

TETRALOGIE: Programa de Vigilancia Científica y tecnológica

Bernard Dousset

Herramienta de inteligencia Competitiva desarrollada en el
contexto ATLES (IRIT)
e-mail: dousset@irit.fr

Key words

Tetralogie, technological watch, scientific watch, bibliometrics, sciencemetrics, infometrics, data analysis, competitive intelligence

Palabras clave

Tetralogie, vigilancia tecnológica, bibliometria, ciencimetrica, infometria, análisis datos, inteligencia 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.

Resumen

Tetralogie es una herramienta informática de inteligencia competitiva que viene con textos cumplidos e información relevando de fuentes bibliográficas, online o en formato Cd/Rom, de Internet o de cualquier otro fuente informática. Con métodos estadísticos y métodos de análisis exploratorio de datos, genera en poco tiempo información estratégica no disponible hasta ahora, como de identidad de los actores, movilidad, surgimiento y evolución de temas y conceptos, terminologías, dominios marcados, clusters, que leer, donde publicar, con quién colaborar. Tetralogie es una herramienta diseñada para la vigilancia científica y tecnológica.

1. Introducción

El programa TETRALOGIE, desarrollado al interno del "Institut de Recherche Informatique de Toulouse" (IRIT) es uno de los elementos esenciales de la estación bibliométrica "ATLAS" elaborada conjuntamente gracias a las ayudas del CEDOCAR y del SGDN, que hacen intervenir numerosos socios nacionales franceses (CNRS, CRRM, IRIT, Ministerios) con el fin de tener en un mismo soporte el conjunto de los métodos que son operacionales actualmente en este ámbito. Este programa está disponible sobre estaciones SUN para los sistemas operativos SUN/OSO y Solaris, y es accesible para evaluación en la red tanto a partir de terminales X como de PC o de Macintosh. Una versión HP está desde hace poco disponible.



Figura 1.: Sede del IRIT en Toulouse

Este software de vigilancia tecnológica es alimentado por bancos de datos bibliográficos, textuales o factuales (en línea o en CD-ROM). Permite, gracias a métodos estadísticos y de análisis de datos evolutivos visualizados en cuatro dimensiones, poner en evidencia: la identidad de los actores y sus relaciones, sus sitios de acción, la emergencia y la evolución de los asuntos y de los conceptos, los elementos estratégicos de propiedad industrial (patentes), los ámbitos virtualmente portadores, qué leer y donde publicar, con quién colaborar.

Cursos teóricos sobre los métodos y trabajos prácticos sobre TETRALOGIE son organizados en el contexto de la formación continuada de la Universidad Paul Sabatier, así como las jornadas de ayuda puntual a los usuarios ya formados.

1.1. Puntos fuertes de Tetralogie

- Disponible para PC, Mac, X y estaciones,
- Permite trabajar a distancia,
- Acepta todos los formatos de bases,
- Administra los cuerpos multi-base, las páginas web, noticias, patentes,
- Considera sinónimos y hace tratamiento semántico,
- Dispone de muchos métodos de análisis cooperativos (ACP,AFC, CAH, CPP),
- Mapas factoriales en 4D,
- Visualización interactiva (relaciones, clases, trayectorias, zoom 3D, Mapas 2D y 3D),
- Estudios basados en la evolución de los entornos,
- Formación, transferencias y control seguros,
- Foros de usuarios en Internet.

1.2. Usuarios destacados

- Ministerios franceses (Defensa, Interior, Finanzas)
- CEDOCAR, DGA, GIAT, SGDN..
- Universidades (Paris, Toulouse, Marseille, Poitiers, 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 Mejicano del Petróleo, TotalFinaElf Sanofi, P. Fabre, L'Oréal, Bourjois, Nestlé, Idelux, Carmeuse, UCB chemical (Bélgica)
- Avant Veille, Delta Veille, Iale tecnología
- AVENIDO, Innovaction, Gfi, Eurespace Pépinière Théogone

2. Estructura de Tetralogie

TETRALOGIE está esencialmente compuesto por dos módulos:

- Un sistema de manipulación de los cuerpos de texto que permite administrar los ficheros procedentes de teledescargas o de interrogaciones sobre CD y extraer, por diversas vías, matrices de cruce teniendo en cuenta las especificidades de cada base y de cada formado.
- Un sistema de análisis de las informaciones contenidas en las matrices, que se articula en torno a una hoja de cálculo 3D específico y que hace intervenir técnicas de análisis de datos clásicas o innovadoras sobre ámbitos estáticos, bipartitos o evolutivos. Este sistema es completado por un módulo de ayuda a la elaboración de los informes de análisis facilitando la gestión de los resultados numéricos y gráficos generados por los diversos métodos que son propuestos.

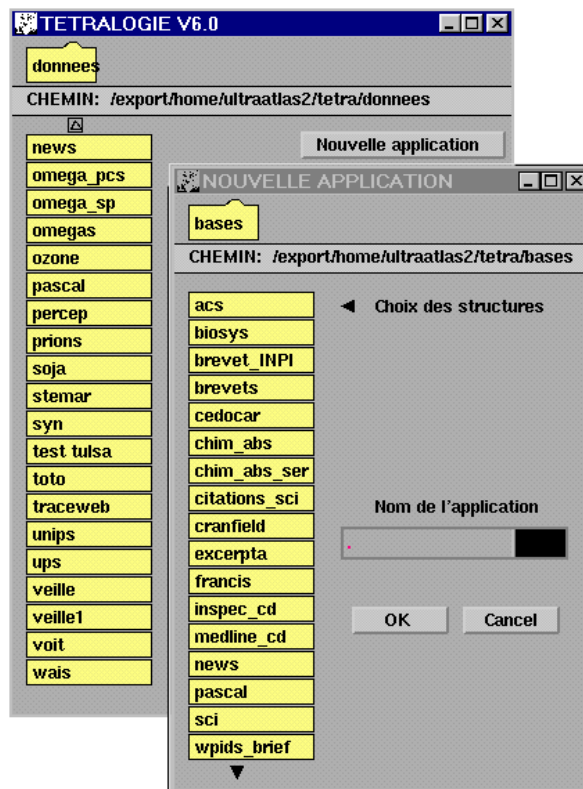


Figura 2.: Ejemplos de interfaz gráfica Tetralogie

2.1. El módulo de manipulación del corpus

Permite, desde el principio, cruzar cada base de datos bibliográfica o cada fuente de información con una aplicación única que efectúa, *on demand*, los cruces entre las informaciones significativas (exógenas o endógenas) contenidas en el cuerpo del texto. Esta herramienta hace intervenir un cierto número de nociones que detallaremos ahora.

- La gestión de las aplicaciones en curso
- Las modalidades de acceso a las bases de datos
- Las estructuras de las bases de datos
- El formato del descriptor de estructura
- La elaboración de filtros y de tesauros
- La elaboración de diccionarios de sinónimos
- Las técnicas de cruce de las informaciones
- Las estrategias de depuración de las matrices

La gestión de las aplicaciones en curso:

TETRALOGIE permite a un equipo de investigadores o a un ingeniero informático administrar, en paralelo, un gran número de asuntos ellos mismos descompuestos en subconjuntos correspondientes, o bien en diversas bases interrogadas, o bien con diferentes estrategias de cruce de la información o a diferentes niveles de afinamiento. Además, los análisis conducen muy a menudo a simplificaciones sucesivas de las cuales el histórico tiene que ser conservado y ser comentado en vista de la elaboración de un informe final. Una estructura arbórea, comparable a la de UNIX, ha sido pues adoptada para organizar lo mejor posible esta gestión.

Las modalidades de acceso a las bases de datos:

Las bases de datos o las fuentes de información son casi todas compatibles con TETRALOGIE, tanto si sueño accesibles por la red conmutada, por las redes especializadas, o disponibles en CD/Rom. Cuando los formatos no son directamente compatibles, los programas UNIX permiten su transcodificación inmediata (PC, MAC).

Las estructuras de las bases de datos:

Cada base de datos posee su propia estructura y nos hemos tenido que adaptar, de manera muy flexible, a los diferentes formatos encontrados. Además, la estructura y el contenido de una misma base dependen de su forma de difusión y del soporte en la cual es implantada (en línea, en CD/Rom). También, hemos concebido un descriptor configurable de la estructura de las bases que se adapta fielmente a cada caso.

El formato del descriptor de estructura:

Este descriptor define los diferentes campos de la base, identificando su bandera (tag), sus separadores, su utilidad y los diversos tipos de información que contienen. Permite también localizar el comienzo de cada nota así como la estructura física del registro.

La elaboración de filtros y de tesauros:

Con el fin de recuperar la información útil de campos no codificados (título, resúmenes, etc.) o de limitar las búsquedas de información en uno subdominio (autores, veces que se ha visto, palabras claves, códigos, años, *pays*...), es útil disponer de filtros o de tesauros que dependen evidentemente del conjunto de las informaciones disponibles. Hemos descrito pues un módulo de ayuda a la constitución de estas herramientas, que permite generar ficheros escogidos, o bien por orden alfabético, o bien por orden decreciente de frecuencia. Estos ficheros, gracias a los editores, facilitan mucho esta tarea preliminar pero esencial.

La elaboración de diccionarios de sinónimos:

Ciertas informaciones son semánticamente equivalentes o jerarquizadas. Así pues también es muy útil disponer de diccionarios de sinónimos jerárquicos, con el fin de reagrupar elementos que han sido juzgados por el usuario como comparables o relacionados, o quizás han sido codificados de forma anárquica por los diseñadores de la base (ex.: años, país, fuentes...). Los diccionarios de sinónimos a diversos niveles son utilizables en TETRALOGIE para encontrar informaciones difícilmente descifrables o para refinar más o menos los análisis.

Las técnicas de cruce de las informaciones

Las informaciones pueden ser cruzadas entre dos campos, bajo campos o grupos de campos con el fin de obtener matrices de frecuencia, presencia - ausencia o coocurrencia sobre las cuales se llevarán a cabo los análisis. Para los datos bipartitos o evolutivos, es posible hacer intervenir un tercer campo con el fin de explotar la matriz en dos, tres o cuatro planos con el fin de analizar las diferencias inducidas (absolutas, relativas, velocidades, aceleraciones, implosiones, explosiones...).

Las estrategias de depuración de las matrices

Niveles de corte, refiriéndose a la frecuencia o la tasa de información, pueden ser aplicados automáticamente con el fin de reducir el volumen de los cuadros obtenidos y así poder simplificar el trabajo de análisis y disminuir, de manera significativa, los tiempos de cálculo y la complejidad de los mapas factoriales y de los árboles de clasificación asociados.

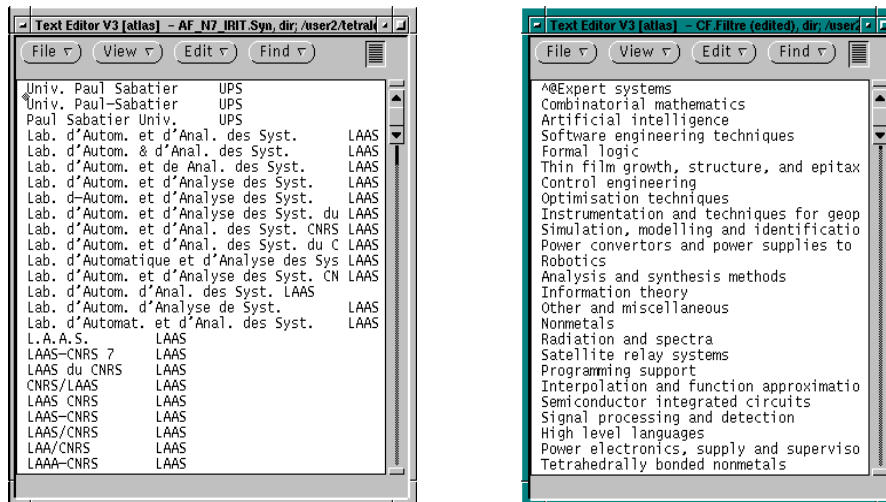


Figura 3.: Ejemplos de interfaz del módulo de manipulación del corpus

2.2. El módulo de análisis de datos

Dicho módulo se estructura en:

- El módulo de análisis de datos se compone de:
- El cuadro 3D de TETRALOGIE.
- Los métodos clásicos de visualización.
- Las técnicas específicas de visualización.
- Los métodos innovadores.
- Las estrategias de análisis.
- La edición de los informes de análisis.
- Las técnicas de análisis multibase

El cuadro 3D de TETRALOGIE:

Sirve para explorar, organizar, transformar, y limpiar las matrices (redundancias, valores únicos, supresión de líneas y/o de columnas...). Las transformaciones matemáticas (normalizaciones) que acompañan a los métodos de elección (alfabético, consistencia en línea o columna) hacen aparecer clusters que podrán ser analizados, entonces, de manera separada. Los histogramas permiten explicar directamente la naturaleza y la intensidad de las relaciones entre líneas y entre columnas. Finalmente, el cuadro 3D autoriza las pre-clasificaciones de los agentes (colores, iconos) y sus ponderaciones (peso).

Los métodos clásicos de visualización:

Son integrados en TETRALOGIE los métodos siguientes:

- el análisis en componentes principales A.C.P.,
- el análisis en componentes principales reducido A.C.P.R.,
- el análisis factorial de las correspondencias A.F.C.,
- la clasificación ascendente jerárquica C.A.H.(4 normas),
- la clasificación por particiones (centros móviles)

Las técnicas específicas de visualización:

Los mapas factoriales procedentes de estos métodos de análisis pueden ser visualizados en 2D, 3D o 4D y animados de manera interactiva o automática (zoomes, rotaciones, eliminación de puntos selectiva o central, elección de los colores y de los iconos...). Para las clasificaciones, un árbol plenario permite escoger un nivel de corte y así identificar las clases que serán entonces importadas y serán visualizadas en los mapas factoriales correspondientes según acimutes óptimos.

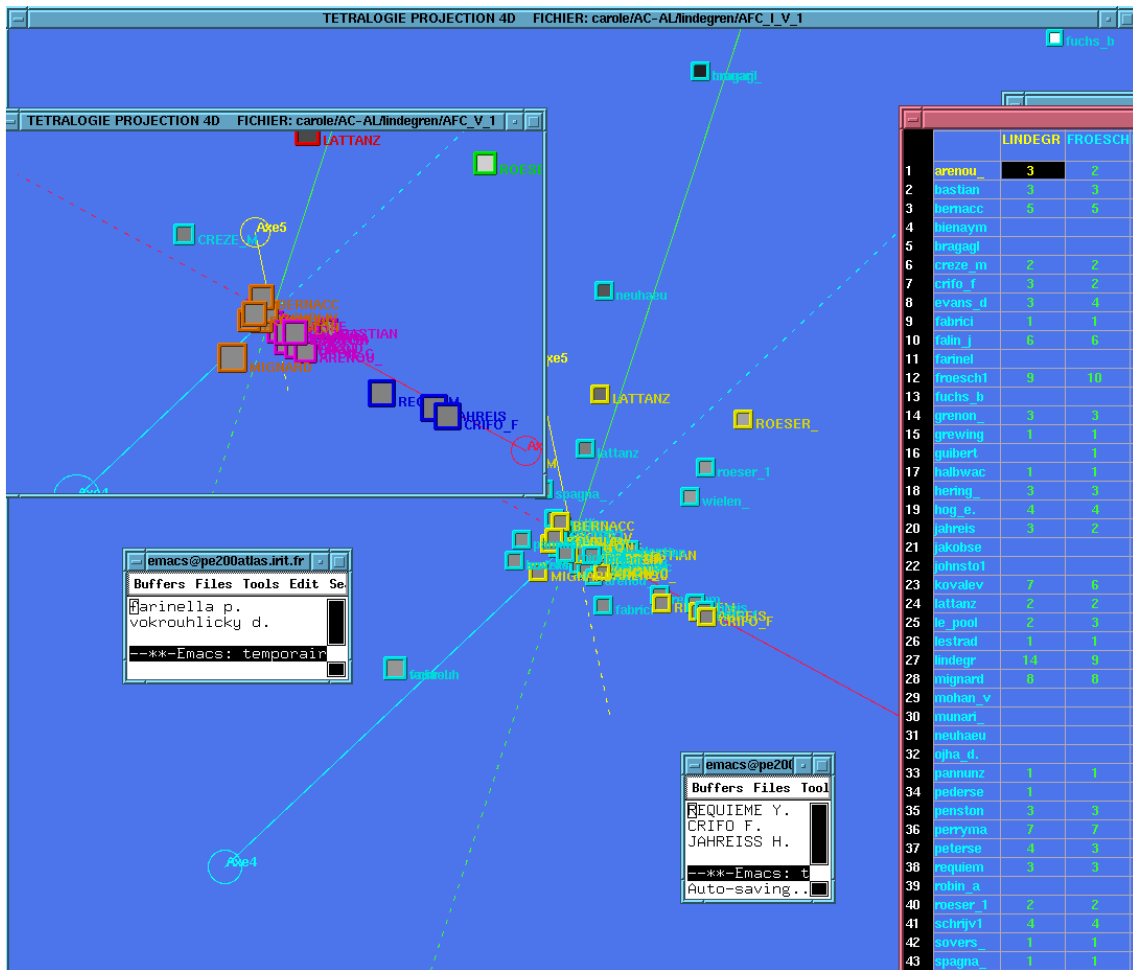


Figura4. : Ejemplo de mapa factorial visualizado en 4 dimensiones

Los métodos innovadores:

En A.C.P.: visualización 3D de las relaciones entre líneas y círculo de correlación de las columnas, clasificación ponderada sobre el círculo de correlación.

En A.F.C.: visualización de las relaciones entre las líneas, entre las columnas y mixtas. Análisis relacional de los datos refiriéndose a las relaciones (secundarios, ternarias...).

Para el análisis de los datos evolutivos: visualización global o selectiva de las trayectorias, análisis de la velocidad de variación y de su aceleración, analiza calidad de la evolución relativa para las rotaciones pro clusters. Además, un método de explicación de la constitución de las clases por sistema experto es propuesto.

Las estrategias de análisis:

Es muy interesante conducir un análisis gracias al sistema multi-ventana y a las técnicas de interconexión de proceso. De hecho, las diversas metodologías propuestas colaboran de forma estrecha a lo largo del análisis y se controlan mutuamente entre ellas para conseguir que las conclusiones no provengan todas del mismo enfoque.

El observatorio de la calidad de la información y de los análisis permite, en todo instante, controlar la validez del paso continuo (decrecimiento de los valores propios, tasa de información, densidad, correlaciones inter-líneas e inter-columnas).

La edición de los informes de análisis:

La edición automática de un informe de análisis es uno de los elementos esenciales para la comodidad del usuario, que no tiene más que escoger y comentar los análisis, las vistas y las clasificaciones que puede desear conservar en su informe final. La edición este, administrada por un módulo, permite la previsualización, la reorganización, el enmascaramiento y la impresión de las diferentes fases de un análisis complejo.

Las técnicas de análisis multibase:

En la mayor parte de los casos, una sola base no es suficiente para cubrir bien un tema de estudio, es pues necesario interrogar más asuntos y agrupar los resultados. Dos soluciones son entonces posibles: reformatear en un formato único o acoplar los campos equivalentes continuando tratándolas como entidades específicas. Es esta segunda solución la que hemos adoptado ya que evita el reformateo y permite sacar la máxima información útil de cada base. Además, toda modificación o actualización del corpus es por eso instantánea. Los campos título, autor, resumen, fuente, y año pueden ser tratados gracias a los ficheros de sinónimos que utiliza ya, por otro lado TETRALOGIE.

3. Los servidores de TETRALOGIE

3.1. Descripción del servidor

TETRALOGIE es un programa que corre bajo UNIX en estaciones de trabajo (SUN, HP, Silicon Graphic, IBM o PC bajo Linux), que están en la mayoría de casos conectadas a Internet. Considerando su construcción, este programa utiliza plenamente las posibilidades de este entorno repartido y multi-usuario. En efecto, una sola estación accesible a través de la red permite a un equipo de vigilancia utilizar, a distancia, el programa TETRALOGIE y colaborar en ciertos análisis, o bien intercambiar herramientas (diccionarios, filtros, sinónimos, descriptores ...), o bien efectuando individualmente las partes del análisis que dominan mejor. El disco del servidor sirve para centralizar los asuntos tratados y los temas más próximos y complementarios, para confrontar los resultados obtenidos y para elaborar estrategias comunes y repartidas para llegar a muchos análisis complejos y multi-expertos. También, las herramientas puestas en común y específicas a cada base o en cada grupo de bases están directamente disponibles para análisis ulteriores. Guiones centrales de análisis son propuestos para las bases de datos más corrientes, permitiendo poner en marcha automáticamente el conjunto de los cruces más significativos y la generación de los diferentes mapas factoriales o de las clasificaciones que se corresponden más.

3.2. Las modalidades del servidor

La conexión en un servidor puede ser realizada en local (módem directo) o a través de Internet vía los servidores red. La primera solución se tiene que tener en consideración en el caso de una estación local y de diversos puestos de trabajo. El segundo conduce a utilizar TETRALOGIE como un sistema federador permitiendo a un equipo de vigilancia repartido poner en común el programa, las herramientas y los análisis sobre un soporte único, que servirá de enlace entre los diferentes servicios y los diferentes actores de una vigilancia científica y técnica llevada a distancia.

Físicamente las soluciones son numerosas:

- estaciones de trabajo bajo UNIX del mismo constructor,
- estación de trabajo de otro constructor (SUN y HP, SUN y IBM ...)
- terminal X conectado a la red,
- Macintosh provisto del emulador X: MacX,
- PC provisto de un emulador X: Pc-Xware, X-Win, Reflection X,

- PC bajo Linux o Solaris,

La conexión puede ser realizada de diferentes maneras:

- en local por la vía serie (módem directo),
- en local por la red local de la empresa,
- a distancia por la red conmutada y módem (un poco lento),
- a distancia por Internet vía un servidor.

4. Conclusiones

TETRALOGIE es una herramienta con reconocida funcionalidad en el ámbito de de inteligencia estratégica como medio para llevar a cabo vigilancia científica y tecnológica. Desde hace más de 10 años, este programa nos permite efectuar análisis estratégicos sobre cuerpos de información textual procedentes de las fuentes más diversas como: las bases en línea, los Cd, la Web visible e invisible, noticias, las patentes, la prensa, los rastros de conexiones a los sitios web, las bases internas ... La información elaborada de donde proviene representa una síntesis del conjunto de los documentos: identificación de los actores y de sus relaciones, bajo asuntos coherentes, señales fuertes y débiles, tendencias, componentes estratégicos y, si se requiere, estudios dirigidos complejos realizados por expertos en análisis. Así pues, la optimización del tiempo en investigaciones complejas es muy relevante.