TETRALOGIE: Software for monitoring Science and Technology

Bernard Dousset

Tool for competitive intelligence developed in the
context of ATLES (IRIT) – University of Toulouse
e-mail: dousset@irit.fr

Key words

Abstract

Tetralogy is competitive intelligence software with full text or factual data from bibliographic data bases, on line or on CD/ROM, from the Internet or any other computer source. With statistical methods and exploratory data analysis methods, it shows strategic information unreachable up to now in a short time. Information such as actors identity, mobility, emergence and evolution of topics and concepts, terminology, strong domains, clusters, what to read, where to publish, and with whom to cooperate. Tetralogy is a tool designed for scientific and technological watch.

1. Introduction

The TETRALOGY program, developed in the "Institut de Recherche Informatique de Toulouse" (IRIT) is one of the essential elements to the bibliometric station "ATLAS". This was a joint development with CEDOC and SGDN, involving many national French Ministries (CNRS, CRRM, IRIT,). This was done to take into account all the methods currently operating in this area. This program is available on Sun operating systems for SUN / OSO and solarium, and is available for evaluation at the network from both X terminals and PC or Macintosh. HP version was made recently available.

Figure 1.: IRIT Headquarters in Toulouse

This technological software is powered by bibliographic databases, factual or textual (online or on CD-ROM). It allows, through statistical methods and analysis of evolutionary data displayed in four dimensions, to show: the identity of the actors and their relationships, their sites of action, the emergence and evolution of issues and concepts, the strategic elements of industrial property (patents), the virtual areas of delivery, what to read and where to publish, and with whom collaborate.

Theoretical courses on methods and practical work on Tetralogy are organized in the context of continuing education at the University Paul Sabatier, and the assistance to users and trained.

1.1. Strengths of Tetralogy

- Available for PC, Mac, X and stations,
- They cane work from far away,
- It accepts all formats,
- Manages multi-based bodies, web-pages, news, patents,
- It takes into account, synonyms, and semantic processing
- There are many methods of co-operative analysis (ACP,AFC, CAH, CPP),
- Map factorials 4D,
- Interactive display (relationships, classes, backgrounds, 3D zoom, 2D and 3D maps),
- Studies based on the evaluation of the environment,
- Training, transfers and insurance supervision,
- User forums on the internet.

1.2. Featured users

- French Ministries (Defence, Interior, Finance)
- CEDOCAR, DGA, GIAT, SGDN...
- Universities (Paris, Toulouse, Marseille, Poities, Marne la Vallée, Strasbourg, Louvain la Neuve, Unam/Mexico, Unc/Colombia)
- Ecole des Mines de Paris, IPC Barcelona
- CEA, CNRS, INRA, INSERM, CIRAD, CTBA
- Airbus, Aérospatiale, PSA, France Télécom
- Instituto Mexicano del Petróleo, TotalFinaElf Sanofi, P. Fabre, L'Oréal, Bourjois, Nestlé, Idelux, Carmeuse, UCB chimical (Bélgica)

- Avant Veille, Delta Veille, Iale tecnología
- AVENIDO, Innovation, Gfi, Eurespace Pépinière Théogone

2. Structure of Tetralogy

TETRALOGY is composed of two modules:

- A system for handling the bodies of text that lets you manage files from downloading or questioning CDs. Also from extracting and crossing referencing and taking into account the specificities of each base and each formed.

- An analysis of the information contained in the matrixes, which is based on a 3D spreadsheet that involves specific and data analysis techniques on traditional and innovative fields, static or evolving bipartite. This system is supplemented by a module to aid the preparation of reports. Reports which provides analysis of the management of the numerical results and graphs generated by various methods
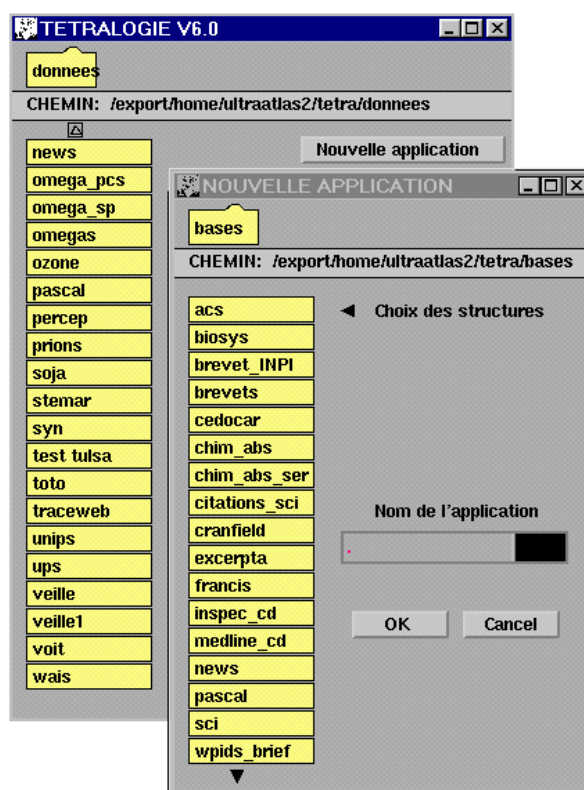
Figure 2: Sample GUI Tetralogy

2.1. Manipulation of the body of the text

IT allows the cross referencing bibliographic databases or information sources with a single application that performs on demand, between meaningful information (exogenous or endogenous) in the main text. This tool involves:

- Managing the running application
- Arranging access to the database
- The structures of the databases
- The format of the structure descriptor
- The development of filters and thesauruses

- The development of dictionaries of synonyms
- Techniques for cross referencing information
- Strategies for the treatment of Matrices

The management of running the application:

TETRALOGY allows a team of investigators or an engineer to manage in parallel, a large number of cases themselves cut-down into subsets corresponding either at various bases interrogated or cross referenced with different strategies of information or different levels. Furthermore, the analysis often leads to successive simplifications of what has historically been discussed in view of preparing a final report. A tree structure, similar to that of UNIX, has been adopted to organize the best possible management.

The arrangements for access to the database:

The databases or sources of information are almost all compatible with TETRALOGY, whether dream accessible by the switched network, the network specialist, or available on CD / ROM. Where formats are not directly compatible, UNIX programs allow his immediate transcoding (PC, MAC)

The structure of databases:

Each database has its own structure and we have to adapt what is found in different formats. In addition, the structure and content of a base depends on its outreach and support in which is implanted (online, CD-Rom). Also, we have designed a configurable descriptor of the structure of the bases that faithfully adapts to each case.

The format of the descriptor structure:

This descriptor defines the different fields of the database, identifying their flag (tag), separators, utility and various types of information that they contain. It can also locate the beginning of each note as well as the physical structure of the record

The development of filters and thesauruses

To retrieve information from non-coded fields (title, abstracts, etc.), or limit the search of information in a sub-domain (authors, who have seen at times, keywords, codes, years, country ...), it is useful to filter through all of the available information. We have described a module, to aid the creation of these tools, to generate files chosen either alphabetically or by order of decreasing frequency. Thanks to the editors, these files enable this preliminary but essential task.

The preparation of dictionaries of synonyms:

Certain details are semantically equivalent or hierarchical. Thus it is also very useful to have a hierarchical thesaurus in order to regroup elements that have been judged by the user as comparable, related or perhaps have been codified in an anarchic way by designers of the database (ex.: years, country, sources ...). Dictionaries of synonyms are used at various levels in Tetralogy for hard to find information, to decode refined analysis.

Techniques or cross referencing information

Information can be cross referenced between two fields in order to obtain matrices of frequency, the presence - or absence co-occurrence on which the analysis is carried out. For the bipartite or evolutionary data, a third field can be introduced. This is to operate the array in two, three or four layers, so as to analyze the differences induced (absolute, relative, speeds, accelerations, implosions, explosions ...).

Strategies to purify the Matrixes

Filters, referring to the frequency or the rate of information, can be applied automatically in order to reduce the size of the pictures obtained. This simplifies the work of analysis and reduces, significantly, the calculation time and complexity of the factorial maps and associated classification trees.
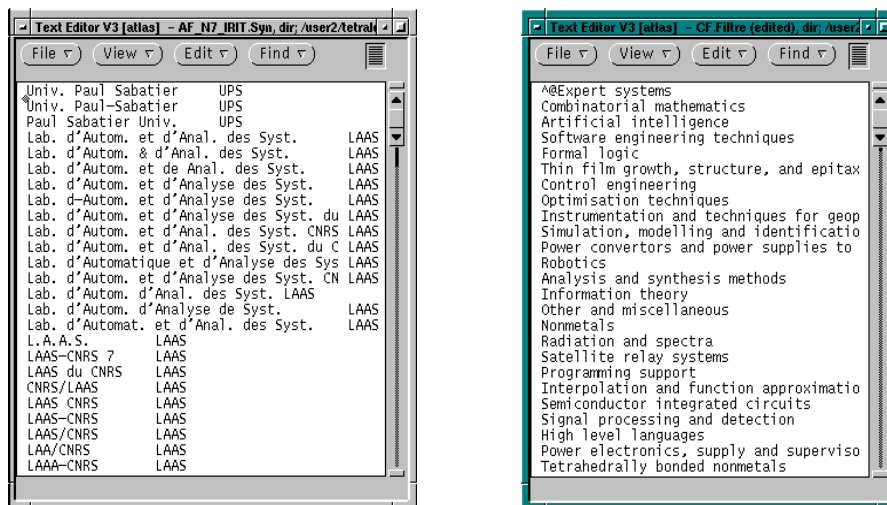

Figure 3: Sample of an interface model which manipulates the main body of the text

2.2. The data analysis module

The data analysis module consists of:

- 3D Table of TETRALOGY.
- The traditional display methods
- The specific techniques of display
- Innovative methods
- Analysis strategy
- Publications of analysis reports
- Multiphase analysis techniques

3D Table of TETRALOGIE:

Allows you to browse, organize, transform and clean matrices (overlap, unique values, and deletion of rows or columns ...). Mathematical transformations (standards) that came with the methods of choice (in alphabetical order, consistency in line or column) clusters do appear to be separately analysed Histograms directly explain the nature and intensity of relations between lines and between columns. Finally, the table allows the 3D pre-classification of the agents (colours, icons) and their weights (weight).

Traditional display methods:

TETRALOGY is integrated into the following methods:
- The Analysis of Principal Component (A.C.P.),
- The Analysis of Reduced Principal Component (A.C.P.R.),
- Factor Analysis of Correlation A.F.C.,
- The Ascending Hierarchical Classification CAH (4 standards),
- Classification by partitions (mobile)

Specific techniques of display:

Factorials maps from these methods of analysis can be viewed in 2D, 3D or 4D animations and interactive or automatic (zoom, rotations, or selective removal of central points, choice of colours and icons ...). For classification purposes, a plenary tree can be used to chose the different filters and then identify classes that will then be imported and displayed in corresponding factorial according to optimal azimuths.
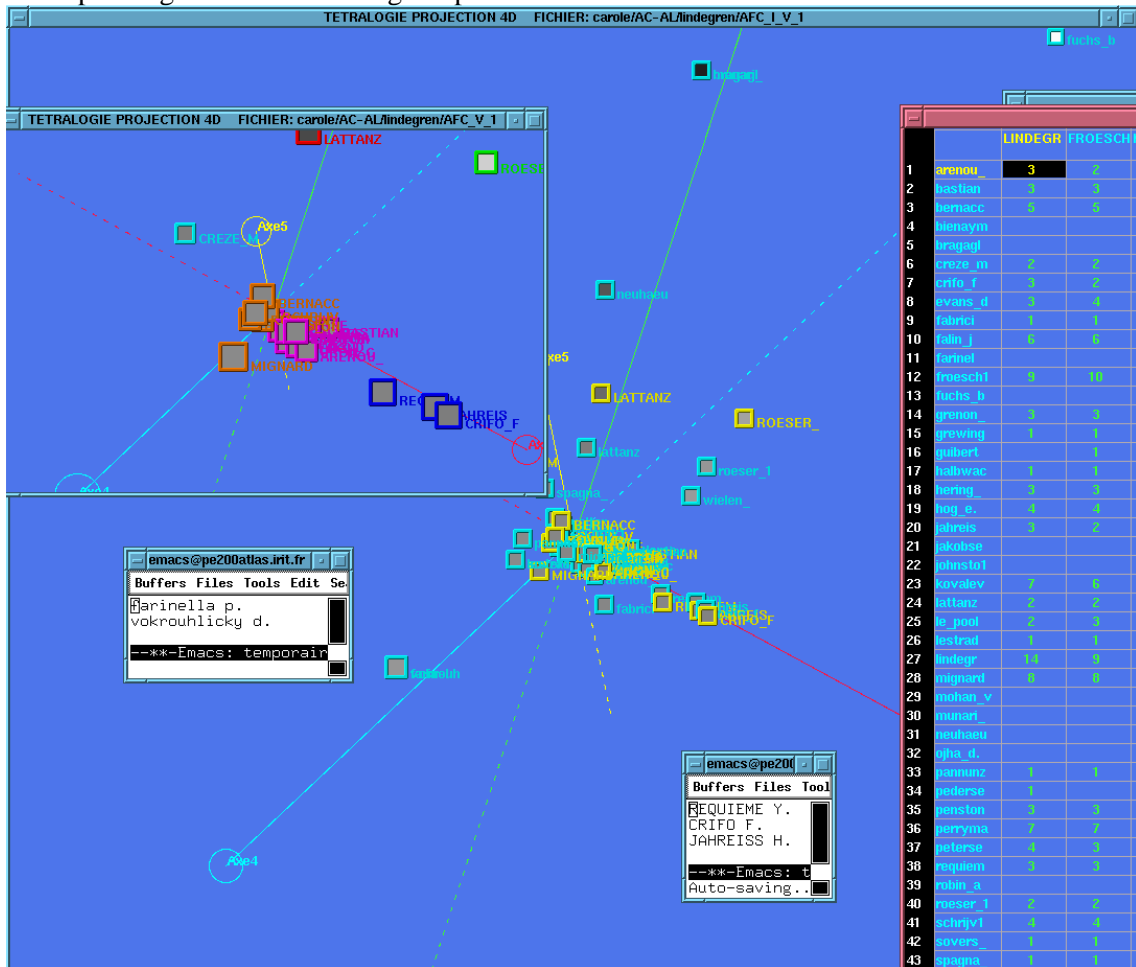

Figure 4: Example of a factorial map displayed in 4 dimensions

Innovative methods:

In ACP: 3D display of the relationships between lines and circle of correlation of columns, sorting the weighted correlation circle.

In A.F.C.: display of the relationships between lines, between the columns and mixed. Relational analysis of data referring to the relations (secondary, ternary ...)

For the analysis of evolutionary data: global or selective paths of visualization, analysis of variance of speed and acceleration, analysis of developments relating to the rotations of clusters. In addition, a method of explaining the constitution of classes by expert system is proposed

Strategies of analysis:

Thanks to a multi-window system and the technical interconnection process, driving is a joy. Different methodologies are proposed to work closely throughout the exercise and control each other, to ensure that the findings do not come out all the same. The observatory of the quality of information and analysis can, at any instant, control the validity of the continuous transition (decrease of the Eigen values, information rate, density, inter-correlations and inter-column lines).

Editing analysis reports:

The automatic editing of an analysis report is an essential element for the comfort of the user, who simply has to choose the analysis and comment, hearings and classifications that they may want to retain in its final. This edition, which is managed by a module, allows the preview, reorganization, masking and printing of the different phases of a complex analysis.

Multi-base Analysis Techniques:

In most cases, a single base is not enough to cover either a topic of study, it is necessary to interrogate more issues and group results. Two solutions are then possible: reformat into a single format or couple the equivalent fields by seeking special status. It is the second solution we have adopted as it avoids reformatting and allows the maximum amount of information to be extracted from each base. Furthermore, any modification or updating of the body of the text is instant. The fields title, author, abstract, source, and year can be treated through the files using synonyms and TETRALOGY.

3. TETRALOGY servers

3.1. Description of a server

TETRALOGY is a program that runs under UNIX on workstations (SUN, HP, Silicon Graphic, IBM PC or Linux), that are in most cases connected to the Internet. Considering its construction, this program uses the full potential of this environment and distributed multi-user. A single station accessible over the network allows a monitoring device, used to distance the program and collaborate with TETRALOGY analyse, or share tools (dictionaries, filters, synonyms, description ...), or well-performing individual parties that dominate the analysis better. The disk server is used to centralize the issues discussed, to compare the results and to develop strategies to reach and spread to many complex and multi-analysis experts. Sharing specific tools to each database or each group of foundations makes further analysis more accessible. Central analysis scripts are proposed for most databases, making the automatic launch all the more significant and closes the generation gap of the corresponding factorial maps of classifications.

3.2. The methods of the server

A server can connect through the local (direct modem) or through the Internet via network servers. In the case of a local station and miscellaneous jobs, the first solution has to be taken into consideration. The second leads to TETRALOGY being used as a unifying system enabling distributed surveillance equipment pooling program, tools and analysis on a single support. These will link the various departments and stakeholders when monitoring scientific technologies from far away.

There are many physical solutions:
- work stations under UNIX of the same manufacturers,
- work stations of different manufactures (SUN y HP, SUN y IBM ...)
- Terminal X connected to the internet,
- Macintosh emulator provided X: MacX,
- PC emulator provided X: Pc-Xware, X-Win, Reflection X,
- PC under Linux or Solario,

The connection can be accomplished in several ways:
- Locally via serial (direct modem),
- Locally by the network of local enterprise,
- Remotely over the switched network and modem (slow),
- Remotely via an Internet server.

4. Conclusions

TETRALOGY is a tool with a feature recognized in the field of strategic intelligence as a means to conduct surveillance of science and technology. For more than 10 years, this program allows us to carry out strategic analysis on bodies of textual information from various sources such as: bases online, Cd, visible and invisible Web, news, patents, press, traces of links to websites, and the foundation internal... The information generated from which a synthesis of all the joined documents: identification of actors and their relations under consistent, strong and weak signals, trends, and strategic components, if required, studies conducted by experts in complex analysis. Thus, the optimization of complex investigations in time is very important.