

# TETRALOGIE: Programa de Vigilància Científica i tecnològica

Bernard Dousset

Eina d'intel·ligència competitiva desenvolupada  
ATLES Bernard Dousset (IRIT)  
e-mail: [dousset@irit.fr](mailto:dousset@irit.fr)

## Key words

Tetralogie, technological watch, scientific watch, bibliometrics, Sciencemetrics, Infometrics, data analysis, competitive intelligence

## Paraules clau

Tetralogie, vigilància tecnològica, bibliometria, cienciometria, infometria, anàlisi dades, intel·ligència competitiva

## Abstract

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

## Resum

Tetralogie és una eina informàtica d'intel·ligència competitiva que ve amb textos complets i informació rellevant de fonts bibliogràfiques, online o en format Cd/Rom, d'Internet o de qualsevol altre font informàtica. Amb mètodes estadístics i mètodes d'anàlisi exploratori de dades, genera en poc temps informació estratègica no disponible fins ara, com d'identitat dels actors, mobilitat, sorgiment i evolució de temes i conceptes, terminologies, dominis marcats, *clústers*, que llegir, on publicar, amb qui col·laborar. Tetralogie és una eina dissenyada per la vigilància científica i tecnològica.

---

## 1. Introducció

El programa TETRALOGIE, desenvolupat a l'interior del "Institut de Recherche en Informatique de Toulouse" (IRIT) és un dels elements essencials de l'estació bibliomètrica "ATLES" elaborada conjuntament gràcies a les ajudes del CEDOCAR i del SGDN, que fan intervenir nombrosos socis nacionals francesos (CNRS, CRRM, IRIT, Ministeris...) per tal de tenir en un mateix suport el conjunt dels mètodes que són operacionals actualment en aquest àmbit. Aquest programa està disponible sobre estacions SUN pels sistemes operatius SUN/ÓS i Solaris, i és accessible per avaluació en la xarxa tant a partir de terminals X com de PC o de Macintosh. Una versió HP és des de fa poc disponible.



**Figura 1.:** Seu de l'IRIT a Toulouse

Aquest programari de vigilància tecnològica és alimentat per bancs de dades bibliogràfiques, textuals o factuais (en línia o en CD-Rom). Permet, gràcies a mètodes estadístics i d'anàlisi de dades evolutives visualitzades en quatre dimensions, posar en evidència: la identitat dels actors i les seves relacions, els seus llocs d'acció, l'emergència i l'evolució dels assumptes i dels conceptes, els elements estratègics de propietat industrial (patents), els àmbits virtualment portadors, que llegir i on publicar, amb qui col·laborar.

Cursos teòrics sobre els mètodes i treballs pràctics sobre TETRALOGIE són organitzats en el context de la formació continuada de la Universitat Paul Sabatier, així com les jornades d'ajuda puntual als usuaris ja formats.

### 1.1. Punts forts de Tetralogie

- Disponible per a PC, Mac, X i estacions,
- Permet treballar a distància,
- Accepta tots els formats de bases,
- Administra els cossos multi-base, les pàgines web, notícies, patents,
- Considera sinònims i fa tractament semàntic,
- Disposa de molts mètodes d'anàlisi cooperatius (ACP,AFC, CAH, CPP),
- Mapes factorials en 4D,
- Visualització interactiva (relacions, classes, trajectòries, zoom 3D, Mapes 2D i 3D),
- Estudis basats en l'evolució dels entorns,
- Formació, transferències i control segurs,
- Fòrums d'usuaris a Internet.

### 1.2. Usuaris destacats

- Ministeris (Defensa,Interior, Finances)
- CEDOCAR, DGA, GIAT, SGDN...
- Universitats (Paris, Toulouse, Marseille, Poitiers, Marne la Vallée, Strasbourg, Louvain la Neuve, Unam/Mèxic, Unc/Colòmbia)
- *Ecole des Mines* de Paris, IPC Barcelona

- CEA, CNRS, INRA, INSERM, CIRAD, CTBA
- Airbus, *Aérospatiale*, PSA, France Télécom
- Institut Mexicà del Petroli, TotalFinaElf Sanofi, P. Fabre, L'Oréal, Bourjois, Nestlé, Idelux, Carmeuse, UCB chemical (Bèlgica)
- Avant Veille, Delta Veille, Iale tecnologia
- ADIT, Innovaction, Gfi, Eurespace Pépinière Théogone

## **2. Estructura de Tetralogie**

TETRALOGIE està essencialment compost per dos mòduls:

- Un sistema de manipulació dels cossos de text que permet administrar els fitxers procedents de teledescàrregues o d'interrogacions sobre CD i extreure'n, per diverses vies, matrius de cruïlla tenint en compte les especificitats de cada base i de cada format.
- Un sistema d'anàlisi de les informacions contingudes a les matrius, que s'articula al voltant d'un full de càlcul 3D específic i que fa intervenir tècniques d'anàlisi de dades clàssiques o innovadores sobre àmbits estàtics, bipartits o evolutius. Aquest sistema és completat per un mòdul d'ajuda a l'elaboració dels informes d'anàlisi facilitant la gestió dels resultats numèrics i gràfics generats pels diversos mètodes que són proposats.

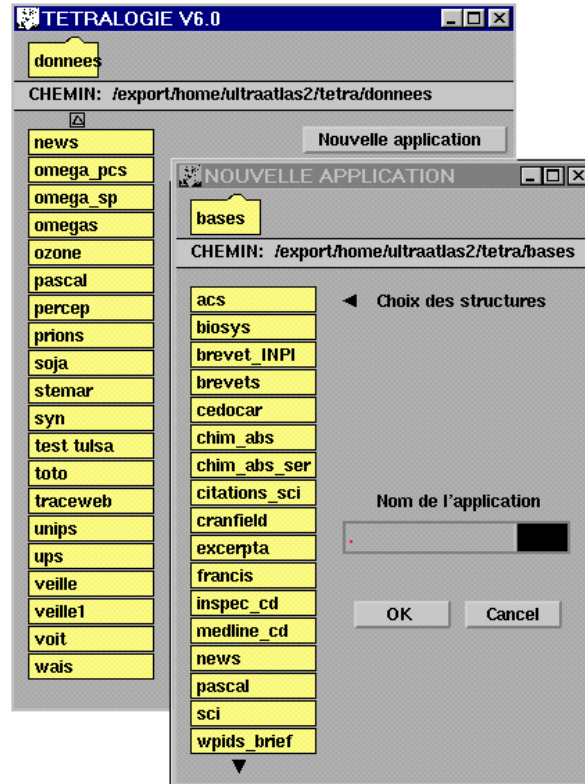


Figura 2.: Exemples d'interfície gràfica Tetralogie

### 2.1. El mòdul de manipulació del corpus

Permet, des de bon començament, crear cada base de dades bibliogràfica o cada font d'informació amb una aplicació única que efectua, *on demand*, els creuaments entre les informacions significatives (exògenes o endògenes) contingudes al cos del text. Aquesta eina fa intervenir un cert nombre de nocions que detallarem ara.

- La gestió de les aplicacions en curs
- Les modalitats d'accés a les bases de dades
- Les estructures de les bases de dades
- El format del descriptor d'estructura
- L'elaboració de filtres i de tesaurs
- L'elaboració de diccionaris de sinònims
- Les tècniques de creuament de les informacions
- Les estratègies de depuració de les matrius

#### La gestió de les aplicacions en curs:

TETRALOGIE permet a un equip de recercadors o a un enginyer informàtic administrar, en paral·lel, un gran nombre d'assumpes ells mateixos descompostos en sota conjunts corresponents, o bé en diverses bases interrogades, o bé amb diferents estratègies de creuament de la informació o a diferents nivells d'afinament. A més a més, les anàlisis condueixen molt sovint a simplificacions successives de les quals l'històric ha de ser conservat i ser comentat en vista de l'elaboració d'un informe final. Una estructura arbòria, comparable a la d'UNIX, ha estat doncs adoptada per organitzar el millor possible aquesta gestió.

### Les modalitats d'accés a les bases de dades:

Les bases de dades o les fonts d'informació són gairebé totes compatibles amb TETRALOGIE, tant si són accessibles per la xarxa commutada, per les xarxes especialitzades o disponibles en CD/Rom. Quan els formats no són directament compatibles, els programes UNIX permeten la seva transcodificació immediata (PC, MAC).

### Les estructures de les bases de dades:

Cada base de dades posseeix la seva pròpia estructura i ens hem hagut d'adaptar, de manera molt flexible, als diferents formats trobats. A més, l'estructura i el contingut d'una mateixa base depenen de la seva forma de difusió i del suport en el qual és implantada (en línia, en CD/Rom). També, hem concebut un descriptor configurable de l'estructura de les bases que s'adapta fidelment a cada cas.

### El format del descriptor d'estructura:

Aquest descriptor defineix els diferents camps de la base, identificant la seva bandera (*tag*), els seus separadors, la seva utilitat i els diversos tipus d'informació que contenen. Permet també localitzar el començament de cada nota així com l'estructura física del registre.

### L'elaboració de filtres i de tesaurus:

Per tal de recuperar la informació útil de camps no codificats (títol, resums...) o de limitar les recerques d'informació a un sota-domini (autors, vegades que s'ha vist, paraules claus, codis, anys, *pays*...), és útil disposar de filtres o de tesaurus que depenen evidentment del conjunt de les informacions disponibles. Hem descrit doncs un mòdul d'ajuda a la constitució d'aquestes eines, que permet generar fitxers triats, o bé per ordre alfabètic, o bé per ordre decreixent de freqüència. Aquests fitxers, gràcies als editors, faciliten molt aquesta tasca preliminar però essencial.

### L'elaboració de diccionaris de sinònims:

Certes informacions són semànticament equivalents o jerarquitzades. Així doncs també és molt útil disposar de diccionaris de sinònims jeràrquics, per tal de reagrupar elements que han estat jutjats per l'usuari com comparables o relacionats, o potser han estat codificats de forma anàrquica pels dissenyadors de la base (ex.: anys, país, fonts...). Els diccionaris de sinònims a diversos nivells són utilitzables en TETRALOGIE per trobar informacions difícilment desxifrables o per refinar més o menys les anàlisis.

### Les tècniques de creuament de les informacions

Les informacions poden ser creuades entre dos camps, sota camps o grups de camps per tal d'obtenir matrius de freqüència, presència - absència o coocurrència sobre les quals es portaran a terme les anàlisis. Per a les dades bipartides o evolutives, és possible fer intervenir un tercer camp per tal de fer esclatar la matriu en dos, tres o quatre plànols per tal d'analitzar les diferències induïdes (absolutes, relatives, velocitats, acceleracions, implosions, explosions...).

### Les estratègies de depuració de les matrius

Nivells de tall, referint-se a la freqüència o la taxa d'informació, poden ser aplicats automàticament per tal de reduir el volum dels quadres obtinguts i així poder simplificar el treball d'anàlisi i disminuir, de manera significativa, els temps de càlcul i la complexitat dels mapes factorials i dels arbres de classificació associats.

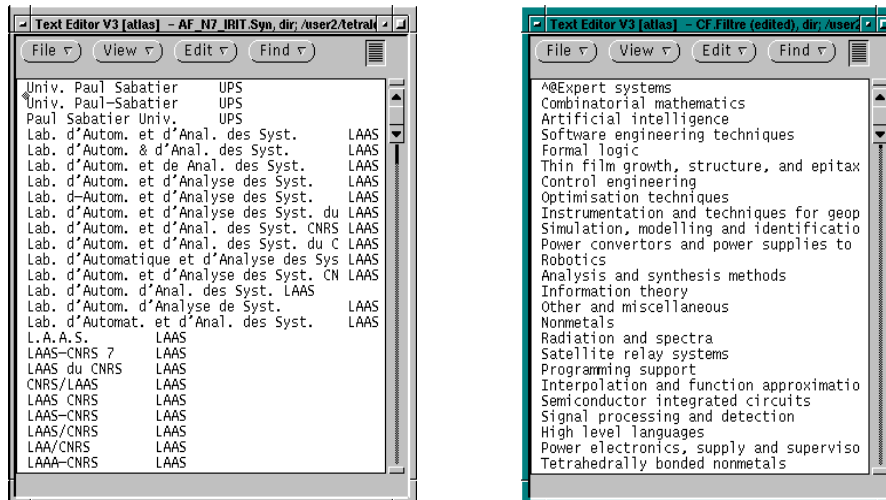


Figura 3.: Exemples d'interfície del mòdul de manipulació del corpus

## 2.2. El mòdul d'anàlisi de dades

El mòdul d'anàlisi de dades es compon de:

- El quadre 3D de TETRALOGIE
- Els mètodes clàssics de visualització
- Les tècniques específiques de visualització
- Els mètodes innovadors
- Les estratègies d'anàlisi
- L'edició dels informes d'anàlisi
- Les tècniques d'anàlisi multibase

El quadre 3D de TETRALOGIE:

Serveix per explorar, organitzar, transformar, i netejar les matrius (redundàncies, valors únics, supressió de línies i/o de columnes...). Les transformacions matemàtiques (normalitzacions) que acompanyen als mètodes de tria (alfabètic, consistència en línia o columna) fan aparèixer clústers que podran, llavors, ser analitzats de manera separada. Els histogrames permeten explicar directament la naturalesa i la intensitat de les relacions entre línies i entre columnes. Finalment, el quadre 3D autoritza les pre-classificacions dels agents (colors, icones) i les seves ponderacions (pes).

Els mètodes clàssics de visualització:

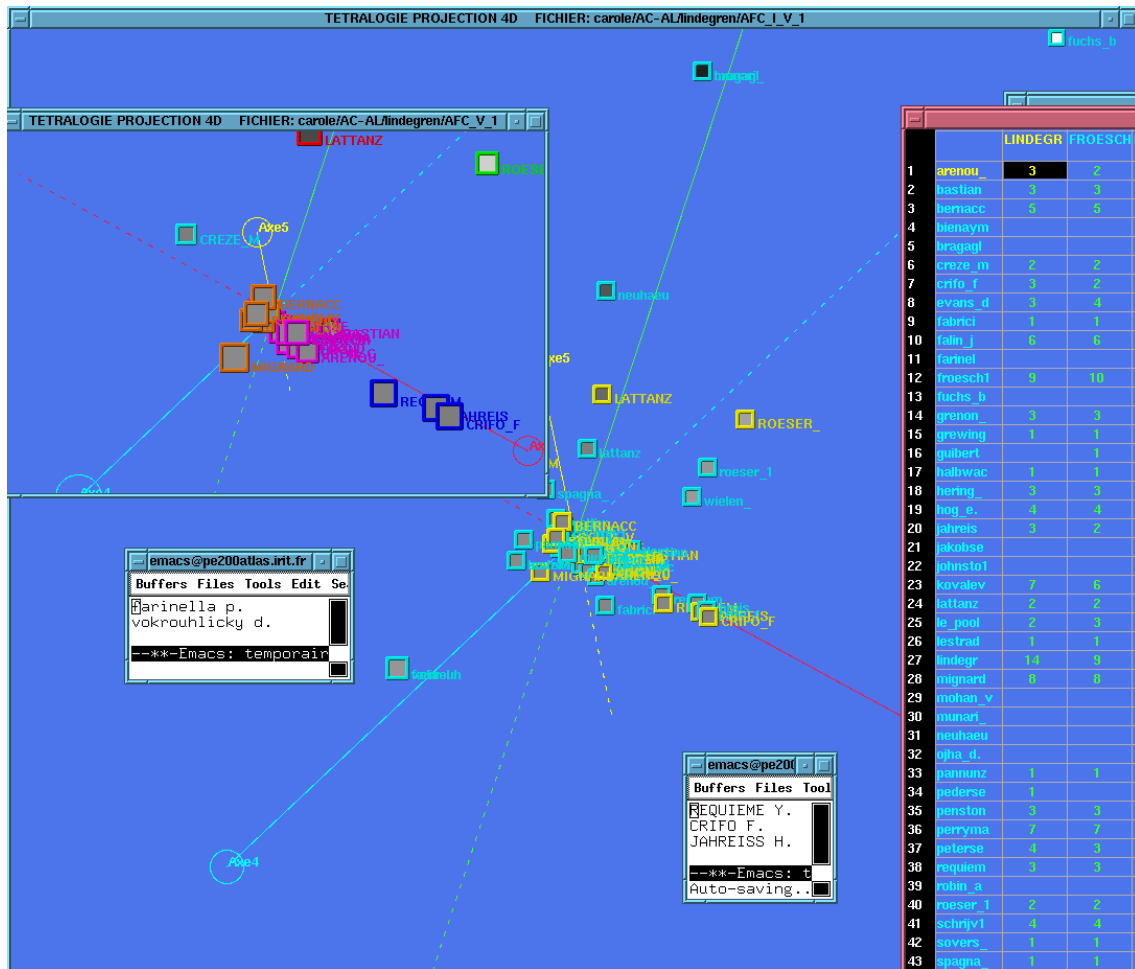
Són integrats en TETRALOGIE els mètodes següents:

- l'anàlisi en components principals A.C.P.,
- l'anàlisi en components principals reduïda A.C.P.R.,
- l'anàlisi factorial de les correspondències A.F.C.,
- la classificació ascendent jeràrquica C.A.H.(4 normes),
- la classificació per particions (centres mòbils)

Les tècniques específiques de visualització:

Els mapes factorials procedents d'aquests mètodes d'anàlisi poden ser visualitzats en 2D, 3D o 4D i animats de manera interactiva o automàtica (zooms, rotacions, eliminació de punts selectiva o central, tria dels colors i dels icones...). Per a les classificacions, un arbre planari

permet escollir un nivell de tall i així identificar les classes que seran llavors importades i seran visualitzades en els mapes factorials corresponents segons azimuths òptims.



**Figura4.** : Exemple de mapa factorial visualitzat en 4 dimensions

Els mètodes innovadors:

En A.C.P.: visualització 3D de les relacions entre línies i cercle de correlació de les columnes, classificació ponderada sobre el cercle de correlació.

En A.F.C.: visualització de les relacions entre les línies, entre les columnes i mixtes. Anàlisi relacional de les dades referint-se a les relacions (secundaris, ternàries...).

Per a l'anàlisi de les dades evolutives: visualització global o selectiva de les trajectòries, anàlisi de la velocitat de variació i de la seva acceleració, analitza qualitat de l'evolució relativa per les rotacions *Pro crustes*. A més, un mètode d'explicació de la constitució de les classes per sistema expert és proposat.

Les estratègies d'anàlisi:

És molt agradable conduir una anàlisi gràcies al sistema multi-finestra i a les tècniques d'interconnexió de procés. De fet, les diverses metodologies proposades col·laboren de forma estreta al llarg de l'anàlisi i es controlen mútuament entre elles per aconseguir que les conclusions no provinquin totes del mateix enfocament.

L'observatori de la qualitat de la informació i de les anàlisis permet, en tot instant, controlar la validesa del pas continu (decreixement dels valors propis, taxa d'informació, densitat, correlacions inter-línies i inter-columnes).

L'edició dels informes d'anàlisi:

L'edició automàtica d'un informe d'anàlisi és un dels elements essencials per la comoditat de l'usuari, que no té més que escollir i comentar les anàlisis, les vistes i les classificacions que pot desitjar conservar en el seu informe final. L'edició est, administrada per un mòdul, permet la previsualització, la reorganització, l'emascarament i la impressió de les diferents fases d'una anàlisi complexa.

Les tècniques d'anàlisi multibase:

En la major part dels casos, una sola base no és suficient per cobrir bé un tema d'estudi, és doncs necessari interrogar-ne més assumptes i agrupar els resultats. Dues solucions són llavors possibles: reformatar en un format únic o acoblar els camps equivalents tot continuant tractant-los com entitats específiques. És aquesta segona solució la que hem adoptat ja que evita la reformatació i permet treure la màxima informació útil de cada base. A més a més, tota modificació o reactualització del corpus és per això instantània. Els camps títol, autor, resum, font, i any poden ser tractats gràcies als fitxers de sinònims que utilitza ja, per un altre banda, TETRALOGIE.



### **3. Els servidors de TETRALOGIE**

#### **3.1. Descripció del servidor**

TETRALOGIE és un programa que gira sota UNIX en estacions de treball (SUN, HP, Silicon Graphic, IBM o PC sota Linux), que estan en la majoria de casos connectades a Internet. Considerant la seva construcció, aquest programa utilitza plenament les possibilitats d'aquest entorn repartit i multi-usuari. En efecte, una sola estació accessible a través de la xarxa permet a un equip de vigilància utilitzar, a distància, el programa TETRALOGIE i col·laborar en certes anàlisis, o bé intercanviar eines (diccionaris, filtres, sinònims, descriptors...), o bé efectuant individualment les parts de l'anàlisi que dominen millor. El disc del servidor serveix per centralitzar els assumptes tractats i els temes més propers i complementaris, per confrontar els resultats obtinguts i per elaborar estratègies comunes i repartides per portar a moltes anàlisis complexes i multi-experts. També, les eines posades en comú i específiques a cada base o a cada grup de bases estan directament disponibles per anàlisis ulteriors. Guions centrals d'anàlisis són proposats per a les bases de dades més corrents, permetent posar en marxa automàticament el conjunt dels creuaments més significatius i la generació dels diferents mapes factorials o de les classificacions que es corresponen més.

#### **3.2. Les modalitats del servidor**

La connexió a un servidor pot ser realitzada en local (mòdem directe) o a través d'Internet via els servidors xarxa. La primera solució s'ha de tenir en consideració en el cas d'una estació local i de diversos llocs de treball. El segon condueix a utilitzar TETRALOGIE com un sistema federador permetent a un equip de vigilància repartit posar en comú el programa, les eines i les anàlisis sobre un suport únic, que servirà d'enllaç entre els diferents serveis i els diferents actors d'una vigilància científica i tècnica portada a distància.

Físicament les solucions són nombroses:

- estacions de treball sota UNIX del mateix constructor,
- estació de treball d'un altre constructor (SUN i HP, SUN i IBM...)
- terminal X connectat a la xarxa,
- Macintosh proveït de l'emulador X : MacX,
- PC proveït d'un emulador X : Pc-Xware, X-Win, Reflection X,
- PC sota Linux o Solaris,

La connexió pot ser realitzada de diferent maneres:

- en local per la via sèrie (mòdem directe),
- en local per la xarxa local de l'empresa,
- a distància per la xarxa commutada i mòdem (una mica lent),
- a distància per Internet via un servidor.

### **4. Conclusions**

TETRALOGIE és una eina amb reconeguda funcionalitat en l'àmbit de d'intel·ligència estratègica com mitjà per portar a terme vigilància científica i tecnològica. Des de fa més de 10 anys, aquest programa ens permet efectuar anàlisis estratègiques sobre cossos d'informació textual procedents de les fonts més diverses com: les bases en línia, els Cd, el Web visible i invisible, notícies, les patents, la premsa, les traces de connexions als llocs web, les bases internes... La informació elaborada de on prové representa una síntesi del conjunt dels documents: identificació dels actors i de les seves relacions, sota assumptes coherents, senyals fortes i febles, tendències, components estratègics i, si es requereix, estudis dirigits complexes realitzats per experts en anàlisis. Així doncs, l'optimització del temps en recerques complexes és molt rellevant.