSS" is an umbrella term for a format that spans several different versions of at least two different (but parallel) formats. The original RSS, version 0.90, was designed by Netscape as a format for building portals of headlines to mainstream news sites. It was deemed overly complex for its goals; a simpler version, 0.91, was proposed and subsequently dropped when Netscape lost interest in the portal-making business. But 0.91 was picked up by another vendor, UserLand Software, which intended to use it as the basis of its blogging products and other web-based writing software.

In the meantime, a third, non-commercial group split off and designed a new format based on what they perceived as the original guiding principles of RSS 0.90 (before it got simplified into 0.91). This format, which is based on RDF, is called RSS 1.0. But UserLand was not involved in designing this new format, and, as an advocate of simplifying 0.90, it was not happy when RSS 1.0 was announced. Instead of accepting RSS 1.0, UserLand continued to evolve the 0.9x branch, through versions 0.92, 0.93, 0.94, and finally 2.0.
2.1.1.1 So which one do I use?

That's 7 -- count 'em, 7! -- Different formats, all called "RSS". As a coder of RSS-aware programs, you'll need to be liberal enough to handle all the variations. But as a content producer who wants to make your content available via syndication, which format should you choose?

<table>
<thead>
<tr>
<th>Version</th>
<th>Owner</th>
<th>Pros</th>
<th>Status</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90</td>
<td>Netscape</td>
<td></td>
<td>Obsoleted by 1.0</td>
<td>Don't use</td>
</tr>
<tr>
<td>0.91</td>
<td>UserLand</td>
<td>Drop dead simple</td>
<td>Officially obsoleted by 2.0, but still quite popular</td>
<td>Use for basic syndication. Easy migration path to 2.0 if you need more flexibility</td>
</tr>
<tr>
<td>0.92, 0.93, 0.94</td>
<td>UserLand</td>
<td>Allows richer metadata than 0.91</td>
<td>Obsoleted by 2.0</td>
<td>Use 2.0 instead</td>
</tr>
<tr>
<td>1.0</td>
<td>RSS-DEV Working Group</td>
<td>RDF-based, extensibility via modules, not controlled by a single vendor</td>
<td>Stable core, active module development</td>
<td>Use for RDF-based applications or if you need advanced RDF-specific modules</td>
</tr>
<tr>
<td>2.0</td>
<td>UserLand</td>
<td>Extensibility via modules, easy migration path from 0.9x branch</td>
<td>Stable core, active module development</td>
<td>Use for general-purpose, metadata-rich syndication</td>
</tr>
</tbody>
</table>

2.2 What does RSS look like?

Imagine you want to write a program that reads RSS feeds, so that you can publish headlines on your site, build your own portal or home-grown news aggregator, or whatever. What does an RSS feed look like? That depends on which version of RSS you're talking about. Here's a sample RSS 0.91 feed:

```xml
<rss version="0.91">
  <channel>
```


<title>XML.com</title>
<link>http://www.xml.com/</link>
<description>XML.com features a rich mix of information and services for the XML community.</description>
<language>en-us</language>

<item>
  <title>Normalizing XML, Part 2</title>
  <link>http://www.xml.com/pub/a/2002/12/04/normalizing.html</link>
  <description>In this second and final look at applying relational normalization techniques to W3C XML Schema data modelling, Will Provost discusses when not to normalize, the scope of uniqueness and the fourth and fifth normal forms.</description>
</item>

<item>
  <title>The .NET Schema Object Model</title>
  <link>http://www.xml.com/pub/a/2002/12/04/som.html</link>
  <description>Priya Lakshminarayanan describes in detail the use of the .NET Schema Object Model for programmatic manipulation of W3C XML Schemas.</description>
</item>

<item>
  <title>SVG's Past and Promising Future</title>
  <link>http://www.xml.com/pub/a/2002/12/04/svg.html</link>
  <description>In this month's SVG column, Antoine Quint looks back at SVG's journey through 2002 and looks forward to 2003.</description>
</item>
</channel>
</rss>

Simple, all right? A feed comprises a channel, which has a title, link, description, and (optional) language, followed by a series of items, each of which have a title, link, and description.

Now look at the RSS 1.0 version of the same information:

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns="http://purl.org/rss/1.0/"
<channel rdf:about="http://www.xml.com/cs/xml/query/q/19">
  <title>XML.com</title>
  <link>http://www.xml.com/</link>
  <description>XML.com features a rich mix of information and services for the XML community.</description>
  <language>en-us</language>
  <items>
    <rdf:Seq>
    </rdf:Seq>
  </items>
</channel>

<item rdf:about="http://www.xml.com/pub/a/2002/12/04/normalizing.html">
  <title>Normalizing XML, Part 2</title>
  <link>http://www.xml.com/pub/a/2002/12/04/normalizing.html</link>
  <description>In this second and final look at applying relational normalization techniques to W3C XML Schema data modelling, Will Provost discusses when not to normalize, the scope of uniqueness and the fourth and fifth normal forms.</description>
  <dc:creator>Will Provost</dc:creator>
  <dc:date>2002-12-04</dc:date>
</item>

<item rdf:about="http://www.xml.com/pub/a/2002/12/04/som.html">
  <title>The .NET Schema Object Model</title>
  <link>http://www.xml.com/pub/a/2002/12/04/som.html</link>
  <description>Priya Lakshminarayanan describes in detail the use of the .NET Schema Object Model for programmatic manipulation of W3C XML Schemas.</description>
  <dc:creator>Priya Lakshminarayanan</dc:creator>
  <dc:date>2002-12-04</dc:date>
</item>
Quite a little bit more verbose. People familiar with RDF will recognize this as an XML serialization of an RDF document; the rest of the world will at least recognize that we're syndicating essentially the same information. In fact, we're including a bit more information: item-level authors and publishing dates, which RSS 0.91 does not support.

2.3 How to access RSS feeds

There are a number of ways to access RSS feeds. You can install a news reader that displays RSS feeds from the Web sites you select, enabling you to view hundreds of headlines at once. After installing the news reader, you can add each feed manually from the Web site by clicking on the "Subscribe" or the "XML" orange button next to the feed.

An alternative to downloading a dedicated news reader is to use a Web-based news reader. For example, My Yahoo! or My AOL users can now add RSS feeds directly to their personal page.

2.3.1 Aggregator

Client software or a Web application which aggregates syndicated web content such as news headlines, blogs, podcasts, and vlogs in a single location for easy viewing.

2.3.1.1 Functions

Aggregators reduce the time and effort needed to regularly check websites for updates, creating a unique information space or "personal newspaper." Once subscribed to a feed, an aggregator is able to
check for new content at user-determined intervals and retrieve the update. The content is sometimes described as being "pulled" to the subscriber, as opposed to "pushed" with email or IM. Unlike recipients of some "pushed" information, the aggregator user can easily unsubscribe from a feed.

Aggregator features are frequently built into portal sites (such as My Yahoo! and iGoogle), modern Web browsers and email programs.

The aggregator provides a consolidated view of the content in a single browser display or desktop application. Such applications are also referred to as RSS readers, feed readers, feed aggregators, news readers or search aggregators. Aggregators with podcasting capabilities can automatically download media files, such as MP3 recordings. In some cases, these can be automatically loaded onto portable media players (like iPods) when they are connected to the PC.

Recently, so-called RSS-narrators have appeared, which not only aggregate text-only news feeds, but also convert them into audio recordings for offline listening.

The syndicated content an aggregator will retrieve and interpret is usually supplied in the form of RSS or other XML-formatted data, such as RDF/XML or Atom.

2.3.1.2 Varieties

The variety of software applications and components that are available to collect, format, translate, and republish XML feeds is a testament to the flexibility of the format and has shown the usefulness of presentation-independent data.

2.3.1.3 Web-based

Web-based aggregators are applications that reside on remote servers and are typically available as Web applications such as Google Reader or Bloglines. Because the application is available via the Web, it can be accessed anywhere by a user with an Internet connection.

More advanced methods of aggregating feeds are provided via AJAX coding techniques and XML components known as Web widgets. Ranging from full-fledged applications to small fragments of code that can be integrated into larger programs, they allow users to aggregate OPML files, email
services, documents, or feeds into a single interface. Many customizable homepage/portal implementations such as iGoogle, Live.com, My Yahoo!, and Pageflakes provide such functionality.

In addition to personal aggregators, *planet* sites are used by online communities to aggregate community blogs in a centralised location. Such sites are named after the Planet aggregator, an application designed for this purpose.

### 2.3.1.4 Client software

Client software aggregators are installed applications designed to collect Web feed subscriptions and group them together using a user-friendly interface. The graphical user interface of such applications often closely resembles that of popular e-mail clients, using a three-panel composition in which subscriptions are grouped in a frame on the left, and individual entries are browsed, selected, and read in frames on the right.

Software aggregators can also take the form of news tickers which scroll feeds like ticker tape, alerters that display updates in windows as they are refreshed, web browser macro tools or as smaller components (sometimes called plugins or extensions), which can integrate feeds into the Operating System or software applications such as a Web browser.

### 2.3.1.5 Client Libraries

Many programming languages have libraries that are able to download, process, generate and upload RSS feeds. Perl for example has several libraries in the XML::RSS name space of CPAN.

### 2.3.2 RSS Readers or "News Aggregators"

- **News Is Free** is an excellent example of a long-standing web-based news aggregator. Using the free service, you can create customized "pages" for different topics, then have headlines from various resources automatically filled into those pages.
- **Radio UserLand**, popularly known as a blog-building tool, is also another long-standing news aggregator. Enter the URL of a news feed, and it will be added to your personal list.
- **FeedReader** is a small, free software-based tool that I downloaded and tested. As with Radio
UserLand, enter the URL of a feed, and headlines will be brought back and made viewable within the application.

- **NewsGator** is an reader that works within Microsoft Outlook, which looked promising to me, since my life largely revolves around Outlook. The only reason I didn't try it is that I work off a modem and needing to download the 20MB of Microsoft .NET framework files required would have taken forever.

- **NewsMonster** was another tool that looked promising, especially the ability to detect spam in RSS and to rate articles for quality and use ratings provided by others. Unfortunately, it's not for Internet Explorer users, working only with Mozilla 1.0 or Netscape 7.0 and higher.

- **Snarf** is supposed to be a download-free RSS reader for Internet Explorer. I could get the Snarf window to open in my browser and add news feeds. Unfortunately, I couldn't see any way to actually view these feeds, once added. Perhaps others will have more luck.

- There are plenty of other news aggregators beyond the ones I've named, and I'd encourage you to check out some comprehensive lists of these programs. Aaron Swartz maintains a nice, short list of RSS readers, and Radio UserLand provides another short list. John Abbe has made a giant list of RSS readers which looks to be regularly maintained. You'll even find readers for PDAs classified on it. A similar list is offered by Haiko Hebig.

- Finally, how about the web's major directories? What RSS readers do they list? LookSmart doesn't provide a category for these, while Yahoo lists only a dozen. So, head ye over to the Open Directory's RSS News Readers category, which has 40 listings.

### 2.4 Applications currently working using RSS technology


  An online utility that can convert RSS feeds into PDF files.

- **RSS: Your Gateway To News & Blog Content**

  This article by Danny Sullivan of Search Engine Watch takes a look at many of the techniques, methods, and tools used for RSS syndication.
2.5 RSS: Uses and Applications No Tradicionals

The popular use of RSS generally is considered only limited to publish and subscribe headlines of news and Blogs, recently the Internet community is finding interesting and news ways to find application for this interesting technology. In this chapter will be explained some of the most interesting projects that are working now.

Could be for business, marketing, RP, competitive intelligence, security protection, vide entertainment or to track downloads of your newest product, RSS is a innovator technology which could be applied to thousands of ideas.

Here there are some of the most popular:

2.5.1.1 Monitoring Newsgroups/Usenet

Google groups is a community and service of group discussion online free that offers the most complete file of the web of the sending Usenet. The Groups of Google can be monitored using feeders RSS. For monitoring specific groups of news simply add "/feed/msgs.xml" at the end of the URL of the Group of Google. The new URL adds the reader of the group of news and receives updates whenever a new sending appears.

Examples:

Other services oriented to the community online and collaboration groups online that are now supporting RSS are:

- **YahooGroups**
  http://www.yahoogroups.com/

- **SocialText**
  http://www.socialtext.net/

- **Near-Time**
  http://www.near-time.com/
2.5.1.2 Shared Calendars:

Construct a calendar for public or private use and receive updates of calendar using an RSS feed.

The shared calendars can be created as personal follow-ups to pay bills or social events. The calendars that can be used for specific groups like schools, sports events or of the industry. A shared calendar can be totally a syndicate and shared using feeders RSS.

Examples:

- **CalendarHub**
  
  [http://www.calendarhub.com](http://www.calendarhub.com)
  
  Example Feed - [http://www.calendarhub.com/rss/21882/month](http://www.calendarhub.com/rss/21882/month)

- **RSS Calendar**
  

- **Zimbra**
  

2.5.1.3 Weather:

Enter a zip code and get weather information for a specific region. Click on the icon and add the orange RSS feed to your RSS reader. Every day, when you specify, receive reports of local regional climate as they are issued. The RSS feeders can be customized to live local weather conditions, local radar, trends pollen forecasts regional video with expert commentary and details of the extended forecast.

Examples:

- **Yahoo Weather**
  
  [http://weather.yahoo.com/rss/](http://weather.yahoo.com/rss/)
• Weather.com
  http://www.weather.com
• RSS Weather
  http://www.rssweather.com

2.5.1.4 Classifieds (Houses/Jobs)

Is intrigued about the availability of specific employment opportunities in industries specific? Receive notifications when vacancies become available. The RSS feeders can be created by region ranges wages or jobs.

Examples:

• Craigslist
  http://www.craigslist.org/
  Craigslist Example Feed
• RSSJobs
  http://www.rssjobs.com
• CareerChannel
  http://www.careermag.com/MKT/RSS/

2.5.1.5 Ego Searches

A egosurfer is someone who surf the Internet in search of his own name, to see if there are articles on. Ego searches are free and simple searches designed to monitor blogs and news websites in search for mentions of your company, product, competitors or other specific keywords. Conducting ego searches not only allows you to stay informed, but also maintain a strategic advantage over competing companies. A number of new Internet services are freely available that make these 'ego searches' are straightforward and uncomplicated.

Many of the services allow you to create a feeder based on keywords. Every time new information containing the keyword appears in the resources sought, the RSS feed is updated.
Examples:

- **Egosurf**
  
  http://www.egosurf.org/

- **Monitor This**
  
  http://uclan.info/depot/monitorthis/

- **Technorati Watchlist**
  
  http://www.technorati.com

- **More ego-searching tools**
  
  http://www.small-business-software.net/ego-searches.htm

### 2.5.1.6 Tracking Shipments:

Receive notifications via RSS feeders when packages are sent. Personalized RSS feeders can be created to monitor the progress of the shipment. Track the status of packages sent through the U.S. Postal Service, UPS, DHL and Federal Express.

Examples:

- **Simple Tracking**
  
  http://www.simpletracking.com/

- **Track2RSS Project**
  
  http://track2rss.sf.net/

- **PackTrack**
  
  http://www.packtrack.com/

### 2.5.1.7 Offers and Special Stamps:

All love a bargain, RSS feeders promoting discounts depend on the time factor or offers are becoming increasingly popular. Restaurants sent special lunches while the travel companies submitted bids last minute to potential travellers using RSS feeders.
Examples:

- **Coupons.com**
  
  http://www.coupons.com

- **DealNews**
  
  http://deanews.com/rss.html

- **HotDealsClub**
  
  http://www.hotdealsclub.com/

2.5.1.8 Monitoring online auction houses:

Items on eBay and other major online auction houses depend on the time factor and few individuals have enough time to make daily checks of specific products being discarded. Create an RSS feed using keywords to monitor products as they are available for auction.

Examples:

- **Bidding Tools**
  
  http://www.freebiddingtools.com

  Example Feeds

- **AuctionMonitor**
  
  http://www.auctionmonitor.net/

- **RSSauction**
  
  http://www.rssauction.com/

2.5.1.9 Monitoring prices:

Monitoring allows you to track prices and monitor changes in price of products and niche markets. Still in its infancy this niche market certainly will become much more demanded.
Examples:

- RSStalker
  http://www.rsstalker.com/
- FuelWatch
  http://www.fuelwatch.wa.gov.au/...

2.5.1.10 Podcasting and Podcatching:

I was thinking in to don't list that here because nobody wants to hear more about podcasting, but then I had to include it here for many applications no so well know, as a reference to those who had not yet been exposed to this widespread and effective use of RSS. "Podcasting is the distribution of audio or video files, such as radio programs or music videos over the Internet using RSS or Atom syndication to be heard on mobile devices and personal computers.

The term podcast, as "radio" can mean both the content and the method of shipment. Web sites podcasters can also offer direct download of their files, but the subscription feed new content automatically sent is what distinguishes a podcast from a simple download or streaming in real time. Generally, the podcast offers a kind of "show" with new episodes either sporadically or planned intervals: daily, weekly, etc.. In addition to this, there is a network that offers multiple podcast shows of the same feeder."

Podcasting is to enable you and me, listening to our favourite programs audio and video wherever, whenever and with the device you want.

Examples:

- iTunes
  http://www.apple.com/itunes/
- Odeo
  http://www.odeo.com/
- ZiePod
  http://www.ziepod.com/
2.5.1.11 Video Edition:

Thanks to SSR ordinary people collaborative video editing may soon become a reality for many. Kent Bye has been working on this idea as well as the guy behind some of the most updated for the new video editing sites that have been launched recently. RSS is a key component of all these tools that help those who are working on an item or a specific project, to keep abreast with new clips sent to work with them.

Examples:

- **Jumpcut**
  
  http://www.jumcut.com/

- **Eyespot**
  
  http://www.eyespot.com/

2.5.1.12 The Discover Discoveries from Its Friends

- **Delicious**
  
  http://del.icio.us/

- **Digg**
  
  http://www.digg.com/

- **Reddit**
  
  http://www.reddit.com/
2.5.1.13 Newsmastering:

Newsmastering emerging is a new application in which a large number of RSS feeders, enhanced by specific filters are combined to create thematic information feeders’ high value. It can be called radars or when covering news new radars content.

Examples:

- **MySyndicaat**
  

- **Newsgator**
  

2.5.1.14 Social Networking for music lovers

Find RSS what their friends are listening, crawled new albums, songs and artists.

**Examples:**

- **Mercora Radio 2.0**
  
  [http://www.radio2-0.com/msearch/index2.jsp](http://www.radio2-0.com/msearch/index2.jsp)

2.5.1.15 Swarming

The term swarming is applied to a fish, birds and insects, and describes a group of animals that their bodies are similar size and orientation, generally moving in the same direction (source: Wikipedia).

New online services are emerging that will enable small large groups of people are kept synchronized even though they were not in physical contact.
They invariably use SMS text messaging, e-mail and RSS through its channel of communication to disseminate more quickly and effectively to their interconnected groups.

2.5.1.16 X-Events

The X-Events are a new and more effective way to conduct live events that extend to well before and long after the event itself. Through the use of new media technologies including RSS, blogs, P2P, personal media aggregators, forums, wikis and other tools, the organizers of the event in the near future will be able to win mountains of money it understands that the name of the game is in the event extended, interactive in which the live event (online or not) is only a part of the various components and services offered.

Examples:

- EventMingle
  http://www.eventmingle.com/

2.5.1.17 Video Search

- Dabble
  http://www.dabble.com/

2.5.1.18 Bug reporting

- Version Tracker
  http://www.versiontracker.com/faq/rss/

2.5.1.19 Crime Monitoring

- Chicago Crime
  http://www.chicagocrime.org/
2.5.1.20 RSS Mashups

Instant messaging using RSS

- (http://battellemedia.com/archives/002176.php)

2.5.1.21 BroadCatching

(RSS + P2P in order to download fast a custom tv)


2.5.2 RSS is still very young:

Both publishers and consumers are actively seeking ways to capitalize on this unique and powerful technology of sending content and further uses of RSS feeders that are emerging daily.

RSS may soon become an invisible technology accionando hundreds, but thousands, of different applications and uses.

2.6 Objectives

In this chapter it will be analyzed the different objectives that want to be achieved with the elaboration of this project and the functional objectives that the program should accomplish to achieve the main objectives.

2.6.1 Main objectives of the project

The proposal of this project is to generate newspapers for the users obtaining news from different sources. It can be very useful for users that usually look for a specific kind of news in different newspapers or for people that want to publish his newspaper for a group of people adding own news and sending it automatically in a easy way.
So the main objective that the project must achieve is to create newspapers with the requirements of the users and obtaining a perfect selection of the news following the criterions of each of the users.

At the same time the user interface has to be intuitive and easy to use for the users in order to not being difficult to use the application and so everybody can use it easily.

Another objective is that the user can have a newspaper library with all the newspapers created for him until the moment with the possibility of sending them or just read when necessary.

2.6.2 Functional objectives of the project

To achieve the main objectives of the project it needs different functionalities that will be briefly explained here. Each big group of functionalities has different smaller functionalities that will be detailed in the specification chapter.

First of all, the application needs a registration system for the users because only registered users could access to the newspaper service. So there will a registration functionality, a modify profile functionality and a remove profile functionality.

Another big group of functionalities is the related with the newspaper creation. Each user will be able to create different kinds of newspapers so there will be a create newspaper functionality where the user will give the criterions for the news he want in the newspaper. A modify criterions and remove newspaper functionality will exist to modify and remove an existing newspaper respectively. Finally, the user will be able to add or remove emails from a list of people that he/she wants to send the publications of the newspaper.

Before creating the finally publication, the user will be able to check the preliminary newspaper given from the website an remove news that he doesn’t want to publish and to add more news redacted by him or taken from other sources.

After that, the user can publish the newspaper in different ways, depending in how the newspaper was created. One way is creating a pdf file or just an html page with the newspaper. Both options can be
selected at the same time and if the user wants, the newspaper will be sent to the list of emails he created.

The last functionality for the users will be the possibility of checking all the publications created until the moment so there will be like a newspaper library for that.
3 Specification

In this chapter of specification it will be analyzed deeply the functional requirements of the project. It will also be presented the non-functional requirements of the application. Those non-functional requirements are some determining factors that the application of the program must accomplish. It will also be analyzed the chosen technologies and the development environment for the elaboration of this project. In this chapter it also will be presented an “UX Model”, that will give us a first global view of the external structure of the application and will allow us to design the internal structure of the application and the database design.

3.1 Functional requirements

The functional requirements make reference to the features of functionality that the application should provide. Although in the objectives chapter the different functional requirements were exposed briefly, now those will be deeply detailed to explain perfectly each one of the program functionalities.

To show graphically the application functionalities it will be used the use cases notation. There it will be shown the different actors that interact with the application.

Functional requirements of the system:

- User registration
  - New profile
  - Modify profile
  - Remove profile

- Newspaper management
  - Create new newspaper
  - Modify criterions
  - Remove newspaper
  - Add/Remove emails
• Modify newspaper
  o Add new
  o Delete new

• Publish newspaper

• Read newspaper library

3.1.1 User registration

This group of use cases contains all the functionalities related with the user registration system and the profile management of each user.

The first functionality will be New profile. As it was explained before, the users that would like to enjoy the News Hunter services will have to register previously in the website. In this use case there will be different steps. First of all the user will introduce his personal details and a Captcha Security Code to be sure that the registration is done by a human person and not a robot. After that the user will receive an email with a link to activate the new account just to check that the email introduced is in use. After that the registration finishes and the user is invited to create his first newspaper.

Another functionality of this group will be Modify profile. There the user will have the chance to change his personal details anytime he wants. That’s useful because the user will be able to change his email or personal data in case that there’s any change.

To finish with this group, it will exist the option for a user to remove his own profile. It will be possible with the use case Remove profile that will allow the users to delete all his newspapers and data from the website.

3.1.2 Newspaper management

This group of use cases contains all the functionalities related with the newspaper management giving to the user the opportunity of create one custom newspaper and decide the style of it.
The first functionality will be *Create newspaper* and will give to the user the chance to create a new newspaper in his account. The user will decide the name of the newspaper and a description for it. After that a list of different sources will be suggested to the user in order to choose the ones that are interesting for him. At the same time it will be possible for him to add new sources if it’s necessary. After choosing the sources, the user will decide the criterions that he wants for the news that will appear in the newspaper. To finish the newspaper creation, the user will choose the format of the newspaper and will be able to define a list of emails that would receive the newspaper’s publication. After that the newspaper will be created.

Another three less important functionalities will appear in this group. First of all, *Modify criterions* that will allow the user to change the news criterions for the newspaper created by him. The second one will be *Remove newspaper* that as its name shows, will allow the user to remove the newspaper that he doesn’t want to receive anymore. To finish *Add/Remove emails* will allow the user to modify the list of emails that will receive each newspaper.

### 3.1.3 Modify Newspaper

This group of use cases contains all the functionalities related on the edition of the newspapers. After News Hunter creates the newspaper, the user will be able to check if the news that appears there are the ones he wants to read. If not the user has two different possibilities.

First of all he would find the functionality *Add new* that will allow the user to write his own news in the preliminary version of the newspaper. That’s useful for users that want to publish own newspapers or just want to add more information in the definitive newspaper.

The second functionality is called *Delete new* and will give to the user the chance to remove the news that he consider not important from the preliminary version of the newspaper. That can be useful to remove new that the user doesn’t want to read or to remove news that by different reasons appears in the newspaper but are not related with the criterions chosen by the user.

After the review of the user to the newspaper using those two functionalities, the newspaper will be prepared to publish.
3.1.4 Publish newspaper

This functionality will allow the user to publish the definitive version of the newspaper after its revision. When the user decides to publish the newspaper, it will be sent in different ways depending on how the user created it. One possibility is to send the newspaper to the list of emails that the user decided at the creation of the newspaper in HTML format, PDF file format or both. The other possibility is that the user receives the newspaper in the same possibilities of format that was commented before.

After publishing the newspaper, it will be stored in the newspaper library and it won’t be possible for the user to modify it anymore.

3.1.5 Read newspaper library

That’s the last functionality of the application. After having done several publications, maybe the user wants to read again an old publication or wants to send it again to the list of people or to another group of people. That will be possible with the Newspaper library functionality that will store all the newspapers created by the users of the website.

3.2 Non-functional requirements

A project of these characteristics not only has requirements about the functionalities to develop. Further than define what the application will do, it must be defined in the specification where it will be done, it means, define the environment where the application will be used and where it will be developed.

The non-functional requirements are those application requirements non-related with the functionalities of it but that conditions the application and its development and execution environment. In this way, it’s defined what will be needed to use and develop the program and what the program will provide besides its functionalities. In this section it will be detailed the compatibilities, incompatibilities, needs and some characteristics of the application that give shape to the group of non-functional requirements of the project.
The first of the non-functional requirements to comment is that the News Hunter engine will be developed to work in a Windows environment as a Windows Service. The database of the application will use SQL Server of Microsoft and the server that will store the application will use Apache HTTP Server to store the web interface that users will use to create their newspapers and more. It means that the server will work under Microsoft technologies and open source technologies but the users will be able to use the application in a web environment regardless of the operating system they use. The decision of using Microsoft software to develop the engine of the application is taken by the programmers because they are experienced in Visual Studio environment and their workplaces are using Windows environment. At the same time they think that .Net technologies are one of the most used nowadays so they think is worth to use them. At the same time it’s been chosen Apache HTTP Server because is the most used web server and the programmers are also experienced to work in this environment.

As it was commented before, the users will be able to use the News Hunter software just having an internet connection and a web browser. That’s the best possibility nowadays because each day there’s more people using different operating systems instead of Windows.

Another thing to comment is the technologies that will be used to develop the application. As it was commented before, it’s been decided that the development of the engine of the application will be done using C# programming language under the .Net technologies. That decision is taken because it’s a powerful language that each day is having more users. The web interface will be programmed using PHP language because it’s a common combination to use PHP and Apache Server and is a language that has a lot of libraries oriented to the web programming.

To finish with this section, just comment that the users that want to receive the newspapers in PDF file format will have to install a PDF reader in their computer.

3.3 Development environment

The development environment where will be done the specification, design, implementation and tests of the application of the project is decided by the needs of the program to develop and in big part by the functional and non-functional requirements of the application. It must be found an environment that allows us to reach the objectives of the application, providing it with the needed functionalities, in
the context that provide the non-functional requirements. In every project it must be controlled two factors, the hardware environment and those related with the software field.

In this project it has no really importance for the development the hardware environment because it’s not needed special or specific hardware, just a current and ordinary computer.

### 3.3.1 Software development environment

In this section it will be detailed the development environment and the software that will be used to make the project. Because the development of the software will be, in part, with Microsoft technologies, the operating system used for it will be Windows. There are several Apache HTTP Server versions to work under Windows so it’s not a problem.

Some of the tools needed would be a C# compiler and a text editor to program the application so it will be used the Microsoft Visual Studio 2005 that includes compiler for .Net languages and lots of libraries for its languages and at the same time an editor. At the same time is one of the most used technologies nowadays to create software.

The other programming language that will be used to program the web interface will be PHP. This languages doesn’t need a compiler because is a interpreted language. To program with PHP it will be used a text editor like EditPlus for Windows and Smultron for MAC.

To make the system information persistent it will be used a database. After using Apache HTTP Server and PHP language to develop the project, the most logical option is to use MySQL database to store all the information related with the users and the newspapers information. That will generate what is called a WAMP System that means creating an internet infrastructure using the tools: Windows, Apache, MySQL and PHP.

To make the design of the project it will be used Rational Software that provides different tools to expand the use of modern software engineering practices and design the structure of the project. Rational Rose will be the Rational Software used for the design of this project.
Another thing to take into account will be the generation of PDF files for the newspapers. In order to achieve that it will be necessary to use an special library that allow us to generate PDF files by PHP language.

At the end the chosen environment for the development of the project is the following:

- Operating system Windows XP
- Visual C# language
- Visual Studio 2005 development platform
- MySQL Server
- Apache HTTP Server
- PHP language
- Rational Rose
- Smultron editor
- Editplus editor
- PDF File generation library

3.3.2 Technologies analysis

In the last section was done a briefly review of the technologies that will be used to develop the project. In the following section it will be analysed deeply each of those technologies.

C#

C# (C Sharp) is an object-oriented programming language developed by Microsoft as part of the .NET initiative and later approved as a standard by ECMA and ISO. Anders Hejlsberg leads development of the C# language, which has a procedural, object-oriented syntax based on C++ and includes aspects of several other programming languages (most notably Delphi and Java) with a particular emphasis on simplification.

The decision of using this programming languages is because is one of the languages that you can use with .Net platform and the programers are more experienced with C# than with another one so it doesn’t require an extra learning. At the same time, the language is enough powerful to reach the objectives of this project.
Visual Studio 2005

Microsoft Visual Studio is Microsoft's flagship software development product for computer programmers. It centers on an integrated development environment which lets programmers create standalone applications, web sites, web applications, and web services that run on any platforms supported by Microsoft's .NET Framework (for all versions after Visual Studio 6). Supported platforms include Microsoft Windows servers and workstations, PocketPC, Smartphones, and World Wide Web browsers.

Visual Studio 2005, codenamed Whidbey (a reference to Whidbey Island in Puget Sound), was released online in October 2005 and hit the stores a few weeks later. Microsoft removed the " .NET " moniker from Visual Studio 2005 (as well as every other product with .NET in its name), but it still primarily targets the .NET Framework, which was upgraded to version 2.0. Visual Studio 2005's internal version number is 8.0 while the file format version is 9.0. Microsoft released service Pack 1 for Visual Studio 2005 on 14 December 2006.

Visual Studio 2005 was upgraded to support all the new features introduced in .NET Framework 2.0, including generics and ASP.NET 2.0. The IntelliSense feature in Visual Studio was upgraded for generics and new project types were added to support ASP.NET web services. Visual Studio 2005 also includes a local web server, separate from IIS, that can be used to host ASP.NET applications during development and testing. It also supports all SQL Server 2005 databases. Database designers were upgraded to support the ADO.NET 2.0, which is included with .NET Framework 2.0. C++ also got a similar upgrade with the addition of C++/CLI which is slated to replace the use of Managed C++.

Other new features of Visual Studio 2005 include the "Deployment Designer" which allows application designs to be validated before deployments, an improved environment for web publishing when combined with ASP.NET 2.0 and load testing to see application performance under various sorts of user loads.

Visual Studio and .Net technologies are one of the most used nowadays for software creation so thats the reason why it will be used to create this project.
MySQL

MySQL is a multithreaded, multi-user SQL database management system (DBMS) which has, according to MySQL AB, more than 10 million installations. The basic program runs as a server providing multiuser access to a number of databases.

MySQL is owned and sponsored by a single for-profit firm, the Swedish company MySQL AB, which holds the copyright to most of the codebase. This is similar to the JBoss model. It is dissimilar to the Apache project, where the software is developed by a public community and the copyright to the codebase is owned by its individual authors.

The company develops and maintains the system, selling support and service contracts, as well as proprietary-licensed copies of MySQL, and employing people all over the world who collaborate via the Internet. MySQL AB was founded by David Axmark, Allan Larsson, and Michael "Monty" Widenius. The CEO is Mårten Mickos.

The MySQL Company also sells another DBMS, MaxDB, which is from an unrelated codebase.

- MySQL was first released internally on May 23, 1995
- Windows version released on January 8, 1998 for Windows 95 and NT
- Version 4.0: beta from August 2002, production release March 2003 (unions)
- Version 4.1: beta from June 2004, production release October 2004 (r-trees, subqueries, prepared statements)
- Version 5.0: beta from March 2005, production release October 2005 (cursors, stored procedures, triggers, views, XA transactions)
- Version 5.1: currently pre-production (since November 2005) (event scheduler, partitioning, plugin API, row-based replication, server log tables)

MySQL is popular for web applications and acts as the database component of the LAMP, MAMP, and WAMP platforms (Linux/Mac/Windows-Apache-MySQL-PHP/Perl/Python), and for open-source bug tracking tools like Bugzilla. Its popularity for use with web applications is closely tied to the popularity of PHP and Ruby on Rails, which is often combined with MySQL. PHP and MySQL are