



HAL
open science

Systeme d'intelligence économique et paramètres sur l'utilisateur : application à un entrepôt de publications

Babajide Afolabi, Odile Thiery

► To cite this version:

Babajide Afolabi, Odile Thiery. Systeme d'intelligence économique et paramètres sur l'utilisateur : application à un entrepôt de publications. Systemes d'informations élaborés : Bibliométrie - Linguistique, Centre d'Études Scientifiques de Défense, Jun 2005, Ile Rousse/France. inria-00000507

HAL Id: inria-00000507

<https://inria.hal.science/inria-00000507>

Submitted on 26 Oct 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Système d'intelligence économique et paramètres sur l'utilisateur : application à un entrepôt de publications

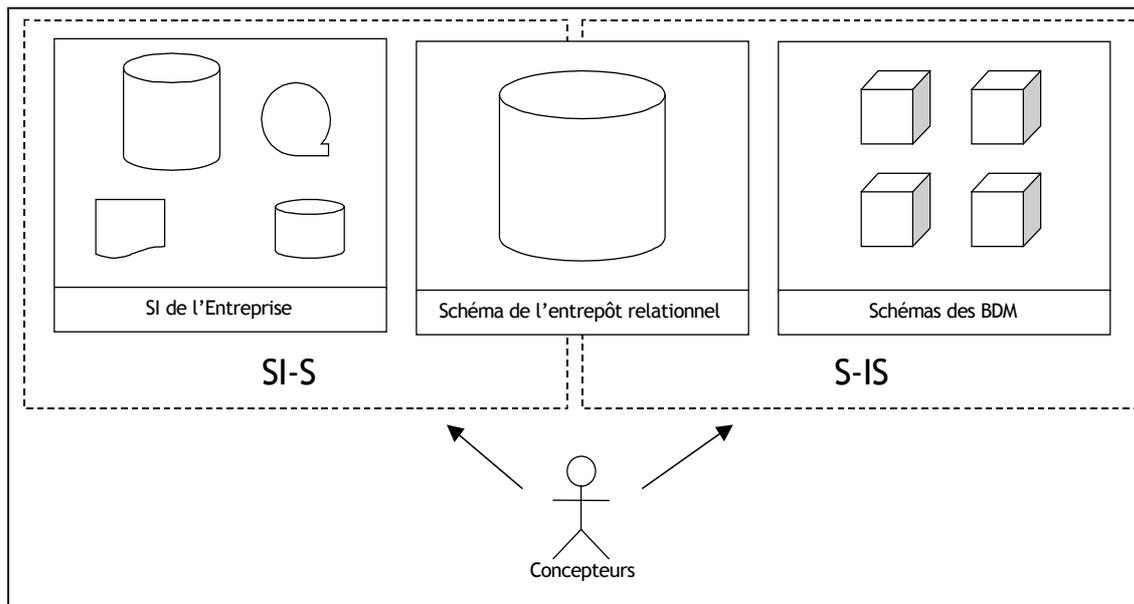
Babajide AFOLABI*, Odile THIERY*
Babajide.afolabi@loria.fr, Odile.Thiery@loria.fr

Mots-clés : Intelligence économique, entrepôt de données, modélisation utilisateur, base métier, besoin informationnel, besoin décisionnel.

1. Introduction

Selon [Revelli, 1998], l'Intelligence Economique (IE) « est le processus de collecte, de traitement et de diffusion de l'information qui a pour objet la réduction de la part d'incertitude dans la prise de toute décision stratégique ». Dans sa forme la plus simple, un système d'information (SI) peut être considéré comme un ensemble des moyens (organisation, acteurs, procédures, systèmes informatiques) nécessaires au traitement et à l'exploitation des informations dans le cadre d'objectifs définis au niveau de la stratégie de l'établissement, des métiers, et de la réglementation. Nous mettons l'accent sur deux types de systèmes d'information stratégiques (SIS). Nous appelons le premier, qui aide à la gestion de l'entreprise au jour le jour, un système d'informations stratégiques opérationnelles (SI-S), il s'agit d'un système d'informations comportant « des informations stratégiques et permettant l'automatisation de l'organisation pour satisfaire au mieux les objectifs stratégiques de la direction ». Par exemple un SI améliorant la gestion des stocks. [Tardieu and Guthmann, 1991]. Le deuxième, un système stratégique d'informations (S-IS) est, d'après nous, un système d'information « qui est consacré aux décisions stratégiques et ne comporte que des informations de type stratégiques ». Par exemple, un SI peut permettre au décideur d'observer des résultats de chiffres d'affaire par pays sur plusieurs années ou encore au veilleur de mettre en exergue les choix qu'il a faits lors de l'analyse des réponses d'un système de recherche d'information sur le Web [David and Thiery, 2003]. Le S-IS agit sur les informations déjà stockées pour aider à la prise de toute décision stratégique. Les décisions prises dépendent des informations qui s'y trouvent, et aussi de l'utilisateur qui a pour fonction l'interrogation et l'appropriation de ce système dans le but de prendre des bonnes décisions.

La figure suivante montre que le *SI de l'entreprise* est le premier à être construit. Le SI est divers et varié. Il comporte des informations *stratégiques*, par exemple des indications de chiffre d'affaires. Il faut en extraire les informations nécessaires à la prise de décision et également leur structure, c'est-à-dire les métas-données. Ceci constitue *l'entrepôt relationnel*, appelé ainsi car il est actuellement géré par un SGBD relationnel. De cet entrepôt sont extraites des *bases de données multidimensionnelles (BDM)*, appelées ainsi car elles permettent de regarder l'organisation sous différents angles ou dimensions (par exemple sur l'axe temps ou quantité vendue de produits ou encore chiffre d'affaires). Ces bases de données multidimensionnelles constituent le *S-IS*. En effet elles ne sont constituées que de données propres à la décision. [Thiery et al. 2004]



Représentation du SI-S et du S-IS

Les décisions prises en utilisant un SI, sont fondées sur l'information qui s'y trouve et elles dépendent aussi de l'utilisateur qui a pour objet l'appropriation de tels systèmes pour son processus décisionnel. Ainsi nous considérons un Système d'Intelligence Économique (SIE) comme un système qui combine les propositions des domaines des Systèmes d'Information Stratégiques (SIS), et la modélisation de l'utilisateur. Le but final d'un système d'intelligence économique est d'aider l'utilisateur / décideur dans son processus de prise de décision.

La figure 2 montre l'architecture d'un système d'intelligence économique proposée par l'équipe SITE¹. Dans cette structure nous distinguons les quatre étapes suivantes.

¹ Modélisation et Développement des Systèmes d'Intelligence Economiques – Laboratoire Lorrain de Recherche en Informatique et ses Applications Nancy, France.

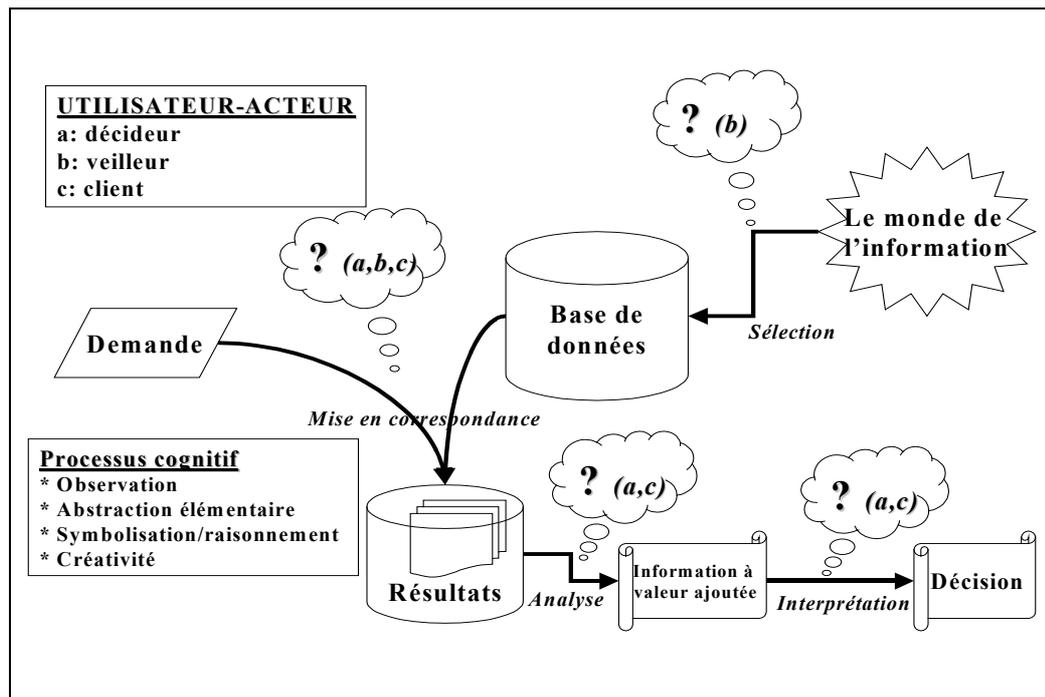


Figure 3 : Architecture d'un système d'intelligence économique

- **Sélection** : Elle permet de constituer le SI de l'entreprise qui peut être (i) la base de données de production (celle qui permet l'exploitation courante de l'organisation), (ii) l'ensemble des informations support d'un système de recherche d'informations (en documentation par exemple) ou (iii) un système d'information stratégique reposant sur un entrepôt de données. Ce SI est constitué à partir de sources de données hétérogènes, à l'aide d'un filtrage de la réalité
- **Mise en correspondance** : La mise en correspondance permet à tout type d'utilisateur d'accéder aux informations du SI. Deux principales méthodes d'accès à l'information sont actuellement proposées aux utilisateurs : accès par **exploration** et accès par **requête**. L'exploration est basée sur la technique d'hypertexte. Les requêtes sont exprimées à l'aide d'opérateurs booléens. Le résultat de la mise en correspondance est un ensemble d'informations.
- **Analyse** : Afin de donner de la valeur ajoutée aux informations trouvées, des techniques d'analyse sont appliquées au résultat. Par exemple, l'assistante d'un chef de service que nous considérons comme un *veilleur* pourra établir des tableaux de bord pour son chef de service.
- **Interprétation** : Il s'agit là de permettre au *décideur* ou en général au client du système de prendre les bonnes décisions. L'idée est que le décideur n'est pas forcément le client du système c'est à dire, par exemple, un *veilleur*. On voit alors tout l'intérêt de capturer des connaissances sur le décideur et de les mettre dans les méta-données de l'entrepôt afin de construire une base métier spécifique à un groupe de décideur ou mieux encore à un décideur particulier. Les deux termes : *veilleur* et *décideur*, seront décrits dans le paragraphe suivant.

Dans ce processus, nous pouvons également identifier trois acteurs principaux :

- **Décideur** : C'est celui qui est apte à identifier et à poser le problème à résoudre en terme d'enjeu, de risque ou de menace qui pèse sur l'entreprise. En d'autres termes, il connaît les besoins de son entreprise, les enjeux et éventuellement les risques et menaces qu'elle peut subir.
- **Veilleur** : celui-ci se spécialise dans la méthodologie de collecte et d'analyse de l'information. Son objectif est d'obtenir des indicateurs ou des informations à valeur ajoutée sur lesquels s'appuieront les décideurs. Après avoir reçu le problème à résoudre exprimé par le décideur, le veilleur doit le traduire en terme des attributs des informations à collecter qui sont nécessaires pour le calcul des indicateurs. Le veilleur doit identifier également les sources pertinentes d'informations et employer son expertise pour obtenir de ces sources des informations pertinentes car la totalité des informations dans une source pertinente d'informations n'est pas forcément pertinente. Cette dernière remarque peut être illustrée par le filtrage nécessaire sur une base de données d'organisation considérée sûre et valide mais dont toutes les informations ne sont pas nécessaires pour le problème à traiter.
- **Client** : est l'utilisateur final du système, il peut prendre la casquette du décideur ou celle du veilleur. Cet utilisateur doit être défini en prenant en compte le niveau du système avec lequel il interagit.

2. Modélisation de l'utilisateur

2.1 Rôle de l'utilisateur

L'utilisateur, en fait pour nous un décideur, a un rôle central à jouer au sein d'un système d'informations stratégiques. Son utilisation efficace du système est étroitement liée à son degré de connaissance du système. La première étape serait donc de pouvoir évaluer ses connaissances, et donc d'en déduire l'importance de son rôle, sa façon de travailler, les données les plus souvent utilisées, etc. Ensuite, à partir de ces informations, une structure personnalisée peut être générée pour améliorer son utilisation du système. Un environnement de travail complet peut améliorer grandement son efficacité. D'autre part, il pourra mettre en évidence des éléments critiques, des failles ou des lacunes du SIS. Pour cet utilisateur – décideur, le processus de prise de décision commence par la prise en compte d'un problème décisionnel, que l'on peut traduire comme son besoin décisionnel (BeD). La résolution d'un BeD comporte l'identification de ses besoins informationnels (BI), ses besoins stratégiques, ses besoins humains etc. Nous mettons l'accent sur ses besoins informationnels. Le BI peut être défini en fonction du modèle de l'utilisateur, son environnement et son (ses) objectif(s). Notre hypothèse est qu'un entrepôt de données (ou data warehouse) est à la base de tout SIS. Un entrepôt de données se définit d'abord comme une « collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision » [Inmon, 1995]. Cet entrepôt donne naissance, par filtrage par rapport à des *profils utilisateurs*, à des bases métiers (ou data marts). Ce sont des sous bases de l'entrepôt de données destinées à une fonction de l'organisme. Elles sont alimentées périodiquement et reposent sur une vue multidimensionnelle de données. Enfin, elles sont non modifiables par les utilisateurs.

2.2 Notre problématique

La plupart des systèmes d'informations existants ont été conçus en prenant en compte un stéréotype de modèle de l'utilisateur, reposant sur le type de son métier. Par exemple un SI autour de la gestion de commandes prendra en compte la façon dont on a l'habitude dans l'entreprise de gérer les commandes. Stock à flux tendu ou non ? Livraison dissociée de la facture ? Paiement à 30,60 ou 90 jours ? etc. Dès lors, ce modèle n'est pas complet puisque chaque utilisateur agit différemment selon ses besoins et sa façon de travailler. Par exemple, un utilisateur / décideur pourrait avoir un besoin spécifique à lui (au niveau de traits personnels, style cognitif, préférences) qui n'a pas été traité dans le SI. Nous essayons de répondre à la question suivante : « Quels sont les paramètres à rajouter au modèle de l'utilisateur, quelles sont les données à rajouter dans l'entrepôt ou la base métier pour répondre au mieux au besoin informationnel de l'utilisateur ? »

2.3 Le modèle proposé

L'objectif de la modélisation de l'utilisateur est de pouvoir personnaliser les réponses du système. La modélisation de l'utilisateur est la façon de représenter un utilisateur et ses comportements. Cela concerne également la façon d'exploiter les connaissances dont le système dispose à son sujet. Trois catégories de modèle sont proposées :

- Le **profil de l'utilisateur** où à un utilisateur est associé sa requête qui exprime son besoin. Dans ce contexte, le besoin de l'utilisateur est relativement stable. Le profil est appliqué aux nouvelles informations afin de lui proposer les informations les plus pertinentes.
- Le **modèle implicite** de l'utilisateur où le comportement et les préférences de l'utilisateur sont déterminés d'une manière implicite. Par exemple, la visualisation d'un document par l'utilisateur peut être interprétée comme une adéquation du document par rapport à sa requête.
- Le **modèle explicite** de l'utilisateur où le comportement et les préférences de l'utilisateur sont également représentés mais selon les spécifications de l'utilisateur. Ici, l'utilisateur ne fait que visualiser un document mais il faut qu'il indique son opinion sur le degré de pertinence du document par rapport à sa requête.

Les travaux par [Thiery and David, 2002] sur la personnalisation des réponses en système de recherche d'information ont adapté les quatre phases cognitives identifiées dans le processus d'apprentissage humain, c'est-à-dire :

- **La phase d'observation** : l'apprenant prend connaissance de son environnement par le processus d'observation ;
- **La phase d'abstraction élémentaire** : l'apprenant désigne les objets observés par des mots, ce qui correspond également à une phase d'acquisition de vocabulaire ;
- **La phase de symbolisation et de raisonnement** : l'apprenant emploie des vocabulaires spécialisés qui relèvent d'un niveau d'abstraction des concepts élevés. Par exemple, quelqu'un à un niveau d'abstraction bas peut dire « je vois un poisson », mais dira pas « je vois un piscivore » (i.e. un oiseau qui mange des poissons). Ou

encore lors de la gestion de commandes la secrétaire engage la commande auprès d'un fournisseur mais ne manipulera pas forcément aisément le compte charge correspondant ;

- La **phase de créativité** : l'apprenant découvre et s'approprie des connaissances qui ne sont pas présentées d'une manière explicite dans le système. Par exemple dans un cadre d'expérimentation du prototype BIRDS que nous avons développé, un élève a repéré le fait que tous les carnivores ont des pattes courtes, ce qui n'avait pas été représenté explicitement dans le système.

Ce modèle a été transformé en un modèle de l'utilisateur dans un cadre de recherche d'information. Les deux premières phases (observation et abstraction élémentaire) sont combinées dans une seule phase **exploration** qui nous donne un modèle :

$$M = \{\text{Identité, Objectif, \{Activité\} \{Sous-sessions\}}\}$$

Où

$$\text{Activité} = \{\text{Type-activité, Classification, Evaluation}\}$$
$$\text{Type} = \{\text{Exploration, Requête, Synthèse}\}$$
$$\text{Classification} = \{\text{Attributs, Contraintes}\}$$
$$\text{Evaluation} = \{\text{Solution du système, Degré de pertinence}\}$$

- **Identité** : L'identité de l'utilisateur. Ce paramètre permet d'individualiser l'historique des sessions de l'utilisateur.
- **Objectif** : L'objectif principal ou *besoin* de l'utilisateur pour la session.
- **Activité** : Une activité de l'utilisateur pour obtenir des solutions à son besoin en information. Une session est composée de plusieurs activités. Une activité est définie par trois paramètres : *type-activité, classification, évaluation*.
- **Type-activité** : Les types d'activité correspondent aux différentes phases d'habitudes évocatives, à savoir ici *l'exploration, la requête et la synthèse*.
- **Classification** : La classification est l'approche que nous employons pour l'accès à l'information. La technique de classification permet à l'utilisateur d'exprimer sa demande en information dans les phases d'habitudes évocatives que nous implémentons. L'utilisateur pourra spécifier les *attributs* des documents à classer et les *contraintes* qui doivent être satisfaites par les documents. C'est ce que nous appelons « classification avec contraintes ». Par exemple, l'utilisateur peut spécifier l'attribut *auteur* et *équipe=SITE*. Notre système donnera à l'utilisateur les fréquences des publications des auteurs, mais uniquement des auteurs de l'équipe SITE. Le système permettra également à l'utilisateur d'accéder aux documents publiés par chacun des auteurs. L'utilisateur peut spécifier plusieurs attributs. Tous les attributs spécifiés peuvent être identiques ou différents. Par exemple (*attribut : auteur, auteur*) permet d'obtenir les fréquences des co-publications des auteurs.
- **Évaluation** : L'utilisateur pourra évaluer la *pertinence* de toutes les *solutions* du système. Notons que l'évaluation de l'utilisateur porte sur le degré de pertinence et les raisons du jugement. Par exemple, un document peut être jugé *non-pertinent* parce que l'utilisateur le *connaît déjà*. Nous ne voulons pas faire abstraction de la raison de la non pertinence du document pour l'utilisateur en tant qu'individu. Nous voyons ici un

élément important de l'individualisation de notre modèle. Nous ne pouvons pas généraliser le jugement de la non-pertinence d'un document par un utilisateur sans connaître les raisons. Ces raisons ne sont pas identiques pour tous les utilisateurs.

- **Sous-sessions** : Une sous-session est représentée de la même manière que la session principale sauf pour le fait que l'objectif sera associé à l'objectif principal et ne constituera pas une session à part.

Le modèle de l'utilisateur a permis de proposer une architecture de SI qui repose sur l'évolution cognitive de l'utilisateur. L'architecture permet à l'utilisateur d'explorer la base d'informations pour découvrir son contenu, formuler des requêtes, effectuer des annotations et lier des activités de recherche à un besoin en information (objectif).

2.4 Le besoin informationnel

Le besoin informationnel d'un utilisateur est un concept qui varie en définition, selon les chercheurs et selon les différents utilisateurs. Il existe des recherches qui ont essayé de lui donner une définition [Campbell and Rijsbergen, 1996], [Xie, 2000]. Pour nous un besoin informationnel est une représentation informationnelle d'un problème décisionnel. Définir un problème décisionnel implique une connaissance sur l'utilisateur et son environnement. Donc, un problème décisionnel est une fonction d'un modèle de l'utilisateur, de son environnement et de son objectif. Nous nous appuyons sur la définition d'un problème décisionnel selon [Bouaka, 2003]. Le problème décisionnel a été défini ainsi :

$$P_{\text{décisionnel}} = f(\text{enjeu, caractéristiques individuelles, paramètres de l'environnement})$$

L'enjeu est ce que l'organisation risque de perdre ou de gagner selon la décision prise. Il est composé de l'Objet, le Signal et l'Hypothèse.

Les caractéristiques individuelles font référence à l'utilisateur, ses comportements et ses préférences.

Les paramètres de l'environnement désignent les apports de la société sur l'organisation. Ils peuvent être globaux ou immédiats.

Donc, un problème décisionnel peut être défini comme

$$P_{\text{décisionnel}} = f((O, S, H), (SC, TP, I), (EG, EI))$$

Où :

O = l'**Objet** de l'environnement détecté par le décideur ;

S = le **Signal** associé à cet objet, c'est-à-dire la signification de l'objet d'après le décideur ;

H = les **Hypothèses**, qui traduisent les résultats possibles ;

SC = le **Style Cognitif** qui représentent les différences spécifiques et individuelles à chaque personne ;

TP = les **Traits de Personnalités** qui sont un ensemble des structures cognitives et affectives conservées dans le temps sur les individus pour faciliter leur adaptation aux événements, aux circonstances, aux thèmes, aux individus et leur perception de l'environnement et enfin aux décisions ;

I = l'**identité** qui est une référence unique pour chaque utilisateur et pour chaque session ;

EG = l'**Environnement Global**, il regroupe l'environnement social, politique, économique. C'est-à-dire l'image de l'organisation.

EI = l'**Environnement Immédiat**, qui a une implication directe sur l'organisation. Par exemple, les clients, les fournisseurs, les concurrents etc.

3. Domaine d'Application

Actuellement nous travaillons sur une version nouvelle de METIORE [David, 1999] (qui est une évolution de STREEMS, une application qui gère des données botaniques et des données d'exploitation sur les arbres) fondée sur la plateforme JAVA qui sera accessible à partir de l'Internet et donc disponible pour tous. La nouvelle version sera utilisable avec n'importe quel fichier valide XML ; ce fichier étant une base pour les différentes analyses souhaitées.

METIORE est un système de recherche d'informations appliqué aux références bibliographiques. Le système est actuellement expérimenté sur une collection de plus de 6000 références bibliographiques du laboratoire LORIA². Le système permet à l'utilisateur de pouvoir évoluer dans ses niveaux d'habitudes évocatives. Cette fonctionnalité est implémentée dans le système par la méthode de recherche d'information que nous avons développée et que nous appelons analyse croisée avec contraintes.

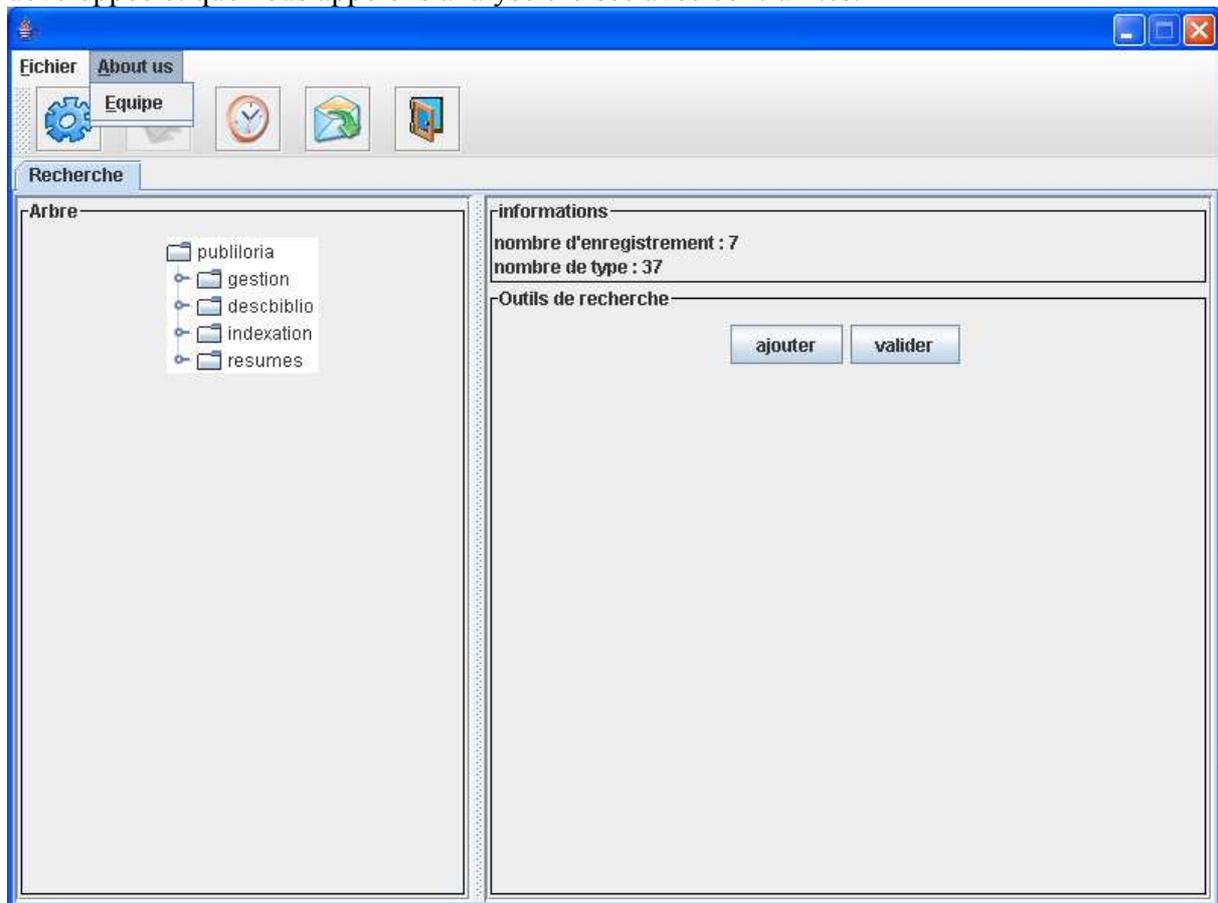


Figure 3 : La page d'accueil de METIORE

Comme toute application qui peut être utilisée dans un environnement de recherche coopérative d'information, METIORE est composé d'une interface graphique. Rappelons que l'interface de l'application permet à l'utilisateur d'effectuer des interactions avec le système.

² Laboratoire Lorrain de Recherche en Informatique et ses Applications.

La méthode de recherche d'information par analyse croisée avec contraintes permet de formuler une recherche classique et d'indiquer des attributs pour une analyse croisée. Une recherche classique est composée de critères combinés par des opérateurs booléens. Par exemple dans METIORE, (author * david) OU (year >=1990) indique que l'utilisateur cherche des références où l'auteur contient le nom «david» ou des publications à partir de 1990. L'interface d'analyse croisée avec contraintes permet d'indiquer trois attributs pour une analyse. Les combinaisons possibles des attributs sont présentées dans le Tableau 1 : (*Types de spécification des attributs pour les analyses*).

Pour le type 1, aucun attribut n'est spécifié par l'utilisateur. Dans ce cas, nous utilisons le champ «référence» comme attribut1.

Le type 2 permet à l'utilisateur de spécifier l'un des attributs d'une référence bibliographique. Le système fournit les fréquences des valeurs de l'attribut.

Les types 3 et 4 permettent à l'utilisateur de spécifier deux attributs pour l'analyse. Dans le cas du type 3, les deux attributs sont identiques. Cela permet de calculer les co-occurrences intra-champ. Par exemple auteur, auteur pour connaître les co-auteurs. Dans le cas du type 4, les deux attributs sont différents. Cela permet de calculer les co-occurrences inter-champ. Par exemple auteur, année pour connaître les publications par année pour chaque auteur.

Les types 5 et 6 permettent à l'utilisateur de spécifier trois attributs pour l'analyse. Pour ces deux cas, l'attribut 3 doit être différent de l'attribut1 et attribut 2. Par contre dans le cas du type 5, l'attribut1 et l'attribut 2 sont identiques et dans le cas du type 6 l'attribut1 est différent de l'attribut 2. En spécifiant l'année comme le troisième attribut, l'utilisateur peut observer l'évolution des co-occurrences des valeurs des deux premiers attributs.

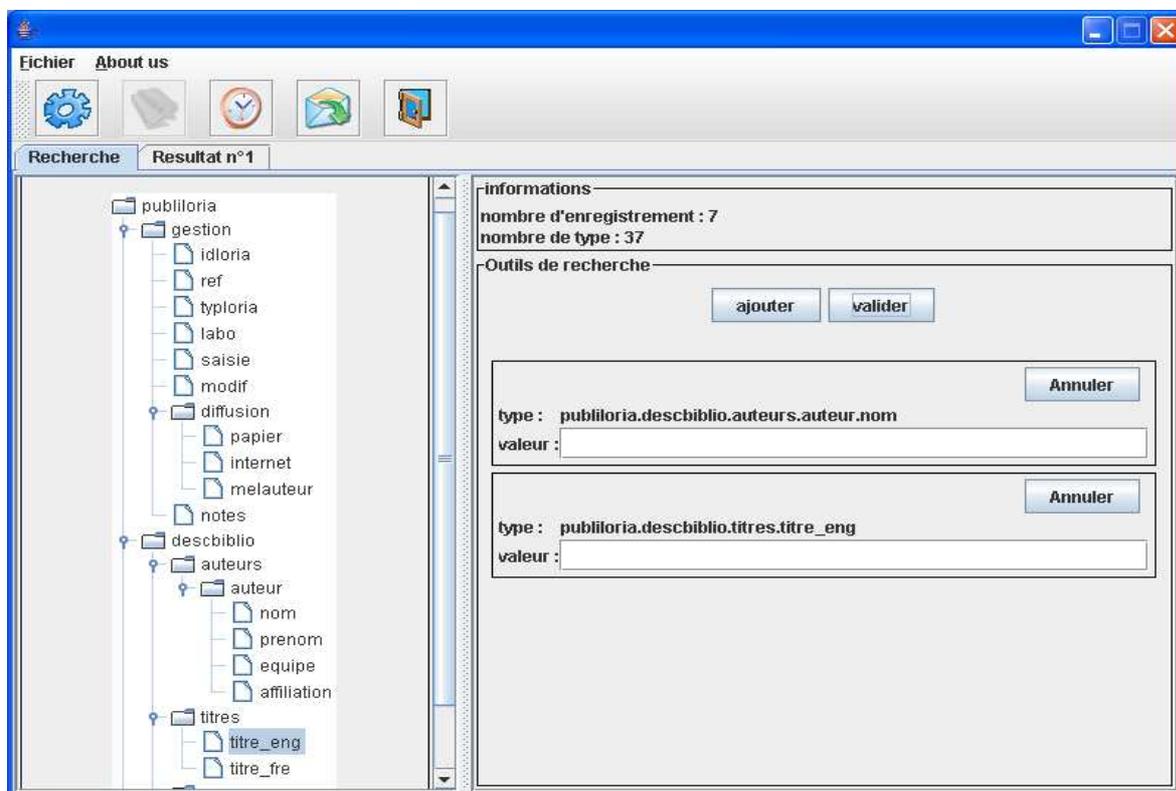


Figure 4 : Une requête en METIORE

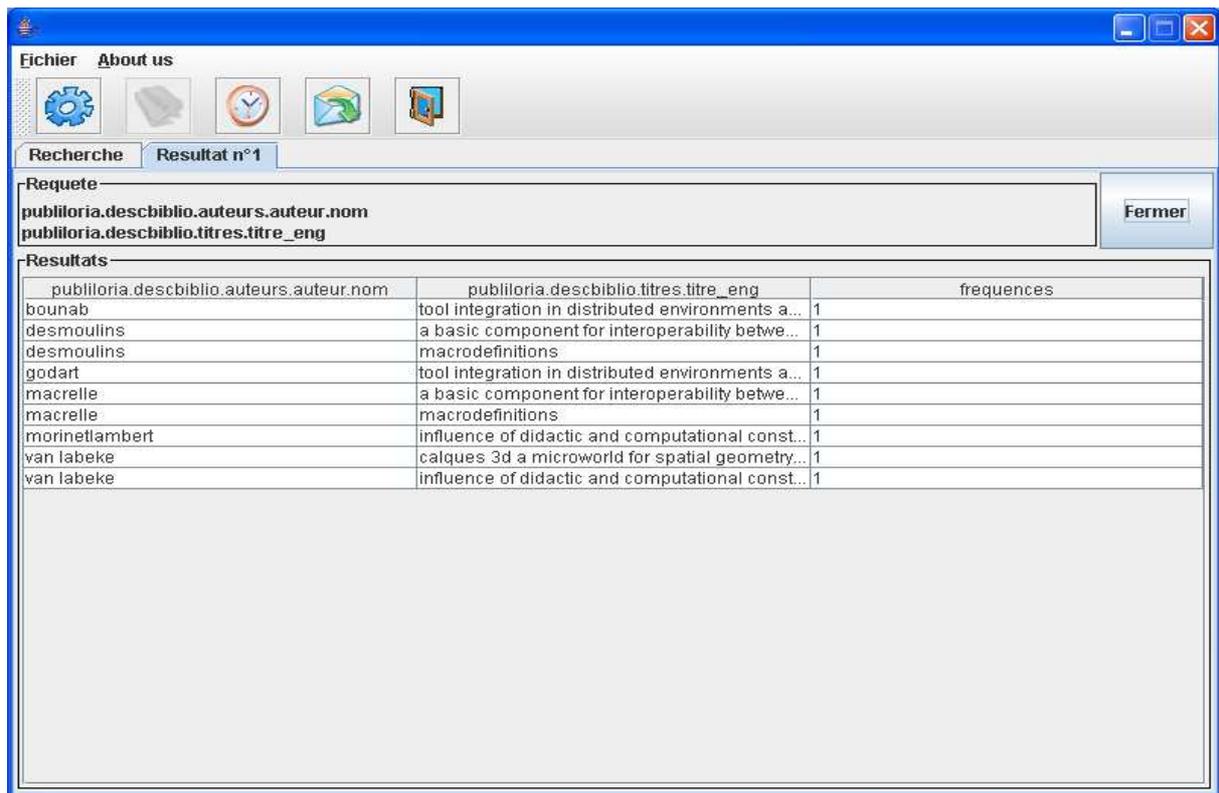


Figure 5 : Premiers résultats

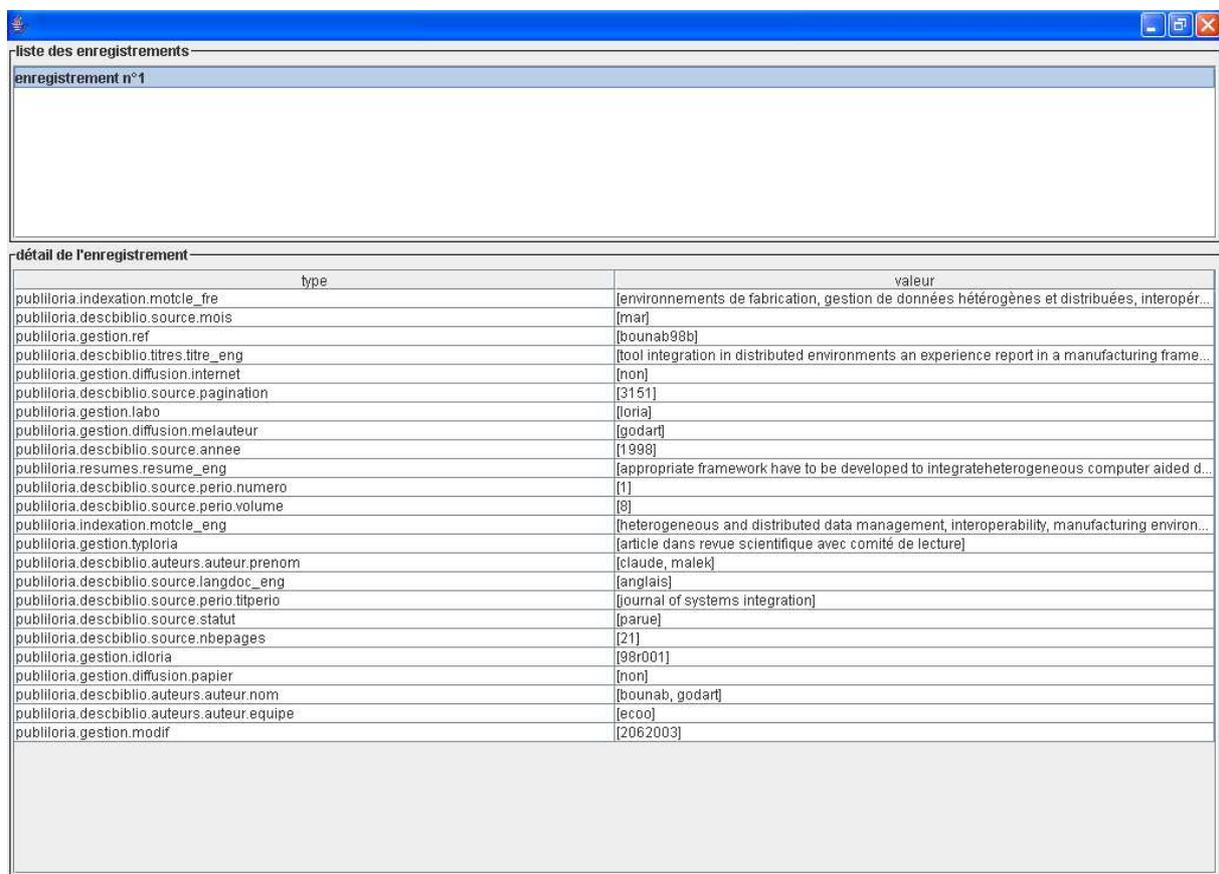


Figure 5 : Une représentation d'une fiche correspondante à un des résultats.

Il faut noter que seuls les attributs sont spécifiés par l'utilisateur. En donnant les fréquences des valeurs des attributs spécifiés par l'utilisateur, le système donne une vue globale de la base à l'utilisateur. Le système fournit également la liste des références correspondant aux résultats des analyses, ce qui donne un moyen à l'utilisateur pour découvrir les valeurs possibles pour les attributs. Cette fonctionnalité facilite donc la recherche par le contenu. Tous ces types d'analyse permettent de suivre l'évolution et l'émergence de concepts.

TYPE	ATTRIBUT1	ATTRIBUT2	ATTRIBUT3	VARIATION
1				Utiliser «référence» comme Attribut1
2	x			
3	X	x		Attribut1 == Attribut2
4	X	x		Attribut1 != Attribut2
5	X	x	x	Attribut1 == Attribut2
6	X	x	x	Attribut1 != Attribut2

Tableau 1 : Types de spécification des attributs pour les analyses

Quand l'utilisateur démarre l'application, nous lui demandons son objectif principal. Nous regardons si l'utilisateur a déjà formulé un objectif similaire. Si oui, le système lui propose les solutions classées selon les types d'évaluation qu'il a données. Il peut changer les évaluations. S'il est satisfait des solutions, alors il peut arrêter sinon il peut modifier son objectif ou effectuer d'autres recherches. Nous introduisons ici la notion de sous-objectif. S'il n'y a pas d'objectif enregistré pour l'utilisateur qui ressemble à celui qu'il vient de présenter, le système regarde s'il y aurait des objectifs similaires qui ont été présentés par d'autres utilisateurs. Si oui alors cette solution servira de point de départ pour le système. L'utilisateur pourra toujours évaluer les solutions du système. S'il n'y a pas d'objectif similaire retenu par d'autres utilisateurs, alors l'utilisateur doit effectuer des recherches pour trouver ce qu'il cherche. Ceci est possible puisque chaque activité (à chaque fois qu'il se connecte au système) de l'utilisateur est enregistrée. Le système alors observe le comportement de l'utilisateur. Par exemple, si l'utilisateur répète une même requête pour le même objectif alors le système peut déclencher un dialogue pour savoir pourquoi. L'intégration du paramètre «objectif» dans le modèle de l'utilisateur est déjà prévue dans l'interface d'accès et d'analyse de l'information par les zones «OBJECTIF» et «Mots clés». La zone «OBJECTIF» donnera le texte complet de l'objectif de l'utilisateur. La zone «mots clés» donnera la liste de mots clés associés à cet objectif.

Ces activités enregistrées pour chaque utilisateur seront stockées dans un fichier XML qui sera analysé afin de pouvoir améliorer les propositions du système.

Comme nous l'avons déjà dit nous nous plaçons dans ce cadre d'applications de « recherche d'information », reposant sur l'entrepôt de documents publiés par des chercheurs du centre de recherche LORIA. Cet entrepôt contient les publications, historisées et classées selon la nomenclature habituelle, des membres du laboratoire. Nous avons déjà travaillé sur la

classification, la normalisation et l'amélioration de telles ressources documentaires électroniques et notre objectif est de constituer un réel entrepôt documentaire depuis lequel nous pourrions susciter tout type d'analyse d'information. En particulier, nous nous proposons de produire des bases métiers selon les différents utilisateurs de l'entrepôt, en utilisant le contenu du fichier XML, produit des activités de l'utilisateur.

Par exemple, voyons les bases métiers obtenues lors d'une analyse de cet entrepôt de documents. Nous ne cherchons plus les documents même mais les informations sur ces documents qui peuvent aider pendant un processus de prise décision. Dans la figure 3, nous pouvons observer les bases métiers pour les 3 types d'utilisateurs essentiels :

1. L'une (i) contient des documents et une représentation de l'évolution des recherches des utilisateurs selon un sujet de recherche, par exemple un thésard qui fait une recherche sur un sujet précis qui veut savoir « qui a écrit quoi ? » ou « quelles sont les publications d'un auteur soit toutes, soit par une année particulière ».
2. La deuxième (ii) concerne les documents qui aident à savoir l'évolution des demandes d'un utilisateur au cours du temps afin de pouvoir le tenir au courant des nouvelles acquisitions dans son domaine de recherche ou de ne pas lui recommander la même chose plusieurs fois, le destinataire est logiquement le documentaliste ;
3. La troisième (iii) concerne des documents selon l'évolution des publications des équipes de recherche dans un centre de recherche au cours du temps. Cette information peut être intéressante pour la direction de ce centre de recherche mais l'est moins pour les documentalistes.

Chacun des utilisateurs aura une vision différente des données de l'entrepôt par le biais des bases métiers et peut-être souhaitera que lui soient proposées uniquement les données utiles pour répondre à ses besoins.

Il faut également noter que la structure actuelle des entrepôts de données n'en permet pas une évolution facile. Ainsi lors d'une interrogation de la base de documents du LORIA, nous voulions suivre l'évolution des publications de chaque équipe. Or en fait, nous nous sommes aperçus que l'attribut « EQUIPE » qui pourrait nous aider à réaliser cette tâche n'était pas inclus dans la base car ce champ n'avait pas été prévu à la construction de la base. Un autre exemple est une question posée par un utilisateur qui voulait en savoir plus sur la visibilité des chercheurs des pays du sud. Ainsi, nous voulions voir si nous pouvons répondre à sa question en utilisant notre base de documents mais nous avons constaté que la balise/ l'attribut « pays d'origine » qui pourrait nous aider à résoudre cette question est manquante.

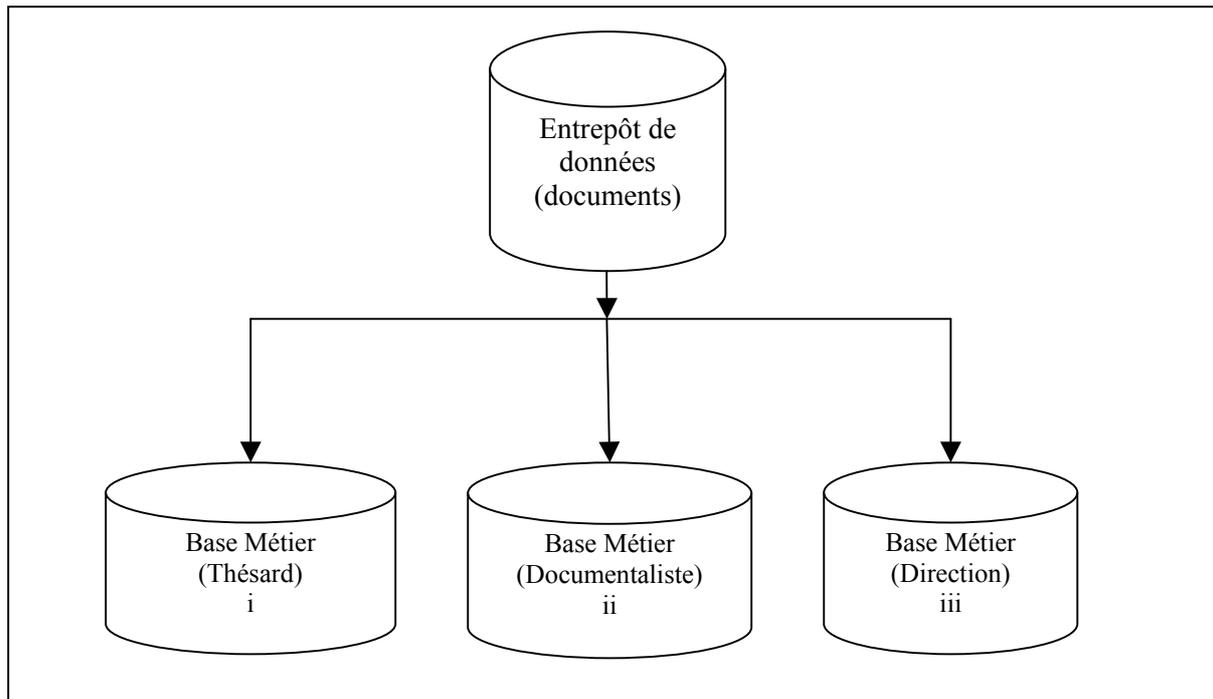


Figure 4 : Les bases métiers de l'entrepôt de documents

Nous avons donc constaté que certains attributs, nécessaires au processus de prise de décision, peuvent être absents. Et c'est un de nos objectifs de détecter ce phénomène à l'usage de la base métier et de trouver des solutions pour y remédier.

4. Propositions supplémentaires

Pour les attributs manquants, nous avons classifié les types d'informations qui peuvent manquer. Il peut y en avoir deux majeurs :

- **Les attributs liés à l'utilisateur :** ce sont des attributs qui décrivent l'utilisateur, ses préférences, ses habitudes, ses besoins informationnels etc.
- **Les attributs liés aux documents :** ce sont des attributs qui décrivent les documents qui se trouvent dans la base. Ils sont nécessaires à la description des documents.

Pour chacun de ces attributs, il y a des données associées qui sont les valeurs associables à ces attributs. Nous sommes d'avis que ces attributs peuvent être représentés dans la base mais sans les valeurs associées. Ainsi il faut remarquer que des attributs ou des valeurs peuvent être manquants.

La solution proposée, pour résoudre le problème des informations manquantes dépend de l'objet considéré : un attribut ou donnée. Une première solution est d'étendre les sources d'informations à toutes les autres bases du centre et de ne pas se fonder seulement sur les bases du centre de documentation. Par exemple, l'annuaire téléphonique du laboratoire LDAP qui contient les noms de tous les membres du centre de recherche avec leur appartenance à une équipe peut être utilisé comme une source de données sur les utilisateurs et leur affiliation respective. Les systèmes de GRH peuvent également fournir des informations intéressantes. Ce que nous proposons ici est d'étendre les sources d'informations à toutes les bases qui

peuvent assister le système pour répondre aux besoins reposant sur des informations manquantes.

On pourrait aussi exploiter les métas données de l'entrepôt. La définition des métas données de l'entrepôt doit être faite de manière à prendre en compte l'utilisateur, ses comportements, et ses préférences pour pouvoir scinder l'entrepôt en bases métiers plus intéressantes pour chaque utilisateur. Ce type de solution pourrait aider l'utilisateur dans son utilisation quotidienne du système.

Ainsi, l'étude de la base doit nous permettre de vérifier si les données peuvent répondre telles quelles à toutes les attentes de tous les utilisateurs. L'étude de l'historique des activités de l'utilisateur peut mettre en valeur ses comportements et donc donner des indications sur ses besoins.

L'idée de base de notre travail s'appuie sur l'utilisation d'une méthode d'apprentissage pour décliner un modèle de l'utilisateur, qui semble être le mieux adapté pour notre problématique. Nous envisageons également de développer un cahier de charges qui intègre les comportements de l'utilisateur. Nous tenterons, ensuite, de comparer les deux approches dans le but d'implanter l'une de deux, suivant ce qui nous paraîtra le plus intéressant.

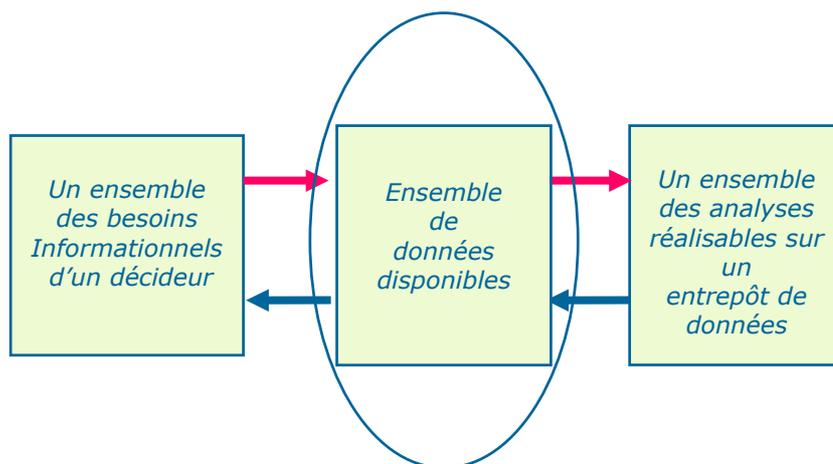


Figure 4 : Entrepôt de données

La figure 4 montre les interactions possibles entre les données disponibles dans un entrepôt (ou une base de données), les analyses réalisables sur un entrepôt de données et les besoins informationnels d'un décideur. Les besoins informationnels peuvent être transformés en requêtes (qui peuvent être des traductions ou des représentations des analyses réalisables sur l'entrepôt). Les données disponibles dans l'entrepôt sont utilisées pour répondre à ces requêtes avant de proposer une réponse au demandeur d'information. C'est-à-dire que, pour les réponses les mieux adaptées aux besoins informationnels, nous nous appuyons sur les données disponibles. Mais, nous nous demandons si les données disponibles sont suffisantes pour répondre à tous les besoins ou s'il y aura la nécessité d'acquérir des données supplémentaires ? Ou bien est-ce que les données, telles qu'elles sont, peuvent répondre à tous les besoins informationnels exprimables par l'utilisateur. Est-ce que les analyses réalisables sur un entrepôt peuvent exprimer tous les besoins ? Existe-t-il un lien entre ces

trois ensembles ? Un des nos objectifs de recherche est d'étudier les effets d'un changement d'un item d'un de ces ensembles sur les autres. Ceci, nous aiderait à constater l'existence ou non d'un mapping (ou un matching) entre ces trois ensembles.

5. Conclusion et perspectives

Dans cet article après avoir rappelé ce que nous entendons par Intelligence Economique, Système d'Intelligence Economique, Système d'Information Stratégique, entrepôt de données, base métier, nous avons explicité notre définition de l'utilisateur, de son problème décisionnel, de son besoin informationnel. Puis nous avons défini le modèle utilisateur que nous avons retenu et nous avons montré comment l'exploiter afin de construire les bases métiers orientées BI dans le but de répondre à la question soulevée par notre problématique. Ensuite nous avons présenté notre domaine d'application et les premières expérimentations que nous avons déjà menées et les solutions que nous avons trouvées pour répondre à notre problématique. Enfin nous avons présenté nos objectifs actuels de recherche.

Une bonne définition d'un besoin informationnel avec un modèle de l'utilisateur dans un contexte de la recherche d'information (cf. la section 2) peut être utilisée pour extraire d'avantage d'informations sur l'utilisateur, en particulier des informations sur ses comportements lors de ses connexions au système. Nous pourrions ainsi en déduire les raisons de ses comportements et par conséquent lui proposer des réponses adaptées à son niveau. Nous travaillons actuellement sur une modélisation du besoin informationnel.

Notre modèle du besoin informationnel doit prendre en compte l'utilisateur, ses comportements, ses préférences etc. Ce modèle nous servira à la définition de métas données à stocker dans l'entrepôt de données.

6. Bibliographie

- [Bouaka and David, 2003] Najoua Bouaka and Amos David, "Modèle pour l'Explicitation d'un Problème Décisionnel : Un outil d'aide à la décision dans un contexte d'intelligence économique" in Proceeding of the Conférence on "Intelligence Economique : Recherches et Applications", 14-15 avril 2003.
- [Campbell and Rijsbergen, 1996] Iain Campbell and Keith van Rijsbergen, "The Ostensive Model of developing information needs" in Proceeding of the Second International Conference on Conceptions of Library and Information Science, Copenhagen, October 1996.
- [David, 1999] "Modélisation de l'utilisateur et recherche coopérative d'information dans les systèmes de recherche d'informations multimedia en vue de la personnalisation des réponses" Mémoire de HDR en Sciences de l'Information et de la Communication. Université Nancy 2. 29 Juin 1999.
- [David and Thiery, 2003]. Amos David and Odile Thiery, "L'Architecture EQUA²te et son Application à l'intelligence économique" in Proceeding on the Conférence on "Intelligence Economique : Recherches et Applications", 14-15 avril 2003.
- [Haynes, 2001] S. R. Haynes, "Explanation in Information Systems: A Design Rationale Approach", PhD thesis submitted to The London School of Economics, 2001.
- [Inmon, 1995] W. H. Inmon, "What is a Data Warehouse?" Prism Tech Topic, Vol. 1, No. 1, 1995
- [Kimball et al. 1998] Ralph Kimball, Laura Reeves, Margy Ross and Warren Thornthwaite, "The Data Warehouse Toolkit: Experts Methods for Designing, Developing and Deploying Data Warehouses", Wiley, 1998.

- [Revelli, 1998] C. Revelli, "Intelligence stratégique sur Internet", Paris, Dunod, 1998.
- [Tardieu and Guthmann, 1991]. H. Tardieu and B. Guthmann, « Le triangle stratégique » Les Editions d'Organisation, 1991.
- [Saracevic, 1996] T. Saracevic, Modeling interaction in information retrieval (IR): A review and proposal. Proceedings of the American Society for Information Science, 33, 3-9. 1996
- [Thiery and David, 2002] Odile Thiery and Amos David, "Modélisation de l'Utilisateur, Système d'Informations stratégiques et Intelligence Economique", Revue Association pour le Développement du Logiciel (ADELI). 2002. n° 47.
- [Thiery et al., 2004] O.Thiery, A. Ducreau, N. Bouaka, A. David, "Piloter une organisation : de l'information stratégique à la modélisation de l'utilisateur ; application au domaine de la GRH". In Congrès Métamorphoses des organisations Nancy Novembre 2004.
- [Xie, 2000] Hong Xie, "Patterns between interactive intentions and information-seeking strategies" in Information Processing and Management 38 (2002). p 55 – 77.