

Valoriser le patrimoine documentaire des entreprises par le prisme des métiers

Article inédit. Mis en ligne le 05 avril 2011.

Caroline Djambian

Caroline Djambian est enseignant chercheur en Sciences de l'information et de la communication. Après plusieurs expériences dans le domaine de l'Intelligence Economique, elle effectue une thèse sous la direction de Sylvie Lainé-Cruzet (Laboratoire ELICO, Lyon 3) au sein de la Division Ingénierie Nucléaire d'EDF, intitulée «valorisation d'un patrimoine documentaire industriel et évolution vers un système de gestion des connaissances orienté métier ». Elle enseigne la Veille technologique durant deux ans à Lyon 3 et aujourd'hui en tant qu'ATER, à la Faculté des Sciences et Techniques de l'Université Aix-Marseille 3.

Plan

Introduction

Se fonder sur l'analyse amont

Harmoniser le système, représenter le savoir

Une méthode orientée métier : des textes aux experts

Conclusion

Références bibliographiques

Résumé

Le patrimoine documentaire des entreprises s'est souvent accumulé sans que ces dernières puissent s'adapter au rythme des évolutions des Tics. La mémoire collective qui ne cesse d'être produite voit sa masse croître, est devenue épars et hétérogène, et nombre d'entreprises aujourd'hui confrontées à des problématiques transverses ont du mal à mobiliser leurs connaissances de façon opérationnelle.

Nous présentons ici le cas de la Division Ingénierie Nucléaire (DIN) d'EDF face à la nécessité de valoriser son patrimoine informationnel. Nous exposons pourquoi un travail amont de contextualisation est essentiel dans des cas comme celui-ci, afin de situer le mode de fonctionnement du système d'information dans une problématique structurelle, les aspects techniques devant être rapidement dépassés pour prendre en compte l'organisation dans sa globalité.

Dans ce contexte où problématiques micro et macro se confondent, les métiers cœurs de l'entreprise s'imposent comme la base de toute réflexion. La documentation qu'ils produisent et utilisent, véhicule les connaissances techniques de l'entreprise, qui y sont exprimées par des concepts propres aux métiers. Leur terminologie est la clé permettant de valoriser les connaissances et de mieux gérer le patrimoine documentaire par lequel elles transitent. A travers l'exemple de la DIN, nous présentons notre approche résolument empirique et qualitative, pour faire évoluer le système existant vers une base de connaissances centrée sur le « sens métier » de l'organisation.

Mots-clés : patrimoine documentaire, document technique, gestion des connaissances, base de connaissances, ontologie, terminologie.

Abstract

“Upgrading an industrial documentary heritage by the prism of jobs”

Documentary heritage of the enterprises has often been accumulated without that they could adapt to the pace of ICT developments. Collective memory never stops being produced sees its mass growing, becomes scattered and heterogeneous and many companies today face transverse problems struggle to mobilize their knowledge operationally.

We present here the specific case of the Nuclear Engineering Division (DIN) of EDF France and the need to upgrade its information heritage. We explain why a job of contextualization ahead is essential in cases like this one, to locate the operating mode of the information system in a largest structural problem, technical aspects having to be quickly overwhelmed in order to consider the organization as a whole.

In this context, where micro and macro issues mingle, the core business of the company emerges as the basis for any reflection. Documentation they produce and use, vehicles the technical knowledge of the company, which is expressed by specific business core's concepts. Their terminology is the key to knowledge enhancement and a better management of the documentary heritage through which they pass. Through the example of the DIN, we present our approach resolutely empirical and qualitative, to evolve the system to a knowledge base centered on the "business core sense" of the company.

Keywords: documentary heritage, technical document, knowledge management, knowledge base, ontology, terminology.

Resumen

« Actualización del patrimonio documental de negocios por el prisma de oficios técnicos »

Patrimonio Documental de negocios a menudo se ha acumulado, sin que puedan adaptarse al ritmo de la evolución de las TIC. La memoria colectiva que está siendo producida constantemente ve su masa cada vez mayor, se ha vuelto escasa y heterogénea y muchas empresas que hoy se enfrentan a problemas transversal tienen dificultades para movilizar sus conocimientos en operativos.

Presentamos aquí el caso específico de la División de Ingeniería Nuclear (DIN) de la EDF Francia y la necesidad de valorar su patrimonio de información. Explicamos por qué un trabajo preliminar de contextualización es fundamental en casos como este, para buscar el modo de funcionamiento de la sistema de información en una problemática estructural más amplia, aspectos técnicos deben ser rápidamente superados considerando la organización como un todo.

En este contexto, donde los temas micro y macro se funden, las profesiones técnicas de las empresas surgen como base para cualquier reflexión. La documentación que producen y utilizan, transmite los conocimientos técnicos de la empresa, que se expresan por los conceptos de estos oficios. Su terminología es la clave para valorar los conocimientos y gestionar mejor el patrimonio documental a través del cual pasan. A través del ejemplo de la DIN, presentamos nuestro enfoque decididamente empírico y cualitativo, para evolucionar el sistema actual a una base de conocimientos respecto al "sentido negocial" de la organización.

Palabra clave : patrimonio documental, documento técnico, gestión del conocimiento, base de conocimientos, ontología, terminología

INTRODUCTION

Force est de constater un problème devenu récurrent dans les entreprises : leur activité est assise sur une production importante de documentation technique et ce patrimoine documentaire, formé au cours de leur histoire, s'est souvent accumulé sans que ces dernières aient eu le temps de s'adapter au rythme des évolutions des technologies de l'information. Cette mémoire collective, clef de voûte de l'activité des entreprises, ne cesse d'être produite, voit sa masse croître inexorablement, et est éparse et hétérogène (Gandon & Dieng-Kuntz, 2005). Or, des problématiques transverses demandent aujourd'hui de mobiliser ces connaissances de façon très concrète et opérationnelle.

C'est le cas de la DIN (division ingénierie nucléaire) d'EDF. Ses besoins couvrent une réflexion large allant de la valorisation du patrimoine documentaire à la capitalisation des connaissances de métier (problématiques qui ne peuvent en réalité être dissociées). L'ensemble de son système de gestion de l'information est concerné. Alors que tout le processus d'ingénierie nucléaire passe par la production documentaire, les flux informationnels sont centralisés dans une GED (gestion électronique des documents) devenue obsolète, surchargée et inexploitable. Contenant à ce jour plusieurs millions de documents, elle est incapable de répondre efficacement à ses tâches premières de gestion ou recherche d'information. Les utilisateurs contournent alors le système en développant de façon anarchique une multitude d'outils annexes qui contribuent toujours plus à faire échapper l'information à tout contrôle.

Le fait est que les entreprises ont tendance à sous-estimer la nature du problème en considérant que seule la dimension technique est en cause, alors que l'origine en est plus fondamentalement organisationnelle. La recherche systématique de solutions auprès d'outils si performants soient-ils, est rarement suffisante pour dénouer des situations parfois devenues très complexes. Dans ce cas, il faut entrer plus amont au sein de l'espace communicationnel où l'information qui y circule et les usagers (qui en sont paradoxalement les émetteurs et les récepteurs) se sont trouvés décorrélés à un moment donné. Nous montrons à travers l'étude de la DIN d'EDF, l'intérêt d'une analyse amont approfondie. Les résultats qui en découlent, ont porté à considérer l'idée de métier technique de l'entreprise, notamment les notions communes qui animent une communauté de métier, comme étant une possible articulation entre les usagers et les flux informationnels, axant la réflexion sur « comment faire ressortir et utiliser le sens métier ? ». Nous illustrons ces questionnements en exposant nos choix méthodologiques pour construire une base de connaissances d'un domaine spécifique de l'ingénierie nucléaire.

SE FONDER SUR L'ANALYSE AMONT

Aller au-delà du système d'information

Qu'il s'agisse de gestion de la documentation ou des connaissances, résoudre à long terme les problématiques complexes qui se sont construites avec le temps dans les entreprises, implique en premier lieu la compréhension de l'espace commun que l'on veut réorganiser. Plus que comprendre, appréhender cet espace, signifie en partant du système d'information, identifier et qualifier son environnement au sens large, définir l'ensemble de ses acteurs et évaluer les besoins et évolutions pressentis à court et long termes. Face à un champ aussi vaste qui est le nôtre, il convient d'adopter une approche systémique et pragmatique visant la décomposition progressive du système en composants élémentaires pour en faire émerger des éléments localisables et traitables (Carlier, 1994) : « ... *diviser chacune des difficultés que j'examinerais en autant de parcelles qu'il se pourrait et qu'il serait requis pour mieux les résoudre...* » (Descartes, 1987). Ainsi, l'identification des différents sous-

systèmes permet de localiser les structures de pilotage, celles d'information ou de nature opératoire.

L'appréhension du cadre empirique auquel nous sommes confrontés (la DIN d'EDF) a ainsi débuté par une phase exploratoire. Elle s'avère indispensable quel que soit le cas traité pour faire naître cette série d'hypothèses de départ qui subiront ensuite des ajustements progressifs. Cette phase a pour nous consisté dans l'observation prolongée en situation naturelle, complétée par des entretiens (directifs ou non) d'un panel sélectionné d'acteurs du système d'information et d'utilisateurs. Le but est d'obtenir une compréhension globale des flux, des possibilités d'évolutions, des contraintes et des points forts et faibles du système. Ces observations et entretiens ont produit une matière très riche qu'il est alors important de structurer. Nous en avons fait reposer l'analyse sur trois niveaux constants d'interrogation : le niveau supra-documentaire (la fonction documentaire dans son contexte global), inter-documentaire (systèmes et acteurs), intra-documentaire (le document). En gardant ces trois strates comme lignes directrices on s'aperçoit rapidement qu'elles évoluent d'elles-mêmes vers des questions de niveau structurel particulièrement intéressantes à considérer.

Cette démarche initiale a permis, en allant au-delà de l'analyse de surface, de commencer à aborder les enjeux et difficultés tels que perçus par le management et parallèlement par les utilisateurs. Elle a en fait dépassé nos attentes en nous menant naturellement à considérer les dysfonctionnements du système comme révélateurs de problèmes d'ordre supérieur (Durampart, 2006). Qu'ils soient culturels, organisationnels, démographiques..., ils forment un jeu complexe qui influence le fonctionnement général des systèmes concrets mis en place et auxquels sont confrontés au quotidien les utilisateurs. Les observations donnent donc dans un premier temps une conscience générale du système et permettent d'enclencher des réflexions encore floues. La confrontation aux entretiens, tant auprès des professionnels du système d'information que des utilisateurs et du management, offre ensuite des opinions diverses et complémentaires. La synthèse de cette matière brute permet de situer les vraies racines des dysfonctionnements. En partant donc de problématiques techniques, on est introduit dans des sphères fonctionnelles, ce qui laisse présager que les réponses sont également à rechercher sur ce plan.

Identifier les modes de fonctionnement

Les résultats de notre analyse du Système d'Information de la DIN permettent d'illustrer cette relation des champs *micro* et *macro* dans l'entreprise. Le mode de fonctionnement informationnel que nous rencontrons ici est en fait clairement identifié (Pintea, 1995). Dit séquentiel (linéaire ou diachronique), il est caractéristique de l'influence directe d'un « management qualité » obligatoire et omniprésent, centré sur la maîtrise des états du document, c'est-à-dire l'identification des acteurs, la traçabilité et la possibilité de restaurer l'état documentaire *ante* (l'ensemble du système documentaire à la DIN d'EDF est assis sur une procédure qualité complexe très structurée). L'avantage en est, outre la traçabilité des actions ou l'identification des responsabilités, de donner la possibilité de répartir une série d'opérations élémentaires entre plusieurs opérateurs peu qualifiés et facilement interchangeables. L'emploi de personnels documentaires diplômés peut ainsi être notablement réduit. Dans notre cadre, le processus documentaire a effectivement été entièrement atomisé, suite à des remaniements de personnels, entre une large palette d'acteurs notamment les utilisateurs eux-mêmes, qui produisent et indexent les documents. La décentralisation de la fonction documentaire vise évidemment à faire l'économie d'agents qualifiés. Mais ce mode de fonctionnement correspond également à des environnements industriels comprenant une ligne d'autorité hiérarchique clairement descendante et une communication se limitant essentiellement au monologue du management : les projets de

systèmes d'informations sont le fait de prises de décision unilatérales hors d'une considération réelle des besoins des usagers. On comprend donc que ce système soit très lourd, puisque privilégiant le cadre des processus et de la sécurité qu'ils apportent, au détriment de la rapidité d'action.

Pour répondre au manque de réactivité des systèmes « officiels » que ce mode met en place, des outils gravitaires moins formels et plus individuels, tels que bases Notes, Access, intranets..., se multiplient à l'initiative des usagers eux-mêmes et de façon relativement désordonnée. Ils échappent à cette logique et vont même à son encontre en se fixant sur un fonctionnement dit simultané (ou synchronique). L'orientation proprement gestion de projet de ces outils les recentre sur des priorités de réactivité, de coopération, de traitement de l'information, bien plus que sur la production des documents, lesquels sont vus comme des produits dérivés des actions documentaires et non plus comme une finalité. Le mode simultané est généralement celui qui donne le plus une impression de confusion documentaire et la multiplication des bases de données et intranets ainsi construits au sein de multinationales comme EDF peut devenir impressionnante. A la DIN, l'outil officiel est une GED contenant environ 10 millions de documents. Malgré la mise en place d'un outil de recherche floue visant à faciliter l'accès à l'information, les usagers sont contraints de détourner le système en créant leurs propres outils annexes. On dénombrait ainsi environ 20 000 bases de données pour EDF et GDF avant leur scission, dont 400 uniquement pour la DIN et plus de 162 intranets EDF. Les formats en sont variés : bases Notes, Oracle, Access, Excel, FileMaker... Tout ceci est en réalité une évolution normale du système officiel linéaire qui naît sous la pression d'une exigence de travail rapide et coordonné de la part des usagers. Dans ce cas on a alors plus souvent affaire à un matériau volatil et mouvant (documents de gestion de projets, emails, annotations...). Ce mode synchronique ne peut s'imposer que dans un contexte d'expertise généralisée, avec un niveau de compétences des différents acteurs assez homogène. Dans notre cas 70% des ingénieurs sont cadres, travaillant essentiellement par réseau informel et mails dont les contenus échappent à la GED. On constate donc la coexistence de deux systèmes l'un officiel et l'autre officieux, évoluant conjointement, dans des directions et à des allures diverses. Cette distorsion qui s'est créée à la DIN, s'est aggravée progressivement, allant jusqu'à engendrer de réels dysfonctionnements dans la gestion informationnelle de l'entreprise.

En réalité, ces systèmes, séquentiels ou simultanés, se doivent pour être harmonieux d'être en relation avec les enjeux et objectifs, mais aussi les moyens et besoins documentaires de leurs membres. Une dissonance avec l'un de ces éléments (enjeux, objectifs, moyens, besoins) engendre automatiquement à terme des carences pouvant affecter le système d'information dans son ensemble et entraîner jusqu'à son rejet par les usagers, comme cela se passe dans l'organisation étudiée. Les besoins réels sont ignorés par le management, les moyens et résultats qu'il met en place par conséquent sont mal adaptés, les objectifs sont peu clairs, car souvent définis sans connaissance approfondie des besoins originaux et ne faisant pas l'objet d'une communication dans l'organisation. Les enjeux visés par l'entreprise, tels que la traçabilité, la possibilité de retour à l'état *ante*, l'identification des acteurs, sont quant à eux développés à l'excès au détriment des autres fonctions du système, afin de couvrir les manquements opérationnels. L'incohérence qui en découle expose l'organisme à des surcoûts de traitement des informations, à l'augmentation du temps nécessaire à leur mobilisation, à leur éparpillement et à la démotivation des usagers. Les informations deviennent moins fiables, se perdent, et sont transformées au cours de leur circulation, etc. Tout le système est à clarifier en partant d'une logique inverse, c'est-à-dire de la base, des usagers.

HARMONISER LE SYSTEME, REPRESENTER LE SAVOIR

Prendre en compte les exigences de l'environnement

Remettre en cohérence les besoins avec les moyens existants, voilà notre visée première. Une meilleure orientation des outils vers les préoccupations des métiers de l'entreprise est incontournable : mal adaptés aux pratiques et aux besoins, ils rendent l'information peu exploitable. Cette information est la référence des métiers cœurs, le patrimoine fondamental de l'entreprise, qui véhicule tout le savoir de conception et exploitation des centrales depuis leur construction. Elle est devenue avec le temps épars (dans des bases de données métier, la GED officielle, la bibliothèque, internet, les boîtes mails ou armoires personnelles...), hétérogène (formats papier ou numérique, image ou texte...), et difficilement accessible dans l'ensemble des documents de gestion courante de l'entreprise, car stockée sans distinction.

Or, les contextes socio-économiques imposent aujourd'hui à nombre d'entreprises comme à la DIN d'EDF, des contraintes transverses qui nécessitent de faire appel aux ressources informationnelles de l'organisation de façon très concrète. Les enjeux sont multiples et cruciaux. Le premier en est le renouvellement des compétences. Le départ massif des experts ayant connu la construction des centrales dans les années 70 ne coïncide pas avec l'intégration des nouveaux arrivants, entraînant une déperdition notable de connaissances techniques entre les deux générations d'ingénieurs. Cette perte est d'autant plus importante dans une entreprise où la tradition de compagnonnage est historique : l'acquisition de la compétence s'est jusqu'ici toujours faite par une formation *in situ* au contact des pairs pendant 1 à 2 ans, le savoir de métier ne s'acquérant qu'au sein de l'entreprise. La question du renouvellement des compétences rejoint directement la problématique documentaire, puisqu'elle est centrée sur la pérennisation des connaissances des métiers cœurs, tout dans l'entreprise passant par la production de documents techniques. Se pose ensuite la lourde exigence de la durée de vie des installations nucléaires : les centrales arrivant au terme des 30 ans de vie prévus à leur construction, la documentation joue un rôle primordial pour soutenir le remplacement progressif des matériels et accompagner la conception de réacteurs de nouvelle génération (EPR : European Pressurised Reactor), ceci sans oublier la question omniprésente de transparence vis-à-vis de l'Autorité de Sûreté Nucléaire. Ainsi, la documentation de référence des métiers cœurs de l'entreprise est le point d'appui de tout le travail d'ingénierie nucléaire et doit impérativement être recensée, pérennisée et valorisée.

La contextualisation du système d'information de la DIN nous a permis de dégager des conclusions au plan global. Nous avons vu qu'émerge la nécessité d'homogénéiser les modes informationnels, de simplifier le système en se dirigeant vers la recherche d'information plus que vers la recherche de documents, et de se rapprocher des pratiques. Ainsi, les objectifs de nos travaux sont de rendre la documentation de référence plus exploitable et cohérente et de toucher tant les aspects de gestion que de recherche d'information, qui font tous deux défaut. Nous souhaitons rendre transmissibles les connaissances techniques que la documentation contient. Elle est la seule trace permanente laissée dans l'entreprise par le passage des ingénieurs (départs en retraite, mutations...) : les artefacts humains créateurs des connaissances passent, les artefacts documentaires porteurs des connaissances restent.

Valoriser les connaissances par le sens collectif

Nous avons choisi comme terrain d'application une compétence spécifique de la DIN. Considérée comme sensible, elle est vectrice d'une réelle culture collective de métier dans

l'organisation et est particulièrement concernée par les problématiques transverses que nous venons d'exposer. Pour schématiser notre réflexion, nous devons considérer qu'un nouvel ingénieur intègre le domaine métier avec un bagage théorique qu'il parfait par la réelle acquisition de la compétence en interne, par l'expérience empirique. Dans notre cas, l'acquisition des connaissances ne se réalise plus au contact des pairs mais via la connaissance portée par les textes, « *sorte de mémoires exosomatiques* » (Polanco, 1999 ; Popper, 1979). La connaissance est dite « *représentée sans sujet connaissant* » (Popper, 1979) car enfouie dans la documentation technique et traduite en langage écrit. Comme B. Lamizet (Lamizet, 1992), nous appréhendons l'entreprise comme un espace communicationnel où circulent les connaissances collectives qui y ont été construites. Elles sont médiatées à travers les documents qui en deviennent la représentation sociale. Le langage propre au métier vient quant à lui animer, actualiser et objectiver ces connaissances collectives. Il est une clé pour valoriser les connaissances techniques. En maîtriser les codes communs garantit une bonne réappropriation des savoirs du domaine.

Afin d'approfondir ces réflexions, il convient de comprendre comment métier et documents s'articulent. Comment se forment et circulent les connaissances dans le domaine particulier que nous avons choisi ? Pour répondre à ces interrogations, nous partons là aussi d'entretiens menés auprès des usagers. Ces entretiens ont confirmé les besoins forts qui s'étaient précédemment dessinés : accès à la documentation par thématiques ou par navigation dans les domaines métier et dans les textes, facilitation à l'entrée dans la compétence pour les nouveaux arrivants, prise en compte des différents profils de recherche. Il est par exemple, directement demandé aux nouveaux ingénieurs de travailler sur un thème. Ce type de recherche étant impossible dans la GED qui n'offre qu'un accès par référence exacte (connue des anciens), ils s'orientent vers des sources externes ou informelles de type Google, récupération de mémoires d'anciens stagiaires, réseau externe...

Nous avons souhaité par nos travaux proposer des orientations opérationnelles en structurant les connaissances contenues dans la documentation technique à l'aide de la terminologie des métiers cœurs de l'entreprise. Pour ce faire, nous avons construit une base de connaissances, à savoir : une ontologie et une terminologie du domaine, ainsi qu'une documentation que nous avons rendue cohérente et ciblée au format texte (documents identifiés, référencés puis océrisés suite aux entretiens), organisée selon les concepts du domaine. Nous souhaitons cette documentation accessible facilement par les termes métier, en langage naturel ou par navigation (plus aisée pour les nouveaux arrivants). Notre souci premier a été d'inverser la logique unilatérale qui soutenait jusqu'alors les projets dans l'organisation en partant réellement des pratiques des ingénieurs. Ainsi, nous considérons que le système le plus exploitable pour un usager est celui qui est construit selon son raisonnement, ses usages et son langage.

Nos postulats de départ sont que les connaissances d'ingénierie nucléaire transitent par les médias que sont les documents techniques et y sont exprimées par des concepts propres aux métiers. Ensuite, c'est par la mise en correspondance de la connaissance produite (émanant des compétences individuelles et collectives) avec la réalité de l'organisation et de ses métiers (que nous avons étudiée précédemment), qu'il est possible de valoriser son patrimoine informationnel. Cette mise en correspondance vise à aboutir à une représentation explicite du domaine, au sein d'une base de connaissances centrée sur le *sens métier*.



Les bases de connaissances sont au cœur de nombreuses applications telles que gestion des savoir-faire et des compétences, veille, recherche documentaire... Leur construction est une opération résolument humaine, intellectuelle et complexe, pour laquelle il n'existe pas de méthode générale unanimement acceptée (Lainé-Cruzet, 2006a). Les approches sont donc souvent empiriques, ce qui fut notre cas. Pour que le domaine soit correctement décrit, l'intervention préalable aux traitements informatisés des corpus de textes a été lourde. L'intervention humaine fut également fort présente en aval pour l'interprétation, la validation et l'affinement des résultats rendus par les outils. Ainsi, face à un système complexe, il est apparu judicieux d'adopter une approche qualitative alliant de façon complémentaire travail avec les experts et extraction à base des textes.

Les travaux du domaine basés sur la linguistique de corpus (animée en France par le groupe TIA (terminologie et intelligence artificielle)) consistent en des approches sémasiologiques, partant des mots (c'est-à-dire des textes) pour aller vers le sens. Les recherches de N. Aussenac-Gilles et A. Condamines proposent une méthode de construction de modèles de domaine se fondant sur une analyse de corpus utilisant des techniques de TAL (traitement automatique des langues) pour aboutir à une Base de Connaissances Terminologique (BCT). Des recherches internes au Groupe EDF utilisent ces techniques pour la mise en cohérence des systèmes documentaires (Boccon Gibod, 2006). Considérant le panorama des travaux actuels et les exigences de notre contexte, où les notions communes qui animent les connaissances de métier occupent une place centrale, nous avons choisi de tirer parti de ces techniques de linguistique de corpus, enrichies par des approches onomasiologiques, centrées sur le concept. Parmi elles les « *ontologies régionales* » de B. Bachimont (Bachimont, 2000) ou les travaux de capitalisation des connaissances menés à EDF par C. Roche (Roche, 2007) apportent un regard distancié en mixant les techniques centrées sur les textes et celles axées sur les experts. Notre méthode utilise également les traitements automatiques en alternance avec l'intervention humaine, l'expert de métier étant au centre des travaux, de façon à équilibrer constamment les deux aspects et réduire les distorsions qui peuvent apparaître entre les résultats des deux approches.

UNE METHODE ORIENTEE METIER : DES TEXTES AUX EXPERTS

La base de connaissances que nous souhaitons construire avait donc pour but de permettre une mise en valeur de la documentation par le langage, c'est-à-dire la terminologie et les concepts propres aux métiers, pour une meilleure appropriation de l'information et des connaissances (Lainé-Cruzet, 2006b). Le travail a été volontairement basé sur un échange constant avec les experts, selon plusieurs étapes proposées par C. Roche (Roche, 2005). Nous nous servons des techniques de linguistique de corpus pour initier la démarche. Ainsi, la première phase consiste en la définition d'un corpus de textes avec l'expert du domaine. Cette phase essentielle détermine tout le travail et ses résultats,

c'est pourquoi nous y avons attaché une attention particulière et avons choisi un corpus très ciblé, composé de huit textes techniques (570 pages) couvrant 90% des concepts du domaine. Nous avons voulu ainsi privilégier la qualité à la quantité des textes considérés comme représentatifs. L'étape suivante fut l'extraction automatique de candidats termes issus de ce corpus sous la forme de deux lexiques différents, l'un de syntagmes nominaux et l'autre d'acronymes, très répandus dans le langage de l'entreprise, aboutissant après un retraitement manuel à 687 termes validés par les experts du domaine.

Nous arrêtons ici de suivre les approches sémasiologiques pour prendre de la distance par rapport aux lexiques. Cet éloignement est primordial dans notre méthode, car il est impératif, pour préparer la construction du réseau conceptuel, de sortir de la langue et d'amorcer l'entrée dans la sphère extra-linguistique qu'elle relate. Pour ce faire, l'étude approfondie des corpus techniques et la consultation régulière des ingénieurs permet d'identifier les notions du domaine. Se détacher des mots pour entrer directement dans le notionnel fait en quelque sorte économiser une étape de travail. Une modélisation partant des lexiques sera sémantique. Or, « *il n'y a pas de concepts dans un texte, mais des traces linguistiques de leurs usages* » (Roche, 2005). Ce travail en amont entièrement humain est donc indispensable pour aller directement vers la construction d'un réseau conceptuel et non sémantique. Rappelons que la terminologie ne s'intéresse pas aux objets linguistiques pour eux-mêmes, mais pour leur relation à un concept. Le fait est que de nombreux travaux, notamment de linguistique de corpus, se recentrent sur les mots et voient donc en cette étape une finalité. Elle n'est pour nous qu'une transition pour aller vers les représentations notionnelles.

Suite à la validation par les experts, le modèle conceptuel doit alors être retravaillé pour préparer le passage à l'ontologie : les relations complexes établies dans le modèle conceptuel doivent être simplifiées en relation de subsomption, ce qui implique une réorganisation totale des concepts. L'ontologie du domaine obtenue finalement est composée de plusieurs vues complémentaires. Ici on peut alors revenir aux lexiques pour les mettre en cohérence avec l'ontologie. La définition de la terminologie consiste à mettre en relation chaque terme des lexiques avec les concepts de l'ontologie. Ce travail minutieux permet l'indexation des documents précédemment recensés lors des entretiens, puis océrisés au format texte pour uniformiser les formats hétérogènes et rendre cette documentation totalement exploitable. Une plate-forme de test a ainsi été mise en place pour la recherche et la gestion de l'information, puis soumise aux usagers. L'ontologie au format XML peut être intégrée à tout type d'outil de gestion et de recherche d'information. L'exploitation de la terminologie permet de gérer les variations en langue, comme les différences d'usages entre communautés de métier au sein du même domaine, acronymes et développés, synonymes, homonymes, multilinguisme...

Toute notre méthode est basée sur l'échange avec les experts du domaine, ainsi que la compréhension réelle des notions du domaine dans la recherche constante de dépasser la partie purement linguistique du travail pour prendre en compte cette part extra-linguistique qui exprime seule les références communes articulant les savoirs. Une approche plus automatisée ne paraît pas viable, du moins par rapport aux objectifs que nous visons, les résultats en étant trop superficiels et tronqués. La première phase d'extraction automatique s'est avérée d'autant plus décevante qu'elle s'appliquait à un corpus très restreint : une construction manuelle du lexique aurait sûrement pu être envisagée, mais aurait-elle abouti aux mêmes résultats ? Notre travail est donc résolument basé sur l'humain et l'intelligence qu'il apporte aux traitements automatiques.

CONCLUSION

L'exemple que donnent nos travaux n'en est qu'un parmi d'autres. Les entreprises sont aujourd'hui nombreuses à se trouver confrontées à ces problématiques. L'intégration massive de l'outil informatique il y a près de 20 ans et la déferlante d'information qu'elle a induite a certes grandement facilité certains aspects des métiers, mais a aussi engendré des dysfonctionnements difficiles à dénouer. Une confiance aveugle en l'outil et un refus sans cesse renouvelé d'investir dans des moyens humains adaptés pour rendre ces outils intelligents (ou tout du moins logiques et maîtrisables), aboutissent à des situations inextricables. En réponse à cela, force est de constater que la tendance s'accroît encore : les effectifs documentaires qualifiés sont réduits et les réponses sont une fois encore recherchées auprès des outils, dans une fuite en avant technologique.

Il est essentiel de ne plus négliger la phase d'étude amont à la mise en place de solutions techniques. Elle permet l'analyse et la prise en compte de tous les paramètres inhérents au projet. Cette phase est souvent éludée pour un gain de temps et d'argent alors que, paradoxalement, passer outre produit *a posteriori* des surcoûts considérables, pour des résultats finaux inadaptés. L'humain est à ce point non pris en compte, que les usagers sont ignorés tant pour leurs besoins initiaux que pour leur utilisation finale de l'outil. C'est à eux qu'il est demandé de s'adapter et non l'inverse. L'information est dénuée de toute valeur ajoutée et de sa fonction de service.

Or, les problèmes de valorisation de l'information et des connaissances techniques des métiers cœurs de l'entreprise vont aller grandissant dans nombre de structures de taille importante. On l'a vu, les contextes techniques, démographiques, concurrentiels, etc., vont porter à un accroissement fort de ces problématiques. Mais peut-on récupérer le temps perdu en reproduisant les mêmes approches et les mêmes erreurs ? Certes non. Un changement d'optique est incontournable et doit prioritairement s'axer sur l'appropriation de l'information et des connaissances par les usagers. Il n'est plus possible de repousser le problème en laissant s'accumuler sans cesse plus d'information. Dans tous les cas, rien de constructif ne pourra être réalisé tant que l'outil et non l'humain, sera mis au centre des réflexions.

REFERENCES BIBLIOGRAPHIQUES

Aussenac-Gilles, Nathalie ; Condamines, Anne (2000), « Entre textes et ontologies formelles : les bases de connaissances terminologiques », in Jean Charlet et al. (dir.), *Ingénierie des connaissances. Evolutions récentes et nouveaux défis*, Paris : Eyrolles.

Bachimont, Bernard (2000) « Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances », in Jean Charlet et al. (dir.), *Ingénierie des connaissances. Evolutions récentes et nouveaux défis*, Paris : Eyrolles.

Boccon Gibod, Henri (2006), *Application de méthodes et outils de Web sémantique pour la gouvernance d'un système d'information industriel*, Clamart : EDF R&D.

Carlier, Alphonse (1994), *Stratégie appliquée à l'audit des systèmes d'information*, coll. Systèmes d'information, 2^e éd., Paris : Hermès.

Descartes, René (1987), *Le Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences*, rééd., Paris : Vrin.

Durampart, Michel (2006), « L'outil donne à voir l'organisation », communication présentée à *DOCSI-ERSICOM*, Lyon : mars 2006.

Gandon, Fabien ; Dieng-Kuntz, Rose (2005), « Ontologie pour un système multi-agent dédié à une mémoire d'entreprise », in *Ingénierie des connaissances*, Paris : L'Harmattan.

Pintea, Jean (1995), *Reengineering des systèmes documentaires*, Paris : Les éditions d'organisation.

Lainé-Cruzel, Sylvie (2006a), « Terminologie et intelligence artificielle », in *Encyclopédie de l'informatique et des Systèmes d'Information*, Paris : édition Vuibert.

Lainé-Cruzel, Sylvie (2006b), « Valoriser l'information : enrichir les systèmes et les interfaces, faciliter l'appropriation », communication présentée à *DOCSI-ERSICOM*, Lyon : mars 2006.

Lamizet, Bernard (1992), *Les lieux de la communication*, Liège : Mardaga.

Polanco, Xavier (1999), « Extraction et modélisation des connaissances : une approche et ses technologies (EMCAT) », in Maniez J. et Mustafa el Hadi W. (dir.), *Organisation des connaissances en vue de leur intégration dans les systèmes de représentation et de recherche d'information*, Lille : Travaux et recherches.

Popper, Karl (1979), *Objective knowledge: an evolutionary approach*, Oxford, Oxford University Press.

Roche, Christophe (2005), « Terminologie et ontologie », *Langages*, n° 157.

Roche, Christophe (2007), « Dire n'est pas concevoir », communication présentée à *IC 2007*.