

Fachhochschule Oldenburg  
University of Applied Sciences Ostfriesland  
Wilhelmshaven



---

# Parsing from RSS feeds to MHP-Television systems

---

**Diplomarbeit**  
Telecommunication (sound & image)

Ignasi Corretger Casas

Erstprüfer: Prof. Dr. Jörg Thomaschewski

Zweitprüfer: Prof. Dr. Gerhard Kreutz



# **Erklärungen**

## **Rechtliche Belange**

- Die Diplomarbeit entstand an der Fachhochschule Oldenburg / Ostfriesland / Wilhelmshaven.
- Soweit meine Rechte berührt sind, erkläre ich mich einverstanden, dass die Diplomarbeit Angehörigen der Fachhochschule Oldenburg / Ostfriesland / Wilhelmshaven für Studium / Lehre / Forschung uneingeschränkt zugänglich gemacht werden kann.

## **Eidesstattliche Erklärung**

Hiermit erkläre ich an Eides Statt, dass ich die vorliegende Diplomarbeit bis auf die offizielle Betreuung selbst und ohne fremde Hilfe angefertigt habe und die benutzten Quellen und Hilfsmittel vollständig angegeben sind.

19.06.2006

---

(Ignasi Corretger Casas)



Ich möchte mich bei allen bedanken, die mir in Emden geholfen haben und mich in meinem Projekt unterstützt haben. Ich bedanke mich sowohl bei den Dozenten, die zu diesem Projekt beigetragen haben \_vor allem bei Prof. Thomaschewski\_ als auch bei den Leuten, des Erasmus-Programms. Der Austausch ist eine wunderschöne und einmalige Erfahrung. Ich bedanke mich auch bei meinen Freunden aus Barcelona und insbesondere bei meiner Familie. [Deutsch]

Voldria agrair amb l'èxit d'aquest projecte totes les diferents persones que m'han ajudat en la meva estància a Emden. Des de els professors que han intervengut en el projecte en especial al Pf. Thomaschewski, fins a totes les persones que han fet que aquest erasmus sigui una experiència única. Als meus amics de Barcelona i en especial la meva família. [Catalan]

I want to thank with this success project to all different people who had been helping in my stay in Emden. Beginning with all the teachers that take part in my project in special Pr. Thomaschewski, until all people who made this Erasmus an unique experience and incredible. To my Barcelona friends and special to my family. [English]

Vorrei ringraziare col successo di questo progetto tutta la gente che mi ha aiutato col mio soggiorno a Emden. Da i professori che hanno partecipato col progetto, specialmente al Professore Thomaschewski, a tutta la gente che ha fatto possibile che questo Erasmus sia una esperienza unica. Ai miei amici di Barcelona e specialmente alla mia famiglia. [Italian]

J'aimerais tout particulièrement remercier toutes les personnes qui ont contribué à la réussite de mon projet. Avant tout aux professeurs qui ont pris part à mon projet, principalement Pr. Thomaschewski. Ensuite aux personnes qui m'ont permis d'avoir une expérience Erasmus unique et incroyable. Et pour finir à mes amis de Barcelone ainsi qu'à ma famille. [French]

Querria agradecer con el éxito de esto proyecto a todas las personas que me han ayudado en esta estancia en Emden. Desde los profesores que han intervenido en el proyecto en especial al Pf. Thomaschewski, hasta todas las personas que han echo de este Erasmus una experiencia única. A mis amigos de Barcelona y en especial a mi familia. [Spanish]



# Contents

1	Introduction.....	9
1.1	Motivation.....	9
2	Fundamentals of the technologies.....	10
2.1	MPEG fundamentals.....	11
2.1.1	Elementary streams.....	12
2.1.2	Transport streams.....	12
2.1.3	Multiplexing process.....	14
2.2	DVB.....	14
2.2.1	Energy dispersal.....	16
2.2.2	Error correction.....	17
2.2.3	Modulation.....	19
2.2.4	Return channel, interactivity.....	20
2.3	RSS.....	22
2.3.1	Extensible Mark-up Language (XML).....	22
2.3.2	Advantages of RSS.....	22
2.3.3	What is RSS?.....	22
2.3.4	The channel elements.....	23
2.3.5	The items elements.....	25
2.3.6	Reasons of the election.....	28
2.4	MHP.....	29
2.4.1	MHP architecture.....	30
2.4.2	Life cycle of Xlet (application manager).....	31
2.4.3	MHP packages.....	32
2.4.4	Graphical behaviour .....	33
3	Parsing system.....	36
3.1	Solution.....	36
3.1.1	First step.....	36
3.1.2	Second step.....	37
3.1.3	Third step.....	38
3.1.4	Final solution.....	39
3.2	Next Features.....	41
3.2.1	PTSN communication.....	41
4	Conclusions.....	44
4.1	Technical conclusion.....	44
4.2	General conclusion.....	44
5	Bibliography.....	45





# 1 Introduction

In the last years is happening one important revolution, the digitalization of the media. One of the most important media around the world is television system, and has not escaped of this phenomenon. It begun with the standard of the MPEG, which can allows to reduce the information in a dramatical way with lossless.

Next step was the DVB standard which puts ways, methods and systems for the transmission of the MPEG chains. This standard is adapted to the different types of transmission existing around the world like DVB-S (satellite), DVB-C (cable), DVB-H (handles) and DVB-T (Terrestrial television).

With those standards we are ready to introduce new contents in digital support, that allows to send audio and video (primitive television) also digital information too . This new information has derived to put information to create powerful interfaces and at least interactive information thanks to a return channel.

This last characteristic in DVB-T has a special name; MHP, and can let us to program many kind of application run after in set-top box or digital television.

## 1.1 Motivation

I have coursed studies of technical telecommunication speciality in sound and image, this studies has a special approach in the old media and new media. One part of the studies is the new services of the digital television.

This new services are almost new and not very developed with all the community of developers, thanks to be very nearly approach for the expert community, and new standard releases grows every day more and more the options that can be developed.

The realization of this algorithms can also be practical to a company, and have a real application if one broadcaster wants to broadcast the application in it own waves to all the population, instead is possible to make test in a almost "real" situation. We have different degrees of reality (emulation, physical set-top box, chain of communication).

Finally in my country Catalunya will be one of the first regions of Spain which will do the analogue switch-off in 2009. This wants to say this is a very strong opportunity of future, new and with a lot of perspective to be developed and implemented in a huge market, general market and local market. I have to say in Spain the most part of population consume television with terrestrial system, but this MHP system is also applied to other technologies used in other European countries like Deutschland.

## 2 Fundamentals of the technologies

Like all the invents all over the world are made by different parts, also this pieces some times can be divides in other points of study. To have a look with grounds in digital television is right to understand some complex technologies, which all together create a very complex system of sound, image and data distribution.

This diplomarbeit does not have the intention to discover or put more light in some of the points I will explain next. The only intention is to explain the most basis technology which all together construct one part of the digital system television.

- **MPEG** (Moving Picture Expert Group) is a part of the ISO enterprise. It has the mission to create new codecs of video, audio, data and meta-data. All the digital television broadcast use MPEG technology.
- **DVB** (Digital Video Broadcasting) is a consortium that makes standards for the transmission in different ways of telecommunication, like satellite, cable, terrestrial or handle gadgets.
- **RSS** (Really Simple Syndication) is a web feed format. Our project feed from RSS files hosted in remote servers, is commonly used in weblogs and in the last times feeds multimedia content.
- **MHP** (Multimedia Home Platform) is a middleware standardised by the DVB consortium to stablish a system who manage all the interactive java-programs broadcast and received by terrestrial or cable television, and processed by set-top box or digital television.

## 2.1 MPEG fundamentals

The new features users will can enjoy new changes between analogue technology and digital technology, specificity this changes are involved in meanings like:

- More channels: Specificity in communication in satellite and cable technology.
- Much better quality of resolution: with the HDTV (high definition television) and without interference.
- Better quality of sound: The inclusion of Dolby Surround and new standards.
- Additional services and new applications (mobiles services).
- Wide screens: Possibility to watch tv in 16:9 format.
- Transmission of multi-language and multi-subtitles.
- EPG: Electronic program guides.
- Interactive applications relational with the program it's emitted in the air.
- Personal television: pay-per-view.
- Play games in an own mode or multi-player.

The best screen relation will be thanks to the new increase of resolution with the change of standard television to the new HDTV. This change is value in a resolution for 1920 x 1080 pixels in compare with 720 x 480 that has the standard television.

Furthermore we can enjoy better and wider services and applications, this is one of the novelty more important to face with star-up of this new system. This new applications we can classify in differents ways.

One example of *non-download* can be de subtitle showed during a film in the television this application let the deaf audience follow one film without audio. In the other hand we have some example of download application like EPG, this application will be *download* and then charge in the set-top box and will show information to the consumer about the programming, trailers, etcetera.

This is the first step of the new technologies that can come in the future, these ones can be developed when are a number of increasing set-top boxes and a more bigger market of consumers, then the logical of market will make appearance set-top box more powerful processing and more powerful with memory storage, in addition there is another responsibility of the broadcasters should put more bandwidth for the interactive applications this also has relation with governments who control de radio electric spectrum and can allow to put more bandwidth in every multiplex.

The change between analogue technology to digital technology is a process that has begun for 10 years ago. A lot of private companies usually have been operated with proprietaries middelware standards, here we can make a big difference between proprietary market and open standards market . The part of proprietary standards is a lot common in satellite transmission and cable operators opposite there is the part of open standards like MHP more implanted in TDT (terrestrial digital television) transmission in the recent years, before there was some beginners enterprise were using some private code for TDT transmissions, at least they are reconverting to the new sectoral standard.

## Parsing from RSS feeds to MHP-television system

### 2.1.1 Elementary streams

An MPEG encoder will produce a single stream of audio or video content. This is known as an elementary stream (ES), the most basic component. At this point, we need to know that each ES is one piece of video or audio content. If we talk in television meanings we will make the distinction of ES video content and ES audio content.[Msm05]

An ES is a stream of information. For this reason, the content of the stream is often split into different packets. Every packet will include:

- Time etiquette.
- Stream ID that identifies the type of content and how it is coded
- Some synchronization information.

There is a variable packet size it depends on many variables, but it is normal a few kilobytes. When an ES has introduced into packets we can change the denomination to a packetized elementary stream (PES).

PES can be multiplexed into a ready form for transmission, the standard used in digital television is MPEG-2 transport stream, and it is different in comparison with the MPEG used in DVD for example. Now our digital television packets will have more information because we are in a difficult way to transmit more than DVD (than it is a QER, Quasi Error free).[Msm05]

### 2.1.2 Transport streams

Probably the biggest difference between the two types of streams is that a program stream will carry just one MPEG-2 program. A program in MPEG is a group of video and audio PES that will be showed as stream piece of content. This may include multiple camera angles or multiple audio languages.[Msm05]

A transport stream, on the other hand can carry several programs. For instance a transport stream will usually carry 4l Digital Television channels, at least it depends on the bandwidth dedicate in each multiplex.

When we multiplex more than one service in transport stream, what we do is multiplex the different PES of the different streams. The easy idea is think that every PES will be multiplexed one after one. But it is not like that, it depends on many variables like the bit rate of every PES, this will determining the order in which the multiplexer put the packets into the transport stream.[Msm05]

Like when we multiplex information and we put some different information in the same way, we need some mechanism to tell the receiver how it has to extract in a good way all the information.

Then we can see there is a problem for the receiver determining and identifying a PES in every multiplex. Also the receiver has to know which PES makes every service

## Parsing from RSS feeds to MHP-television system

To identify which packets belong to which PES, each PES packets is labeled with a packet identifier (PID) when it is multiplexed into the transport stream. This is numeric ID that uniquely identifies that a PES within that transport stream.[Msm05]

To tell the receiver how theses PES should be organized into services, we also include some non-audio and non-visual information in the transport stream. This is called *service information*, which is basically a meta-data encoded in MPEG-2 packets and broadcast inside the PES of video and audio data. [Msm05]

It tells the receiver which streams belong to which services, and include information for the viewer about the channels. Although some service information is defined by the MPEG standard, then DVB standard has the option to put more information about the service in a more higher level than MPEG.[Msm05]

The final difference between a program stream and a transport stream is the way data is split into packets and encoded. Transport streams are used in environments which errors are likely to happen (is not Quasi Error Free), in the other hand program streams are used in environments in which errors are much less common like CD or DVD and we can talk of QEF (Quasi Error Free environment). This means that transport stream include much more information to be ensure the result of the recovery information will be in a good way, and will not affect the final recovery PES.[Msm05]

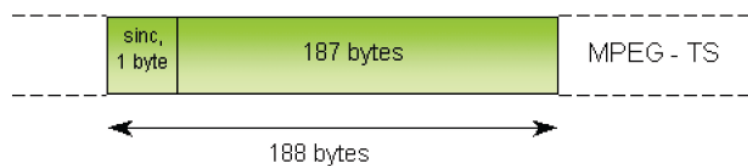


Figure 1: packet of 188 bytes; unpublished lessons [Gir05]

Data in transport streams is split into packets of 188 bytes, which contain some header information (such as the PID of the PES the packet belong to, as well as time stamp information). In addition to the time stamp information encoded is PES packets, each service in a transport stream will include an extra time stamp called the program clock reference (PCR).[Msm05]

MPEG has an stable clock that will be checked every time the receiver has to decode the packets in the correct time. We need to be sure the clock of the receiver is synchronized with the clock of the broadcaster. If we do not have this caution the receiver can decode the data in a bad way, with some problems in the audio and more probably with bad effects in the video de codification, making bad screens.

The solution comes with the PCR which cats like a master clock flor given the service, the receiver can use this to make sure that its clock is running at the correct rate, and will save problems in the decoding process.[Msm05]

### 2.1.3 Multiplexing process

At the simplest level, multiplexing involves to insert and split the various PES into 188-byte transport stream packet, assigning a PID to each PES, and transmitting the packets so that each PES gets its assigned share of the total bandwidth.[Gir05]

Normally, a network operator or broadcaster would assign a fix bit rate to each PES in multiplex, set the MPEG encoder to generate a constant-bit-rate MPEG stream, and let the multiplexer do its work. This traditional use of the bandwidth can include a loss of spectrum, because simple scenes can use less spectrum but there are also other scenes no than simple that needs a more higher bandwidth, then the next step is to make a variable assigner of bandwidth to take more performance of the little bandwidth assign in every case.[Gir05]

The solution of that, adopted by many operators use a technique called statistical multiplexing. The multiplexer (in the emission centre) uses information about the complexity of the scene and will make some processing to determine how many bandwidth more or less is needed to be in the air. Next step will be determine and put in a right order the complex scenes which will take borrow from the light scenes of bandwidth. The result of that will be a make up system of scenes, complex and non-complex.[Gir05]

## 2.2 DVB

By the moment the only thing we have are different packets understood by the multiplexer and the receiver (demultiplexer). What we should look in the transport chain, and it is the more complex part of all the system.

It depends on the way to transmit we should put more controls or not, for the instance cable environment has few errors than terrestrial communication. DVB then has different standards for every way.

One part of the DVB is FEC, this will put iin the digital television tools to make sure the most important part of the information will be received in a good way to be demultiplexed. To make a general idea this is the general chain of a receiver:

## Parsing from RSS feeds to MHP-television system

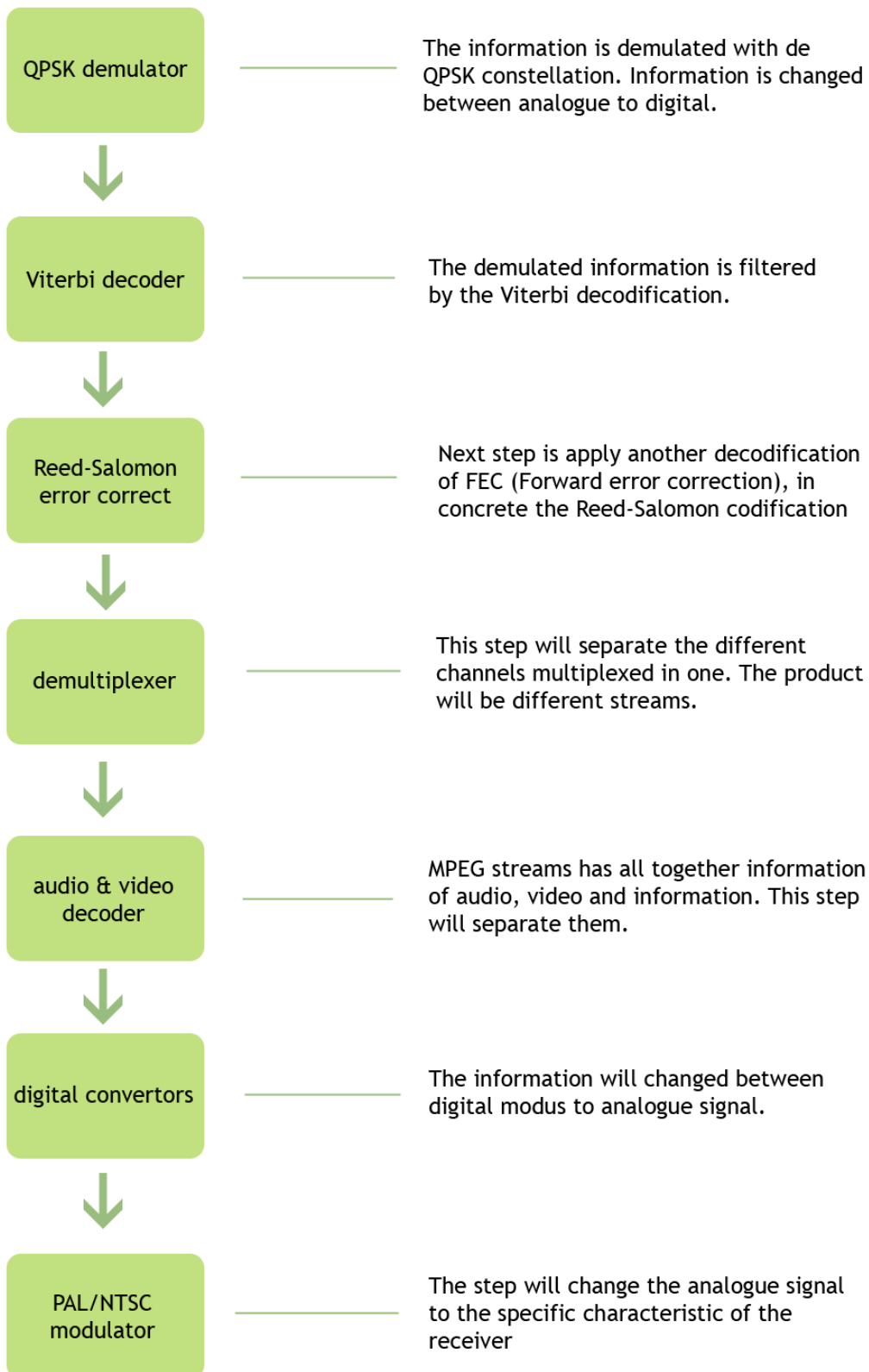


Figure 2: Digital television receiver steps

## Parsing from RSS feeds to MHP-television system

### 2.2.1 Energy dispersal

The first problem we face is one of signal quality. Signals are not transmitted on a single frequency. Typically, they use a range of frequencies, and the power of the signal will vary across the frequency spectrum used to transmit that signal. For a completely random stream of bits, the power would be equal across the entire spectrum but for non-random data such as MPEG-2 transport streams this is not the case. This can cause problems, because some parts of the spectrum may have very high power (and cause interference with other signals), in the other hand others may have a very low power. [Gir05]

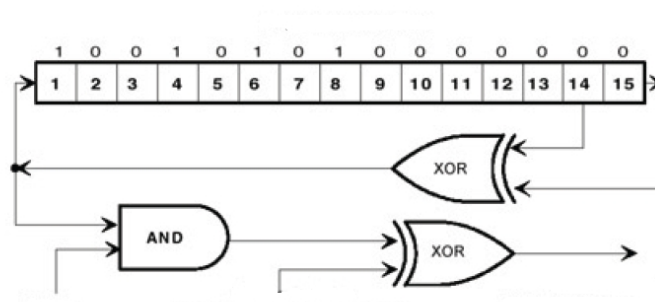


Figure 3: channel codification: unpublished lessons [Gir05]

The transmitter carries out a process called randomization (energy dispersal). This takes the transport stream and uses a pseudo-random number generator to scramble the stream in a way the receiver can easily descramble. The purpose of this is to avoid patterns in the signal and to spread the energy equally across the entire spectrum used to transmit the signal. DVB systems use the randomization algorithm shown.[Gir05]

It is not possible to put working the randomiser all along the signal because we need to make some synchronization with the receiver. Then inside the signal will be a sync byte at the beginning of the signal, and periodically will reset the randomiser in our case it will do it every eight transport packets.[Gir05]

The transmitter will invert the sync byte of the first packet in each group of eight, and when the receiver sees the inverted sync it knows that it should reset the randomiser. Other sync bytes are not randomized, then the de-randomiser should know what it is, sometimes is information to reset in other ways (more than 8 transport packets).[Gir05]

Like comment before every standard of DVB can have a different randomiser, it depends on statistical data. In this case we comment the terrestrial case of randomiser, but cable can has another and different one.



## Parsing from RSS feeds to MHP-television system

### 2.2.2 Error correction

Due to the nature of television transmission systems, every transmission will contain some errors because of radio frequency interference like is comment before. For an analogue system, this is usually not a big problem because the quality of the output will reduce gradually, but we are in the digital version we have little pieces of content then if there is a problem with this little piece the problem will be more bigger than the analogue version.[Gir05]

In an IP network, we can do this using checksums and acknowledgements to retransmit packets if one of them is corrupted. In a broadcast network, this is not possible. There may not be any way for the receiver to tell the transmitter about an error, and if there is, it is no realistic for the communication. [Gir05]

This means that this system needs a mechanism which the receiver can correct errors in the data without needing to resending the corrupt packets that are lost or arrive in a corrupt way. To change that is used a technique called forward error correction (FEC), which builds some redundancy into the transmitted data to help detect and correct errors. Most digital television systems use a technique called Reed Salomon encoding to encode this redundancy and to correct errors where necessary.[Gir05]

The version used in digital television systems adds 16 bytes of data to transport stream packet. This gives a total packet size of 204 bytes for a transport stream that is ready for transmission. And in the receiver can work in a statistical good way.

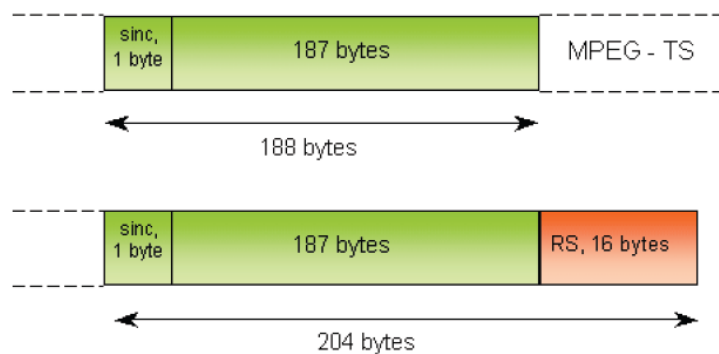


Figure 4: Packet size with Reed Salomon codification; unpublished lessons [Gir05]

Reed-Salomon encoding alone is not enough to solve all of our data corruption problems. Reed-Salomon cannot handle larger errors that corrupt several bytes of data. In the case of 16 bytes of redundancy, it can detect but not correct errors affecting more than 8 bytes. These could be caused by variety of phenomena, such as electrical storms or interference from other electrical equipment.[Gir05]

## Parsing from RSS feeds to MHP-television system

To solve these problems, we use a process called interleaving. This reorders the data before we transmit it, so that adjacent bytes in the stream are not transmitted together. There are several ways of doing this, but one of the most common approaches is to use a two-dimensional buffer to store the data before transmission. By writing data to the buffer in row order, but reading it in column offer, we can easily recorder the content of the stream and reconstruct the original stream at the receiver.[Gir05]

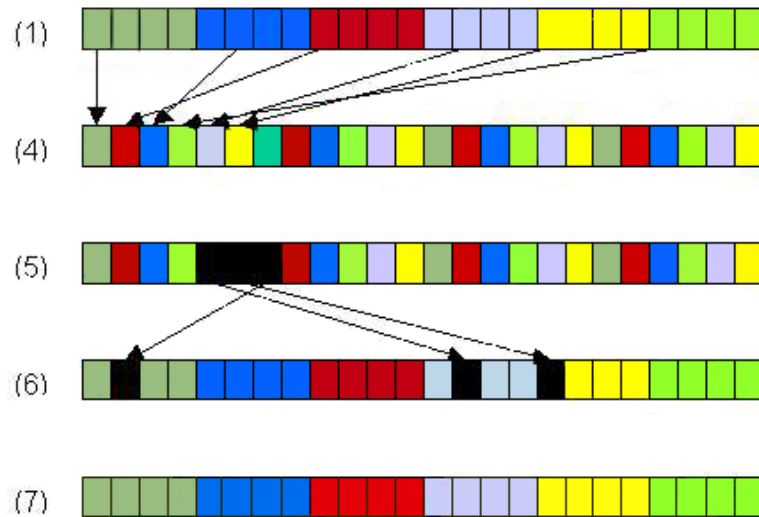


Figure 5: Interleaver representation; unpublished lessons [Gir05]

This process can be modified by introducing an offset between the different rows, so that the algorithm will read byte  $n$  of one row followed by byte  $n-1$  of the next row. This may use less memory than the more straightforward interleaving previously mentioned. Interleaving makes the data much more resistant to burst noise. A burst of corruption in the transmission will be spread across several packets in the reconstructed stream. [Gir05]

The type of coding at the superstructure levels is known as outer coding. To prevent noise from reducing the error-correcting capability of interleaving a layer of coding (called inner coding) is added.

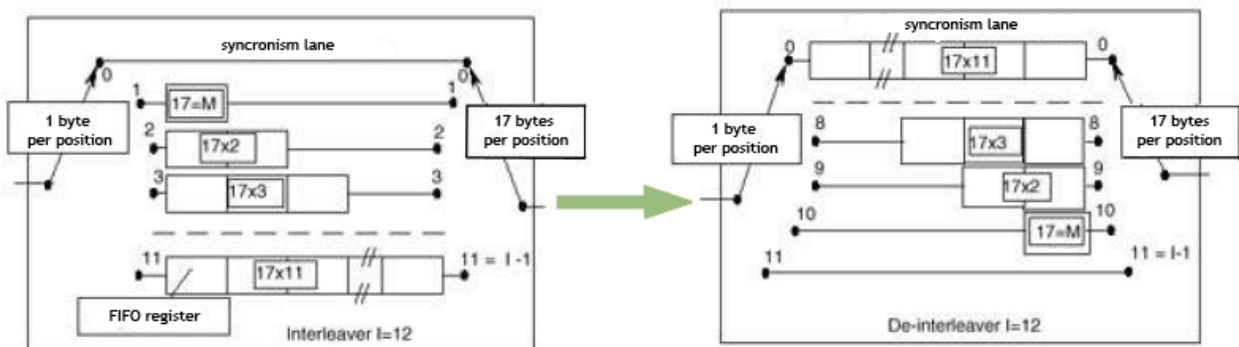


Figure 6: inner coding; unpublished lesson [Gir05]

## Parsing from RSS feeds to MHP-television system

Most inner coding techniques are based on a technique called trellis coding, in which symbols are grouped to form “trellises”. For a group of three symbols, a modulation scheme that stores 8 bits per symbol can store 512 separate values. By using a subset of these as valid values, the network operator can introduce some extra redundancy into the signal. The effect of this is that each symbol may carry fewer bits of data, but for every group of three symbols it is possible to correct one erroneous symbol by choosing the value for that symbol that gives a valid trellis. [Gir05][Msm05]

DVB system are Viterbi coding instead, which is a slightly different algorithm to finding the trellis that most closely matches the received data. Viterbi coding is an extremely complex topic beyond the scope of this book. By default, a Viterbi encoder will generate two separate output bytes for every byte or input, with each output representing the result of a different parity check on the input. This is known as  $\frac{1}{2}$ -rate encoding, because one symbol of actual data is represented by two symbols in the output.[Msm05]

Because this is inefficient, most application use a technique called puncturing, which involves transmitting one for the output bytes at a time in a well-known ratio. This allows the network operator to trade a reduction in error correcting power of an improvement in bit rate. Different networks and different applications will use different puncturing rates, depending on the bit rates they need to achieve and the number of uncorrected errors that can be tolerated.[Gir05]

### 2.2.3 Modulation

Once we have applied inner coding to our bit stream, we are ready to transmit it. The last stage in the process is modulation: converting the digital bit stream into a transmittable analogue waveform.

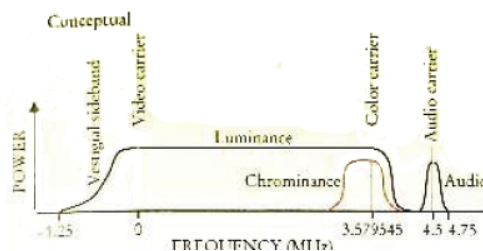


Figure 7: Spectrum of analogue television. [Gir05]

If we used a simple approach to modulate the signal, the amount of bandwidth we would need to transmit each signal would be extremely high. To solve this problems, modulation techniques will normally squeeze several bits of data into each symbol that gets transmitted. This symbol can be coded either by using different amplitude levels to indicate different values or by changing the phase of the signal.

## Parsing from RSS feeds to MHP-television system

The 8-VSB modulation scheme used by terrestrial systems and the QPSK and 8PSK systems use phase modulation to encode either 2 or 3 bits per symbol (by using phases that are 90 degrees apart or 45 degree apart, respectively). These two approaches can be combined to squeeze even more data into a single symbol. This modulation combines QPSK modulation with multiple amplitude levels to support up to 8 bits of data in a single symbol. [Msm05]

QAM and QPSK use two different carriers with a 90-degree phase difference between them (one is in quadrature with the other, and hence the name). In both of these schemes, symbols are represented by combination of both signals. This allows each symbol to carry 4 bits of data. [Gir05]

An alternative modulation scheme the signal across several carriers simultaneously, using a scheme such as QPSK or QAM to modulate data on to each of these. This scheme called COFDM, is used by DVB for terrestrial broadcasts. The inner coding of the bit stream is often closely tied to the modulation technique used, and thus these are usually specified together. [Gir05]

### 2.2.4 Return channel, interactivity

The capacity to offer addition contents in the television shows, allowing the viewer see extra information of the show. It does not matter the way the receiver gets the extra information. In the case of terrestrial television we can get information in two different ways, DSM-CC or carousel information is getting in the ways and thanks to return channel (PTSN or Ethernet connections).

Return channels come in several different ways, depending on the cost of the receiver and the type of network the return channel is connected to, some of the return channel technologies available. Many of these have been tried and tested, although not all are in current use.

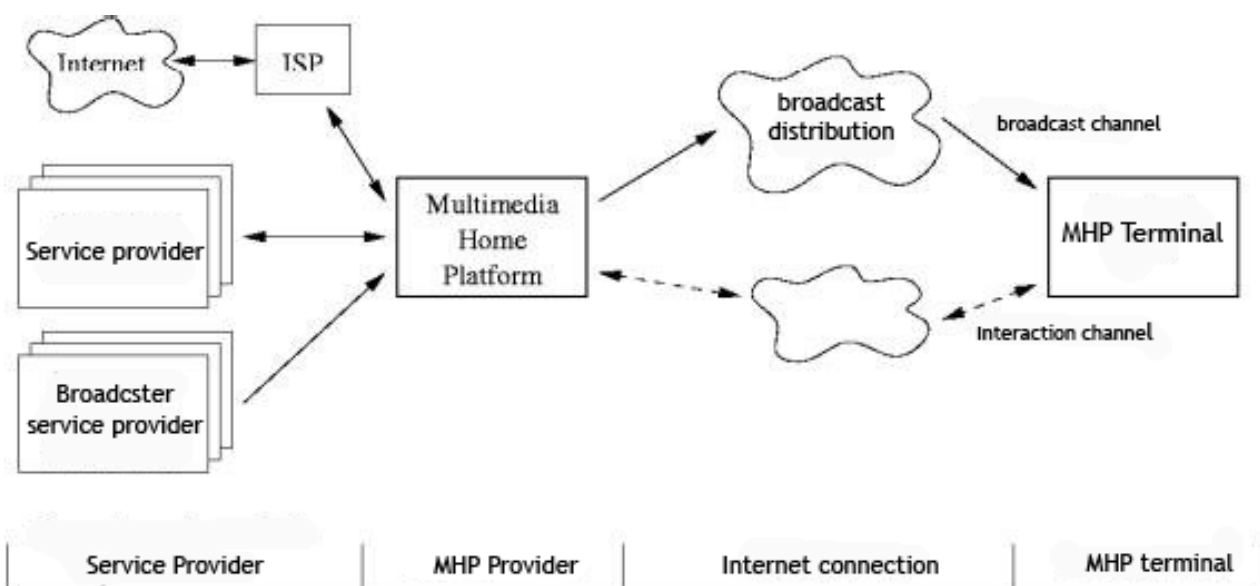


Figure 8: Connection schema of the MHP middleware [Dar06]

## Parsing from RSS feeds to MHP-television system

Decisions about the type of return channel are not just technical ones: they also influence the cost of the receiver and the type of applications that are possible. A receiver that only has a PSTN modem, for instance, will never be able to support some application that need broadband.

Other interesting aspects to take into account is the different kinds of interactivity present nowadays. There are two different: the local and the remote ones. The local interactivity the audience interact with the information save in the receiver. The viewer can access to interactive contain like digital teletext, this is the profile MHP 1.0) and one example of that can be the basic EPG.[Tel06]

There is another option the remote interactive the viewer interact with an external provider of service. The receiver has to be connected with this provider with a return channel (it does not matter which type of connection), this carry out the MHP 1.0 profile and examples of applications can be:

- Advanced EPG: with more options and new formats.
- E-commerce and e-government: Connection with e-shops or institutions.
- Information in carousel: Like flies, stock exchange.

We can also have more types of service, like:

- Information services: Gives independent information of the different programming broadcasted.
- Programming services: Give more information of the broadcasted show.
- Commerce services: Gives the opportunity to send and receive personal information.

<i>Type of service</i>	<i>Local interactivity</i>	<i>Remote interactivity</i>
Information services	<ul style="list-style-type: none"> <li>• EPG: electronic Program Guide.</li> <li>• Digital Teletext.</li> <li>• Games.</li> </ul>	
Programming services	<ul style="list-style-type: none"> <li>• Sport statistic</li> <li>• Advanced information of the show.</li> <li>• Extended news.</li> </ul>	<ul style="list-style-type: none"> <li>• Take part in quiz show.</li> <li>• Polls</li> <li>• Answers and questions for television guests.</li> <li>• Pay-per-view.</li> </ul>
Commerce services		<ul style="list-style-type: none"> <li>• Buy tickets for shows.</li> <li>• Financial transactions.</li> <li>• E-government.</li> <li>• E-commerce.</li> </ul>

Table 1: Kind of service [Tel06]

## 2.3 RSS

RSS is a XML formats for messages and other information that is updated frequently in the websites, and do not have relation with the content that makes syndication. The documents are written in these formats are called feeds.

### 2.3.1 Extensible Mark-up Language (XML)

All RSS feed conform an XML standards. This extensible Mark-up language or XML is a specification developed by the W3c. XML is pared-down version of SGML or Standard Generalised Mark-up Language, designed especially for Web documents XML uses tags that describe the contents of the field. It is very similar to HTML but unlike HTML allows designers to create their own customised tags, enabling the definition, transmission, validation, and interpretation of data between applications and between organizations.

### 2.3.2 Advantages of RSS

1. Great acceptance all around the world & internet community.
2. XML support with a variety of language.
3. Easy way to make programs with RSS.
4. Easy way to understand the feed (automatic or not).
5. Clear documents, sight by specific programs and browsers.
6. Formal way to produce and process documents.
7. Minimum time to create and process RSS file.

### 2.3.3 What is RSS?

RSS is another sub-language rise of the application of the meta-language of XML. Therefore a RSS file is compost document of tags delimited between symbols bigger and smaller similar used in XHTML.

RSS (Rich Site Summary or Really Simple Syndication) and it is designed for the distribution of news or information like news that are hosts in the webs or weblogs.

The RSS files are called commonly RSS feed or RSS channel and contain a summary that has published in the origin web site. The structure in one or more items. Every item has a title, a resume of text and a link of the original source in the website where is all the complete text. Furthermore can include additional information like the name of the author or the date of publishing. Therefore, every source of information that can be cut in items (the message of a forum for example) can be distributed in RSS.

Reading a RSS file of one website is possible to know if it is refreshed or no with exactly the news or text, but without need to access to the website. The RSS file contain one specific link for every item that has the feed that redirects to the website with all the text.

## Parsing from RSS feeds to MHP-television system

### 2.3.4 The channel elements

#### Required

Element	Description	Example
Title	The name of the channel.	CNN.com - Technology
Link	The URL to the HTML website corresponding to the channel.	<a href="http://www.cnn.com/rssclick/TECH/?section=cnn_tech">http://www.cnn.com/rssclick/TECH/?section=cnn_tech</a>
description	Some describing sentence for the channel.	CNN.com delivers up-to-the-minute news and information on the latest top stories, weather, entertainment, politics and more

#### Optional

Element	Description	Example
Language	Specification of the language used in the RSS information. This element has to be described with a list of international values.	Catalan is "ca"
copyright	Evaluation of the copyright situation	Copyright 2005
ManagingEditor	Email address from the editorial responsible.	Editor@cnn.com
WebMaster	Email address from the technical responsible person of the web page.	Webmaster@cnn.om
PubDate	The publication date for the content in the channel. This has to be in the RFC 822 specification of dates.	Sat, 07 Sep 2005 16:00:03 GMT
Lastbuilddate	The last time the content of the channel changed.	Sat, 07 Sep 2005 09:00:31 GMT
Category	Specify one or more categories that the channel belongs to	<category>Newspaper</category>
Generator	A string indicating the program used to generate the channel.	MightyInHouse Content System v2.3
Docs	A URL that points to the documentation for the format used in the RSS file.	<a href="http://blogs.law.harvard.edu/tech/rss">http://blogs.law.harvard.edu/tech/rss</a>
Cloud	Allows processes to register with a cloud to be notified of updates to the channel,	<cloud domain="rpc.sys.com" port="80" path="/RPC2" registerProcedure="ping/Me" protocol="soap"/>

## Parsing from RSS feeds to MHP-television system

### Optional

Element	Description	Example
Ttl	It's a number of minutes that indicates how long a channel can be cached before refreshing from the source.	<ttl>123</ttl>
Image	Indicates the way can be displayed a GIF, JPEG or PNG image.	
Rating	The pics rating for the channel	
Textinput	Specifies a text input box that can be displayed with the channel	
Skiphours	A hint for aggregators telling them which hours they can skip.	
skipDays	A hint for aggregators telling them which days they can skip.	

Table 2: tags from channel way

In this point we can realize there are two different types of elements. The request (very few) and the optional. This tags will describe only the aspects of the channel reference, that is important when we have some different channels of input. For example when we manage a new agregator software we can have information of different feed web sites in the same program, like CNN, MSN and vilaweb.

Can be important for the programmer show information of the different reviews, for example to get information about what hour can the consumer get a new edition of the news that are consulted.

Furthermore we can have meta-data not showed in the interface of the software, but important if we want to go inside of the how periodical is the refresh of the information, o the different authors of the news that we want to read.

Also we have the opportunity to make a hierarchy with some important tag, this tags i item. The *item* tag is used to put only the information of the news, not only the description but we can put some meta-data very similar like what we can put in the channel information.



## Parsing from RSS feeds to MHP-television system

### 2.3.5 The items elements

The next board, will show which are the different tags or elements, that are possible to put in the RSS files:

Element	Description	Example
Title	The title of the item.	Your dusty old PC may hold the cure for cancer
Link	The URL of the item.	<a href="http://www.cnn.com/rssclick/2006/TECH/06/05/research.at.home.ap/index.html?section=cnn_tech">http://www.cnn.com/rssclick/2006/TECH/06/05/research.at.home.ap/index.html?section=cnn_tech</a>
Description	The item synopsis.	Read full story for latest details
Author	Email address of the author of the item.	
Category	Includes the item in one or more	
Comments	URL of a page for comments relating to the	
Enclosure	Describes a media object that is attached	
Guid	A string that uniquely identifies the item.	
PubDate	Indicates when the item was published.	Mon, 05 Jun 2006 10:33:59 EDT
source	The RSS channel that the item came from.	

Table 3: tags from item way

In our project the main information will be in the item element. Then it is so important to pay more attention in this tags than others because is the most important part of all the code of .rss file. This is important for any different reason for the instance of the processing time of the parsing it should exist is important.

In the next page is showed the schema of a normal .rss file to parser in our particular case, then we can see the most part of the code is in the structure of the item (in our case will be the different news of the e-newspaper).

## Parsing from RSS feeds to MHP-television system

Then it's important to take a look exactly in the web site were we will take the fees, in this project we will look for the [www.vilaweb.cat](http://www.vilaweb.cat). One of the most important e-newspapers in Catalonia. This electronical newspapers have always on-line one RSS file which contains the differents news showed in the main page of all the site.

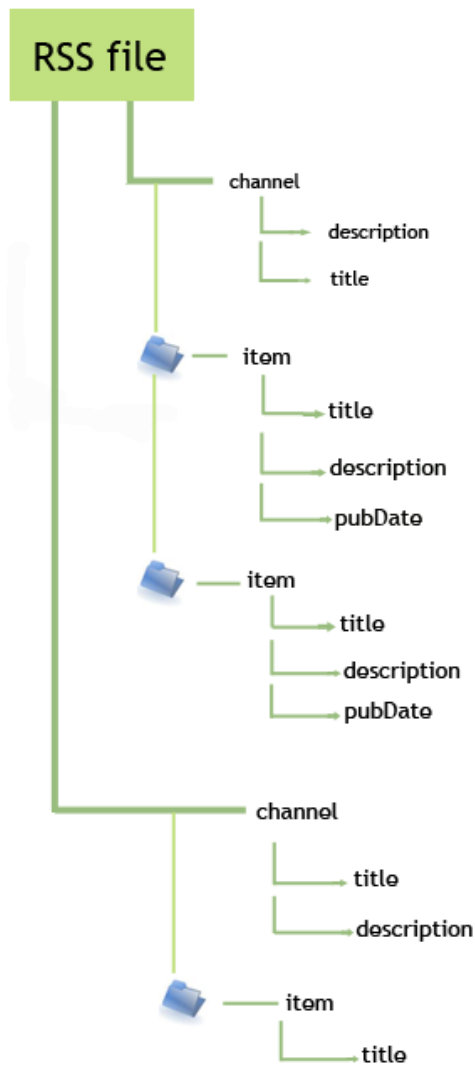


Figure 9: RSS file diagram

We can also found differents RSS files in all the site, but in differents sections like technology or breaking news. How is this translated in our case? When we should develop the algorithms we should take care which is the web page where we will take the information. This is the only caution we will take because all the process will be directly and automatic.

The configuration of the RSS file is so much important to be understood. Before in this brief is explained there are some tags requested and other optional. But it is important to look in the exactly case of Vilaweb. The schema to follow is showed in the adjoining picture.

In the normal case will be only one channel (our channel of news), this channel has the following tags:

- Title
- link
- description
- language
- lastBuildDate

Dependent from the channel tag there are the items tags, every item has the information of one notice, and will contain the following tags:

- Title
- Description
- pubDate
- guid
- link

## Parsing from RSS feeds to MHP-television system

```
<item>
  <title>
  La Fundació Gala-Dali confirma que no participarà en la gestió de la casa natal
  del pintor
  </title>

  <description>
  El president de la <a href="http://www.salvador-dali.org/">Fundació Gala-
  Salvador Dali</A>, Ramon Boixadós, ha confirmat, en una entrevista al
  setmanari <a href="http://www.emporda.info/">Empordà</A>, que no
  participarien en la gestió de la casa natal del pintor. (Ara s'hi fan obres de
  rehabilitació de la façana, que s'haurien d'acabar el mes vinent.) Tot i amb
  això, el batlle de Figueres, Joan Armangué, assegura que quan l'ajuntament
  tingui enllestit el projecte sobre l'ús de l'immoble, la Fundació s'hi
  implicarà. Aquests últims dies ja es comença a percebre el resultat de la
  rehabilitació, projectada per l'arquitecte Josep Azemar l'any 1898. Les obres
  de recuperació de Can Puig, d'un gran valor històric, s'han d'acabar el mes
  entrant.
  </description>

  <pubDate>Wed, 7 Jun 2006 11:17:00 +0200</pubDate>

  <guid isPermaLink="true">http://www.vilaweb.cat/1926439</guid>

  <link>http://www.vilaweb.cat/www/noticia?p\_idcmp=1926439</link>
</item>
```

Figure 10: Item part of the RSS document

- *Title* will contain the notice titular.
- The *description* will contain the news body and can include some external links, for instance if there is a notice affecting some person that has a particular web page, we can make a relation to redirect, in the case of our project we will not pay attention this can be a way to wider a future project.
- The *pubDate* will show the date and hour of publication in the RSS file. The *guid* we will affirm that is a true notice and the consequent link.
- Finally link where we can make a direct connection with the web page and take only the HTTP page where is hosted the exactly notice.

In the schema picture is showed another external channel the reason is the general schema of a RSS file, in our project we always treat one channel, the root of the channel will be: [http://www.vilaweb.cat/media/rss/noticies\\_edicio\\_333\\_rss.xml](http://www.vilaweb.cat/media/rss/noticies_edicio_333_rss.xml).

### 2.3.6 Reasons of the election

The election of the RSS system is because some points:

1. **Universal system:** RSS is a “new” system of tags drifting from a more general system called XML. The RSS history begun in 1997 when was developing Netscape 4.0 and some developers thing in one feed system of tags, specially for Internet. Since this date it was increasing very slowly the web sites who content syndication stuff, nowadays this web feed is very extended all around the net.
2. **Large number of parsing project.** Java and XML are two technologies very closed in one word, parse. A parse is a program (in any language) what makes a translation between to language. In our project the translation is between RSS (xml) to Java.  
In all the life of xml there were a lot of projects with same doubt/problem. Can we translate information in xml specification to Java? The answer is yes, there are three basic “parsers” in Java who make this thing, JDOM, SAX and . Every project depends on different enterprise or non-profit projects, like Sun Microsystem or W3C. This companies make the specification and they put it on the net to share with all the community.
3. **ROME:** We know that RSS depends from XML and it makes that we have a lot of precedents to compare and have a open view, but RSS is different than XML in consequence the parsing would be different too. The solution of this problem is a little project inside Sun Microsystem which is called Rome. This project is based on JDOM parsing (for XML) but there also more libraries that makes more optimal and easy convert RSS to Java.
4. **Large future:** Now RSS is very often to be founded in pages of blogs or news. But like other technologies multimedia is the next step. Then are new concepts more powerful than text (what contains traditional RSS files), this new technologies can be like:
  - **Podcasting:** is the method of distributing multimedia files, such as audio programs or music videos, over the Internet using either the RSS format, for playback on mobile devices and personal computers
  - **Vodcasting:** is an emerging term used for the on-line delivery of video-on-demand video clip content via RSS format.
  - **MP3 blog:** is a type of weblog in which the creator makes music files, normally in the MP3 format, available for download. They are also known as musicblogs or audioblogs.

## 2.4 MHP

MHP is a middleware optimized and created by the DVB consortium. The entire process of developing DVB standards is driven by organizations that make DVB consortium, then this is a reason why the part of commercial issues is so important as well as technical ones. The composition then of the different parts that developed all the standard were:

- **The Steering Board.** This group is elected by the DVB members every two years and is responsible for defining the overall direction of DVB. The Steering Board is also responsible for final approval of DVB specifications.
- **Commercial.** The Commercial Module is responsible for discussing the commercial issues of one work in concrete. Each DVB standard starts from a set of commercial requirements. Once a technical specification is complete, the Commercial Module will make sure that this specification has a viability in the global market of the telecommunication.
- **Technical.** Creates the text of the standard is carried out in the Technical Module. This is where the technical discuss about the different aspects of the new standard, this means that commercial module gives instructions about the viability of one project and then technical module will develop these concrete parts. When the specification has been finalized in the Technical Module, it is passed back to the Commercial Module for approval before going on to the Steering Board.

MHP uses the application model defined by JavaTV, and defines some additional service information to tell the receiver what applications are available. To provide access to the files needed by the application MHP uses DSM-CC object carousel.[Msm05]

MHP defines a number of API that allow an application to access the various features of the receiver. MHP application can be written in Java or HTML, but Java is the most used in the programmer context. Then MHP can let us to access to information, draw objects on the screen, controlling the audio and the video, make communication with other remote servers. [Msm05]

Nowadays there are different version of the same MHP standard, and there exists a migration between different versions. This is important to know because every version can include new features for the developers, but the park of receivers sometimes does not develop at the same velocity. But these are more market aspects not related with this report.[Msm05]

## Parsing from RSS feeds to MHP-television system

### 2.4.1 MHP architecture

When is programmer a MHP application is allowed the access to different information flow. This streams can be processed or not, it depends on what application and which purpose is search.

The different streams can be input or output stream of video and audio data, associate of the broadcast channel and the return channel.

In a clear MHP system we can distinguish three different layers:

- Hardware resources, software resources (depending to the hardware).
- Independent software system from the hardware.
- Applications.

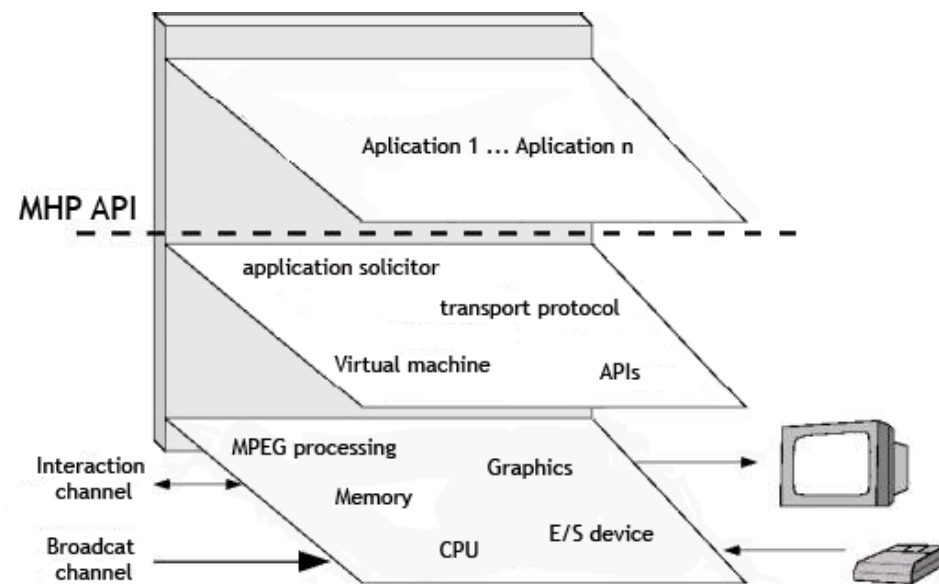


Figure 11: Morphology of the MHP middleware [Mar05]

Like is showed in the graphic beyond, the middleware (MHP) is between the second and the first step. Then we can control the virtual machine, the transport protocol, the application solicitor and some many application residing in the receiver.

When it was develop the first MHP specification, it was a first point to leave, and this point was the specification of the JAVATV APIs from Sun Microsystem. This has an important consequence, and this is the acceptance for the Java Virtual Machine. This would means when we compile [Mar05]

### 2.4.2 Life cycle of Xlet (application manager)

Xlet can be write in two differents and standard ways, the principle is with Java language but now it begin to be developed in HTML. Java has particularities that make more interesting and with more options the algorithms that allows to work the differents applications in the set-top box.

This forms the basis for all systems based around JavaTV, including MHP and OCAP. Like applets the Xlet interface allows an external source to control the life cycle of an application and provides the application with a way to communicate with the interaction inputs.

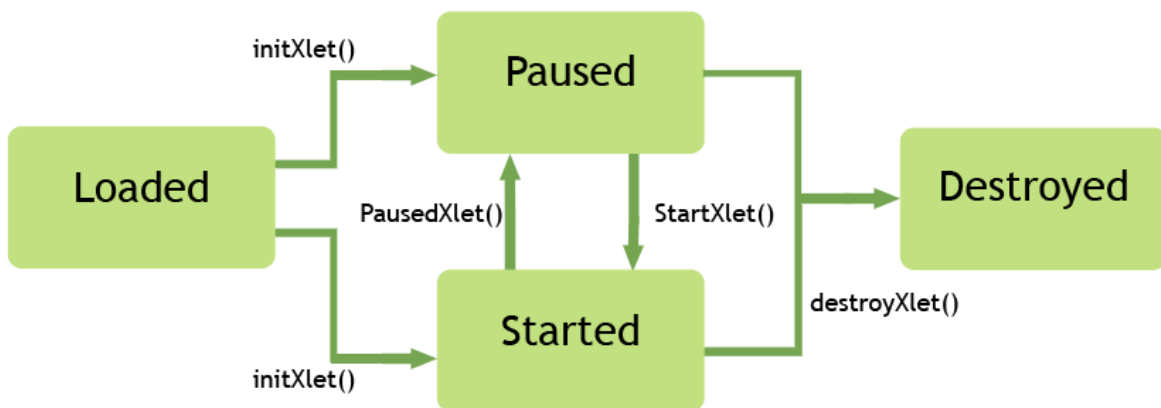


Figure 12: Life cycle of MHP application

There are a lot of similarities then between Xlet an applets, but there are also differents. One of the biggest is the execution system. Xlet have options to be paused and resumed, and applets no. The reason we can find in the hardware of digital television, by the moment the set-top box does not have a lot of memory to process differents instance or run differents applications at the same time, the consequence it is can only be processed one or two Xlet in the same time. Then it is need tools to control the process of paused, start or stop.

Like is showed in the figure 4 we have differents methods to change between the differents states can be the application. For example to change the state between Started to Destroy the application should execute the `destroyXlet()` method, this example when is running the application (because is in the start state) will stop the application and also destroy from the memory (remember the memory is insufficient in the actual set-top box in the actual market (2006)).

### 2.4.3 MHP packages

MHP is composed from different packages, each one makes different aspects related with the general application. All together will make the general application showed on the television.

- Version 1.1 of Java and Personal Java environment give the core and the basic environment for the execution.
- Java Media Framework, extend capacity for multimedia Java, allowing to put audio, video and other format of information for be available for Java application.
- Java TV gives functionalities for interactive television, like EPG (Electronic Program Guides), pay-per-view or other services.[Föt02]
  - *javax.tv.service* Provides mechanisms for accessing the service information (SI) database and APIs representing the SI elements it contains.
  - *javax.tv.service.guide* Provides APIs to support electronic program guides (EPGs), including program schedules, program events and program ratings.
  - *javax.tv.service.navigation* Provides APIs to navigate through services and hierarchical service information.
  - *javax.tv.service.selection* Provides a mechanism to select a Service for presentation.
  - *javax.tv.service.transport* Provides additional information about the transport mechanisms that deliver the content described by the service information data.
  - *javax.tv.xlet* Provides interfaces used by applications and the application manager to communicate.
- DAVIC (Digital Audio Video Council), interfaces which gives interoperability with the audiovisual interactive information and multimedia communication.[Föt02]
  - *org.davic.awt* The DAVIC AWT package contains only a Color class. This class adds the notion of alpha. The alpha value provides support for transparency between Graphics planes and Video planes.
  - *org.davic.media* This package provides various extensions to the Java Media Framework API (JMF) for the control of TV oriented audio/video content.
  - *org.davic.mpeg* This provides access to Transport streams.
  - *org.davic.mpeg.dvb* Provides utility classes for common MPEG concepts as used in DVB.
  - *org.davic.mpeg.sections* Provides access to MPEG-2 section filtering.
  - *org.davic.net* Provides general content referencing.
  - *org.davic.net.ca* Provides an interface to various features of a conditional access system for those applications which need it.
  - *org.davic.net.dvb* Provides DVB specific content referencing.



## Parsing from RSS feeds to MHP-television system

- *org.davic.net.tuning* Provides access to tuning - MPEG multiplex selection.
  - *org.davic.resources* Provides a framework for scarce resource management.
  - *org.dvb.application* Provides access to lists of applications (AIT) which are available in this context and the ability to launch those applications.
- HAVI (Home Audio/Video interoperability Architecture User-Interface), graphic interface with a specially interoperability with television systems.
    - *org.havi.ui* This package implements the HAVi (Home Audio Video Interoperability) user interface.
  - Extra interfaces for user preference, interface extension and more. [Föt02]

### 2.4.4 Graphical behaviour

One of the most important think to be carefully with this project is about the graphical issues. The viewer of the television has a perception of what is seeing and it has to be comfortable and pleasant for him. Also be should put a lor of attention in how different works a television against a PC computer.

This is a typical source of problems when is programmed a digital television application. One reason for this problem is the palette colour accepted in the PAL system (in case of Europe) and NTSC (in the case of U.S.A. and other countries). This television systems had the peculiarity to accept a restrict number of colours, showed in the `picture beyond.

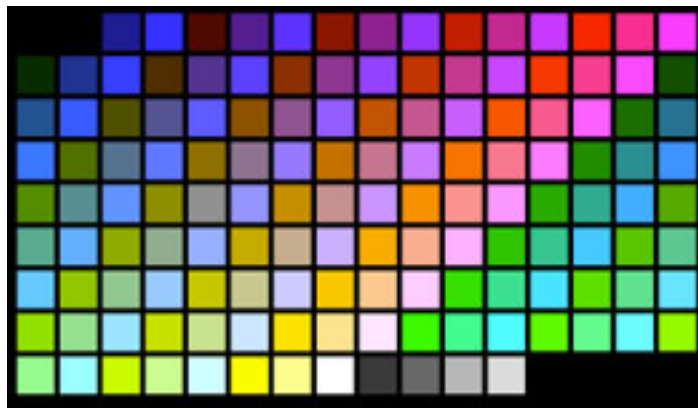


Figure 13: MHP colour palette [Int06]

This means there is not a wide palette to be used, also there are some colours that can not be used because are used in the subtitles menus and then are not available for other uses like backgrounds or other things to be printed in the television screen.

At the beginning is important to look the differences of layers that can be used in the mhp standard, the understanding of this things can increase the different possibilities to make.

## Parsing from RSS feeds to MHP-television system

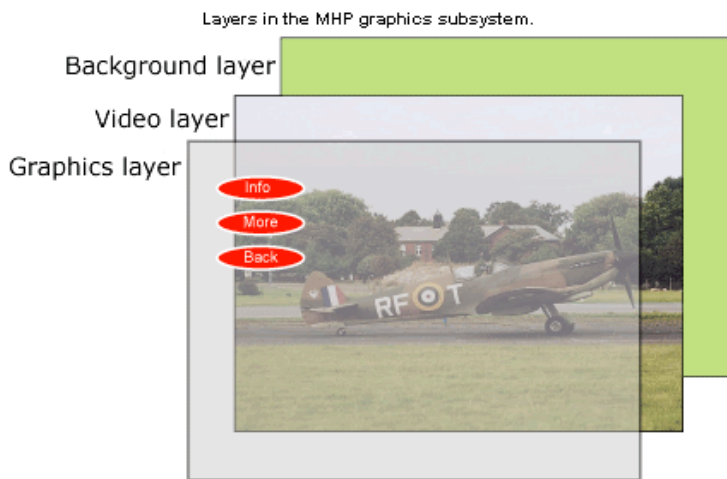


Figure 14: Layers of an MHP application [Int06]

Then we should put attention of concern the graphical view of the application and this has to be done in different steps.

The structure of the layer system is divided in three parts as the picture beyond shows it, the first graphical layer found it is the graphical layer, here is were the developers put all the content of the application. This layer is over from the video layer were is reproduced the MPEG-2 stream of video.[Int06]

Then application has the chance to show some different parts of the video thanks to the alpha factor. This factor let the programmers to put more or less degree of transparency to the layer in question. Also there is the chance to make mini-windows and put MPEG-2 stream there, this applications are often when the broadcaster do not want the audience miss the life video, for example in a football match the broadcaster can show life statistic also is important to show all time the match.[Int06]

To be implement, it has to be used the DAVIC platform, here there are the tools to put the information on the screen. This tools are grouped in the Hscreen class, also this class has subclasses inside for every layer we want to change. Like is showed in the picture 15

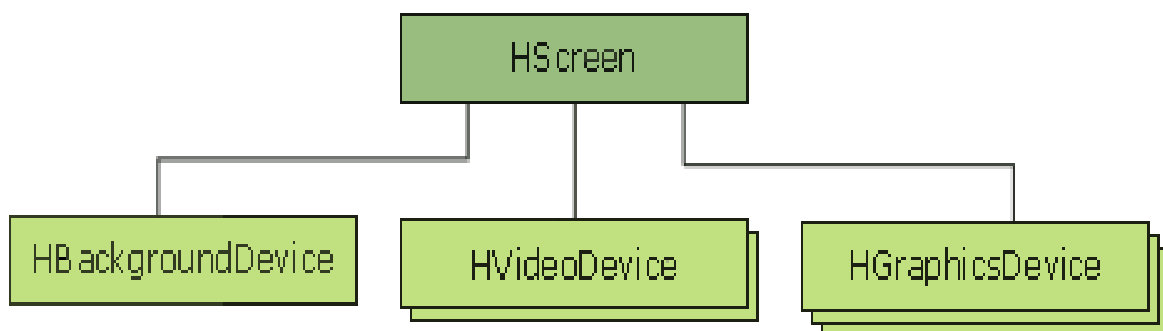


Figure 15: Schema of Hscreen class

## Parsing from RSS feeds to MHP-television system

Then if is necessary to make some changes in the HbackgroundDevice, there are some class that makes the configuration, in this concrete class it comes thanks to HbackgroundConfiguration with that we can make this can of changes centring in the example.

- PIXEL\_RESOLUTION
- FLICKER\_FILTERING
- PIXEL\_ASPECT\_RATIO
- SCREEN\_RECTANGE

These are example of what it can be changed in the background. There are other can be only applied to the graphic layer, or other to video layer like GRAPHICS\_MIXING functionality.

Until here is made a description of how can be changed the graphical configuration of all the layer (in some cases the screen). But also in our project is important to the control of the text, because the composition of the screen will be text and one image in the background.

The text in MHP can be showed in two different ways, one way is with normal screen commands of java, like DrawString(). In this case is possible to change some basic configuration of the letters like bold, italic, type of letter, size of the letter, colour. But this has a really huge limitation, and comes when you are printing in the screen a long string of characters. Then will appear really big problems of non-apparition of text in the screen. This has no solution immediately because the DrawString() is limited to serve general algorithms, and no specially to digital television services.

There is a solution of that, because maybe we want to make a rectangle to put the sms inside and showed in the television, for example. This application only allows to put text inside the rectangle and no more. Then the DVB library offers the DVBTextLayoutManager.

Here we can find the tools to manage this disadvantage of the other java library (DrawString). With functions like getlinespace, setTextwrapping, setLineSpace, or others we can take an absolutely control of all the things me can make in the MHP standard. In the case of this project is used this high level programmer to take a better result in the screen, also the body of the notice (the description variable in our case) is a very long part of the constitution of the news.

# 3 Parsing system

## 3.1 Solution

In this section I am going to explain the different steps that I take to the success result. The first step to take was collect information about the two sides of the parser, one is RSS files and the other is MHP.

The illustration (16) shows in a very general way which are the two general parts of the parser. This box then will make a translation between the tags of RSS (specified in the past chapter).

In other words to make a translation is to understand the tags contained in the file, know everyone which kind of information can contain. And finally transfer the information in to variables that can be understood in java language.

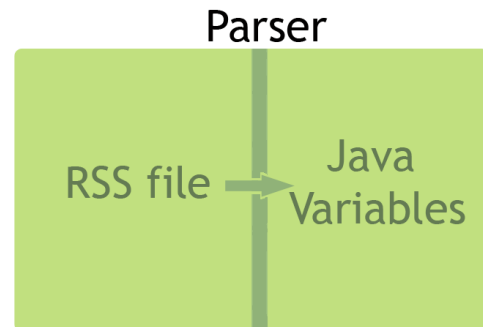


Figure 16: schematic draw of parsing

Of course this is a very general explanation about the work of one parser, but in the next steps I will explain more deeply which is the evolution of my concept about the parsing. Explaining the difficulties founded in all the process and how I can resolve to get the final solution.

Also in this next lines, it is explain the different steps to get success in other parts of the project, like television emulation or some programming aspects.

### 3.1.1 First step

The first task to make is distinguish two different (and in the first time not connected) parts of the project, one part is RSS and the other is MHP. This differentiation is necessary because at the moment we do not know how are the ways to interact RSS and Java, not yet.

The conclusion of this research were the following:

- RSS are files hosted in web pages, where is founded information inside elements and tags. It is important to make an important look which kind of tags contain every variable. To sum up, the date information contained in tags are not the same than a description tags because the space of the memory is not the same or because the way we put the information is not the same too. Then is so important to became aware which kind of tags are in the RSS files where the project will

## Parsing from RSS feeds to MHP-television system

check every time.

This was another point to look, in there moment I search information about all the possible tags (included in the figure 9) this information was important to get in touch with RSS style. It is important too get familiar with all the possibilities that offer RSS standard. But in there moment was important to check if the web page where the parser will connect has all this amount of information, the answer was simply, no.

That's an important thing to know, an important too for a good parsing and optimize the different variables in the main project that will be running in a set-top box, hardware without a lot of memory to use.

- In the other hand was mhp a complex middleware and difficult to know where and which are the entrance to put information in. It takes time in the develop of the project know how we can and the way, for that is explained in the next corresponding step of the memory.

In the firsts weeks this was the task to resolve. The conclusion of all is that XML can be processed by java thanks to different parsers, and this is the gate to communicate both different language. The next tasks are to investigate about how can we make a simple parsing, and how can we make a TV emulation.

### 3.1.2 Second step

The first thing was to develop or find an application which can show on the television a Hello world message. The result is in the image beyond (17).

This result is simple and important, because the result of the search was an Xlet that please the specifications of the problem and show the internal constitution of an Xlet.

Xlet as mentioned before is like a Java applet with difference to see the first Xlet shows the specific part of the different code.

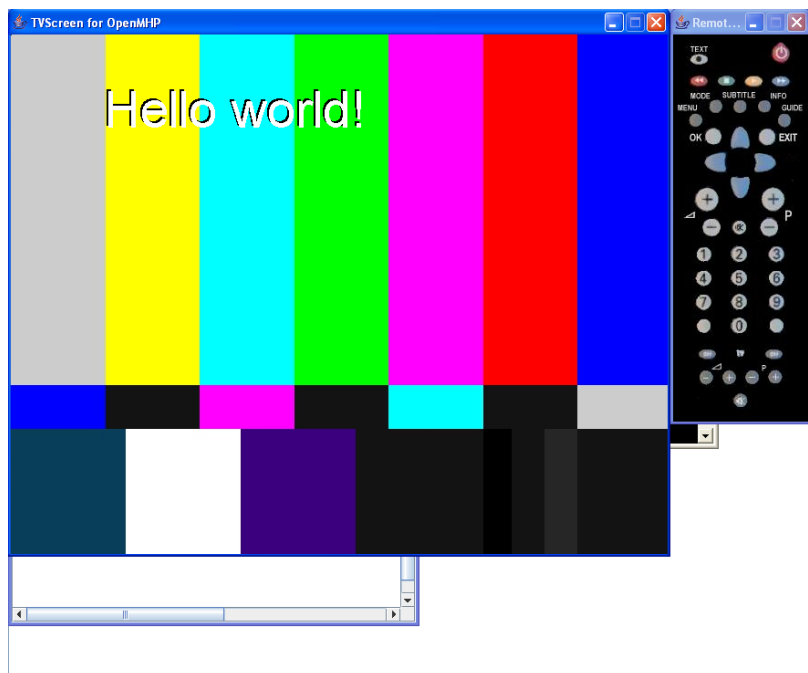


Figure 17: First result of the MHP application

Also is important to began familiarized with the graphic functions (the most difficult ones and it control) and to know how this message (Hello world) can be change to another one.

## Parsing from RSS feeds to MHP-television system

The other part of the work was found some basics about the relation between Java and RSS. At beginning was important to find the relation with two big and extended language, then the search of information was directed to Java and XML.

In this stadium there is a lot of information and projects made it before and with an optimal result and with a good look in internet it is not problem to found it. Instead there are some important aspect to be mentioned:

Is an important decision the election of the application programming interface (API) in the field of parsing. There are a lot of possible parsers to be elected in one project, to choose is one election of optimal result.

There are two principles parsers in Java one is SAX (Simple API for XML) and the other one is DOM (Document Object Model). There a lot of literature and it is not the last intention of this diplomarbeit. One important thing is to mention some other and not main parser existing in the community. Projects like JAXP, Dom4j, ElectricXML, XMLPULL and in special JDOM (for the future).

In this step of the project it was important to make another exercise like the television emulation did it before. In these case the task was to pass from one XML file containing the string Hello world to one Java string variable to save it. In this case was really hard and without success, the suitable code to make was too much waste of time for the rhythm of the project, in consequence was rule out this way.

Then next step to think about was the direct RSS parsing, and one preliminary schema of the project.

### 3.1.3 Third step

In this step the objective it is to find some good parse between RSS files and Java, and implement a primitive schema for the operations of the algorithm.

The first part was succeed when is founded one project inside the java (Sun Microsystem community) which makes some directly parsing between RSS and Java. This project is very recently and have some bugs, but for simple parses works well. The name is Rome is founded in 2004. The work of that was to understand and know the application in one simple algorithm it is tried to make the same like before, one algorithm to parse some string (located in the description tag) to a java variable. It was succeed.

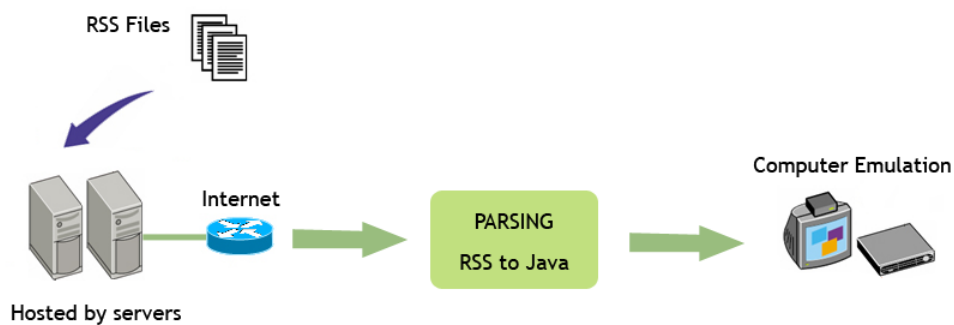


Figure 18: project chain communication: [Dar05]

## Parsing from RSS feeds to MHP-television system

In the other hand it was important to think about one chain of processing information. The first idea is take from one URL (where is located the RSS file) directly by internet. The computer will take the file with an always-connection (Ethernet connection), next of this will be a parsing inside the computer the result of that will be in Java language, understand by the television emulator. This chain is represented in the figure 18.

Then the tasks for the last step were complete with correct and implemented all the chain of the process, from the feed of the document to the show in the “television”. It is important too, to get practice and agility with the *Rome* parser.

### 3.1.4 Final solution

The final step of the project was the programming of the result. The final solution is thinking in 3 different java classes called, `news.java`, `FeedReader.java` and `parser_internet.java`. Like showed in the graphic beyond.

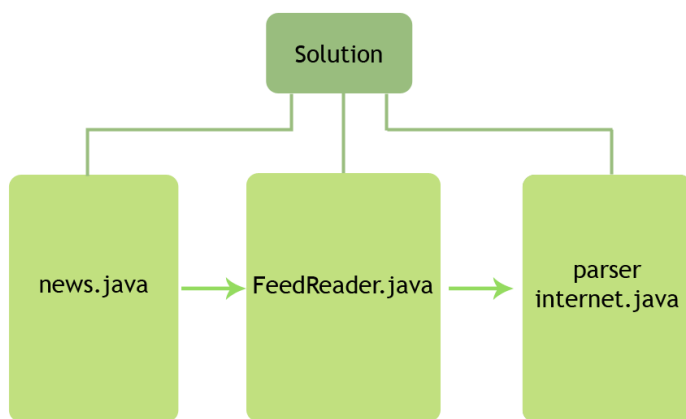


Figure 19: Schema of the application

The main idea to make this architecture is to divide the different tasks in different classes. The result of that is 3 different class `news.java`, `FeedReader.java` and `parser_internet.java`.

Specificity in the *news* class we only declare the different type of variable we will use next, in the future are used to make the parsing and to save the different strings or Date parse. In concrete we will declare 4 variables, 3strings and 1 date.

The next class in the succession is `FeedReader`. In this class we should do 2 important things. One is to create an array of news, every position will contain information of title, description, date of publication and link of each notice parse. Then we will access to this vector called `news` like that: `news[1].description` the result is the description access in the notice number 2.

The most important thing to develop inside the class is the main parse of the RSS to java variables, here is used libraries from `jdom` and `Rome` project. First of all is created a `SyndFeedInput` which has a class `build` that allows to make the parsing to `SyndFeed` class which will contain all the information of the RSS file. Later is created a list of all information of every entries we take of the RSS file.

## Parsing from RSS feeds to MHP-television system

This is made in this lines of code:

```
//Creation of the feed and fill up.  
SyndFeedInput input = new SyndFeedInput();  
SyndFeed feed = input.build(new XmlReader(feedUrl));  
  
//Creation of our entries list.  
List entries = feed.getEntries();
```

Figure 20: Part of the FeedReader code.

Then the algorithm is ready to take everything of every notice. In our case what is wanted is the title, description, date of publication and link of the notice. This will be save in the array of 3 positions created before.

Is the last step is the parser\_internet class. The structure of this class is the similar one like Xlet. There is a part of Start(), Stop(), Destroy() and also paint(). This last is the most important because is the most custom-made since the most important thing is the graphical issues since all the other processing is did it in other classes.

The are some special characteristic about the paint() function in parser\_internet class. The first one is the use of DVB lay out characteristic. This special things are implemented because the large strings of descriptions characters. The body of the description is a string sometimes longer to be displayed in only one line, then that is a problem to make show over the screen. Is a problem using simple java instructions like DrawString() that allows to put on the screen a string.

Then there is a solution making a valid rectangle were can be impressed the text. That effect is made by setBounds(100,0,200,30) function. We enter the coordinate and the extension of height and width, of every text we want to show in the screen of the television. This solution will let to put and organize all the text along the screen without problems of conceal text because every time the line text find the limit will go to the next line, like an editor text.

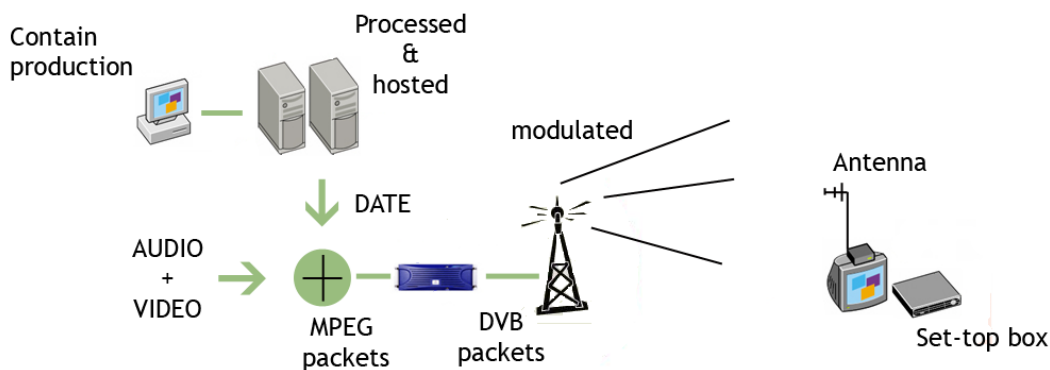


Figure 21: Chain of communications



## 3.2 Next Features

The first conclusion of the project; it is not conclude at all. There are many expansions can be implemented, I will give some ideas and basis for one of these, but for sure one student can follow these first steps.

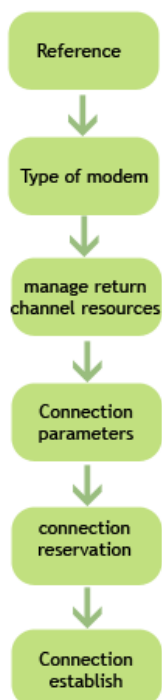
### 3.2.1 PTSN communication

To make interactive the digital television is basically one return channel to send information out of home. In this project (Parsing from RSS feeds to MHP television system) we connect with internet in concrete with one URL because the parser needs to take from this URL the RSS file (in the project a news RSS file). The parser then has an interaction but virtual communication, because the set-top box where has to be put up usually does not have an Ethernet connection like the laptop where is developed and runs.

The normal way to make interaction nowadays (year 2006) the most part of the set-top box only has PTSN (Public Telephone Switched Networks), and always-on interface connections are not very developed by the moment. There many reasons of this delay against Ethernet connection, fundamental is the recent technology of MHP 1.1 standard which is the only specification that allows always-on interface, little and rising market of interaction television, and other many reasons.

The emphasis has to focus in two different aspects:

- Connection aspects.
- Changes of the keyevents (remote control).



The connections aspects can resume in the figure 21:

- I. **Reference** In the first step the application will take conscience about the necessity of to implement the functions of the return channel, here will take reference about what has to do.
- II. **Type of Modem** the application need to know which is the communication interface, here examples can be PTSN, ISDN, LMDS (different types of communications).
- III. **Manage return channel resources** Like all long project resources are very limited inside set-top box in consequence the communication application has to be careful with this.
- IV. **Connection parameters** this class defines the most important parameters needed to set up a connection, such as phone number or user name and password.

Figure 22: Communication process.

## Parsing from RSS feeds to MHP-television system

**V. Connection reservation** the application should interrupt some other tasks to put memory and process calculator to make possible the communication system, because the information can be received in a faster way.

**VI. Connection establish** if all the other steps are took in good way, the connection with a remote server will be successfully.

In every step related to the before steps, there are classes allowing to make an efficient implementation, the most part of these classes are in the DAVIC project (explained before in the MHP chapter).

Now we can define the whole process of the rss feed integrating the way we receive the information, how the set-top box process this request and adopt a internet solution. My vision is divided into different solution, and both are compatible and successful, but different in the way to take the feed.

The first idea is to make a polling parser. This parser has an independence from any broadcaster, then the internal memory of the set-top box has an always instruction to make a polling (a call every pre-fixed period). In this call will be checked if there is or not an update Rss file.

If the answer is affirmative will begin a parsing process, connection to internet extract all the file and later ask the viewer if he/she want to impressed the latest new on the television screen.

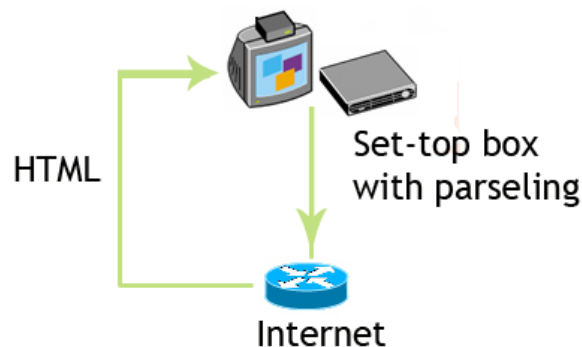


Figure 23: Polling solution

This system has in favour the less use of the spectrum, because the set-top box is never conditioned to the broadcaster. Against there are aspects like the periodicity (if is to elevated can derive in a low process of the set-top box).

## Parsing from RSS feeds to MHP-television system

The second idea is make a parser conditioned on the broadcaster instructions. This one will send by the waves an RSS file to all its viewers, in the carousel way (DSM-CC solution).

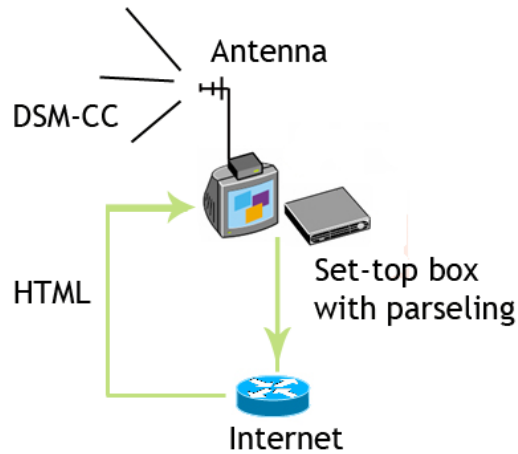


Figure 23: Broadcaster solution

Later the viewer receive and advise in the television, that can be or not accepted. If it is the case will establish a connection and show the information in the screen. Also will be an option to connect direct to the HTML page, and see the notice with picture like a internet browser on the television screen.

## 4 Conclusions

I want to reflect my conclusions in two different ways, one explain which are the specific conclusions for my own project and the other explain the different technical obstacles confront all along project development

### 4.1 Technical conclusion

First of all, all along the project there were difficulties from the classes used in the project in concrete classes related with the Openmhp project. This project is an open source in consequence sometimes the implementation is no the best, and in other cases there are not implementation. The reasons we can find in the sense of MHP standard, this is standard is like MPEG all the functionalities are written but the way can work is a thing to be developed by the different companies in the market place. Then the Openmhp implementation is sometimes very poor.

This project also has an extension and it is like a module in a modular project, then the part I made is the core of the application and some other person can continue the work began. Other parts to be developed can be communications part, better interface with the viewers or extensions of the new MHP standard called DVB-HTML were the television will be more closed to internet.

### 4.2 General conclusion

I can conclude this project making some go over. I have to admit this project has spend a lot of time increasing basic knowledge of some trivial technologies like java. This can be thought like a fatal error for the waste of time, but I think I make a real progress understanding, getting deeply and trying to resolve the different java problems I had to confront.

When I began this project I know there were more disadvantage than advantages: Non-idea of java technologies, non-idea of MHP technologies, non-idea of RSS and feed contents. It was a real challenge.

With the success at the end of the project I can affirm that this project has a really good outcome for me. I can learn many languages and deal with a high-perspective technologies like feeds (RSS) or MHP. Getting the opportunity to get in touch for the first time.

## 5 Bibliography

- [Cha01] Chase, Nicholas: XML and JAVA from scratch. Version: 2001. Safari Books.
- [Hme03] Rusty Harold, Eliotte; Means, W.Scott: XML in nutshell. Version: 2003. Safari Books.
- [McL01] McLaughlin, Brett: Java and XML. Version: 2001. Safari Books.
- [Cra01] Cranford, Jason: DHTML and CSS. Version: 2001. Safari Books.
- [McL02] McLaughlin, Brett: Java and XML data binding. Version: 2002. Safari Books.
- [Msm05] Morris, Steven; Smith-Chaigneau, Anthony: Interactive TV standards a guide to MHP, OCAP and JAVATV. Version 2005. Focal press.
- [McL01] McLaughlin, Brett: Java & XML. Version: 2001. O'Reilly Media.
- [Bud03] Burd, Barry: Java and XML for dummies. Version 2003. Willey-VCH.
- [Rus02] Rusty Harold, Elliotte: Processing XML with Java: A Guide to SAX, DOM, JDOM, JAXP, and TrAX. Version 2002. Addison-Wesley Professional.
- [Gra06] Gràcia, David: Investigació i desenvolupament d'aplicacions interactives en MHP. Version: 13<sup>th</sup> February 2006, Barcelona. ETSETB (UPC).
- [Mar05] Martín, David: Desarrollo e integración de aplicaciones MHP para la TDT. Version: july 2005, Barcelona. ETSETB (UPC).
- [Val00] Valdivielso Poza, Yolanda: Desarrollo de una aplicación interactiva mediante OpenTV. Version: 2000, Barcelona. ETSETB (UPC).
- [Föt02] Fötsch, Peter: Interactive application for the multimedia home platform. Version: December 2002, leee.
- [Pey95] Peya, Liu: Technological Basis forMultimedia Applications. Version 1995. EEE computer society .
- [Bur06]  
<http://www.feedburner.com/fb/a/aboutrss>
- [Gir05]  
[http://gps-tsc.upc.es/imatge/\\_Xgiro/teaching/courses/comav/index.html](http://gps-tsc.upc.es/imatge/_Xgiro/teaching/courses/comav/index.html)

## Parsing from RSS feeds to MHP-television system

### [ETSI]

ETSI EN 301 192: DVB Specification for Data Broadcasting  
ETSI TR 101 202: DVB Implementation Guidelines for Data Broadcasting.

### [ATSC]

ATSC A/90: The ATSC data broadcast standard.  
ATSC A/91: implementation guidelines for the data broadcast standard.

### [TRDVb]

TR 101 154 - "Digital Video Broadcasting (DVB) - Implementation guidelines for the use of MPEG-2 Systems, Video and Audio in satellite, cable and terrestrial broadcasting applications".

### [Xle06]

<http://sourceforge.net/projects/xletview>

### [Open06]

<http://www.openmhp.org>

### [Dom06]

<http://www.jdom.org/>

### [Rome06]

<https://rome.dev.java.net/>

### [Int06]

<http://www.interactivetvweb.org/>

### [Hav06]

[http://movii.et.fh-osnabrueck.de/~kamo/projekte/mhp/mhp\\_tut/mhp/html/](http://movii.et.fh-osnabrueck.de/~kamo/projekte/mhp/mhp_tut/mhp/html/)

### [Jatv06]

<http://forum.java.sun.com/forum.jspa?forumID=36&start=0>

### [MHP06]

<http://www.mhp.org/>

### [Jav06]

<http://java.sun.com/>