Kosmopolis: digital content in Greek language

Gavrilis Dimitrios, Georgiou Panagiotis, Papadatou Fieroula, Tsakonas Giannis*

Library & Information Service, University of Patras, Greece. {gavrilis, panos, fiori, john}@lis.upatras.gr

[* Corresponding author]

1 What is Kosmopolis?

1.1. Project description

"Kosmopolis" is a digital collection of twenty (20) Greek periodicals from the mid–19th up to the beginning of the 20th century. "Kosmopolis" is being implemented by the Library & Information Service, University of Patras (UoP) under the project title *"Digitization and disposal of scientific material in the Greek language"*. The project partners are the Department of Literature (DoL), UoP and the Hellenic Literary and Historical Archive (best known as ELIA), Athens.

The basic objectives of "Kosmopolis" are

— to extend the public and scientific community access to rare and valuable content,

— to enrich the digital content in Greek language,

— to contribute to the preservation efforts of historical and valuable material,

— to create and promote new research challenges in Greek literature and history, through the scientific processing of the digital content and

— to create and to develop new digitisation practises in the Greek informational landscape, through methodological and standardized processing of the digital content.

1.2. Collection description

The collection consists of more than 85.000 pages. The first issue of the collection was published in 1834 (Ionios Anthologia). The last issue was published in 1931 (Noumas). All periodicals focus on humanities and specifically on literature, sociology, philosophy and history. By now three (3) periodicals are available, a sum of more than 9.000 pages. A considerable number of others are under processing. A complete and detailed list of the periodicals is presented in Table 1.

The content is mainly textual, but as in any periodical the texts are accompanied by pictures, photographs and other forms of images. The periodicals have two different kinds of folding, quarto and octavo, and they vary in height from 21 to 32 cm. This raised the issue of pictures' resizing when digitised, in order to have a consistent form of presentation and user interaction. This resizing is done "on the fly" depending on the resolution of the user's screen.

Table 1

Title	Period	Volumes	Issues	Pages
Poikili Stoa	1881–1914	16	—	7.290
Euterpi	1847–1855	8	192	4.610
Chrysallis	1863–1866	4	96	2.934
Ionios Anthologia	1834–1835	1	5	1.218
Apothiki ton Ofelimon Gnoseon	1837–1844	8	96	1.565
Apothiki ton Ofelimon & Terpnon Gnoseon	1847–1849	1	19	316
Nea Zoi	1904–1927	14	108	4.000
Grammata	1911–1921	6	47	3.040
Alexandrini Techni	1926–1930	5	60	2.086
Argo	1923–1927	4	19	726
Melissa	1886–1887	—	2	392
Chloe	1869	—	—	400
Eklekta Mythistorimata	1884–1894	11	—	7.200
Imerologio Skokou	1886–1918	39	—	15.000
Asmodaios	1875–1880	—	342	1.428
Hestia	1876–1895	—	996	19.228
Deltion tis Hestias	1877–1889	—	808	
Ioniki Melissa	1850–1852	—	34	388
Mi Chanesai	1880–1883	—	598	4.574
Noumas	1903–1931	—	803	10.720

2 Periodicals selection criteria

As in several other digitization projects, "Kosmopolis" was governed by a methodological approach concerning the selection of the periodicals that would be digitized. The responsibility for the selection was shared between the ELIA and the DoL teams. The criteria applied were the following:

Availability and completeness of periodicals. As long as these periodicals are maintained in the original printed format, they suffer impairments, due to the wide use and the passing of time; issues and even volumes are missing; paper deteriorates; pages are torn. These periodicals are rare and hard to access, because their full series are dispersed in public and private libraries in different areas in Greece. Researchers usually encounter problems of traveling time and costs, while when they finally locate the material they are cannot reproduce it for further studying. ELIA is holding complete series of these periodicals, being one of the most distinguished agents for the preservation of Greek literacy and press.

Good physical state for convenient digitization. "Kosmopolis" aims to be a long–term project. It is planned in a flexible form in order to extend its developing further than the formal temporal limitations. A serious consideration of the developing team is the transition to html format, through optical character recognition. Thus it was decided to select journals in good physical condition that would be easily and quickly digitized, in an attempt to minimize the scanning time and that would support a high contrast scanning for a later OCR transition stage.

The prominent role they play in the Greek periodical literature of 19th century. Apart from their historical and cultural value all twenty periodicals are of vital importance for research and education purposes in Greek literature. "Kosmopolis" aims to be the publishing vehicle of an underestimated, yet important, part of the Greek literature and press. "Kosmopolis" merges the different backgrounds and attempts to resolve the needs of each partner, such as the preservation of cultural heritage, as it is expressed through ELIA, the research and educational needs of the Dept. of Literature, UoP, and the need for provision of useful and quality digital collections to the interested user community for LIS.

3 Cooperation framework

There are three partners in the project. The first partner is Library & Information Service, Univ. of Patras (LIS), the second is Hellenic Literary and Historic Archive (ELIA), and the third partner is Department of Literature, Univ. of Patras (DoL). All partners share a different background and a need for the establishing of a cooperation framework emerged.

LIS is the project management partner. LIS has specified the project requirements, regarding digitization, and is responsible for system development and hosting. After a comprehensive testing with Greenstone DL Software, LIS decided to move forward by addressing the problems of the project with the development of an "in–house" collection management software. Another task of LIS is the supervision of cataloguing. The DoL partner makes the cataloguing and LIS is supervising their work, by providing guidance whenever is needed. Within the frames of this guidance and support the LIS contribution is extended to the provision of a set of administration tools, produced by the Automation, Research & Development Department of LIS, to the personnel of DoL. This set of administration tools is described in section 4 .

The ELIA team was responsible for the content provision, while both other teams participated in discussions. The most important task of ELIA is the content digitization, according to LIS specifications. ELIA is digitizing each periodical page separately, on a 1:1 proportion and they store the pages in DVD–ROMs. The stored material is delivered to LIS to be further processed.
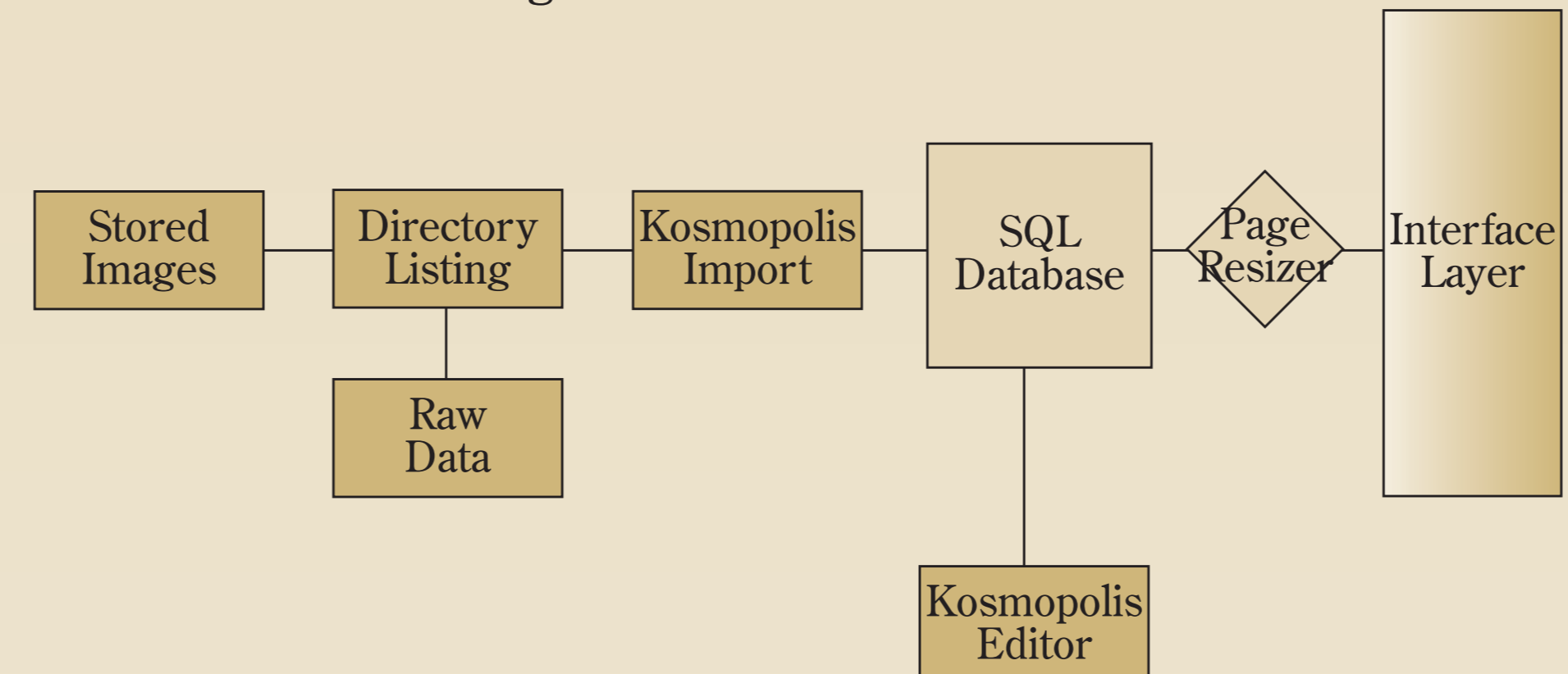
The DoL team consists of two professors and four postgraduate students of the department. They all have good knowledge of the content and they are ensuring the scientific integrity and the user–centered design of "Kosmopolis". Both professors are responsible for the scientific supervision and they participated actively in content selection. The four postgraduate students are the cataloguers of the material. Finally the DoL partner is providing scientific documentation on the digitized material.

4 Administration tools

The personnel of Dpt. of Literature are responsible for the cataloguing of the digitised material. Their information management skills were a serious consideration. In order to skip this problem with minimum cost of time and effort, the Automation, Research & Development Department of LIS provided a set of two tools for the import of metadata in an easy form and their editing, *"Kosmopolis Import"* and *"Kosmopolis Editor"* respectively.
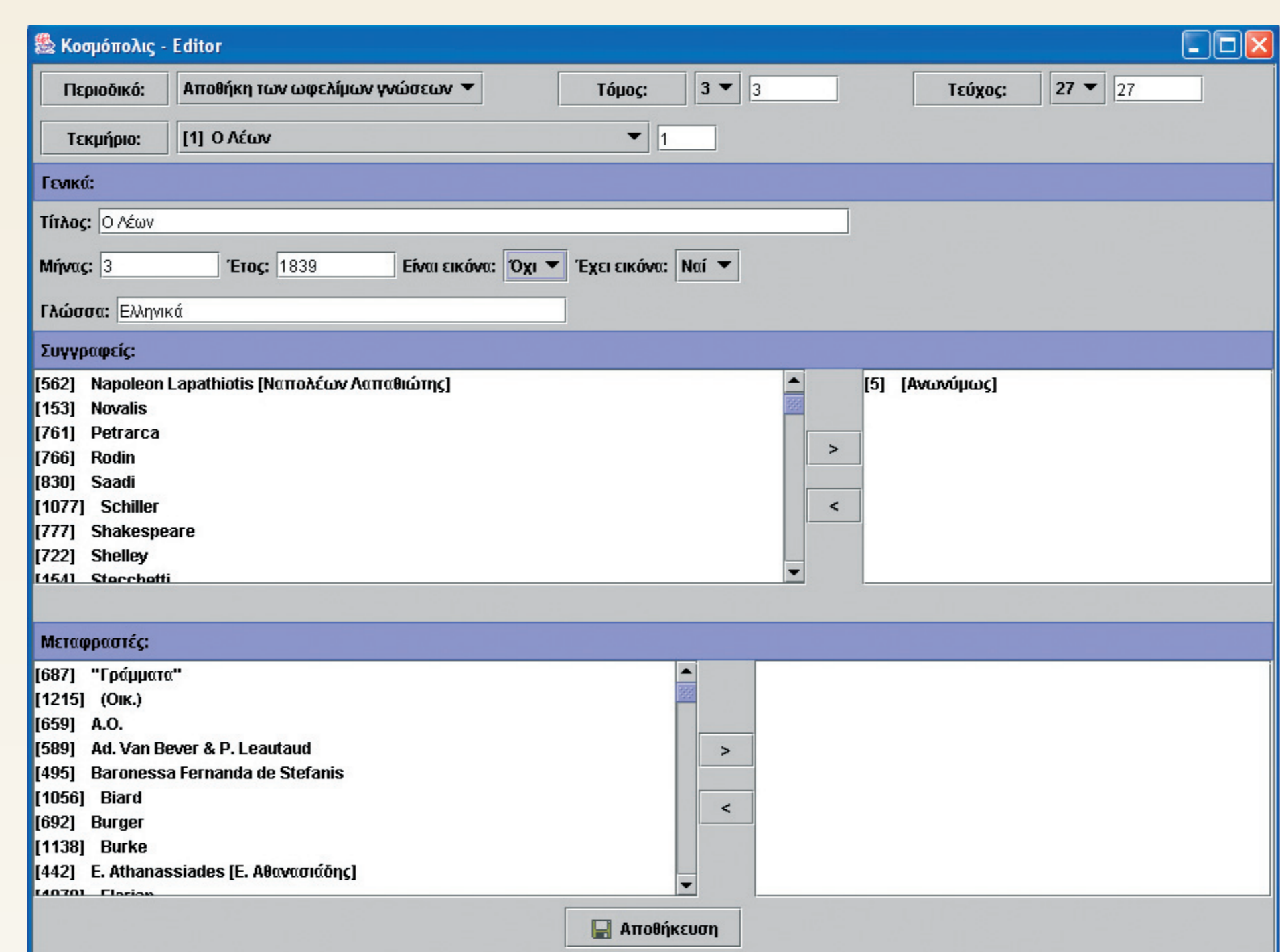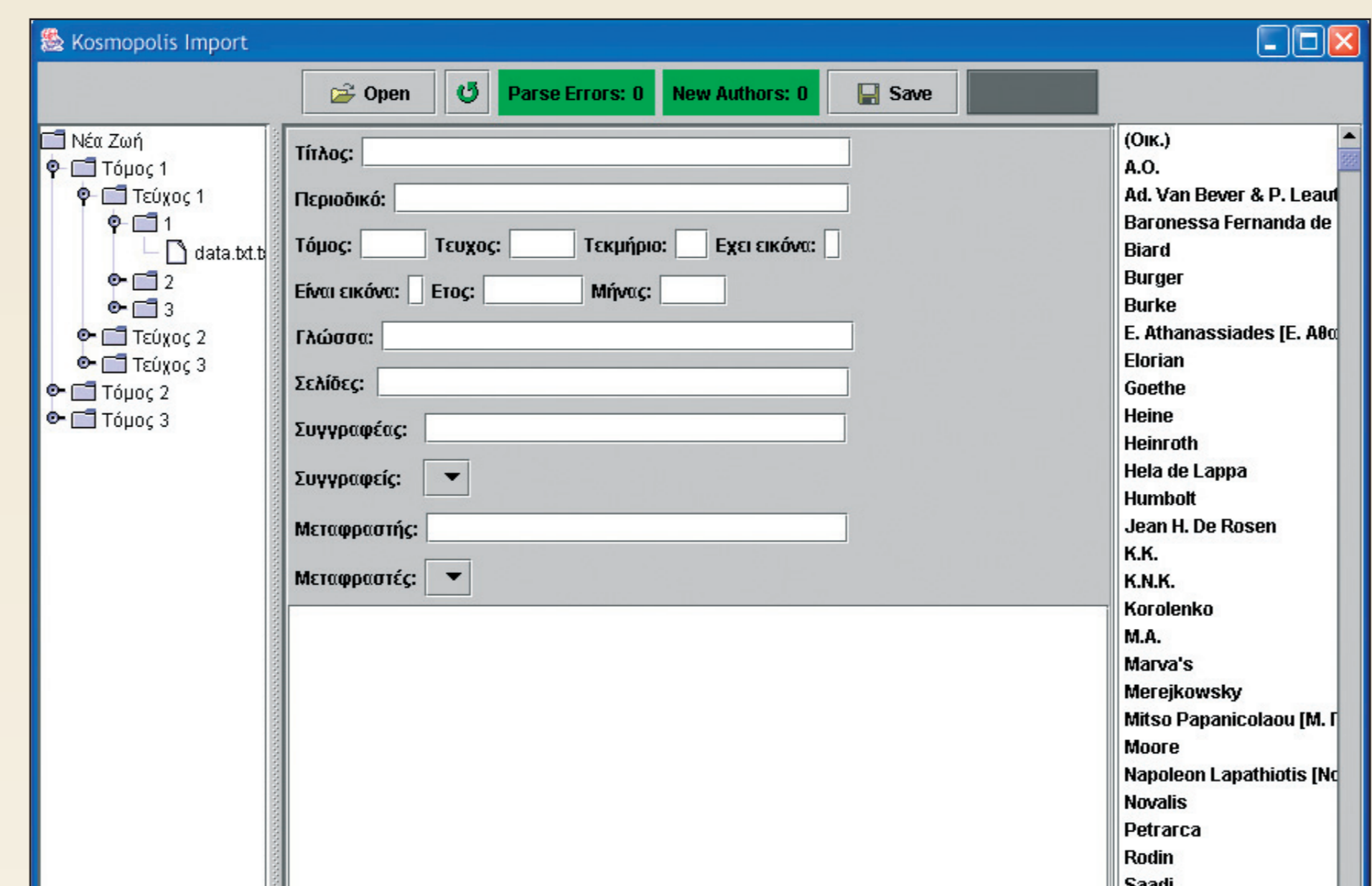
A short description of the cataloguing process in sequential order is the following:

The DoL team is separating the stored images and place them in their relative directory, which is structured by periodical title, volume and issue. The next step is the cataloguing of the material in raw format (txt). Each periodical is assigned to a simple .txt template, according to the metadata that are required.

The metadata are imported in the SQL database with the support of an application, *"Kosmopolis Import"*, which parses the raw data and sorts them in proper predefined fields. In parallel the application is connecting the metadata to the images in the relevant folder of the directory. The user of this application is able to trace in an easy way every possible mistake, e.g. the number of pages isn't corresponding to the actual number of pages in the folder, and to connect the author of the article with an authorized record of the author's name.

The data in the SQL database can be edited with the help of another application, *"Kosmopolis Editor"*.
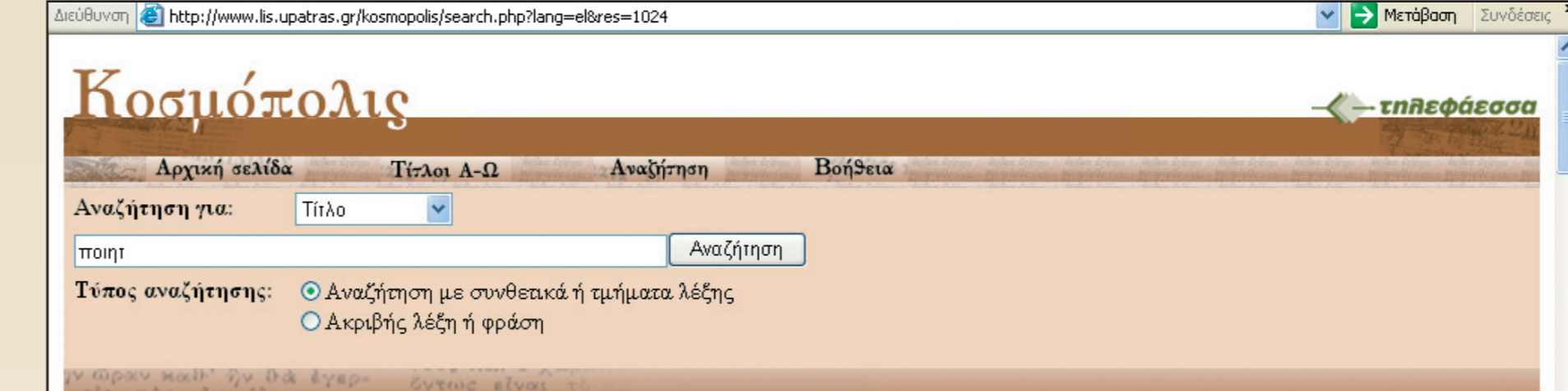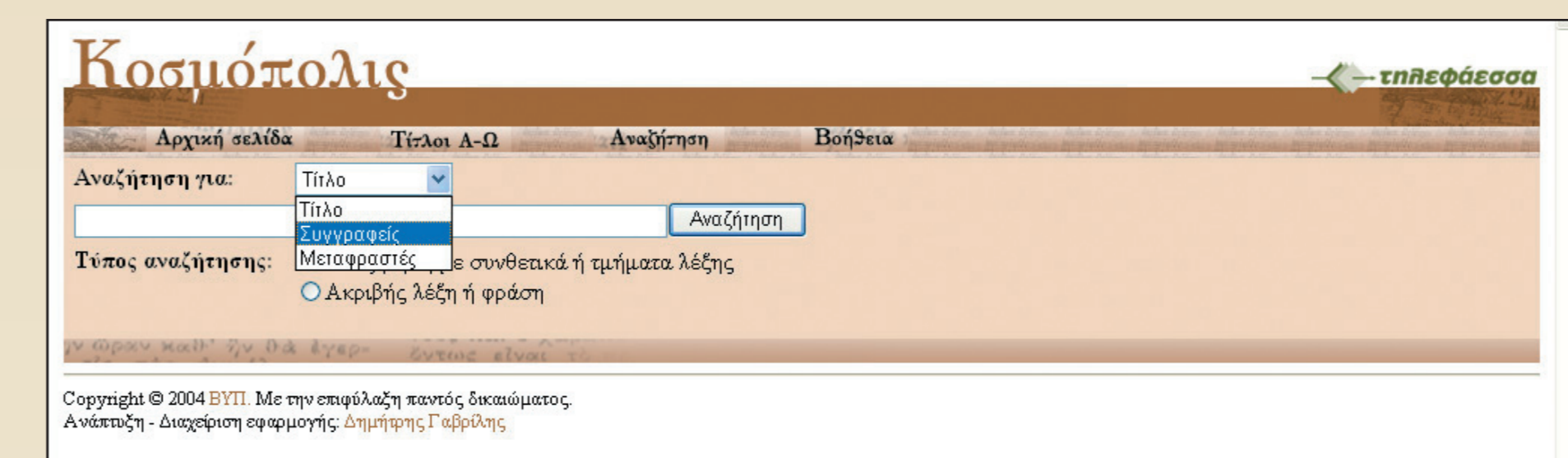

5 User centered design

There was given major importance on the design of a usable interface that would promote a rational and successful interaction. Apart from the user centered design culture that governs the *"in–house"* applications built, this decision was dictated by the features of the user community and the aim was to simplify the interaction elements and to constrain their number to the necessary ones.

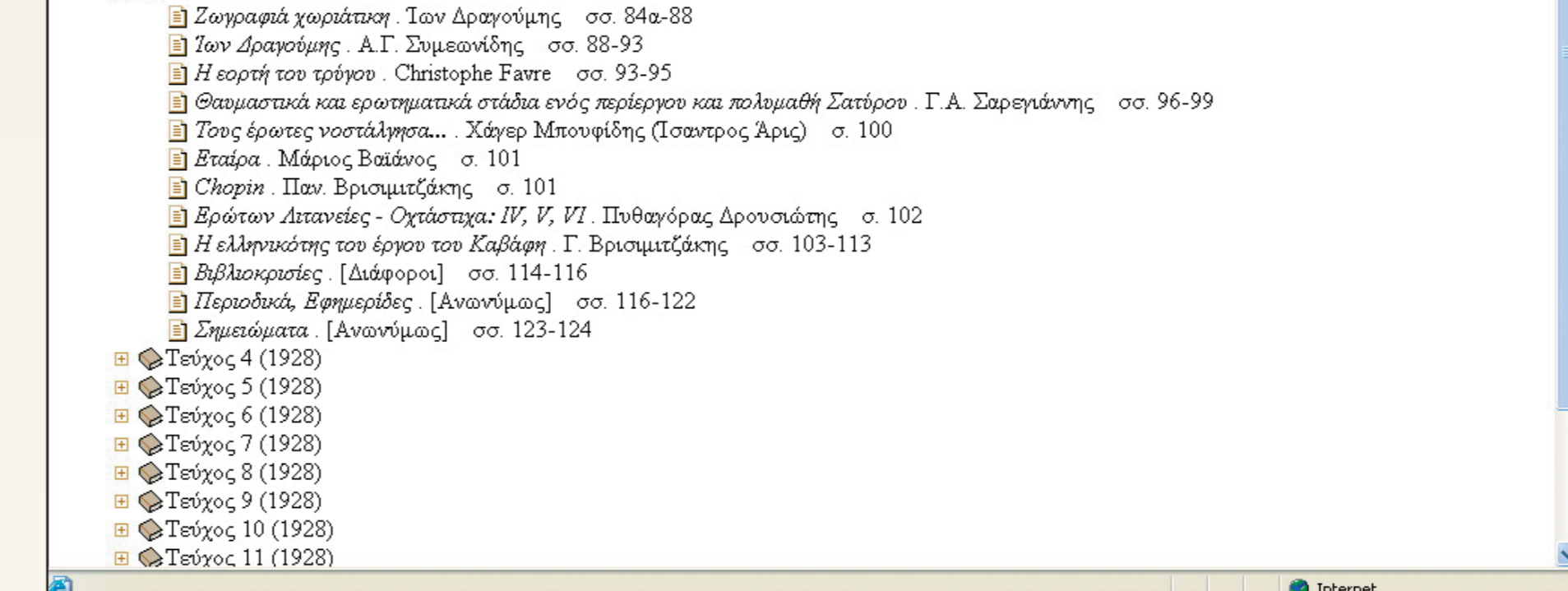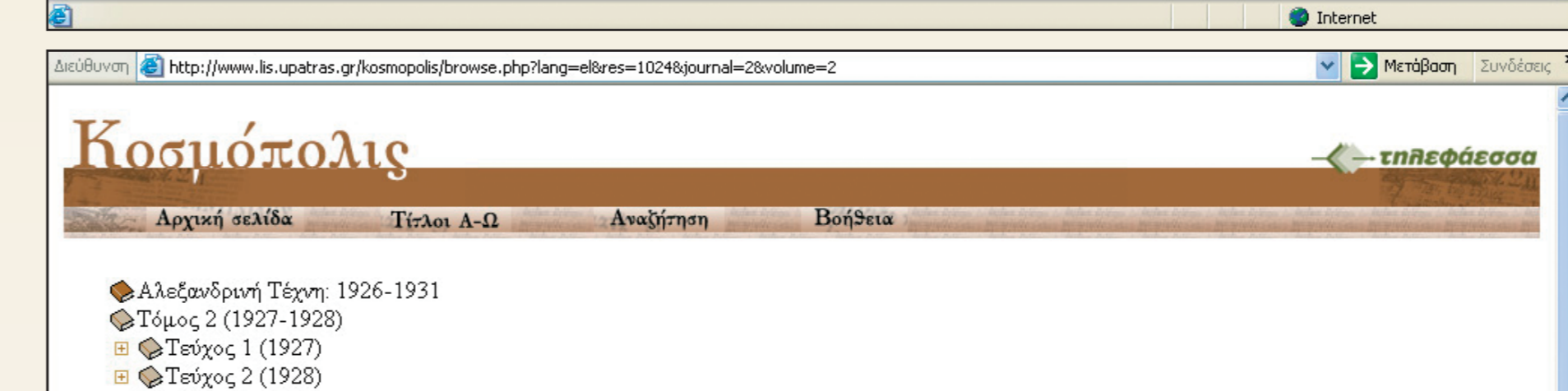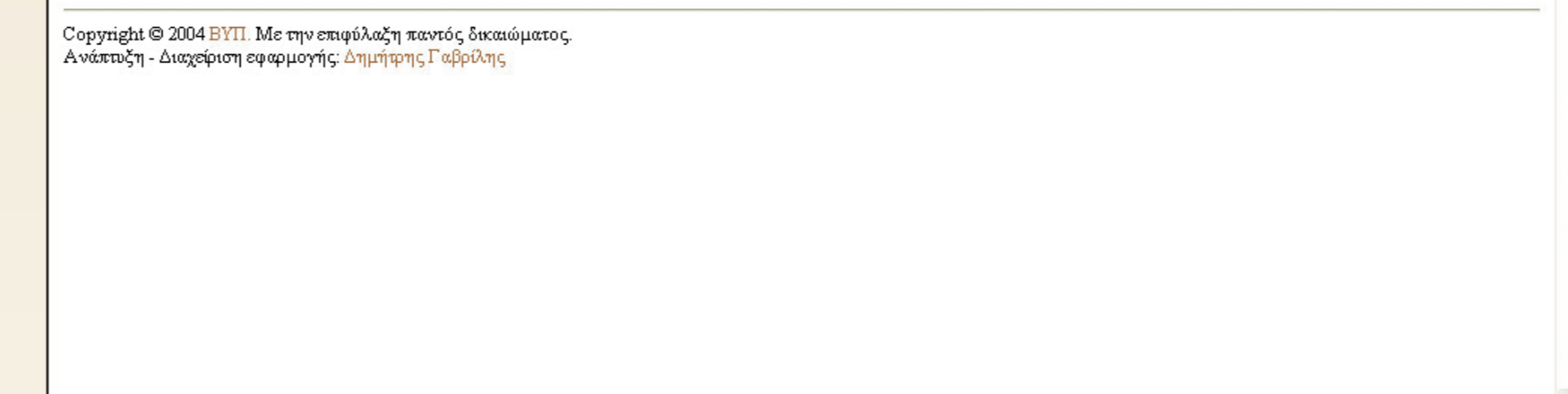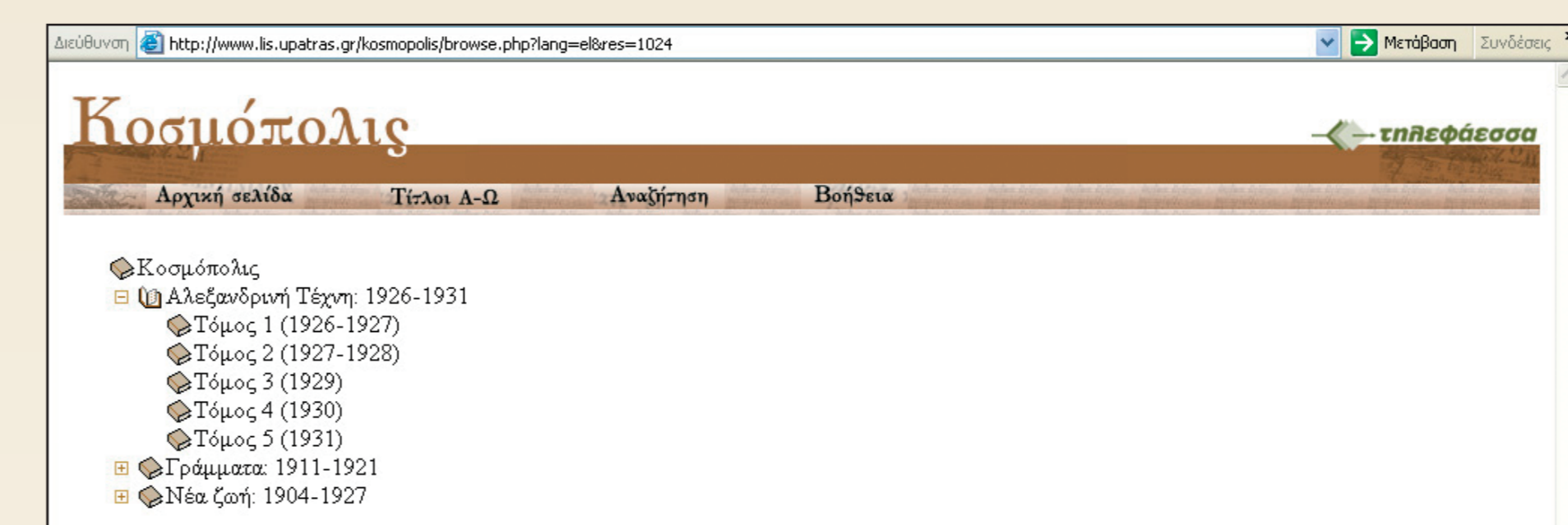"Kosmopolis" supports both techniques of information retrieval, searching and browsing.

Search:

Searching in the collection of "Kosmopolis" can be done by entering the exact word or phrase or a part of it. Truncation is supported without entering any special character or symbol.

Browse:

The user can browse the collection with the help of a hierarchical index. The first level of information is the periodical title and the second level is the volume of each selected title. The hierarchical index is refreshed when the desired volume is selected in order to minimize the constant extension and to shorten the webpage height. The new screen presents all the volumes of the periodical. When the user selects a volume, then the issues are presented and consequently the articles of each issue with the addition of a short set of metadata, namely article title, author name and pages.

Page view:

The user can view the selected article in a new page. The upper part of the page is presenting the complete set of metadata in a simple format. This set of metadata is article title, periodical title, volume and issue numbers, date of publication, author name and page number.

The lower part consists of the preview index and the image presentation area. The preview index is a list of thumbnail images with the exact number of the each page, as it is in the periodicals' page sequence. Each selected page is highlighted by a different shade of colour and is presented in the area on the right. On the upper part of the preview index there is a navigational aid in the form of back/next buttons, helping the user to see the article pages in a linear form.

The viewing area is presenting each page of the article. Above this area is a toolbar, with three possible selections. The user can return to a previous level, depending on the way that he retrieved the article. If retrieved by searching, the user may return to the results webpage and if retrieved by browsing he can return to the table of contents of the issue. The user can request a magnified version of the image and retrieve it in a 1:1 proportion copy. Finally, the user can request a PDF version of the full article for storing and printing purposes.

Final remarks

"Kosmopolis" is an OAI compliant collection and you access it through the URL *http://www.lis.upatras.gr/kosmopolis*. The current version doesn't support bilingual web interaction, but an English version of the web interface is going to be implemented in the near future.