

***L'APPROCHE MEMORIELLE UNE METHODE POUR LA VEILLE
DOCUMENTAIRE SCIENTIFIQUE ET TECHNIQUE : APPLICATION
DANS LE DOMAINE DE LA RECHERCHE EN MEDECINE
REGENERATIVE.***

Elisabeth Scarbonchi (*) , Christian Krumeich(*), Dr Vincent Mallet(), Daniel Laurent(*)**
elisabeth.scarbonchi@univ-mlv.fr

* : Université de Marne-la-Vallée, Institut Francilien d'Ingénierie des Services (I.F.I.S.), Laboratoire
Ingénierie des Systèmes d'Informations Stratégiques et Décisionnelles (ISIS)
Cité Descartes, Champs-sur-Marne, 77454 Marne-la-Vallée Cedex 2.

** : Hôpital Cochin Service de physiologie exploration fonctionnelle – INSERM

Mots-clés : plan de veille, banque d'informations élaborées (BIE), langage mémoriel pivot, situation d'interprétation.

Résumé

L'objet de cette communication est de proposer un système de production d'information élaborée dans le domaine médical appliqué à la médecine régénérative. Ce système s'appuie sur un ensemble de méthodes et de technologies utilisées dans le domaine de la veille scientifique et technologique.

Notre démarche de recherche est fondée sur la mise en œuvre d'une approche dite **mémorielle**, qui permet de traduire au mieux les interrogations et les préoccupations du chercheur exprimées par son plan de veille.

Nous proposons d'utiliser l'atelier MEVA (Mémoire Événementielle d'Actualisation) qui autorise une « traduction mémorielle » d'un plan de veille. Le plan de veille s'élabore avec l'expert et consiste à prendre en compte les connaissances associées aux « contextes » ciblés du domaine étudié. Ces éléments permettent de déterminer des axes de recherche et de choisir les corpus à analyser. Toutes les associations de contextes établies seront traduites en « langage pivot » permettant une capture automatique d'informations pertinentes. Ce système permet de traiter des corpus volumineux acquis sur Internet et / ou dans les Banques de données.

Une fois le corpus pertinent obtenu par l'indexation mémorielle du plan de veille, la mise en œuvre d'outils statistiques permettra de créer de l'information élaborée concernant les différents acteurs (médecins, chercheurs, etc.), les organismes auxquels ils appartiennent (laboratoire, service, etc.) pourront être reliés aux thématiques de recherche ou d'activités.

Le développement de ces processus de filtrage mémoriel pourrait trouver des applications diverses :

- Utilisation des résultats obtenus pour la création de banques d'informations élaborées (BIE) multimédia.
- Travail collaboratif (réseau de santé, co-laboratoires,...) souhaitant mettre en œuvre un plan de veille partagé.
- Création de modules d'enseignement dans le cadre du E-learning.

Introduction

Ce travail a été réalisé au laboratoire d'Ingénierie ISIS, dirigé par le Professeur Daniel LAURENT - Université de Marne la Vallée - en collaboration avec le service de physiologie respiratoire, dirigé par le Professeur Josette DALL'AVA - SANTUCCI et l'INSERM – hôpital Cochin - Paris. Les résultats obtenus, à chaque étape, ont été validés du point de vue médical par le Dr Vincent MALLET.

De nombreux travaux dans le domaine de la recherche documentaire et de la production d'information à valeur ajoutée ont été menés. L'importance de la collecte, du tri et de l'analyse de l'information est incontestablement une priorité afin d'obtenir une information qualitative et quantitative la plus pertinente possible.

En effet, du point de vue qualitatif l'objectif est d'obtenir une information personnalisée en relation avec ses activités de recherche et/ou son domaine de compétence (analyse sémantique). L'aspect quantitatif (bibliométrie) permet de proposer des informations issues de traitement mathématique (statistique, analyse de données). L'approche proposée a pour objet de structurer les connaissances générées par un ensemble de 2 modules d'analyse s'appuyant sur un module d'exploitation de l'information textuelle fondée sur l'analyse mémorielle et un module fondée sur les méthodes d'analyse mathématique.

Dans ce but une plateforme d'analyse et de veille a été mise en place. La plateforme **PIETRA** - Plateforme d'Information Elaborée et Technologies en Recherche Appliquée - comporte un ensemble d'outils de traitement, **MEVA** - Mémoire Événementielle d'Actualisation -, **MATHEO** et les modules d'enseignement à distance (e-learning), qui utilisent le service du campus numérique de l'Université de Marne la Vallée.

Les sources d'information explorées et analysées peuvent être de nature textuelle (Internet...) ou structurée (banque de données...). Dans le cas de cette étude nous avons utilisé la banque de données Medline, référence dans le domaine médical, accessible sur Internet via Pubmed développé par le National Center for Biotechnology Information NCBI à la National Library of Medicine (NLM).

L'équation de recherche utilisée est la suivante :

Stem cell therapy

Cette équation très générique permet de capturer un ensemble conséquent de références de façon à pouvoir explorer un corpus le plus large possible.

Dans cette étude nous nous sommes limité à la source pubmed, mais nous pouvons dans une perspective d'analyse exhaustive, à l'aide de métacrawler et dans l'ensemble du web, obtenir ainsi aussi du texte (html, pdf, document Word...).

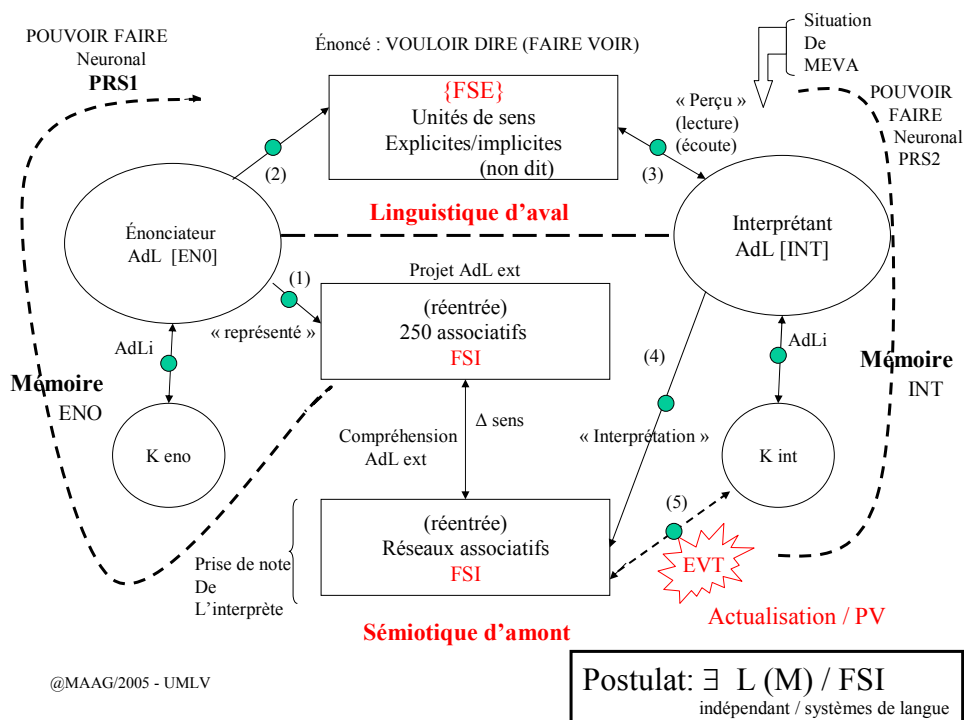
L'approche mémorielle MEVA

L'approche mémorielle MEVA s'appuie, notamment du point de vue théorique, sur les travaux de recherche effectués dans le domaine de la traduction simultanée de Danica SELESKOVITCH et Marianne LEDERER (Langage, langues et mémoire) à l'Ecole Supérieure d'Interprètes et de Traducteurs (ESIT) - Université Sorbonne Paris 3 – [16] et un ensemble de travaux concernant les neurosciences – Antonio DAMASIO - et sciences cognitives référencés par Gérard SABAHA [15]

Les orientations proposées par ces équipes s'appuient sur l'analyse suivante :

Toute compréhension d'une parole, d'un texte ou d'une image dépend essentiellement d'un savoir. Plus précisément, en situation d'acte de langage, l'interprétant fait appel à ses connaissances, son acquis mémoriel, pour progresser d'unités de sens en unités de sens et effectuer ainsi une interprétation possible de l'énoncé (ou d'une partie de l'énoncé) ; cette progression du processus cognitif d'interprétation n'est pas linéaire, une forme sémiotique externe (formule chimique, expression formelle, mot ou expression de langue etc.) peut provoquer un retour arrière pour satisfaire une liaison ou encore une anticipation afin de détecter la présence (ou l'absence) d'une autre forme sémiotique externe, à laquelle il faut ajouter des connaissances implicites requises pour la compréhension de l'énoncé (par exemple parler « d'injection ARNi (Acide RiboNucléique interférent) » en médecine ou en biologie sous-entend une stratégie thérapeutique appliquée à un modèle « souris » et fait sans doute allusion à une recherche concernant les hépatites gravissimes). Cet ajout de contexte est directement issu des capacités mémorielles de l'expert et de son vécu.

Pour MEVA, l'énoncé n'est qu'une suite de formes sémiotiques externes qui correspondent à ce qui est exprimé par l'énonciateur ou l'auteur. Certaines de ces formes sémiotiques externes, empruntant fréquemment des formes synonymes, favorisent des liaisons mémorielles amont, aval ou implicites et déterminent ainsi un (ou plusieurs) contextes d'interprétation possible(s). Un contexte d'interprétation, une fois appréhendé se représente sous forme sémiotique interne à laquelle nous associons un descripteur mémoriel (voir schéma ci-dessous).



Acte de langage (figure 1)

L'acte de langage met en jeu 2 personnes, un énonciateur (eno) et un interprétant (int), tous deux disposant d'une mémoire et de mécanismes cognitifs propres, soit à la production d'un énoncé, soit à l'interprétation d'un énoncé. Chacun d'eux présuppose chez l'autre un ensemble partagé de connaissances acquises.

Cette déclaration permettra d'insérer l'index mémoriel kni/3 dans le bloc index attaché à une partie de l'énoncé (par exemple le paragraphe d'un article scientifique).

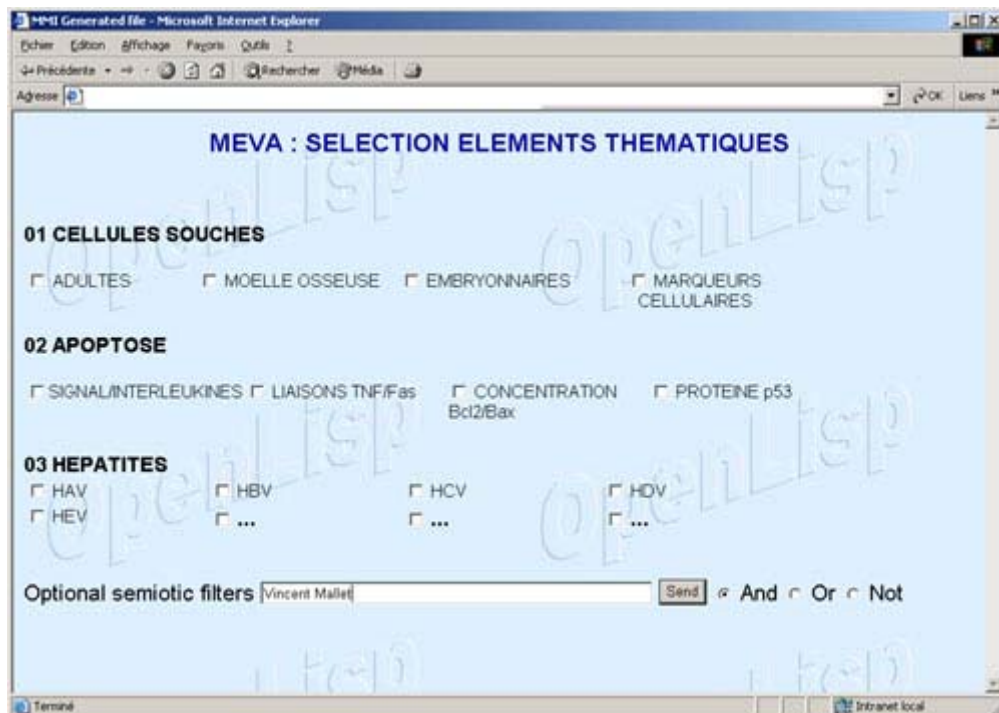
Un tel mécanisme déclaratif, fixé par le chercheur, va permettre de corrélérer les documents vérifiant une thématique du plan de veille, à un certain niveau de connaissance (un protocole thérapeutique par exemple), ce qui évite le « bruit » de réponses se situant à des niveaux disparates de connaissances.

Les résultats obtenus entrent directement dans le management de la connaissance (KM) qui intervient après une phase d'analyse automatique associant le chercheur.

Présentation des résultats obtenus par MEVA

La présentation des résultats de l'atelier MEVA se fait grâce à une interface utilisateur simple et sécurisée, utilisant un navigateur quelconque (Internet Explorer, Netscape, Mozilla Firefox...). L'utilisateur peut également avoir un accès aux éléments de son plan de veille, ce qui lui permet de croiser des contextes (back to selection) afin de constituer des thèmes de recherche, de créer de nouvelles combinaisons et d'obtenir de nouveaux documents pertinents.

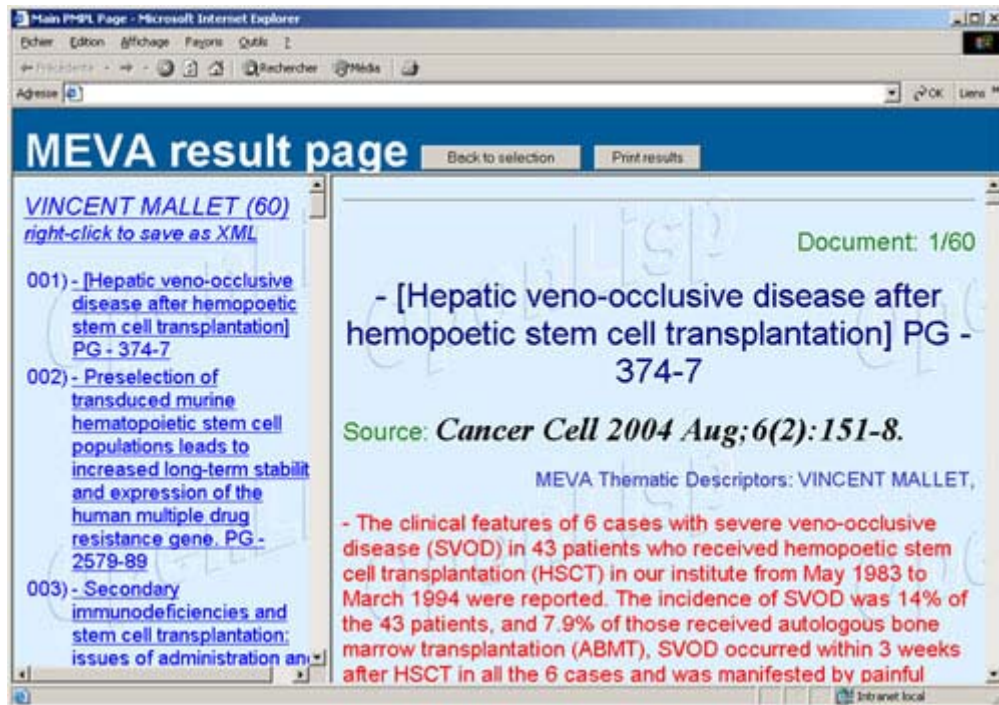
Les thésaurus mémoriels associés aux connaissances, mis en jeu par ses choix, ont été automatiquement chargés et génèrent des informations correspondant à un plan de veille générique qui aboutit à 478 références. Par combinaison d'éléments thématiques, nous pouvons obtenir de nouveaux documents pertinents correspondant à la sélection opérée.



(figure 2)

La recherche thématique peut éventuellement s'affiner, en renseignant le filtre sémiotique, en utilisant des mots clés accompagnés d'opérateurs booléens. Par exemple (figure 2), la combinatoire des éléments thématiques du plan de veille (associations de contextes remarquables fournies par l'expert) se réalise par l'entrée de l'expression : « Vincent Mallet ».

Nous obtenons 60 réponses pertinentes sur les 12 648 références, issues de la capture sur Pubmed, qui se présente de la manière suivante (figure 3) :



(figure 3)

L'analyse mathématique

Le deuxième outil d'analyse permet d'effectuer des traitements statistiques afin d'obtenir de l'information à valeur ajoutée sur les différents acteurs (médecins, chercheurs, etc.), les organismes auxquels ils appartiennent (laboratoires, services, etc.) et tout autre champ exploitable.

Le logiciel utilisé pour ce type de traitement est Matheo Analyzer, issu d'une coopération entre le Centre de Recherche Rétrospective de Marseille (CRRM) de l'université de Marseille Aix-Marseille III et la société Information Management Consulting et Solution (IMCS). Il permet, à partir d'un document structuré, du type banques de données bibliographiques, présenté sous forme de champs type : auteurs, affiliations, mots clés ..., de créer des indicateurs et de générer automatiquement des réseaux d'entités. Les résultats sont présentés sous forme d'histogrammes, de réseaux et de matrices de corrélation.

Après une première phase d'analyse du corpus MEVA, indexation mémorielle appliquant un plan de veille, le système génère de façon automatique, un sous-ensemble de documents qui contiennent un ou plusieurs contextes pouvant satisfaire des associations thématiques remarquables (478 notices). Un premier corpus d'information enrichi (**Corpus1** – format XML) est ainsi établi en implantant un champ supplémentaire. Ce dernier décrit les thématiques de recherche obtenues selon le plan de veille - ce champ a été nommé THM (THématique Mémorielle) :-

THM : ...cs/foie/transplant signifie un **contexte de recherche** mettant en relation des contextes de «cellules souches», de «foie» et de «transplantations».

île Rousee 2005
Journée sur les systèmes d'information élaborée

Exemple de banque de données Medline sur Pubmed (format texte)

```
PMID- 15633661
OWN - NLM
STAT- MEDLINE
DA - 20050106
DCOM- 20050127
PUBM- Print
IS - 0017-8594
VI - 63
IP - 11
DP - 2004 Nov
TI - Case report: a common presentation of a rare disease-hepatosplenic T-cell
    lymphoma.
PG - 341-3
AB - Hepatosplenic T-cell lymphoma is a rare neoplasm characterized by systemic....
AD - US Army, Medical Corps, Tripler Army Medical Center, Honolulu, Hawaii
    96859, USA. jamalah.munir@haw.tamc.amedd.army.mil
FAU - Munir, Jamalah
AU - Munir J
FAU - Preston, Glenn
AU - Preston G
FAU - Polish, Roger
AU - Polish R
LA - eng
PT - Case Reports
PT - Journal Article
PL - United States
TA - Hawaii Med J
JID - 2984209R
SB - IM
MH - Diagnosis, Differential.....
EDAT- 2005/01/07 09:00
MHDA- 2005/01/28 09:00
PST - ppublish
SO - Hawaii Med J 2004 Nov;63(11):341-3.
```

Intégration du premier champ : THM

THM - thm/pvmedreg/csa/foie/therapy;thm/pvmedreg/cs/foie/transplant

Le champ THM fait apparaître dans la notice ci-dessus, deux thématiques séparées par un « ; » :

- l'une réfère au contexte de thérapie cellulaire
- l'autre fait état d'un contexte plus précis de transplantation.

Les champs choisis pour notre analyse sont : AU auteurs, DA (affiliation), LA (langage de la publication), TA (Type de journaux), MH (mots clés du MESH), DP (date de publication), THM (THésaurus Mémoirel).

L'ensemble des éléments enrichis est mis en jeu dans l'analyse statistique effectuée par Matheo Analyzer.

Les entités d'information accessibles peuvent être classées par ordre de fréquence décroissante ou par paire d'entité homogène (intra-champ) ou hétérogène (inter – champ).

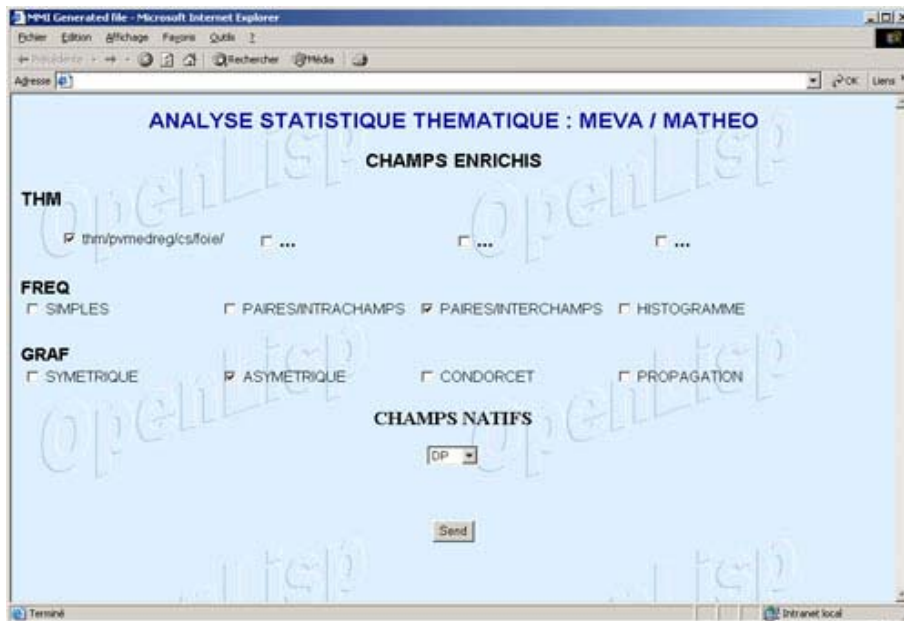
Les résultats obtenus donnent lieu à des représentations graphiques (histogrammes, réseaux, cartes...).

Un deuxième enrichissement des champs de la banque de données peut alors avoir lieu par l'intégration de deux autres champs supplémentaires **FREQ** et **GRAF**.

île Rouse 2005
Journée sur les systèmes d'information élaborée

De la même manière que précédemment l'interface utilisateur permettra une sélection dynamique des champs **FREQ** et **GRAF** pouvant être croisés avec les thématiques (champs **THM**) (figure 4) .

Par exemple, la figure suivante présente l'ensemble des nouveaux champs (THM, FREQ et GRAF) qui peuvent subir des combinaisons par l'expert et qui donne les résultats présentés plus bas (figure 5) :



(figure 4)

Intégration du deuxième champ : FREQ

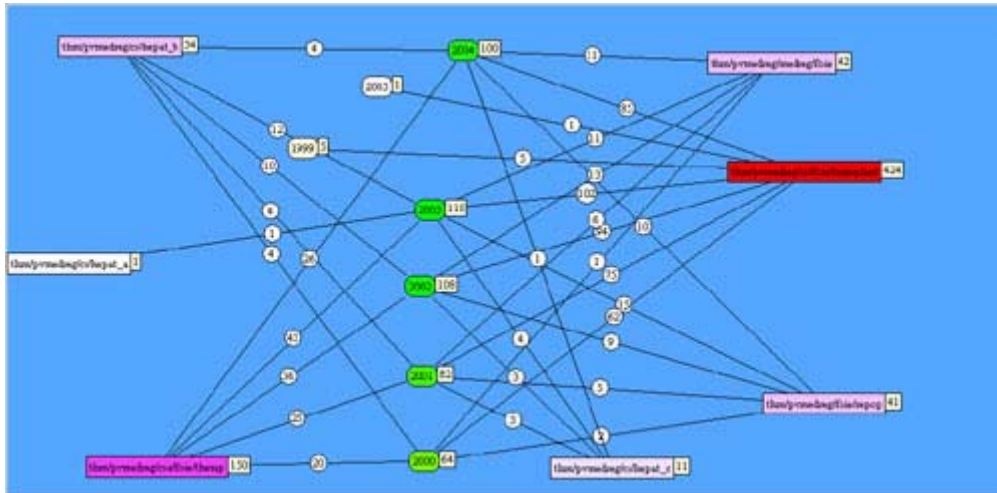
Exemple de résultats obtenus après la sélection « thm/DP »

THM	DP	Fréquence
thm/pvmedreg/cs/foie/transplant	2003	102
thm/pvmedreg/cs/foie/transplant	2002	94
thm/pvmedreg/cs/foie/transplant	2004	85
thm/pvmedreg/cs/foie/transplant	2001	75
thm/pvmedreg/csa/foie/therap	2000	62
thm/pvmedreg/csa/foie/therap	2002	43
thm/pvmedreg/csa/foie/therap	2004	36
thm/pvmedreg/csa/foie/therap	2001	26
thm/pvmedreg/csa/foie/therap	2000	25
thm/pvmedreg/foie/repop	2003	20
thm/pvmedreg/medreg/foie	2002	15
thm/pvmedreg/cs/hepat_b	2003	13...

(figure 5)

Intégration d'un troisième champ : GRAF

Exemple de résultats obtenus ci-dessous après la sélection « thm/DP » (figure 6)



(figure 6)

Dans notre exemple, l'expert nous a demandé de faire une recherche sur ses thématiques de recherche par rapport aux dates de publication. Nous avons constaté une forte mobilisation du contexte « thm/pvmedreg/cs/foie/transplantation », c'est à dire tout ce qui concerne les cellules souches, le foie et la transplantation en 2004, qui est un des axes importants de recherche de notre expert (figure 6).

Nous avons ainsi créé, à partir de l'ensemble des résultats obtenus et de l'enrichissement de la banque de données initiale (Medline), une banque dérivée intégrant des informations à forte valeur ajoutée (BIE). Les caractéristiques de cette nouvelle banque sont :

- l'évolutivité liée à l'adaptation aux différentes stratégies du chercheur
- l'interactivité par l'évolution en temps réel des résultats selon les combinaisons choisies
- l'intégration de résultats graphiques

La création de ce type de banque (BIE) doit prendre en compte les nécessités temporelles des mises à jour qui doivent être accomplies selon les phases suivantes afin de tenir compte des apports de production de l'information scientifique et les besoins de l'expert.

- la capture est permanente
- les traitements MEVA sont hebdomadaires
- les traitements mathématiques sont effectués tous les 3 mois.

Ce type de choix dépend du domaine étudié et des contraintes fixées par l'expert.

Les modules d'enseignement à distance (e-learning) de la plateforme PIETRA

Actuellement nous trouvons sur le marché un grand nombre d'outils permettant l'enseignement à distance (EAD). L'université de Marne la Vallée a acquis par l'intermédiaire du service du campus numérique, l'outil WebCT (Web Course Tools).

La plateforme WebCT est actuellement la plus utilisée dans le monde de la E-formation des universités. Créée en 1997 au département informatique de l'Université de Colombie britannique et commercialisée en 1999, on dénombre à ce jour plus de 2500 universités dans 80 pays qui l'utilisent. En France, un grand nombre d'universités utilisent déjà cette plate forme.

L'accès à WebCT s'effectue par le Web, aussi bien pour les enseignants que pour les étudiants, chacun disposant d'un compte personnel protégé par un mot de passe de sorte que seuls les étudiants inscrits à un module ou à une formation peuvent accéder aux cours mis en ligne.

En raison de son caractère modulaire et interactif, la mise en oeuvre de la plateforme **PIETRA** au campus de l'université à travers WebCT offrira aux étudiants la possibilité de se former et de travailler en ligne, **notamment de créer des plans de veille**, en communiquant d'une manière interactive avec des experts et les enseignants des filières Master spécialisés en information scientifique et réseaux de santé.

Les outils principaux de WebCT :

Informations générales, calendrier, forum, E-mail, Chat, quiz – test, résultats, recherche, syllabus, glossaire, bibliographie, liens, banques d'images, index, notes.

Pour les enseignants :

Ce dispositif permet de proposer des contenus pédagogiques sur le Web sans aucune connaissance de programmation préalable grâce à une palette importante d'outils d'aide mis à disposition tel que le glossaire, les outils de recherche, etc ...

Parmi les fonctionnalités principales on note :

- 1. La production de cours**
- 2. Les outils d'évaluation**
- 3. Le travail en groupe**
- 4. Le tutorat**
- 5. Les évaluations**

Les modules de communication peuvent être utilisés dans le cadre du travail à distance avec les différents acteurs favorisant ainsi le travail collaboratif. Pour cela nous avons utilisé deux outils intégrés dans Webct : Horizonlive - classe virtuelle - (figure 7) et Eduvoice de la société Wimba. Ces outils de communication permettent d'intégrer de la voix aux outils de communications.

Communication synchrone (en direct) :

Voice Direct : « chat vocal »

Horizonlive : classe virtuelle (groupe d'étudiants à distance)

Présentation d'une classe virtuelle d'horizon Live

Communication Asynchrone (en différé) :

Mail vocal

Forum vocal

Conclusion

La plateforme PIETRA constitue un ensemble d'outils cohérents permettant l'accès à un grand nombre de sources d'information, d'en assurer une sélection pertinente, d'effectuer des analyses de contenus dont l'objectif est l'obtention d'informations à valeur ajoutée. Les résultats obtenus sont destinés à alimenter et enrichir les banques de données traditionnelles par la création de nouveaux champs d'information ajoutant aux références bibliographiques des données textuelles bibliométriques et factuelles.

Ces nouvelles sources d'information (BIE) contribuent à structurer et à normaliser les activités de gestion de connaissances (KM) et peuvent trouver des applications dans les domaines scientifiques, techniques économiques

En outre, les capacités fonctionnelles de cette plateforme constituent un dispositif opérationnel directement exploitable et utilisable dans le cadre du e-learning.

Bibliographie

- [1] ANDREÛ (Pascal), Elaboration et traitement d'information complexe pour l'aide à la décision stratégique, thèse de doctorat, université de Marne la Vallée, 1997
- [2] CALLON (Michel), COURTIAL(Jean-Pierre), LAVILLE (Françoise), Co-word analysis as a toll for describing the network of interactions between basic and technological research: the case of polymer chemistry Scientometrics, volume 22, 1991
- [3] CHANGEUX (Jean-Pierre), CONNES (Alain), Matière à penser, Odile Jacob, Paris 2000
- [4] DOU (Henri), ROSTAING (Hervé), LEVEILLE (Valérie), DOU (Jean-Marie), MANINNA (Bruno), systèmes d'analyse automatique de l'Information plates-formes de création de connaissance et travail à distance, Réseau TOPIK CNRS, 2003
- [5] DOUSSET (Bernard), GAY (B), analyse par cartographie dynamique de l'effet de l'innovation sur la structure des réseaux d'alliances dans l'industrie des biotechnologies : application au domaine des anticorps thérapeutiques, 41^{ème} journée VSST, Toulouse, octobre 2004
- [6] DUTHEUIL (Christian), Les techniques bibliométriques, Revue Française de bibliométrie volume 5, 1989
- [7] DUTHEUIL (Christian), Bibliométrie et Scientométrie en France : état de l'art, Documentaliste, Sciences de l'information volume 29, 1992
- [8] ELDELMAN (Gérard) Biologie de la conscience, Odile Jacob, Paris 2000
- [9] LAPLACE (Colette), Théorie de la traduction, Paris, 1994
- [10] LE COADIC (Yves), usages et usagers de l'information, collection 128,2004
- [11]LEITZELMAN (Mylène) DOU (Henri) et KISTER (Jackie), Modélisation de connaissances et fouille d'informations par la cartographie dynamique : application de veille technologique avec Matheo Analyzer, 2004
- [12] MUCCHIELLI (Alex), Pour une « approche communicationnelle » des TIC, CERIC, mai 2005
- [13] PRAX (Jean-Yves), Le guide du Knowledge Management, Dunod, 2000
- [14] RICOEUR (Paul), La mémoire, l'histoire, l'oubli, Paris 2000
- [15] SABAH (Gérard), Intelligence artificielle, linguistique et cognition, La linguistique cognitive, Catherine Fuchs (Ed.), Cogniprisme, Ophrys, Paris, 2004.
- [16] SELESKOVITCH (Danica), Langage, langues et mémoire, Lettres modernes, Paris 1975