DIGITAL PRESERVATION: THE RARE AND UNIQUE'S LONGEVITY

Antonia Arakova & Eleni Mammi
National Library of Greece
Harlem Ministry of Foreign Affairs

Using the period of recent rapid changes in the information environment, the outstanding existence of new technologies makes its appearance and acts with the main aim to support, assist and demonstrate information services' content. The principal goal of the project is to present an important piece of the work of two libraries, National Library of Greece which create digital objects and conducts and the management and the utilization of valuable information resources. On the first hand, National Library of Greece, the biggest public information entity and on the other hand, Lamberts Press S.A, which consists one of the biggest papers publishing houses in Greece with a great number of publications in its catalog. What poster shows is the implementation of two new softwares that enforce the full management of old newspapers and magazines using technologies such as OCR and digital documentation. These software deal with problems, which have to deal with the reality and the decay of old and valuable printed material. Furthermore, the data can be implemented to different technological platforms and they have the capacity to estimate with multilingual tools which need to be preserved and accessed by different kind of users from all over the world.

Step 1: Scanning

The creation of the digital collection from the printed material and the microfilm is performed. Images are in jpeg and tiff format through high digital analysis are produced. Later, these images are coded and archived as a product of this specific work.

Step 2: Article clipping

Article clipping segment is the creation of individual article images with the aim to be retrieved easily one by one from the whole page. During this phase, all articles are marked and catalogued especially in order to be used afterwards through a specific search program and the user interface, to be inserted in the digital collection in the whole or part of the requested article.

Step 3: Metadata

Essential aspects of the newspapers and magazines digital library technical process

Digitalization of the material includes six high level steps, made through the consequent quality and security control of data.

Step 4: Optical Character Recognition (OCR)

Using suitable and standard optical character recognition software, three libraries that have already been tested and automatically recognized. For this purpose, the user will be offered the segment relationships that exist in the page layout of the newspapers and the newspapers pages.

Step 5: Search

The search approach is to provide users with a variety of search tools which will help them to browse in an easy and effective way their information needs. The search, when made in a digital environment, produces a document set that can be searched electronically, in a combination of full-text and Boolean search and is a means to produce a single document set or a set of results. For the achievement of all these above mentioned, the search environment offers some characteristics such as:

1. The most extensive coverage of the Greek Language - various types known.
2. The largest support of various levels of searching sophistication.
3. Provision of simple ways as in the results of searching to be presented.
4. Greek and multilingual support.

Step 6: Access

The end users have the capability of accessing to the whole content and look to the requested article, which is masked in an appropriate way.

Technical Description

Scoping: Data: National Agricultural Text

Method: Strategic

OCR: Optical character recognition

Hypertext: Hyper text

Database: Databases in different types of writing

Access: Access to digital material under the management of the database

Chances - Possibilities

1. Mixture Greek and Latin characters
2. Automatic translation of words
3. Creation of specific symbols
4. Location of change of word, department of word and line of characters
5. Comment on automatic layout
6. Simultaneous production of different cases - segmentation of lines
7. Creation on each printed LSC file that will contain comments of beginning, code operators etc.
8. Batch Job Page images
9. Automatic generation of page in text and graphic
10. Automatic localization of line errors
11. Creation of control and index concrete forms (eg forms that include 1 article in 1 column)
12. Comment on automatic layout
13. Thumbs Through Electronic Pages of Form

Importance

1. Extension of additional functional tags for the recognition of numerals, characters, symbols, or similar notations.
2. Special chronologically database
3. Caching of electronic information services
4. Recognition and distribution of content of Modern Greek forms

The FIRST MODERN 11th (9th & 10th) and 15th CENTURY MAGAZINES AND NEWSPAPERS DIGITALLY METADATA METHOD PRESERVED

MYLAUS (1810 - 1813)

EUROPEAN PRINTS (1840 - 1847)

LIAFET (1848 - 1848)

ATRAKOS (1849 - 1852)

POLYTHIA (1890 - 1891)

THA (1892 - 1895)

TO MPA 1895 - 1902

To VIMA 1952 - 1954