



**HAL**  
open science

# Problématique de la veille informationnelle en contexte interculturel: étude de cas d'un processus d'identification d'experts vietnamiens

Philippe Lambert, Sahbi Sidhom

► **To cite this version:**

Philippe Lambert, Sahbi Sidhom. Problématique de la veille informationnelle en contexte interculturel: étude de cas d'un processus d'identification d'experts vietnamiens. Concepts et outils pour le management de la connaissance - ISKO-Maghreb 2011, May 2011, Hammamet, Tunisie. pp.N/c. hal-00677794

**HAL Id: hal-00677794**

**<https://hal.science/hal-00677794>**

Submitted on 12 Mar 2012

**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.

# Problématique de la veille informationnelle en contexte interculturel : étude de cas d'un processus d'identification d'experts vietnamiens

**Philippe LAMBERT** (IJL, Nancy Université – ASiES, INALCO), **Sahbi SIDHOM** (LORIA/KIWI & Nancy Université)

e-Mails : philippe.lambert@ijl.nancy-universite.fr, sahabi.sidhom@loria.fr,

## **Abstract** —

*The topics of this research cover all phases of expert identification process in a context of information monitoring.*

*Identify scientists for technological expertise is relatively easy through infometric analysis. However, lots of scientists residing in developing countries still remains a challenge. How to proceed then since the largest part of them does not appear in bibliographic databases ? This question raises two problematics : firstly, how a firm has the necessity to find such experts and what is the utility of this ? Secondly, cross-cultural factor must be taken into account and represents a key of success to integrate foreign experts from developing countries into internal expertise teams of the firm.*

*We take the case of Vietnamese experts as example, pointing out cultural characteristic features and difficulties we encountered relating to language and data treatment. We propose a computer treatment of a database, using Natural Language processing to identify experts and get a visual rendering of experts localization.*

*A corollary aspect of our approach is to redefine the role and the responsibilities of information watcher regarding a cross-cultural environment.*

## **Index Terms** —

Information Design (ID), Weak Signal (WS), Natural Language Processing (NLP), Data Visualization, Information Indexing, Knowledge Management (KM), Economic Intelligence, (multilingual) concept, cross-cultural, multicultural teams.

## **I. INTRODUCTION**

La veille stratégique est un des outils de l'intelligence économique par lequel l'information est récoltée, stockée puis traitée dans une analyse dynamique. Son objectif principal est d'asseoir un « *dispositif informationnel par lequel l'entreprise détecte et traite les signaux annonciateurs d'évènements susceptibles d'influer sa pérennité* ». Sa finalité « *est de permettre... de réduire son incertitude... en anticipant les ruptures pouvant se produire dans l'évolution de son environnement socio-économique et technologique* » (Lesca, 2007)

Depuis les années 1990 où le concept d'intelligence économique commença à percer en France, la veille stratégique

a fait l'objet de nombreuses publications. On citera comme référence les ouvrages et les positions des précurseurs du concept tels Humbert Lesca, Henri Dou, François Jacobiak qui édictent des recommandations de bonnes pratiques dans le cadre de la mise en place des activités de veille informationnelle quelque soit leur nature (veille stratégique, technologique, scientifique, etc.). Rappelons toutefois le distinguo qu'il convient d'opérer entre intelligence économique qui, *in fine*, apporte une modification de l'environnement de l'organisme qui l'utilise et les activités de veille qui gèrent exclusivement l'information.

On ne développera pas davantage le concept de veille pour nous concentrer sur une des problématiques de ce début de siècle : les relations qu'entretiennent les activités de veille par rapport à un environnement interculturel.

La dichotomie géopolitique Nord-Sud et les programmes de coopération qu'elle a impulsé apporte de nouvelles stratégies à adopter en matière d'information. Les pays dits « émergents » qui ont pu bénéficier d'aides au développement infrastructurel ont également développé des politiques nationales en matière d'information scientifique et technique. La nouvelle structuration des ressources d'IST de ces pays (archivage, numérisation, professionnalisation des cadres documentalistes) sont de nouvelles sources à exploiter dans la cadre du développement d'un projet dans ces pays.

Cet article posera dans une première partie le cadre conceptuel de l'identification d'experts : quelle utilité et pour quels types de résultats attendus. Une étude de cas sera présentée dans la deuxième partie : notre objectif ayant été d'identifier les thématiques de recherches émergentes au Vietnam et l'identification d'experts vietnamiens. Enfin, nous discuterons dans une troisième partie des résultats obtenus des résultats obtenus et de leurs perspectives.

## **II. L'IDENTIFICATION D'EXPERTS: UNE PORTE D'ENTREE VERS L'INTELLIGENCE CULTURELLE**

### *A. Pourquoi identifier des experts étrangers ?*

L'identification d'experts est une pratique courante dans les activités de veille technologique. Il s'agit d'effectuer un ciblage des spécialistes d'un domaine précis à partir de sources documentaires. Le plus souvent, c'est un traitement

statistique de bases de données bibliographiques qui permet l'identification.

Le recours à un spécialiste externe s'inscrit en général dans une démarche de conseil permettant de combler le vide de connaissances internes de l'organisme commanditaire.

On assiste de plus en plus à la création de groupes de travail pluridisciplinaires constituant une Expertise Scientifique Collective (ESC). Cette pratique est depuis longtemps utilisée dans le cadre d'études menées par des organismes publics. Le but d'une ESC est d'établir un état des lieux des connaissances disponibles sur une problématique définie (table 1).

<b>Élaboration et formulation de question(s)</b>	En partenariat avec le demandeur. L'objectif est de répondre à un questionnement en le traduisant en langage et application scientifique. Extraction de mots-clés « scientifiques » à partir de la question.
<b>Recherche bibliographique</b>	Détection d'articles scientifiques pertinents. Repérage des experts scientifiques et techniques du domaine concerné (sur la base de leurs publications scientifiques).
<b>Constitution du groupe d'experts</b>	Transfert de la bibliographie au groupe d'experts. (Auto) organisation du groupe d'experts (constitution de groupes thématiques - désignation d'animateurs). Travail des groupes thématiques : réunions et échanges par courrier électronique.
<b>Élaboration d'une synthèse par groupe</b>	Élaboration d'une synthèse générale. Production d'un document de communication. Diffusion et restitution de l'expertise, sous la forme d'un séminaire, d'une plaquette d'informations ou de conférence de presse, etc...

**Table 1 : Le déroulement d'une expertise scientifique collective (ESC)(Baligand, 2006)**

Les experts identifiés par la méthode précitée sont la plupart du temps des chercheurs dépendant d'une structure institutionnelle bien intégrée dans le réseau de la documentation de l'Information Scientifique et Technique et qui appartiennent pour la plupart à des pays développés ou à des laboratoires des pays du Nord.

On peut cependant supposer que les chercheurs de laboratoires situés dans les pays du Sud ne bénéficient pas forcément d'un tel positionnement. Dès lors, c'est toute une masse d'informations et de recherches détenue par ces acteurs scientifiques ignorés qui passe inaperçue.

## B. Appréhender le facteur culturel comme dimension stratégique

### 1) Culture nationale et culture organisationnelle

En prenant comme fondement la définition de la culture proposé par Kroeber et Kluckhohn (Kroeber & Kluckhohn, 1952)<sup>1</sup>, notre logique replace l'expert en tant qu'individu

<sup>1</sup> Kroeber et Kluckhohn : « La culture est un ensemble de modèles (les uns explicites, les autres implicites) qui décrit le comportement passé ou détermine le comportement à venir; que l'individu acquiert et transmet par le

culturel, permettant l'accès à un savoir tacite et implicite dans une dimension interculturelle.

Le savoir tacite est une notion créée par le philosophe hongrois M. Polanyi au milieu des années 60 (Polanyi & Sen, 2009). Considérant que tout l'ensemble du savoir ne peut s'exprimer seulement au travers de textes. Son adage « *we can know more than we can tell* » très largement repris sous-tend l'idée que le savoir tacite est difficilement exprimable en mots ou par d'autres canaux d'information du fait de sa subjectivité et de son caractère personnel. La notion fut ensuite reprise dans le domaine de l'innovation internationale dans une logique plus pratique pour les organisations (Nonaka & Takeuchi, 1995) avec le passage du tacite à l'implicite dans le cadre d'échanges de groupes intra entreprise.

L'autre élément de contexte à analyser concerne les relations entre culture et organisation. Cette thématique a été très largement traitée dans les sciences de gestion. On se référera notamment aux travaux de Geert Hofstede (Hofstede, 1984) qui ont donné naissance à une grille dimensionnelle d'analyse très répandue<sup>2</sup> ou à ceux de Davenport et Prusak (Davenport & Prusak, 1997)<sup>3</sup>.

Ce cadre théorique permet d'arriver à un niveau d'abstraction pouvant suggérer des stratégies nouvelles tant dans la récolte de l'information que dans sa validité. La compréhension de systèmes culturels cibles (i.e. le système culturel que le veilleur doit prendre en compte afin d'effectuer une veille informationnelle) est devenu une démarche incontournable de tout projet.

### 2) Dimension culturelle et traitement de l'information : le traducteur-veilleur

On l'a mentionné précédemment, le veilleur assure un processus itératif de valorisation de l'information à destination de commanditaires (décideurs, experts internes à un organisme, etc.). Dans le cadre d'un projet international, deux types de sources d'information sont exploitables : celles dites secondaires qui ont déjà fait l'objet de traduction de d'analyse et celles dite primaire en langue locale. Ces derniers gisements d'information dans des langues dites « peu dotées » restent problématiques à traiter. Soit le veilleur maîtrise une langue étrangère rare soit il peut utiliser des outils sémantiques de traduction dont la pertinence des résultats reste aléatoire.

*biais de symboles; qui constitue la marque distinctive d'un groupe humain, y compris les objets ouverts (ou artefacts) par lesquels ce groupe s'exprime. Le noyau essentiel de la culture est composé d'idées traditionnelles (c'est-à-dire transmises historiquement puis sélectionnées) et particulièrement des valeurs qui y sont attachées. On peut considérer les systèmes culturels tantôt comme les éléments conditionnant d'une action à venir »*

<sup>2</sup> Hofstede a élaboré une grille de lecture constituée de dimensions binaires permettant d'appréhender comment une culture nationale se répercute sur une organisation : 1/ individualisme vs. collectivisme, 2/ masculinité vs. Féminité, 3/orientation à long terme vs. Orientation à long terme. Pour Hofstede, la culture opère comme une sorte de programmation mentale. De l'empreinte de la structuration psychologique laissée sur les individus de l'organisation découlent les comportements. Contrairement à la culture institutionnalisée qui donne des valeurs, au niveau organisationnel, les pratiques et les usages dominent.

<sup>3</sup> Davenport et Prusak proposent une grille d'analyse à quatre dimensions pour comprendre une culture organisationnelle : 1/ouverte vs. fermée ; 2/ orientée sur les faits vs. basée sur les rumeurs et les intuitions ; 3/ centrée sur ce qui se passe en interne vs. en externe ; et 4/ contrôlant vs. donnant du pouvoir. Ces outils d'analyse ont l'avantage d'offrir une base de réflexion à l'analyse d'une organisation en contexte interculturel.

Les technologies sémantiques permettent actuellement à un veilleur de traiter de l'information dans une langue dont il ne maîtrise pas toutes les subtilités. Cependant, le coût temps-efficacité et le degré de technicité reste encore trop élevé pour être véritablement rentable.

Le traducteur-veilleur représente une adéquation à ces problèmes. La définition qu'en donne Mathieu Guidère est « *l'activité de traduction sélective employée comme aide à la prise de décision* » (Guidère, 2008). Le traducteur-veilleur doit pouvoir donner des traduction-analyses et revêt ainsi également un rôle de décideur (i.e. il décide par des sources documentaires à traiter, des informations à récolter ou pas). Il intervient en amont pour l'analyse et la traduction des requêtes en langue étrangère, en aval pour l'analyse et la traduction des retours et *in situ* pour l'analyse et la traduction de la documentation extraites d'après des critères de pertinence arrêtés (Guidère, 2008).

### III. ETUDE DE CAS : L'IDENTIFICATION D'EXPERTS VIETNAMIENS DANS LES THESES DE DOCTORAT AU VIETNAM (2008-2010)

#### A. Les particularismes vietnamiens

Situé au carrefour des civilisations chinoise et indienne puis, à celui des cultures française et américaine de par son histoire plus récente. Si le berceau de la nation vietnamienne se situe aux confins du sud de la Chine, la descente vers le Sud pour coloniser les territoires de l'actuel delta du Mékong a laissé un pays culturellement bipolaire.

Dans un pays étiré d'environ 1750 kilomètres, les projets se mènent différemment que l'on soit au Nord ou au Sud pour des raisons géopolitiques et culturelles.

Les cadres théoriques précités peuvent donner un éclairage de l'environnement culture/organisation. Si l'on prend l'analyse de Hofstede pour comparer les scores entre le Vietnam et la France :

	Vietnam <sup>4</sup>	France
<b>Distance hiérarchique</b> ( <i>Power Distance Index</i> ) <sup>5</sup> :	70	68
<b>Individualisme / Collectivisme</b> ( <i>Individualism / Collectivism</i> )	IND - 20	IND - 71
<b>Masculinité/ Féminité</b> ( <i>Masculinity / Femininity</i> ):	MAS - 40	MAS - 43
<b>L'orientation à long terme</b> ( <i>Long-Term Orientation</i> )	30	/
<b>Aversion pour le risque</b> ( <i>Uncertainty Avoidance Index</i> )	80	86

**Table 2 : comparaison des indicateurs Hofstede**

<sup>4</sup> Valeurs estimées à partir des scores par pays donnés par Hofstede : [http://www.geert-hofstede.com/hofstede\\_dimensions.php](http://www.geert-hofstede.com/hofstede_dimensions.php) (date de consultation : 28 mai 2011)

<sup>5</sup> Le PDI est l'acceptation de la distribution inégale du pouvoir ; le MAS/FEM est le degré d'activité pour chaque sexe ; le LTO est la promotion sélective de la promotion de valeurs hiérarchiques de type confucéennes ; l'UAI est le rapport au niveau d'ambiguïté volontairement supporté par les individus.

La comparaison des indicateurs de Hofstede pour les deux pays montre des profils relativement semblables : les deux pays arborent un PDI autour de 69 ce qui signifie une grande acceptation du degré d'inégalité de pouvoir dans la société. Ce score dénote un haut niveau d'ordre social hiérarchique. D'une façon générale, on ne sera pas surpris du parallélisme existant entre les deux pays : plusieurs facteurs organisationnels comme par exemple l'appareil législatif basé sur le code Napoléon issu de la longue présence française sur place peuvent expliquer cet état. Seul l'indicateur Individualisme / collectivisme subit un écart important. La nature différente des régimes politiques (république vs république socialiste) explique cet écart.

On retrouvera dans le profil organisationnel vietnamien, les grandes orientations du modèle chinois tel que le conçoit Redding (Redding, 2002) : « *un mode rationnel propre issu d'un système de valeurs terminales et substantives associé à un mode particulier de raisonnement interactionniste et pragmatique* »<sup>6</sup>

#### B. Traitement de la base de données des thèses vietnamiennes

##### 1) Déchargement des notices sélectionnées

L'objectif principal de ce travail est la constitution d'une base de données d'experts vietnamiens capables, selon leur champs d'expertise, d'apporter un avis consultatif à la faisabilité d'un projet. Corollaire à cette logique, cette base de données doit également nous permettre de déterminer quelles sont les thématiques les plus exploitées dans les thèses de doctorats soutenues au Vietnam et ainsi de donner une image 1/ des thématiques les plus exploitées dans le cadre des institutions de recherche vietnamiennes 2/ la localisation des centres les plus dynamiques du Vietnam en matière de recherche et 3/ la détection des thématiques ou de technologies émergentes.

Notre analyse sera effectuée à partir de la base de données des thèses de doctorat soutenues au Vietnam proposée par la Bibliothèque des Sciences générales de Ho Chi Minh-ville (*Thư viện Khoa học Tổng hợp Thành phố Hồ Chí Minh*) (<http://gslhcm.org.vn/libol/search/index.asp>).

<sup>6</sup> Cité in Hardy-Zhang & Agostineli, 2008 : « *les valeurs terminales [sont comprises comme] l'ordre et l'harmonie sociale, le contrôle, la protection contre l'insécurité, la hiérarchie, la réciprocité, le collectivisme familial et la connaissance, et des valeurs substantives comme le contrôle étatique, l'identité familiale, la confiance dans les relations personnelles (les « guanxi »), l'apprentissage continu et le goût du secret* ».



**Figure 1 : Portail de la bibliothèque des Sciences générales de HCMV**

Cette Bibliothèque offre divers services documentaires (recherché, demande de prêts, numérisation, etc.). L'accès à la consultation des thèses déposées se fait librement à partir de la page d'accueil du site web.

La consultation des notices des thèses de doctorat a été effectuée avec deux critères de recherche : la date de soutenance et la langue. Seules les notices des thèses déposées entre 2008 et 2010 et rédigée en vietnamien ont été prise en compte. Cet intervalle de temps relativement court permet d'avoir le cliché actuel de l'état de la recherche sur place. Seules les notices en vietnamien ont été prises en compte car nous nous sommes aperçu que les thèses de co-tutelle engageant un établissement universitaire vietnamien et un autre étranger et dont le titre était le plus souvent en anglais avait une soutenance dans le pays de coopération. De plus, ce critère linguistique pouvait laisser supposer de thématiques typiquement vietnamiennes.

Le butinage de la base de données a permis de récolter 1016 notices répondant aux critères précités. Une des difficultés a concerné le téléchargement des notices sélectionnées. Le système offre en effet plusieurs possibilités de format de téléchargement professionnel (XML, UniMarc 21, etc). Cependant, les résultats présentés sont difficilement exploitables en l'état et requiert un reformatage du texte.

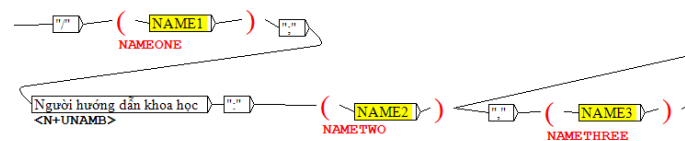
## 2) Processus d'extraction des informations

NooJ est un environnement linguistique développé par Max Silberstein (2005) de l'université de Franche-Comté (France). NooJ est fondé sur la technologie .NET et reconnaît un grand nombre de formats de documents. Outre cet avantage, l'utilisation de l'outil est facile avec une prise en main relativement rapide. Dans l'optique d'un repérage terminologique pour notre corpus bibliographique, NooJ offre la possibilité de créer des grammaires locales (ie. automates

finis) complètement paramétrables pour l'extraction d'informations. Les ressources de NooJ sont principalement constituées de dictionnaires (de la langue) et de graphes syntaxiques de type transducteurs à état fini, permettant le repérage d'expressions complexes, l'extraction de lemmes et l'annotation automatique de ressources textuelles.

Le module de vietnamien de NooJ possède une trentaine de grammaires ainsi que des dictionnaires généralistes et thématiques (Lambert & Fournié, 2008; Lambert, Fournié, & Ho Dinh, 2010).

Les données pertinentes du corpus téléchargé étaient structurées autour de plusieurs balises sémantiques ou signes de ponctuation facilitant leur extraction. NooJ fonctionne par étiquetage en cascade et offre la possibilité d'encapsuler des données dans des variables. Deux graphes spécifiques ont donc servi en premier lieu à reconnaître les termes spécifique à étiqueter (directeur de recherche, titre de la thèse ( *nhan đề* ), lieu de soutenance ( *xuất bản* ), et date ( *năm* ). Dans un deuxième temps, ces termes candidats ont été extraits puis restructurés dans une base de données. Puis un graphe a spécifiquement été créé pour extraire les données encapsulées dans les variables (\$Directeur) (FIG. 2).

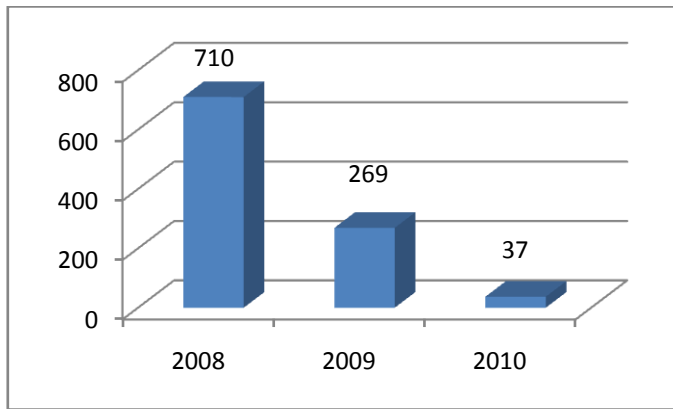


**Figure 2 : Exemple de graphe d'extraction de NooJ**

Les données du Titre ( *nhan đề* ) ont fait l'objet d'un traitement spécifique pour extraire les syntagmes nominaux vietnamien. On obtient alors en sortie de traitement un fichier structuré avec les éléments suivants : a. numéro de la référence bibliographique, b. directeurs de recherche, c. syntagme nominal du titre.

## 3) Analyse infométrique

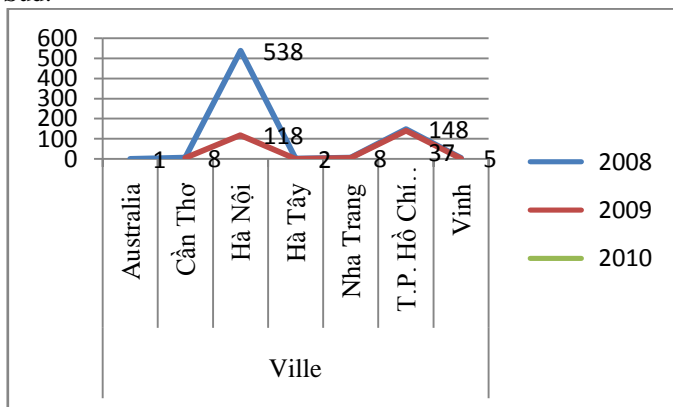
L'analyse de la base produit les informations suivantes : On assiste à un décroissement « normal » du nombre de thèses enregistrées dans la plateforme de la Bibliothèque des Sciences générales.



**Figure 3 : Répartition du nombre de thèses par année**

On peut supposer que le décroissement constaté pour 2009 et 2010 correspond au laps de temps nécessaire à la saisie des données entre le dépôt effectif et leur apparition sur l'application.

Le découpage par centre régional fait apparaître un découpage bipolaire correspondant aux facteurs de géographie humaine nationale : Hà Nội, la capitale au Nord du Vietnam se démarquant du reste du pays, suivie par Ho Chi Minh-ville, au Sud.



**Figure 4 : Répartition des notices par année**

On remarque aussi que si l'on prend 2008 comme année de référence, la capitale vietnamienne domine largement avec 538 thèses déposées contre 118 à Ho Chi Minh ville. La diminution drastique des nombres pour les années suivantes laissent supposer que toutes les thèses n'ont pas encore été enregistrées dans l'application de la bibliothèque.

Institut de Santé des Armées (Học viện Quân y)	118
Université de Médecine de Hanoi (Trường Đại học Y Hà Nội)	89
Université des sciences sociales et humaines de Hanoi (Trường Đại học Khoa học Xã hội và Nhân văn)	76
Université de Médecine et de Pharmacie de HCM-Ville (Đại học Y dược T.P. Hồ Chí Minh)	68
Université Polytechnique de Hanoi (Trường Đại học Bách khoa Hà Nội)	65
Institut Technique des Armées (Học viện Kỹ thuật Quân sự)	37

Institut Politique et d'Administration national Ho Chi Minh (Học viện Chính trị - Hành chính Quốc gia Hồ Chí Minh)	35
Université des Sciences naturelles (Trường Đại học Khoa học Tự nhiên)	33
Ecole Normale de Hanoi (Trường Đại học Sư phạm Hà Nội)	26
(Institut sanitaire et d'épidémiologie central (Viện Vệ sinh Dịch tễ Trung ương)	24
Université de Droit de Hanoi (Trường Đại học Luật Hà Nội)	21
Université d'Architecture de Hanoi (Trường Đại học Kiến trúc Hà Nội)	16
Univesrité de Can Tho (Trường Đại học Cần Thơ)	13

**Figure 5 : Classement des organismes par nombre de thèses déposées**

#### IV. APPLICATION DU DESIGN D'INFORMATION POUR FAVORISER LA DECOUVERTE DE SAVOIR

Le concept de Design D'Information (*Information Design*) est relativement récent en France. Privilégier une information spatialisée plutôt que linéaire, schématisée plutôt que purement alphabétique permet de mettre en exergue certains spécificité de l'information. Eppler [6] donne six principales raisons de privilégier ce canal qui 1/ motive le récepteur, 2/ présente de nouvelles perspectives, 3/ développe la mémoire, 4/ encourage le processus d'apprentissage, 5/ capte l'attention du récepteur et 6/ permet de structurer et de coordonner la communication.

Un grand nombre de définitions présentent le D.I. comme un art. L'art de mettre en scène l'information pour créer du sens. Il est vrai qu'à regarder les productions graphiques qui vont avec, une part non négligeable est laissée à la créativité avec des mises en forme de l'information colorées, animées et multiformes. Cependant, outre l'aspect purement esthétique de cette démarche, le D.I. recèle intrinsèquement une nouvelle façon de penser l'information et pourrait se résumer, comme le fait Dino Karabeg, en « une nouvelle approche de l'information » [7]. Karabeg explique ce qu'est le Design d'Information en proposant l'image d'un bus doté des phares à chandelles. Le bus représente la culture moderne (*modern culture*) tandis que les chandelles symbolisent la tradition informationnelle (*tradition informing*). On constate ce que l'auteur appelle une incongruence (ie.un dysfonctionnement) qui peut être dépassé par le design de l'information. Par ailleurs, la culture moderne étant productrice et consommatrice d'informations à grande échelle, le D.I. via le développement des technologies de l'information, interviendrait comme un remède au problème d'une infobésité<sup>7</sup> chronique.

Partant de ces théories, notre objectif principal est de pouvoir aboutir à une carte de type réseau sémantique permettant à l'utilisateur final de pouvoir avoir accès rapidement aux

<sup>7</sup> L'infobésité fait référence à une surcharge d'informations

directeurs de recherche des études doctorales, leur université d'appartenance et leurs thématiques de recherche.

Un objectif secondaire de cette technique est également de permettre une épuration de l'information non pertinente ou du moins d'importance mineure pour l'utilisateur final. Cette démarche qui est aujourd'hui plus connue sous le terme de curation s'avère particulièrement utile pour de très grands corpus de données.

La visualisation des données extraites dans la première étape du travail sera réalisée via Gephi, un logiciel sous Licence publique générale (GPL3) (<http://gephi.org>) permettant la visualisation de réseaux complexes.

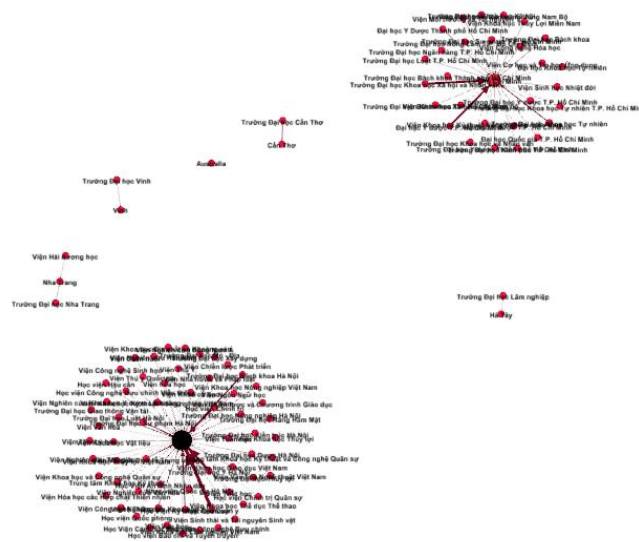
Outre l'aspect esthétique de la visualisation de l'information, cette technique a l'avantage de favoriser la découverte de savoir. Spatialiser les données revient en fait à demander à l'utilisateur final de changer son mode d'appréhension de l'information. On passe ainsi de la linéarité d'une lecture classique à une spatialité se rapprochant davantage des processus de création. On dépasse alors les contraintes

syntaxique et grammaticales données par le texte lui-même pour arriver au cœur de l'information susceptible de créer le savoir.

Outre cette création d'un nouveau savoir, la technique permet de raffiner la masse d'information jusqu'à obtenir des éléments primaires de compréhension. Les syntagmes nominaux sont un bon exemple de cette logique de « curation ».

*A. Processus de visualisation de l'Information et base de données des thèses vietnamiennes*

A partir du fichier sortie de NooJ, on crée un fichier .net structuré comme suit : une section « nodes » composée d'une série de paires clé chiffrée/ intitulé du nœuds et une deuxième section « vertices » composée d'une série de paire de clés pour identifier les relations entre les nœuds. Le fichier est alors ouvert dans Gephi.



**Figure 6 : Graphe de la localisation des thèses déposées**

Le graphe (FIG.6) fait apparaître la localisation des thèses déposées. On retrouve les deux centres les plus productifs Hanoi et Ho Chi Minh-ville qui étaient mentionnés lors de l'analyse infométrique. On remarque également que l'Australie apparaît en tant qu'électron libre. Une thèse comportant un titre vietnamien mais déposée en Australie expliquant cela.

Comme on l'a expliqué, ce type de traitement ne prend sa véritable utilité que dans une démarche interculturelle. Il nous est apparu que la dimension traductologique était incontournable à cette démarche. Nous avons donc ajouté à nos dictionnaires vietnamiens les équivalents en français. Ces ressources bilingues implémentées dans NooJ ont permis d'avoir en sortie de graphes les syntagmes nominaux traduits en français (FIG.7).

Cette fonctionnalité permet donc a un analyste non vietnamophone d'avoir accès directement à des ressources primaires dans une langue qu'il ne connaît pas forcément.

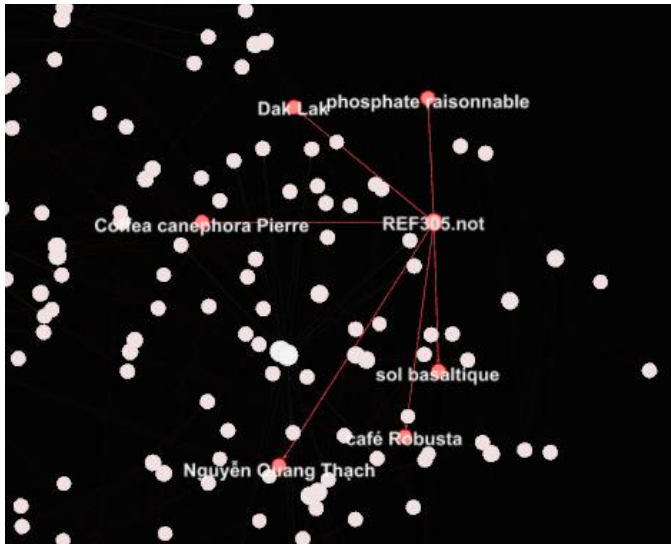


Figure 7 : Exemple de nœuds du graphe traduits en français

## V. DISCUSSION ET CONCLUSION

Cette expérience montre que les techniques que nous avons utilisées demandent une automatisation pour arriver à un état de performance, de robustesse et d'efficacité acceptable.

Pour un processus d'identification d'experts étrangers, la visualisation des informations issues de sources non structurées s'avère utile aux acteurs d'un projet stratégique sur plusieurs aspects :

Premièrement, la visualisation permet de faciliter l'exploration de larges bases bibliographiques. Elle permet en outre de détecter si des thématiques sont plus ou moins exploitées et de pouvoir détecter, le cas échéant, des signaux faibles. Cela est d'autant plus vrai que pour la plupart des pays du Sud, du fait des sauts technologiques rapides et du faible coût des matériaux, certaines technologies innovantes intéressent les pays européens (on pensera par exemple au monopôle de la Chine et à ses ressources naturelles de terres rares dans la production de panneaux photovoltaïques).

Deuxièmement, on peut accéder grâce aux techniques de visualisation de l'information à une information non seulement raffinée mais également issue de corpus en langues étrangères. Notre test concernait une démarche bilingue et la composition de dictionnaires multilingues laissent entrevoir des possibilités non négligeables pour une démarche interculturelle.

Troisièmement, sur un plan plus documentaire, cette logique permet de faire émerger les références potentiellement les plus pertinentes. Sur ce point, il convient d'affiner ces résultats par une analyse statistique, avec par exemple, l'utilisation de certains indicateurs bibliométrique comme le TF-IDF (*term frequency-inverse document frequency*).

Quatrièmement, par rapport à l'attente d'un analyste, notre approche met en évidence la nécessité de prendre en compte le format de dissémination des résultats d'une fouille de

données. La présentation des résultats sous forme d'une base de données structurée sera vraisemblablement plus utile pour mettre en exergue des relations uniques alors qu'une présentation neuronale permettra la fouille de relations multiples et connexes. On peut alors ajouter la notion de besoin fonctionnel de présentation au processus de dissémination des résultats de veille.

Enfin, Dans un monde où les échanges avec les pays du Sud tendent à s'amplifier, le traducteur – veilleur sera amené à avoir une place privilégiée dans la logique de l'intelligence économique. Il devra être capable non seulement de mettre ses connaissances linguistique et interculturelle au service d'une démarche d'intelligence économique mais il devra également pouvoir utiliser les outils de cartographie pour apporter au commanditaire de son travail une information traduite, concise et pertinente.

## VI. REFERENCES

- [1] H. I. Ansoff. (1975). "Managing strategic surprise by response to weak signals". *California Management Review*, vol. 18, no. 2, pp. 21–33, 1975.
- [2] H. I. Ansoff. (1985). "Strategic response in turbulent environments". *Handbook of business strategy*, 1985.
- [3] B. Coffman. (1997). *Weak Signal Research, Part I: Introduction*, *Journal of Transition Management*, MG Taylor Corporation (1997).
- [4] B. Coffman. (1997). *Weak Signal Research, Part II: Information Theory*, *Journal of Transition Management*, MG Taylor Corporation (1997).
- [5] H. Lesca. (2001). "Veille stratégique: passage de la notion de signal faible à la notion de signe d'alerte précoce," in *Colloque VSST*, pp. 98–105, 2001.
- [6] M. J. Eppler and R. A. Burkhard. (2007). "Visual representations in knowledge management: framework and cases". *Journal of Knowledge Management*, vol. 11, no. 4, pp. 112–122, 2007.
- [7] D. Karabeg. (2002). "Designing information design". *Information Design Journal*, vol. 11, no. 1, pp. 82–90, 2002.
- [8] Benjamin Jotham Fry. (2004). *Computational information design. bfa Communication Design*, in *Computer Science Carnegie Mellon University. Massachusetts Institute of Technology*, April 2004..
- [9] P. Lambert, S. Sidhom. (2010). "Knowledge extraction and visualization: case study on ChroniSanté project in France". in *proceedings SIIE'2010, Sousse (Feb. Tunisia)*. IHE Edition.
- [10] S. Sidhom. (2008). "Approche conceptuelle par un processus d'annotation pour la représentation et la valorisation de contenus informationnels en intelligence économique (IE)," 2008. in *Proceedings SIIE'08 - Systèmes d'Information et Intelligence Economique*, IHE Edition vol.1, Hammamet (fév. 2009).
- [11] C. H. Goh, S. Bressan, S. Madnick, and M. Siegel. (1999). "Context interchange: new features and formalisms for the intelligent integration of information," *ACM Trans.*



- Inf. Syst., vol. 17, no. 3, pp. 270-293, 1999.
- [12] L. S. Elting, C. G. Martin, S. B. Cantor, and E. B. Rubenstein. (1999). "Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures," *BMJ*, vol. 318, no. 7197, pp. 1527 -1531, Juin. 1999.
- [13] J. Wyatt. (1999). "Same information, different decisions: format counts," *BMJ*, vol. 318, no. 7197, pp. 1501 -1502, Juin. 1999.
- [14] M. Silberstein. (2005). "NooJ: a linguistic annotation system for corpus processing," in *Proceedings of HLT/EMNLP on Interactive Demonstrations*, p. 11, 2005.
- [15] S. Sidhom. (2002). "Plate-forme d'analyse morpho-syntaxique pour l'indexation automatique et la recherche d'information: de l'écrit vers la gestion des connaissances". Thèse de doctorat, Université Claude bernard Lyon1. 2002.
- [16] M. Le Guern. (1989). "Sur les relations entre terminologie et lexique," *Meta*, vol. 34, no. 3, 1989.
- [17] H. Lesca, S. Kriaa-Medhaffer, and A. Casagrande. (2009). "Veille stratégique: Un Facteur d'échec paradoxal largement avéré: la surinformation causée par l'Internet. Cas concrets, retours d'expérience et piste de solutions". in *Proceedings SIIE'09 - Systèmes d'Information et Intelligence Economique*, IHE Edition. Hammamet (12-14 fév. 2009).
- [18] G. Salton, A. Wong, and C. S. Yang. (1975). "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613-620, 1975.
- [19] S. Card, J. Mackinlay, and B. Shneiderman. (1999). *Readings in Information Visualization: Using Vision to Think*. 1999. Morgan Kaufmann Ed.
- [20] A. Harbaoui, M. Ghenima, and S. Sidhom. (2009). "Enrichissement des contenus par la réindexation des usagers : un état de l'art sur la problématique". in *Proceedings SIIE'09 - Systèmes d'Information et Intelligence Economique*, IHE Edition. Hammamet (12-14 fév. 2009).
- [21] P. Kislin. (2007). "Modélisation du problème informationnel du veilleur dans la démarche d'Intelligence Économique". Thèse de doctorat : université de Nancy2. 05 nov. 2007.