**Introduction**

The way of recording information in these days it’s still a complex issue. There are several competing formats, such as JSON (JavaScript Object Notation) or XML (eXtensible Markup Language), but for this work, was imposed to us the obligation of using XML.

The XML is the basis for another file formats, such as XHTML (eXtensible Hypertext Markup Language), because of that, it became a language highly powerful.

Thus, we were proposed creating an XML properly validated by an XSD (XML Schema Definition), and is capable of managing a Sports Journal, which can be converted to other file formats, including HTML (HyperText Markup Language) and also with some filtering (for classification table for club, etc.)..

**Analysis**

The need to feel informed about the news of the world, meant that we had to find a viable alternative. For this, we created the generalist newspaper where all "news" would be printed on paper. After that, there were those who decided not to make a general newspaper, but by theme (sports, financial) in order to have more news related to that subject and to attract the target audience. With the evolution of technology, the paper starts to become into disuse and increasingly the supporters of the newspapers use the tablet or computer to access their favourite news site.

In this context, have emerged a necessity to have an organized information in order to provide improved tools, either filtering or research.

Each sports news, is written in a certain language may or may not be translated into several languages ​​(usually happens only in international newspapers), will also have its author news, which is usually a journalist associated with the newspaper, although there are also those who compose opinions public and the same appear in the newspaper. In case the date, it was decided to put the date of issuance of the news and not the date of the written notice.

Regarding the content of the news, it can be said that each club has its news, which in turn has a rating associated with that participates in the league which is a particular sport, and the club an office in a particular country. Still on the club, it can be said generally that a club has an acronym associated, for example, when we say SLB, clearly referring to the Sport Lisboa e Benfica, based in Portugal.

To facilitate research and to be easier to relate a story with each other, we chose to assign an identifier to the news.

**Language**

For this problem, each news has an id attribute (in this format “AAAAA-9999-9999”, the AAAAA can be any letters A-Z with a variable size from 2 to 5, the 9999 can be any number from 0-9, with fixed size of 4, the first set of 9999 we choose to enter the year of publication of the notice). This same id will also be used to refer the related and similar news.

Each news will have the following elements: date / time, author, source, title, abstract, photos, club, sports, championship, news and other news body, which will be explained below.

The date / time is divided into 5 elements of type integer, as follows: year, month, day, hour and minute. We believe that there was no need to insert the seconds since sometimes that was omitted in the main information sources.

The author has two elements, namely the name and email address. The name is the type text, while the email contains a validation to verify if that email entered was correct.

The source contains two elements they being the source name and the URI of the source. The name is the type text and URI is the type URI.

The title, abstract and body of the news, are all of the same type. As we have been asked that news could be in several languages, these elements have sub elements that specify what language they are written. Was only possible to define news in Portuguese, English, French, Italian and Spanish. To add a new language, you have to add the type of languages the respective element.

The photo contains two elements diversified, the main image and other images, but with the same attributes, namely the href (will report the location of the image), width and height (the appropriate size of the image). The difference lies in the number of images that they can be placed, and there can only be one main image and an unlimited number of other pictures.

The club contains four elements, namely the name, short name, the country and the current position in the championship. The name element is the type text and element position is the type integer. The name small has a limitation that it can only accept 3 to 5 upper case letters only, and the country only accepts three capital letters.

The sport element is the type text, and there is no restriction.

The element championship has an attribute location, which should be placed the code of the country in three capital letters, and the element name is the type string.

The element other news, was subdivided into 2 equal elements, the related and similar news. In both cases, they accept the element related or similar (as the element we're talking about) with an attribute, the id of the news that we are referring to, and complies with the id of the news.

**Transformations**

**Conclusion**