

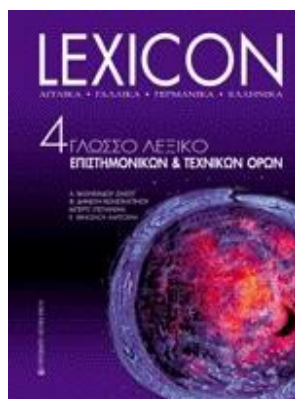
The LexiGraf web page

Dear Sir/Madam, thank you for visiting the LexiGraf web page. This page aims at providing information on a multilingual science lexicography project, currently taking place at the Aristotle University Thessaloniki Greece, along with the tools developed revolving around it. We hope that you will find the information contained herein useful to your research and commercial activities.

Abstract: The LexiGraf 2.1 software base has been developed under MS Windows to assist the creation of multilingual and bilingual terminology dictionaries in print. Any language combination supported by MS Windows can be accommodated to a maximum of 8 languages (single byte language sets only for the 16bit version LexiGraf). The system offers direct transition from terminology collection to the printing press; provides terminology management services, LAN use and excellent quality output to typesetters for subsequent printing-press reproduction. Multilingual dictionaries produced by LexiGraf deploy a term reference index technique, which allows their direct implementation in 2 volumes, regardless of requested number of languages. By combining available languages by 2 bilingual dictionaries are automatically produced ready for the printing press. Recently introduced features include process integration with OCR products for updating existing terminology resources and connection layers with popular DTP software products. The resulting database files can be utilised for electronic media publishing undertakings. Currently this product is being used in the Aristotle University - Thessaloniki Greece on a 4 language (Engl/ French/ German/ Greek ~ 50,000 terms/language) natural sciences dictionary + 10 bilingual ones + electronic versions.

Keywords: Desktop publishing, multilingual lexicography, bilingual lexicography, science terminology, electronic dictionaries

If you need to talk to someone on the described projects, try y dot hatzopoulos at mycosmos dot gr

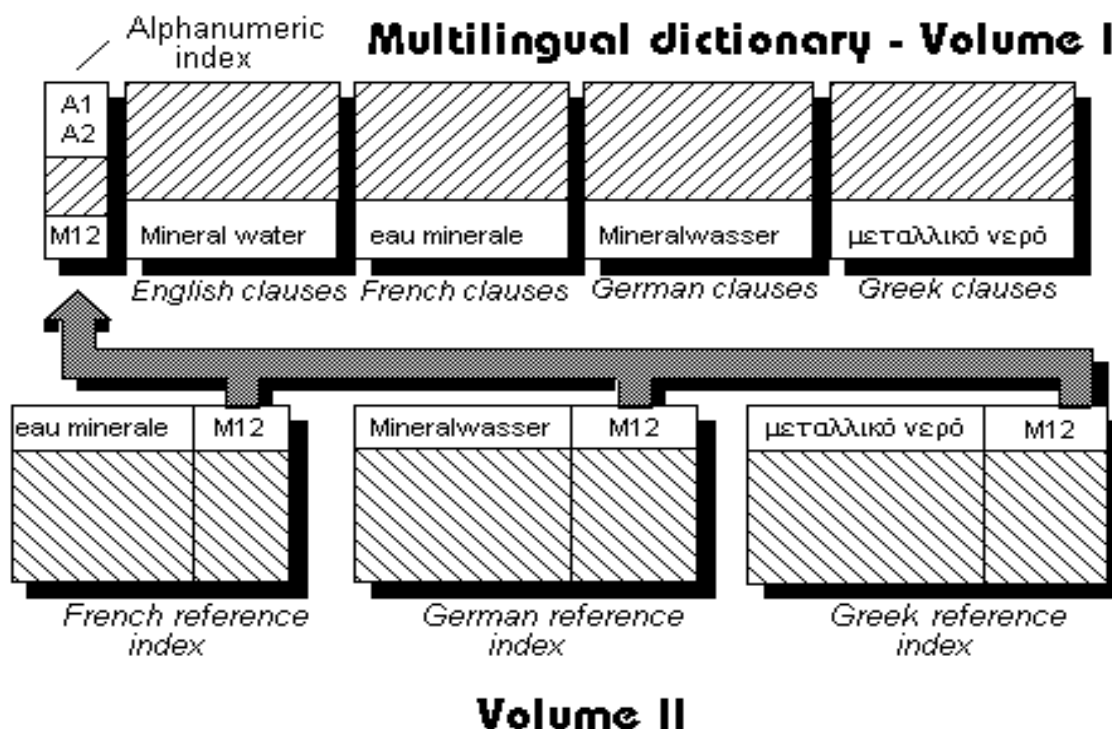


The project's background

LexiGraf 2.1 is a software base developed to automate the implementation process of multilingual and bilingual terminology dictionaries. Work on the LexiGraf project, was initiated under a software consultation assignment from the Foreign Languages Teaching Centre of the Aristotle University Thessaloniki Greece. A group of foreign language philologists from the above department is working on a multilingual (English- French-German-Greek / 50.000 terms per language) natural sciences dictionary. (c/o Mr Berts Stepanian internal postal box 160 Aristotle University Thessaloniki Greece - tel +3031- 99.78.62 / fax +3031-99.84.19)

LexiGraf was developed to serve the role of a 'smart' Desktop Publishing tool. What differentiates LexiGraf from other DTP tools is that it incorporates relational database management functions and is oriented at multilingual dictionary development. In other words, the users enter their terminology material into LexiGraf databases and from that point on, LexiGraf automatically takes care of dictionary processing chores, page layout, synonym indicators, type facing etc., thus delivering dictionary pages to the end-user, ready for subsequent printing-press reproduction.

The finalised printed dictionary product follows a layout structure based on reference indexes, indeed quite common in multilingual dictionaries published world-wide. In the Greek market, such kinds of dictionaries are commonly referred to as dictionaries with alphanumeric indexing. Any multilingual dictionary can be produced in a two volume layout structure through use of computer generated term reference indexes.



Volume I contains term translations in all of the available languages sorted according to English equivalents; volume II contains access reference

catalogues for the remaining languages other than English, indicating each respective term's location in the multilingual dictionary. The result of the process is a dictionary with good use of printing space that allows access to and from any of the available languages - all on paper.

Printed bilingual dictionaries can be automatically produced out of the multilingual database material by combining the languages by two - taking into account, if requested, the disciplinary content of the terms (glossary or focused dictionary implementation). Automated pagination is performed accordingly, thus delivering a product which presents a sound marketing asset with no additional effort. The process can lead directly from terminology collection to the printing-press, with significant savings in time, labour and expenditure.

Output configuration

The desired output of the terminology material has been the following:

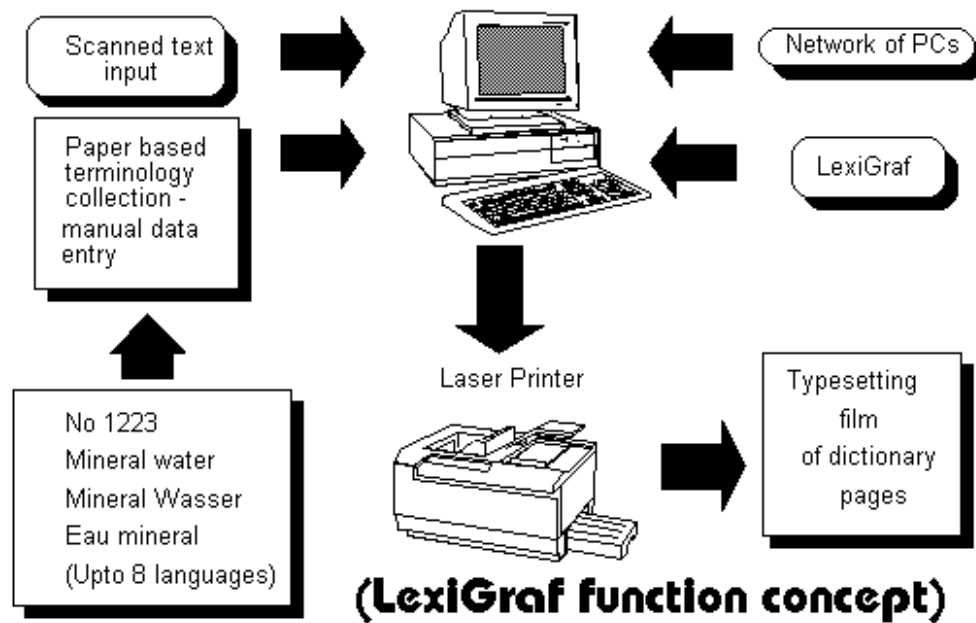
1. Multilingual dictionary in print (using a reference index technique as described below)
2. Bilingual dictionaries in print (full listings - combinations of languages by two) E.g. Engl/Greek + Greek/Engl dictionaries, German/Greek + Greek/German dictionaries, French/Greek + Greek/French dictionaries.
3. Bilingual dictionaries in print (restricted output according to discipline content)
4. Electronic version of the multilingual dictionary
5. Secondary markets - licensing the content to firms involved in machine translation as source for appending their respective knowledge databases.
6. Extension of the available terminology resources into additional European languages and replication of output forms (1)-(5)

Technical info

To this date, two versions of LexiGraf have been developed. One under DOS, which has been rendered obsolete by the advancement in PC technology, and one under Microsoft Windows.

The development tools in the latter case were Borland's Turbo Pascal for Windows 1.5 + Paradox Engine 3.101 + minor libraries from independent vendors. Currently there is a 32 bit UNICODE compliant version under development, using Borland's Delphi32 Client/Server for increased performance. (Windows'95 and NT native code)

The LexiGraf system operates under Microsoft Windows, allows up to 20 concurrent network users and uses Paradox database files which can accommodate a maximum of 800.000 terms for each language. It supports several independent terminology sets for terminologists working on dictionary development projects evolving in parallel. To function, each LexiGraf workstation requires 4MBytes of RAM and 5MBytes of free hard disk space.



Customised versions of LexiGraf can be delivered -within a 2 month notice- in any language combination for which Ms Windows TrueType fonts and related keyboard drivers are available, ranging to a maximum of 8 languages. Currently this option applies for single byte language sets.

OCR input

Currently, development work is being performed on combining the LexiGraf tools with optical character recognition software to recycle terminology material already existing on paper for updating purposes by the respective copyright holders. With the currently available technology the process does require human intervention, some programming work and subsequent proof-reading to be effective; Every existing dictionary presents its own intricacies and requires individual handling. Nevertheless, a cost reduction of 50%-60% compared to manual data-entry is quite often feasible for good quality texts.

The same applies for the time duration of the data import. After examining the given text's OCR performance, suggestions are made on the most cost effective data-import configuration, namely:

- Manual data entry (typing).
- In-house OCR and subsequent proof-reading.
- Assigning specialised third party associates to perform the task; or combinations of the above.

Final stage of the process is importing the data into LexiGraf databases for processing, updating, appending and subsequent repagination. Although LexiGraf supports a high speed internal printing interface and an excellent quality connection layer to MS WORD, an interface between LexiGraf and Quark XPress is under development using XTags. This feature enhances the DTP performance of LexiGraf, due to the flexibility and operation speed it offers. Diagrams and pictures can be easily imported to Quark XPress and integrated with the dictionary text, a feature useful to scientific publications. The publisher will have the ability of easily personalising the dictionary

product by altering typefaces, adding footnotes, pictures, comments etc. Quark Xpress will also serve as a bridge between LexiGraf and the Macintosh systems - which are standard in the prepress and publishing business. The Xpress documents can be easily ported from one platform to the other, even in a cross-platform PC - Mac LAN.

Additional Development

- **UNICODE support**
- **CLIENT/SERVER architecture**
- **Automated electronic dictionary application engine**
- **SGML import capabilities**

First market application of LexiGraf

The current terminology content being applied on LexiGraf consists of the following in mathematics, chemistry, biology, physics, geology etc., in English, French German and Greek:

Acoustics, advanced mathematics, aerospace engineering, algebra, analytical chemistry, analytical mathematics, astronomy, astrophysics, atomic physics, biochemistry, calculus, chemical engineering, climatology, communications, computer science, control systems, crystallography, electricity, electromagnetism, electronics engineering, fluid mechanics, food engineering, geochemistry, geodesy, geography, geology, geometry, geophysics, hydrology, inorganic chemistry, mapping, material technology, mechanical engineering, mechanics, metallurgy, meteorology, mineralogy, mining engineering, nuclear physics, nucleonics, oceanology, optics, ordnance, organic chemistry, paleobotany, particle physics, petroleum engineering, petrology, physical chemistry, plasma physics, quantum mechanics, relativity, set theory, solid state physics, spectroscopy, statistical mechanics, statistics, systems engineering, thermodynamics, vectoring.

As mentioned, apart from a two-volume multilingual dictionary, a set of bilingual ones is intended to be published in the following language compositions, containing varied combinations of the above stated discipline listings, according to Greek scientists' needs.

- English- Greek & Greek -English dictionaries
- French - Greek & Greek - French dictionaries
- German - Greek & Greek - German dictionaries

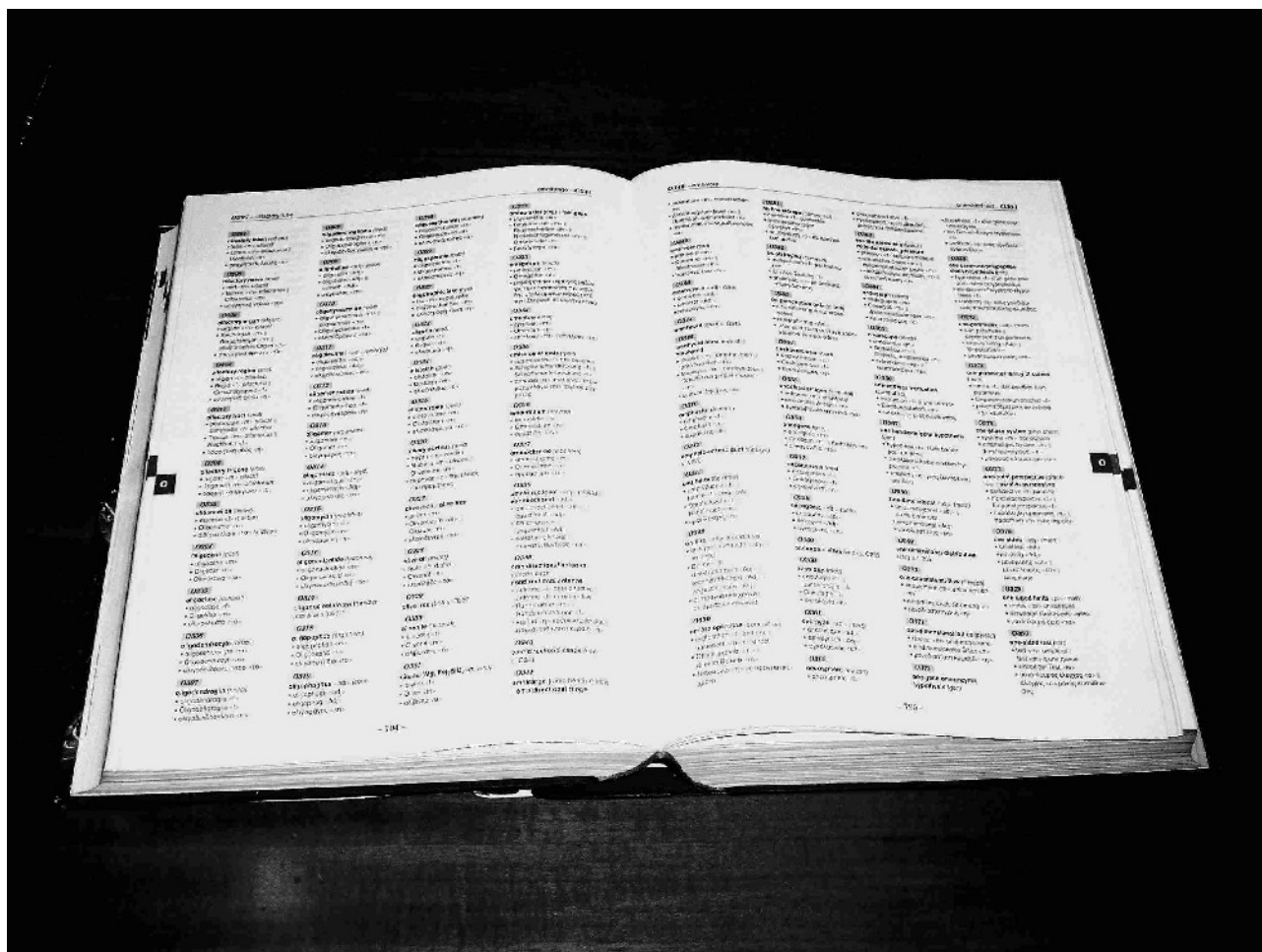
assay	English	French	German	Greek	ασσο
A1	assay balance (eng)	●balance </> d'essai ♦balance </> de precision	●Feinwaage </> ♦Analysenwaage </>	●αναλυτικός ζυγός <ο> ♦ζυγός <ο> ακριβείας	
A2	assemble <vt> (sci tech)	●monter <vt>	●zusammenbauen <vt>	●συναρμολογώ ♦μοντάρω	
A3	assembler (com sci)	●assembleur </m> ♦program d'assemblage	●Assembler </m> ♦Assemblieprogram </m>	●συμβολομεταφραστής <ο>	
A4	assembling (com sci)	●assemblage </m> de montage	●assemblieren <vt>	●συμβολομετάφραση <η>	
A5	assimilate <vt> (physio)	●assimiler <vt>	●assimilieren <vt>	●αφομοιώνω <μτρ>	
A6	association (chem)	●association </f>	●Association </f>	●συζευξη <η>	
A7	associativity (math)	●associativité </f>	●Assoziativität </f>	●προσεταριστικότητα <η>	

Alphanumeric index

Multilingual dictionary sample produced by LexiGraf

LexiGraf can produce any bilingual dictionary which results by enumerating the combinations of the available languages by two. The multilingual dictionary under development is mainly a reference tool for library use, and use by professional translators. After all, it is the outcome of an academic project. Bilingual dictionaries are the ones that would probably bear a market potential, since they will be far cheaper and will respond to the needs of a much wider group of scientists. A number of European firms have already expressed their interest in extending the language base. Evaluations of possible multimedia applications are also being made. The first commercial results on LexiGraf produced dictionaries reached the Greek market in 2004.

[A snapshot of this dictionary can be seen below](#)



LexiGraf was demonstrated at the Current Research Information Systems 98 event, hosted by the DGXIII of the European Commission in Luxembourg on March 12-14, 1998.

LexiGraf's presentation in C.R.I.S 98



Yiannis Hatzopoulos at a CRIS98 demonstration booth

LexiGraf was demonstrated on March 12-14 1998 at the *Current Research Information Systems 98* Event organised by the European Commission in Luxembourg. The event took place at the Batiment Tour complex, at the Kirchberg plateau. The Directorate General 13 of the EC provided the funding and the logistical support. Yiannis C Hatzopoulos was invited to demonstrate LexiGraf at a parallel demonstration session dedicated among other things to the provision of multilinguism in research information systems.



In the above snapshot the entrance to the demonstration facilities of the event is depicted.



Participants were invited to test the web-based demonstrations at the event's cybercafe



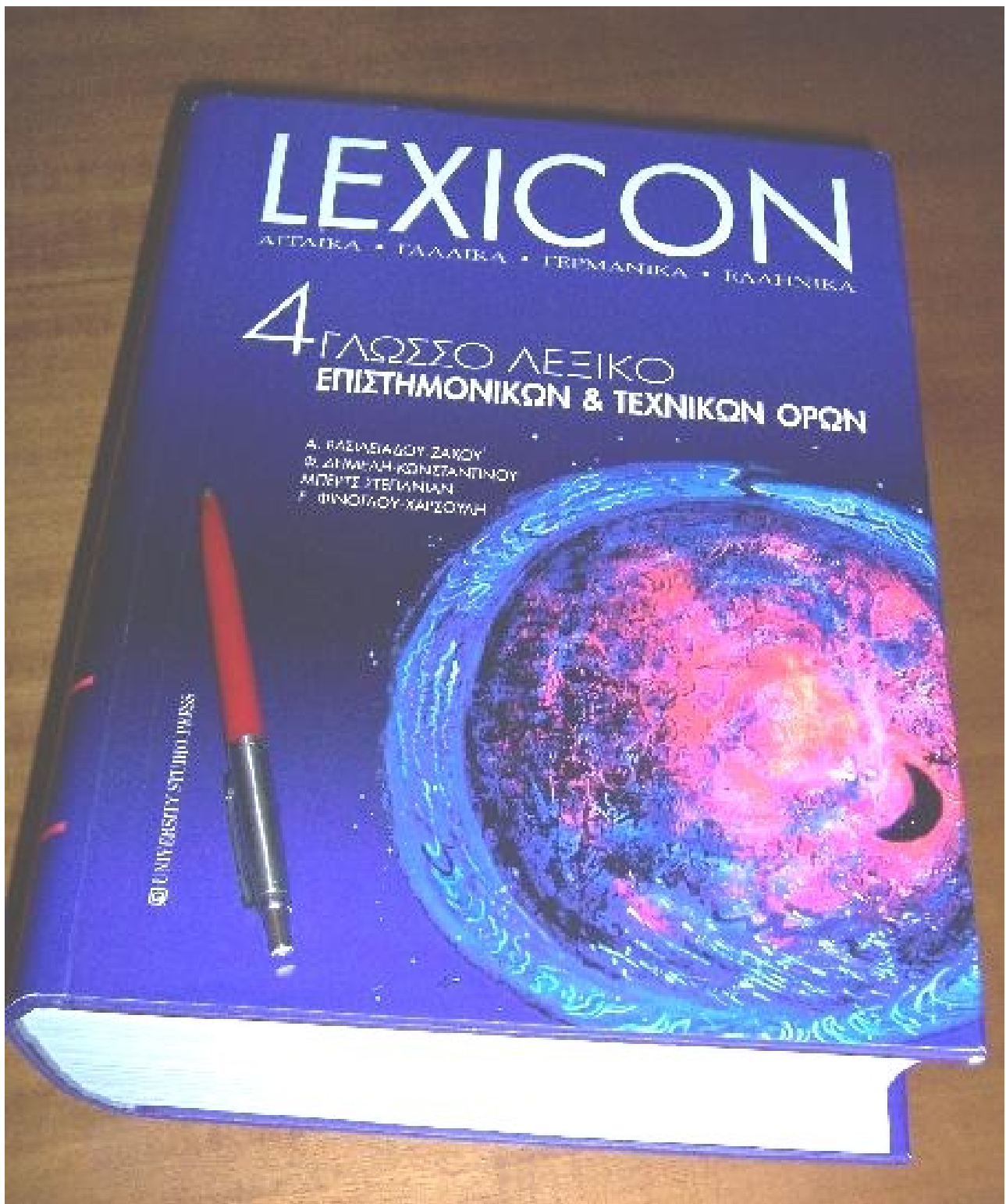
CRIS98 was completed by a panel discussion

The European Commission should receive every credit for the excellent organisation of the event, the quality of the selected speakers' presentations as well as the very interesting applications demonstrated.

Further information on CRIS 98 can be traced by browsing the CRIS'98 web page at www.cordis.lu/cris98

Lexicon: A four language dictionary of Scientific & technical terminology

A 2000 page, 4 language dictionary, developed using lexigraf has been published by University Studio Press SA with ISBN : 960-12-1276-0



[University Studio Press web page www.scientificbooks.gr](http://www.scientificbooks.gr)