Abstract: - This article presents the Filimage system designed for the images automatic extraction and their textual comments available on the web. The problematic, the architecture overview, and processing steps are described. The implementation consists in three sequential processing modules. The decomposition of textual and visual objects is the proposed solution allowing a separate treatment. The contextual exploration method is used for the textual extraction. Text semantic analysis is based on hierarchical textual indices. Automatic Matching and Associating (AMA) strategy produce a short multimedia documents from the full-length web documents. We present the final web site results in the navigator, performance results management and the conclusion.


1 Introduction

In this paper, we propose a new approach based on a semantic method to determine the relationship between digital images and textual segments in web documents by using a natural language processing NLP techniques. The overall goal is to establish automatically the correspondences between the images and their associated texts. It is necessary to provide a powerful system, implementing semantics resources and offering advanced features so as to produce short documents from downloaded web pages. Our research is extending text extraction through image skills. The most important reasons for which users need supports are:

- Information quantity is high point expanding on Internet;
- Information variety forms available electronically is also increasing…

Indeed, a multimedia document is a combination of different components (texts, images, sounds, videos…) in order to ensure the optimal use of any heterogeneous resources. On the right, Fig.1 shows a web site example with diverse informations (title, navigation bottom, link, icon, logo…). There are different levels of use according to each support and user’s requests. For example, the professional information searching user is in front of a continual attempt to upgrade his knowledge. The Internet-based technologies are suitable tools to undertake automatic tasks.

1. By “images”, we understand the visual objects like, pictures, figures, photographs, tables, diagrams… integrated in document and accompanied by a textual legend.

Fig.1. Mars Exploration Rover Mission site web contains several objects (texts, images…).

Networks supports within an organization like (Senate, OECD, United Nations, UNESCO…) facilitate knowledge transmission and offer access to relevant information like reports, forecast statistics. A number of issues only allow filtering the digital image however, one of the most important justifications specified for the establishment of image matching and associating with textual comments computer’s networks is to provide a foster comprehension. Our system allows summarizing high information volume, which involves time-
consuming or is difficult to reread quickly; for instance, the full-length documents include images (diagrams, tables, photographs, mathematical formulas...) from various domains like astronomy, geography, economy, medical images management… The corpus is composed of full text reports, electronic newspapers, scientific documents, and any resources available on-line. We present the problematic, the system implementation, processing steps, and performance results. We conclude with the research current status.

2 Problematic

Numerous advances in automatic text-based extraction have been made [3], [5], [10]. However, further relevant aspects of image relationship and text are also to be observed. The computer science includes NLP, image processing, pattern recognition, and computer vision. The images and videos extractions have been based upon such information like color, shape, and texture. Many multimedia-indexing systems use textual data, such as captions and content descriptors. The images/videos collections are annotated with descriptive texts [8], [4], [9], [11]. Others search engines rely on keyword matching.

How can the whole relevant text semantic contents be captured? How can the textual and visual information extracted be organized? The textual component associated with a non-textual must transmit the same relevant informational content. The Fig.2 web page example makes the problem more understandable.

3 Solution

In this section, we explain our approach to solve the correspondence problem between images and textual comments[1]. The solution used for the web page extraction is based on its structural analysis.

Fig.2. In this web page, the relevant components according to user researches are the big Saturn image and its textual reference.

We must determine what the relevant information on this page is. Which image is, according to the user’s point of view, the most relevant to be extracted? How can the textual segment directly related to it image be determined and how can the extracted components be reconnected?

3.1 Design and implementation strategies

We attempt to develop a system capable of processing different modalities. It is designed to permit the organization and management of several types objects. The extraction result gives a certain order to the “space of the web”. Nevertheless, there are two important issues need to be dealt with the architecture. The system complexity is the automatic identification and the extraction. It must be realized for both components. The solution to this problem is to split up the components and to create connecting mechanisms with the results supplied. It is made up of operations set. We present an overview of system designed displaying each objects processing. Fig.4 shows on the next page, the treatment architecture. It involves three sequential modules:

- pre-processing,
- extraction,
- automatic matching and associating (AMA).

Fig.3. This figure represents the source code structure of the Mars rover web page (fig.5).

In this approach, the web document subsequent analysis is a solution. It is decomposed through the exploitation of the HTML data structure. The “Filimage” system treatment overview is oriented-texts and oriented-images. It describes the basic procedures, which have to be applied step by step to specific text and images markups. After the images extracting, we have to obtain their textual comments from the text semantic analysis. Then starts the automatic relevant images association with extracted textual segments in order to combine each relevant component. For this, requirements specification must be defined.

4 A Java code implementation.
Fig.4. The Filimage architecture.

We see in the extraction module how the images are identified (1") (Fig.4). The first text transition step is the pre-processing module (1). It is in charge of the reading of the source code, and then returning a result to the last module (2). It converts document content via data structure to a text any more markup (3). The textual segments are coming from semantic analysis (4). A connection between the two modules allows automatic matching and associating operation (5). After the document is decomposed, the objects such as images and texts are then stored into various databases. Another important issue needs to be dealt with is the images storage location so as to obtain the final site result. The extracted textual comments are also stored.

3.2 Semantic filtering

In this section, we focus on the most critical resources for text semantics analysis. The textual segments are obtained by Contextual exploration method [7]. This computational linguistics method is based on hierarchical textual indices research. The acquisition of the linguistics resources is realized thanks to indices, rules, and databases. Analyzing the linguistic expressions in the context allows relevant markers to be exploited automatically. The contextual exploration called CE enables extraction based on semantics (meaning) including a wide variety text’s specific relationships (time, space, causality, definition, etc.). We consider all the various textual information types from the diverse sources used in the Language Processing (title, text, legend, source indication…). In general, the problem is compounded by the fact that a word usage in any language is full of ambiguity; a same word can have many senses depending on the context. The localization of the markers is used in a context to resolve ambiguity. The analysis aims is identifying the relationship between indices co-presents in the same context. The texts semantic analysis also is done to understand which indices type are more precisely corresponding to visual component and must be taken as the relevant markers [7]. According to this point of view, an extracted segment related to image can be seen as it textual comment. The markers are localized in the context thanks to the released rules [6]. We see a context analysis example (Cf. Fig.5).

Fig.5. We see in this preceding screen-capture, the context including the relevant image, it legend and the textual comments. The indices used by CE semantic analysis are underline.

The relevant indicator is “A picture” and the associated indices: “shows” is on the indicator’s right. A rule allows localizing the relevant indicator “image” and the associated indices is: “are used by”.

A following, CE rule example:

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of the rule</td>
<td>RSBXXX</td>
</tr>
<tr>
<td>Released task</td>
<td>Filimage</td>
</tr>
<tr>
<td>Comment</td>
<td>type scheme: table, argument: following table shows…</td>
</tr>
<tr>
<td>Indicator Class</td>
<td>E1= Creating space</td>
</tr>
<tr>
<td>Condition1</td>
<td>C1= Indices Class</td>
</tr>
<tr>
<td>Condition2</td>
<td>C2= Indices Class</td>
</tr>
</tbody>
</table>

This Filimage rule is including, the indicator class, the research space, the conditions and attributed tag to the sentence.

5. Professor J.P. Desclés developed Method.
6. LaLICC laboratory ContextO platform.

3.3 Automatic matching and associating module (AMA)
In this section, we describe the third module. The system designed must provide the response to the following questions: how the objects to be organized and how the objects must be communicating each other? The visual component use is ambiguous if it is isolate treated. Indeed, a system combining two information types is more efficient than addressing one of them. The automatic operation is AMA (Automatic matching and associating) the final design module. It allows solving the integration issues of a visual component (image) and a textual component (semantic content). This module is based on the object’s extraction and is use to visualize the combination result. The solution is the CE semantic-based retrieval joining together the extracted images into a unique posting result. The system deals of a table containing the components, and uses those components to design a new result web page. After the extracted images are stored, the extracted textual segments are use for being match with them. Such a process is appropriate to the sequential organization of the document space. This key success depends on the taken decisions at the extraction task resulting time. We must note that the linguistics markers are participating to held coherence between the various textual and the visual components. The image matching strategy is based on the context constraints. The last design relies is the fusion between the extracted images and the textual comments.

3.4 Management options
We must note that after the analysis, the extracted components of the initial web page are transferred to another new page [2]. The system includes a database organizing the extracted images with an elementary interface to permit user access to the system. Therefore, it opens automatically extra results window displaying. The final web site in the navigator includes the posting page "result" with two visualization management options:
- "Visual_Image" window, images can be visualized by automatic scroll. As such, it allows an image to be used in many different ways (indexing…).
- "Plan" is another option built by using the relevant information is stored (titles, subtitles, images, plan...). The extracted images are organized according to the initial page structure. The results pages include the title and the subtitles, which are accompanied by images helping fast user’s comprehension.

Fig.6. This screen-capture is a part of “Plan” window and scrolling "Visual/Image" window integrated

4 Results
All the images and the textual comments are extracted as shown following (Fig.7). These two examples facilitate understanding the off-line results analysis. The first screen-capture below is a new web page built according to the initial page. This animated image is also extracted by the Filimage system.

Fig.8. A representative result (on the right) is obtained by Filimage system from the initial page on the left.

The following screen-capture on the next page shows an initial French senate report, on the left. The performance automatic association result (on the right) is carried out by images and textual comments.

8 www.senat.fr/rap/r98-44/r98-414_mono.html
9 www.ccrs.nrcan.gc.ca/ccrs/rd/ana/ultrafine/overview_f.html
5 Conclusion

In this paper, we presented an innovative “image-text oriented” approach. This system is building an automatic extract, thanks to the AMA module capable of summarizing any digital document. It is based on a combining automatically image identification and NLP techniques by using text semantics.

This work contributes to extend text automatic extraction system to the image information retrieval. The solution proposed is designed and implemented under the postulation; such association provides further communication through the relevant and higher qualities information.

The image full expansion on the net, and especially multimedia contents generate new requirements for effectiveness access to information. Indeed, the images and their textual comments constitute a semantic entity. The interaction activity engenders a deeper cognitive understanding. The key point of this system is to provide a high level of objects integration according to the user’s point of view. By creating an automatic visualization and understanding support, we research to promote not only more effective information access system, indeed, we also include a fast image exchange information where the users are able to integrate relevant information required. We must recall the use interest of image-text combination for e-learning... The text semantic filtering thanks to the CE method linguistic markers makes appear the relevant aspect of web’s textual information and independently from any application domains. The extraction results deliver a certain order to the “Web space entropy”. The two options present the possibility of web pages accurate summarizing and the management results allow using the images, the texts together or separately. It provides the ability to associate images with any textual entities. The results show using this approach lead to exploit the advanced comprehension of the image and the text skills.

Knowledgements
Thanks are due to Professor J.-P. Desclés (Paris-Sorbonne University), the authors of the treated web sites.

References: