



HAL
open science

Indexation des demandes formulées dans un contexte d'intelligence économique

Sahbi Sidhom, Stéphane Goria, Amos David

► **To cite this version:**

Sahbi Sidhom, Stéphane Goria, Amos David. Indexation des demandes formulées dans un contexte d'intelligence économique. Colloque Indice, Index Indexation, CERSATES, UMR-CNRS 8529 (Equipe anthropologie des savoirs) et GERICO, Nov 2005, Lille/France. <inria-00000886>

HAL Id: inria-00000886

<https://inria.hal.science/inria-00000886v1>

Submitted on 1 Dec 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.



HAL Authorization

Indexation des demandes formulées dans un contexte d'intelligence économique

Auteurs :

Sahbi SIDHOM
MCF & Chercheur
de l'équipe SITE du LORIA

Stéphane GORIA
Doctorant
de l'équipe SITE du LORIA

Amos DAVID
Professeur & Responsable
de l'équipe SITE du LORIA

LORIA - Université Nancy2, BP 239, 54506 Vandoeuvre Cedex

Sahbi.Sidhom@loria.fr

Stephane.Goria@loria.fr

Amos.David@loria.fr

Communication In :

Colloque international : « Indice, Index et Indexation », Lille, 3 - 4 novembre 2005

Résumé :

Que nous ayons à faire à des contextes purement documentaires ou dédiés à l'Intelligence Economique, un certain nombre de points communs peuvent être mis en évidence. Tout d'abord en Documentation comme en Intelligence Economique (IE), nous sommes en présence de systèmes, d'utilisateurs, de concepteurs, de fournisseurs et de demandeurs. De plus, dans ces deux contextes il question de fournir des documents pertinents en fonction des besoins des demandeurs. Dans le cadre de la modélisation des systèmes d'IE, il nous a semblé intéressant d'inclure un objet particulier correspondant à « la demande » d'information ou de documents. Or, les demandes d'information ne sont pas uniquement propres aux contextes d'IE ; c'est pourquoi, il est important de positionner aussi cette approche avec ses particularités dans un contexte dédié à la documentation sous l'angle particulier de l'indexation.

Dès lors, nous considérons que la modélisation d'un système d'information peut être plus performante si elle utilise des représentations sur : le système lui-même (son fonctionnement, sa conception, les modifications effectuées, ...), les documents et données qu'il contient, les sous systèmes d'exploitation qui le composent (thésaurii, systèmes d'acquisition de connaissances, procédures d'indexation manuelles, ...), les utilisateurs directs (informaticiens, documentalistes, veilleurs, ...) ou indirects (demandeurs d'informations, gestionnaires de centre de documentation, entreprise,) du système, et les demandes formulées ou anticipées. De cette manière, si nous considérons principalement l'apport de la représentation d'un objet « demande » dans la modélisation des interactions système-utilisateur et utilisateur expert-demandeur pour caractériser un système de recherche d'information, certains problèmes qui sont plus difficilement représentables, sinon, peuvent apparaître. Ainsi, par exemple, du point de vue système-utilisateur, l'indexation des demandes sous une forme formelle, permet de disposer d'une ressource supplémentaire de connaissances, pour faire évoluer le système et notamment influencer sur la modélisation du système d'accès aux informations. En effet, toute modélisation d'un système, produit une certaine représentation de ce système à travers l'interface d'interrogation ou de gestion du système. Si les demandes formulées ne sont pas aisément traductibles en requêtes systèmes, il est claire que : soit l'interface doit être modifiée, soit qu'il faille proposer un didacticiel à l'utilisateur ; ce qui

revient aussi en quelque sorte à modifier l'interface du système. En outre, du point de vue expert- demandeur, les problèmes d'expressions de demandes rencontrés par les documentalistes ou veilleurs qu'ils géraient implicitement à travers la formalisation d'une demande peuvent apparaître plus clairement. De la sorte, des éléments non utilisés ou abandonnés lors de la "résolution" d'une demande, de même que les reformulations éventuelles qui ont pu amener l'expert en recherche à traduire le véritable besoin du demandeur peuvent être sauvegardées et pourront aider à mieux comprendre le système lors d'une éventuelle re-conception. Un autre avantage de la formalisation et de la sauvegarde des demandes, est aussi la possibilité de disposer d'un listing des besoins insatisfaits en termes de documents présents dans le système. Par exemple, si un livre est souvent demandé, mais n'est pas disponible dans le centre de documentation ; plutôt que de ne compter que sur les facultés de mémorisation des personnels pour fournir cette information au bon moment (la réunion annuelle d'achats des nouveautés par exemple), cette information pourra remonter automatiquement lors d'une action de reporting programmer en fonction des réunions stratégiques correspondantes.

Dans une première partie de ce travail, nous allons vous présenter notre approche dans la conception d'un système d'information orienté vers l'analyse et l'indexation des demandes dans un contexte d'IE, par analogie à la gestion des entreprises selon Takeuchi et Nonaka (1997) et selon une approche d'acquisition de connaissances, qui part des documents et des méthodes d'indexations (intellectuels et automatiques) que nous envisageons.

Dans une deuxième partie, nous allons vous présenter une approche linguistique computationnelle limitée aux syntagmes nominaux (SN) pour aborder le problème de l'indexation automatique. Cette limitation n'étant pas restrictive, mais accorde un intérêt au statut du descripteur : Le syntagme nominal est en effet « *l'unité minimale du discours qui permet de désigner un objet du discours* ».

En terme de réalisation, nous serons amenés à évoluer vers un système de gestion des connaissances autour du SN qui s'applique aux demandes formulées dans le contexte de l'étude. Le mécanisme d'indexation automatique des demandes s'est concrétisé par la conception d'un noyau d'indexation automatique. Le noyau d'indexation se scinde en trois modules :

- la conception de différents outils automatiques servant à l'analyse morphologique du langage naturel. Il s'agit des ressources linguistiques composées essentiellement par les dictionnaires électroniques et la grammaire de réécriture des éléments syntagmatiques ;
- l'implémentation de l'analyseur morphosyntaxique par la compilation des règles de réécriture, pour l'analyse des corpus sur les « demandes » ;
- l'extraction des SN à partir des arbres syntagmatiques dans les demandes et leur visualisation. L'architecture de l'analyseur s'est fondée sur les automates à transitions augmentées et en cascade (CATN) de W. Woods (1980-1997).

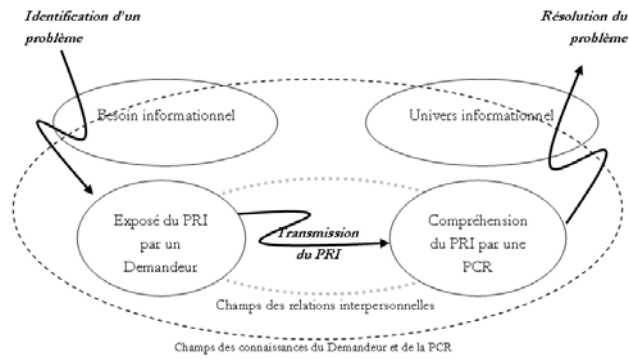
C'est à partir des objets de la « demande » d'information ou de documents que le processus de « rétro-filtrage » des connaissances s'opère pour un ajustement des demandes formulées et la prise en compte de nouveaux paramètres pour améliorer les résultats du système.

Mots-clés :

Formalisation de la demande, indexation, syntagme nominal, intelligence économique, système de recherche d'information, analyseur morphosyntaxique.

Présentation :

Modélisation systémique de la résolution d'un PRI par un tiers

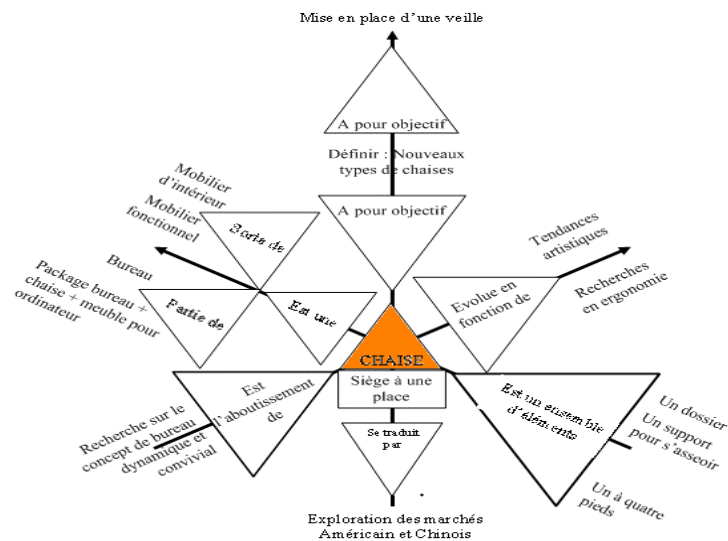


PCR : Personne Chargée de Résoudre le problème

PRI : Problème de Recherche d'Information

2

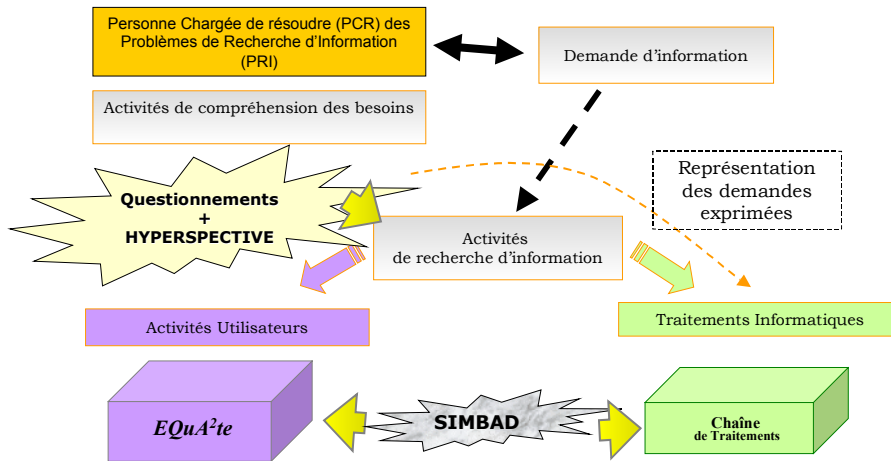
Représentation de l'interprétation d'un sujet de recherche



HYPERSPECTIVE

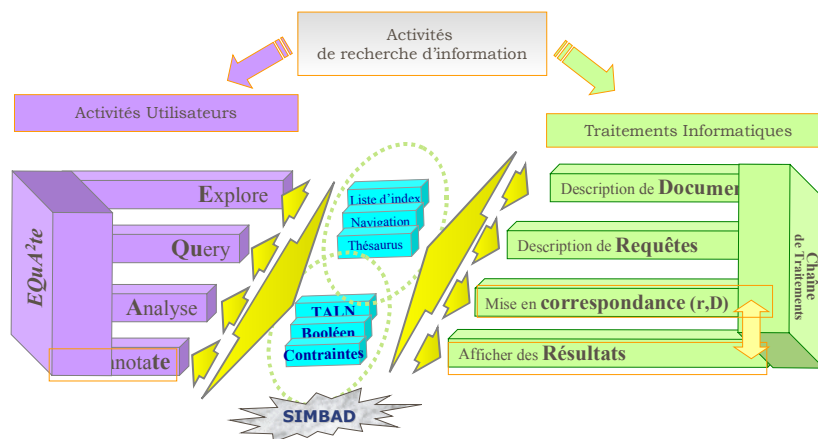
3

Architecture fonctionnelle : Demandeurs - Utilisateurs - Système



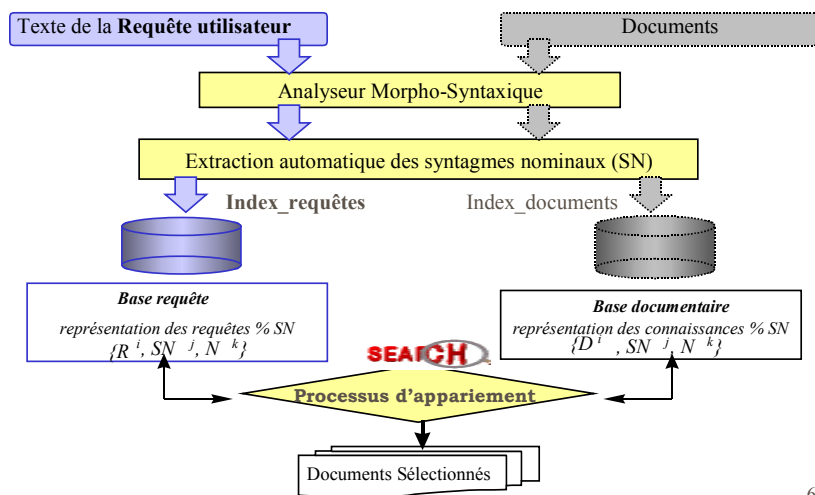
4

Architecture fonctionnelle : modèles EQA²te et SIMBAD



5

SIMBAD pour la recherche d'information



6

SIMBAD pour le réseau de connaissances

