

Le projet RAMCESH et son application aux tunnels

N. Faure - Solem France SA (Archamps)
R.M. Faure - Centre d'Etude des Tunnels (Lyon)
G. Hémond - Solem Expertises SA (Genève)

Résumé :

Cet article présente les résultats du projet Ramcesh (Recueil Assisté et Maniement des Connaissances des Espaces Souterrains Habités) portant sur le recueil et le maniement des connaissances. La première partie situe le contexte de cette réalisation vis à vis de la problématique du recueil de connaissances et du domaine d'application (le construit en Génie Civil) dont la spécificité sera précisée. (Parties 1, 2 et 3). Les parties 4, 5 et 6 concernent une théorie de la connaissance qui conduit, pour tenir compte des spécificités du domaine concerné, à utiliser une ontologie de poids léger et une structure appelée "granule de connaissance" afin de stocker de façon efficace la connaissance. Les parties 7 et 8 portent sur la mise en œuvre de l'outil, son introduction dans un service et présentent ses possibilités.

Abstract:

The RAMCESH project and its application to tunnels

This paper presents the main results of the Ramcesh project achieved by the CETu and Solem SA. Ramcesh is for Knowledge Management about Living Underground Works. The specificity of underground works is first established, that leads to the use of light weight ontology and a new structure of knowledge called "granule de connaissance". With this approach it is possible to store any kind of knowledge and use it in a very efficient manner. The large possibilities of Ramcesh are detailed and its use in a company is presented (installations, downsizing of worksites).

1 - Introduction

Le projet Ramcesh (Recueil Assisté et Maniement des Connaissances des Espaces Souterrains Habités) est une recherche conduite conjointement par la société Solem Expertises SA et le Centre d'Etude des Tunnels, avec l'aide de l'Université Lyon 3 (Laboratoire Modeme) pour stocker sans redondance et retrouver instantanément la connaissance correspondant à un contexte donné.

Ramcesh est un projet fondé sur les collaborations d'experts d'un domaine, impliqués dans un processus cumulatif de stockage de bribes de connaissance pour finaliser une base capable de transmettre ainsi les connaissances d'un domaine et d'aider efficacement tout utilisateur dans son travail de conception ou de recherche. Ceux qui alimentent la base en seront les premiers utilisateurs et apprécieront les possibilités innovantes et puissantes de l'outil.

L'article retrace la genèse du projet, présente la théorie de la connaissance utilisée et décrit le processus conduisant à une base de connaissance réactive fondée sur le stockage de la connaissance avec son contexte d'usage.

2 - Contexte du projet Ramcesh

Le projet Ramcesh est le fruit d'un long processus de réflexion puis de maturation, aiguillonné par l'évolution des technologies informatiques et répondant à un besoin d'accès rapide à la connaissance.

2.1 - Le recueil des connaissances

Depuis quelques années la nécessité du recueil des connaissances est admise par

tous dans le but principal de transmettre savoir et savoir-faire entre générations. Devant la difficulté du processus, les experts ne sont pas enthousiastes. Beaucoup d'applications de "management des connaissances" produisent une cartographie des connaissances de l'entreprise. Cette cartographie est alors utilisée par le Directeur des Ressources Humaines qui peut ainsi faire une adéquation entre les moyens humains et les domaines de l'entreprise. Le "livre des connaissances" produit par la méthode MASK (Ermine, 2003) est un processus de mise à plat de la connaissance dont l'appropriation par un nouveau venu reste classique, par la lecture. Nous avons voulu aller plus loin et faire que la connaissance puisse "réagir" à un contexte donné pour, par exemple, détecter une situation à risque demandant une action particulière. Cependant, le choix de cette action revient à l'utilisateur et l'outil Ramcesh apparaît comme un bon compagnon qui avertit, mais ne décide pas.

2.1.1 - La transmission des savoirs

La transmission des savoirs fut d'abord orale, puis livresque et notre enseignement est encore basé sur le livre. Mais comment retrouver dans un livre la connaissance dont on a besoin pour traiter le problème posé par un contexte particulier ? On a cru arriver à ce but avec l'aide de moteurs de recherche. Ceux qui utilisent 'Google' en connaissent les limites, car la contextualisation des connaissances n'y est pas incluse.

Dans Ramcesh la contextualisation est centrale, si bien que le savoir est toujours pertinent, et Ramcesh tourne les pages des livres pour apporter la bonne connaissance, celle qui est utile au problème décrit.



Cette transmission du savoir, si précise soit-elle, ne donne pas de solution, mais les bons éléments pour que l'utilisateur énonce une solution. La justesse de cette solution dépend de la précision du contexte décrit. Et quel avantage de voir sur son écran tout ce qu'il faut savoir sur ce contexte. L'art de l'ingénieur n'est que facilité.

2.1.2 - Travaux antérieurs au projet Ramcesh

Le projet Ramcesh est dans la continuité des projets XPENT et WASSS, dont les enseignements ont été pris en compte.

XPENT (Faure et al, 1988) (Mascarelli et al, 1992) a été un projet de coopération avec l'université Laval à Québec pour la mise au point d'un système expert en stabilité de pentes. Cela fut l'occasion de découvrir un monde ouvert (du point de vue informatique) et de saisir le comportement de l'expert par les interviews que la méthode KOD (Knowledge Oriented Design) imposait. L'expert ne développe pas son expertise avec grande cohérence, c'est un parcours fait d'aller-retours, quasiment fractal, avec uniquement le but d'arriver à une solution à l'aide de nombreuses références, pas toujours judicieuses, mais servant de support à cette pensée tumultueuse. A la fin du projet XPENT, la nécessité d'une base de cas d'où les experts tireraient des parties de raisonnement a été mise en évidence (Mascarelli, 1994).

Le projet WASSS (Faure et al, 1999) (World Area Slope Stability Server) a suivi le projet XPENT dans le cadre du groupe technique pour les glissements de terrain du Comité international de mécanique des sols et des fondations. Un glossaire multilingue a été développé et les exigences souvent contradictoires des pays membres ont été intégrées dans un serveur de cas, pour les pentes (Faure et al, 2001), et pour les tunnels (Faure et al, 2002). Nous retiendrons de ce projet la difficulté du dialogue entre communautés et la simplification extrême qu'il faut apporter aux manie-

ments informatiques et aux théories mises en œuvre. Complicé à la rigueur mais surtout pas complexe.

2.1.3 - Le but fondamental du projet Ramcesh

Le but fondamental à l'origine de Ramcesh est de retrouver instantanément ce que l'on sait, parce que l'on a lu cette information ou que l'on nous en a parlé. Qui n'a pas cherché dans sa mémoire, l'article, le livre qui parle de ce sujet ?

On fait souvent appel à sa mémoire parce que la recherche documentaire usuelle produit trop de documents et est peu productive au niveau de la connaissance élémentaire, cachée dans un des documents.

Dans Ramcesh, la recherche se faisant par contexte, elle peut être aussi précise que possible, l'utilisateur affinant le contexte. Il est assisté en cela par le glissement sémantique (recherche par proximité) que permet l'ontologie.

2.1.4 - Chercher la connaissance dans le document

Comme il a été dit, la recherche par document est souvent peu productive car trop de documents sont sélectionnés. Il faut les lire, ou bien le document est très complet et la connaissance cherchée est cachée dans le volume des informations du document.

Ramcesh stocke les éléments de la connaissance et garde leurs références au(x) document(s), car la connaissance peut avoir été trouvée dans plusieurs documents. Cette connaissance élémentaire peut donc être retrouvée par son contenu (approche sémantique), par son contexte et par ses références. C'est cette combinaison des approches qui rend Ramcesh très efficace.

Pour stocker les éléments de connaissance, Ramcesh propose un traitement simple et assisté des documents pour en extraire cette connaissance et il ne la stocke qu'une seule fois. (voir chapitre 8)

2.2 - Les enjeux

Les enjeux du projet Ramcesh sont importants par le fait que le projet fournit un moyen de stocker la connaissance, de façon non redondante, permet ainsi de la transmettre ou de la retrouver et surtout rend cette connaissance réactive en relevant automatiquement les contradictions d'un discours ou d'un projet, ce qui permet ainsi d'améliorer celui-ci vers une meilleure qualité. On sait que l'oubli d'une information, au sein d'un projet, peut conduire à des surcoûts parfois importants. En matière d'aménagement et de construction les textes règlementaires sont très nombreux et complexes, leur appropriation est longue, et ce n'est pas le renouvellement permanent des équipes qui facilite l'accumulation systématique des savoir-faire, qui sont dans la théorie de la connaissance développée dans Ramcesh considérés de la même façon que les savoirs. Ainsi, la complexité des textes n'est plus une difficulté.

2.2.1 - Qualité de la production

En ingénierie le projet est une production essentielle et tendre vers de meilleurs projets dans un monde où le temps est compté est une gageure. Le système Ramcesh peut aider à augmenter la qualité d'un projet.

La détection des incohérences et des oublis au sein d'un projet est une étape vers la qualité du projet. C'est une des conséquences bénéfiques du système Ramcesh. Comme la connaissance est toujours rattachée à un contexte, la description du contexte rappelle la connaissance qui est simplement proposée à l'utilisateur, car une définition insuffisante du contexte peut amener une connaissance apparemment contradictoire. En effet dans le monde ouvert de la construction (voir plus loin), la vérité dépend de la précision du contexte : ce qui est faux dans un contexte donné peut devenir vrai dans un contexte mieux précisé. Et cette précision demandée



Figure 1
Bandeau de
la version
de Ramcesh 2005

est nécessaire pour lever les incohérences apparentes, et de fait, améliorer la qualité du projet.

2.2.2 - Le traitement d'un flot d'information

Dans tous les domaines, et particulièrement en sciences de l'ingénieur, l'information est abondante, revues, livres, internet. Mais pour une information pertinente, celle dont j'ai besoin, combien de lectures inutiles et de temps passé ! Un service documentaire qui stocke et classe des documents en sélectionne beaucoup en réponse à une question. Le demandeur lit les premiers documents et abandonne parfois sans trouver la réponse. De même la fouille sur Internet n'est pas toujours productive. La lecture de documents comme des articles est parfois besogneuse car on retrouve de nombreuses fois la même connaissance sous des aspects à peine variés. Pourtant une très grosse partie de la connaissance est déjà écrite dans des textes et c'est pourquoi la démarche de Ramcesh s'appuie essentiellement sur l'écrit.

Le stockage de la connaissance dans le système Ramcesh se fait sans redondance car c'est la connaissance exprimée sous une forme simple (le granule de connaissance) qui possède les références des ouvrages dans lesquels se trouve cette connaissance. Ce travail de mise en forme est fait à la création de la base de connaissance qui est collaborative et évolutive, il suffit que le texte soit lu une fois pour que la connaissance soit efficacement stockée.

Le but élémentaire est la recherche d'une connaissance qui s'applique à un contexte et vice-versa. A partir de cela bien des possibilités sont envisageables.

2.2.3 - Une idée de l'originalité et de l'antériorité

La formalisation de la connaissance et son repérage dans plusieurs documents permet d'en connaître l'antériorité et sa réutilisation d'un document à l'autre. Les redondances multiples, de livre en livre, d'article en article, sont détectées et enregistrées et l'on pourra rendre à César....

3 - La spécificité du construit

Parce que les tunnels sont dans un sol, le projet Ramcesh porte initialement sur la géotechnique qui s'occupe du sol, milieu naturel créé par une lente évolution, plusieurs milliards d'années, de la croûte terrestre. Nous ne connaissons jamais les

détails de cette évolution, si bien que le monde de la géotechnique est un monde ouvert.

Un monde ouvert au sens informatique est un monde où toutes les règles ne sont pas connues, il peut toujours arriver quelque chose de non prévu. Dans le monde des tunnels c'est une réalité quotidienne, qui est aussi vraie pour tous les construits en liaison avec le sol.

Un monde fermé est un monde parfaitement connu, où tous les éléments sont recensés. Le monde de l'industrie est souvent de ce type : dans un processus industriel, sur une chaîne de montage par exemple, seules les pièces prévues peuvent arriver.

Ceci explique en partie pourquoi les systèmes experts des années 80 ont eu de belles réussites dans le monde industriel, mais jamais dans les sciences de la terre. La découverte d'un gisement par le système expert "prospector" semble bien n'avoir été qu'un heureux concours de circonstances.

De la géotechnique à l'acte de construire, la problématique reste sensiblement identique, aussi le projet Ramcesh peut être appliqué à tout domaine d'investigation et de réalisation dans un domaine vaste et partiellement connu.

Ramcesh œuvrant dans un monde ouvert utilise une approche élargie par rapport à celle des systèmes experts, par l'ajout des apports des ontologies et des univers de modèle.

Dans le paragraphe suivant nous précisons le monde de la géotechnique, mais le lecteur comprendra que la spécificité de cette science peut être transposée à bien d'autres domaines.

3.1 - Définition de la géotechnique

La géotechnique est la science qui se préoccupe des interactions d'un sol et d'un construit, route, bâtiment ou ouvrage d'art. A ce terme de géotechnique est parfois substitué celui de géosciences, mais cette dernière appellation apparaît moins éclairante sur la nature empirique de l'activité géotechnique, c'est-à-dire essentiellement fondée sur la relation à un contexte donné. Toute démarche géotechnique ne peut se concevoir qu'en fonction d'une situation réelle ou supposée, d'une contextualisation d'un problème, et ce généralement par analogie avec une situation similaire déjà connue. Cela traduit, de fait, l'activité géotechnique comme un

ensemble de savoir-faire et de techniques, comparables au tour de main d'un artisan (Magnan 2002). Le terme de géosciences est pourtant utilisé parce qu'il est aussi révélateur d'une des deux caractéristiques fondamentales de ce domaine, à savoir son hétérogénéité de pratiques. De fait, les géosciences constituent l'ensemble hétérogène des sciences, ou éléments scientifiques, mobilisées dans la pratique géotechnique.

3.1.1 - Hétérogénéités techniques

La géotechnique est en effet le point de rencontre de spécialités distinctes : géologie, mécanique des milieux continus et discontinus, chimie, hydraulique, etc... Chacune de ces spécialités implique une conception a priori des éléments géotechniques, une terminologie et des pratiques particulières. Par exemple, le concept, aussi fondamental soit-il, de sol, ne sera pas forcément compris a priori de la même manière par un géologue ou un spécialiste de stabilité des pentes. Pour le premier, un sol est un ensemble de roches situé sous la couche meuble de la surface. Pour le second, en revanche, cette couche meuble, parfois appelée sol détritique et située au-dessus du substratum rocheux rigide, est l'objet de son étude, de "son" sol.

3.1.2 - Hétérogénéités régionales

Un deuxième élément fonde l'hétérogénéité de la géotechnique, à savoir sa disparité régionale ; les types de roches et les configurations de sol sont liées à des régions, ce qui donne donc naissance à des terminologies et des concepts régionaux. En France, le granit du Mont d'Ambin est un exemple et évoque une forme particulière de granit et certaines de ses propriétés. Le Toki granite, autre exemple, est évoqué dans certains articles traitant de projets géotechniques au Japon et définit un autre type de granit, non caractérisé autrement que par sa provenance. Dans l'un et l'autre cas, les spécialistes peu familiers de la géologie alpine ou japonaise sont confrontés à des données incomplètes, car la dénomination particulière du granit implique l'importance de propriétés particulières. La généralisation des caractérisations régionales n'est pas nouvelle, et transparait assez facilement avec l'échelle des temps géologiques, le "calendrier géologique", où les étages et sous-étages géologiques portent les noms dérivés des régions typiquement représentatives des sols des époques correspondantes : Maestrichtien (Maestricht), Hettangien (Hettange), Coniacien (Cognac),



etc... Ces travaux de dénomination et de calibrage correspondant sont souvent compliqués par le caractère évolutif de ces éléments qu'on pourrait croire figés : l'échelle des temps géologiques est ainsi remise à jour tous les quatre ans. Ces quelques exemples permettent de broser le tableau d'un environnement professionnel où les usages sont importants. Dans le cas de l'échelle des temps géologiques, cela correspond à l'enrichissement d'une colonne "autres dénominations", aux limites parfois mal identifiées, mais d'une manière générale cela correspond à l'enrichissement informel d'une polysémie déjà rendue redoutable par l'étendue du domaine géotechnique.

Afin d'illustrer la combinaison de ces éléments régionaux et terminologiques, citons enfin le cas typique de l' "argile de Mexico", qui, si elle est effectivement propre à Mexico, n'est pas de l'argile, mais un résidu détritique de roches volcaniques : synecdoque accolée à une régionalisation témoignant de la pérennisation d'une dénomination abusive mais couramment utilisée.

3.1.3 - Un monde artisanal

Il y a comme une impossibilité de la connaissance pour l'abord d'un projet en géotechnique. On ne sait que bien peu de choses et tout est possible. Cet incertitude est aussi source de contradictions, souvent apparentes et qui seront levées par une meilleure connaissance montrant que le processus de conception est itératif avec de nombreux retours sur le terrain.

Cette impossibilité de la connaissance n'est pas seulement le fait de la double hétérogénéité du domaine qui vient d'être décrite, mais aussi de la nature même des projets géotechniques. Ceux-ci comment généralement par la conception d'un construit inexistant à implanter dans un environnement inconnu. Le sol, en effet, est un ensemble qui échappe aux méthodes d'observations usuelles, et notamment visuelles ; on ne peut en connaître les caractéristiques qu'au travers de tests et de sondages qui, pour aussi précis qu'ils soient, ne permettent qu'une approximation de la réalité. Par ailleurs, un sol donné n'est pas reproductible en laboratoire, et les sondages ont un coût élevé. De fait, un sol est un système complexe dont l'étendue physique elle-même est soumise à hypothèse, puisque ses caractéristiques peuvent être générées via des interactions à partir d'éléments dont il est a priori impossible de déter-

miner l'emplacement, la proximité ou même la nature. Quand à l'approche géostatistique elle est illusoire, le volume testé du sol, donc connu, correspondant dans les meilleurs cas à une partie par milliard. (Magnan, Pouget, 1998)

On trouve là les raisons fondamentales de l'empirisme mentionné en premier lieu dans cet article, les causes de l'importance de l'expertise et la motivation principale pour le raisonnement par analogie, avec tous les risques qu'il comporte pour les projets de conception. C'est pourquoi la démarche pragmatique (au sens premier du terme, c'est-à-dire dont la validité est déterminée par les résultats) est très souvent utilisée en géotechnique.

Cet état de fait peut être résumé en une approche artisanale où chaque projet est une œuvre unique.

3.2 - Tentatives de formalisation

Paradoxalement, la communauté géotechnique est une communauté fortement mondialisée et très communicante, dans la mesure où pléthore de documents d'une nature ou d'une autre sont produits, dans des revues spécialisées ou à l'occasion des nombreux congrès et symposiums internationaux. Un premier constat est que l'on retrouve dans la forme générale documentaire ce qui vient d'être dit : on distingue les documents à haute technicité, généralement relevant d'une spécialité (le plus souvent en physique des matériaux) et les rapports de chantier, essentiellement descriptifs, décrivant situations rencontrées, mesures prises en fonction de ces situations et résultats de ces mesures. De manière comparable, et pour ce qui concerne la pratique géotechnique en France, le recours à des systèmes d'aide informatique se limite bien souvent à des logiciels de calcul, consacrés à des points spécifiques : dimensionnement de tunnels en fonction d'un sol donné, calcul des confortations nécessaires à la stabilisation d'un talus, estimation des trajectoires de chutes de blocs en fonction de certains paramètres, etc... La communauté géotechnique française a connu un vaste engouement pour les systèmes à base de connaissance dans les années 1980 à 1990. De nombreux systèmes experts ont donc été mis à l'étude (CESSOL, XPENT, CASTOR, SOUT, etc...), mais rarement utilisés en pratique. Les causes de ce fait sont multiples et ont parfois été analysées (Magnan 1992), mais deux constantes apparaissent :

- le besoin pour l'utilisateur de comprendre l'outil qu'il manipule, au moins ses prin-

cipes fondamentaux, et le rejet des intermédiaires pour le traitement de la connaissance,

- l'investissement souvent trop important, au moins en temps et en énergie, pour alimenter une base de connaissances en mesure de fournir des réponses satisfaisantes.

Les tendances les plus récentes concernant les systèmes d'information visent à faciliter le travail collaboratif, avec notamment l'utilisation de plus en plus accrue des réseaux et de systèmes d'annotations et commentaires concernant des documents de travail, ou de mise à disposition de cas pour constituer une base de références communes. Ce dernier type de système rejoint ce qui était dit plus haut concernant la littérature du domaine et la distinction entre articles à haute technicité et descriptifs de cas. On constate du reste que la communauté géotechnique véhicule à travers ces descriptions nombre d'histoires qui augurent favorablement de ce que des techniques comme le *story telling* pourraient apporter au domaine.

4 - Les formes de la connaissance

La connaissance est une chose difficile à définir. Un effort important a été fait lors de l'élaboration du projet Ramcesh pour comprendre la nature de la connaissance. Nous présentons une approche ingénieur (intuitive et pragmatique) de cette connaissance, puis les résultats des efforts faits sous la forme d'une théorie de la connaissance.

4.1 - Approche intuitive de la connaissance

Pour l'ingénieur la connaissance semble prendre quatre formes :

- La dénomination des objets et des concepts. En effet savoir l'existence et la définition de quelque chose (un objet, une théorie, une action, un état...) est une connaissance.
- La seconde forme de la connaissance est l'ensemble des relations et des faits reliant deux ou plusieurs concepts. Un savoir faire comme la composition d'une formule de béton fait partie de ce second type de connaissance.
- La troisième forme est l'algorithme qui permet une traduction numérique des relations entre objets.
- La dernière forme est la donnée, généralement chiffrée. Savoir qu'un tunnel a une

largeur de 10 mètres est aussi une connaissance.

Un des *challenge* du projet Ramcesh est la recherche d'une formalisation unique permettant le maniement facile (et intégré) de toutes ces formes de connaissances.

4.2 - Une théorie de la connaissance

La connaissance n'existe pas en elle-même, elle ne se manifeste que par la réalisation de quelque chose. Faire un projet nécessite de la connaissance, mais laquelle ? Il y a forcément une dimension téléologique dans la connaissance, elle est exhibée par un but. Elle est même quantifiable, un bon projet est un projet pour lequel la connaissance utilisée n'entraînera pas d'erreur, alors que la connaissance d'un mauvais projet procurera beaucoup d'ennuis et de retards.

C'est pourquoi, faire du recueil de connaissance dans l'absolu a peu de sens. Dans notre domaine, il s'agit de bien construire, notre but. Mais il s'agit aussi de communiquer notre projet, d'où une dimension maïeutique à la connaissance utilisée.

Nous avons déjà rencontré la nécessité de contextualisation pour des raisons de vérité du discours.

Pour bâtir l'approche de Ramcesh, en tenant compte de ces contraintes, nous allons approfondir les notions de concepts et de relations entre ces concepts.

Le modèle du signe va être très utile, mais aussi la théorie de la représentation du discours pour comprendre, puis définir une technique de représentation de la connaissance. Les structures ontologiques permettront la mise en œuvre, mais de façon incomplète, de cette approche, ce qui amènera à définir le granule de connaissance. Cette théorie de la connaissance a été développée par Nicolas Faure (Faure N., 2007).

5 - Le concept dans le projet Ramcesh

5.1 - Ce qu'est un concept ?

Intuitivement la notion de concept correspond à une représentation mentale possédant un nom. Un objet (voussoir), une théorie (la méthode convergence confinement), un lieu (le tunnel de Fourvière) peuvent être qualifiés de concepts. Un ensemble de lettres, le nom, identifie un concept et sa compréhension induit une image mentale. Du fait de l'imperfection et de l'évolution permanente du langage, ce

n'est pas aussi simple car synonymes, homonymes, formes linguistiques brouillent cette correspondance que l'on voudrait bi-univoque.

Des théories linguistiques et sémiotiques, que nous ne développerons pas ici, nous viennent en aide pour lever flou et ambiguïté, et permettent de bâtir une représentation du monde à l'aide de deux nouveaux composants : une ontologie de domaine et le granule de connaissance.

5.2 - L'approche conceptuelle étendue à l'ontologie

Les théories évoquées ci-dessus induisent la nécessité de la spécification (choix d'un identifiant unique pour un concept donné), et de la conceptualisation (définition de l'ensemble des traits conceptuels pour chaque concept spécifié). Ces deux mécanismes permettent de délimiter un volume dénotatif explicitement borné, même si des variations mineures sont toujours possibles. Chaque concept identifié par un syntagme (groupe de mots) unique dans l'ontologie sera accompagné d'une définition accessible à tous (Faure N., 2006).

5.3 - Les relations entre concepts

Les relations entre concepts permettent de décrire le monde et peuvent être très complexes. Dans le projet Ramcesh, le souci d'être compliqué mais pas complexe (au sens de l'approche systématique (Lemoigne)) conduit à séparer la formalisation des relations, entre l'ontologie et une forme proche du langage usuel, le granule de connaissance.

Seulement deux relations seront présentes dans l'ontologie, pour qu'elle reste simple et facilement gérable (ontologie pragmatique de poids léger) ; les autres relations, que nous appellerons relations métier, serviront à construire le granule de connaissance.

Les deux relations retenues pour l'ontologie sont la relation de subsomption (est un) et la relation de voisinage (est proche par le sens).

Le chapitre suivant développe ontologie et granule de connaissance.

6 - Ontologies et granules de connaissance

La base de connaissance générée dans le système Ramcesh comporte deux bases de données. L'une représente l'ontologie du domaine et la seconde l'ensemble des granules de connaissance, dont les syntagmes sont ceux de l'ontologie. La

liaison entre ces deux bases est très étroite, ce qui permet continuité et glissement sémantique.

6.1 - Ce qu'est une ontologie

L'approche la plus généraliste pour décrire ce qu'est une ontologie est de la relier à d'autres outils de même type et plus largement connus du grand public. Divers auteurs ont déjà établi cette relation, généralement en décrivant une gradation formelle des outils les plus simples jusqu'aux ontologies. Nous commentons ci après les principales structures de collection de mots.

- L'index est un ensemble de termes non structurés (sinon par ordre alphabétique, mais il s'agit là d'une structuration morphologique et non sémantique), généralement extraits d'un document (ou d'un corpus documentaire s'il s'agit de plusieurs documents), et éventuellement reliés à une ressource (classiquement, un numéro de page).
- Le glossaire est un index assorti de définitions, ce qui introduit la notion de sens, de sémantique, que ne comporte généralement pas l'index.
- Le thésaurus est un glossaire assorti de relations entre les termes. Généralement, ces relations sont de type terme spécifique / terme général (par exemple : " baleine à bosse " est plus spécifique que " baleine ", et " cétacé " est plus général que " baleine "), de synonymie ou de voisinage (qui mettent en relation des termes comme "Titanic " et " iceberg ", par exemple). Les thésaurus sont généralement éminemment pragmatiques (c'est-à-dire qu'ils sont construits avant tout pour pouvoir retrouver des documents dans un ensemble documentaire et supportent rarement une volonté de modélisation) et fonctionnent avec des mots contrôlés (des descripteurs) qui représentent des catégories auxquelles on rattache les concepts lorsqu'ils sont rencontrés dans de nouveaux documents. En raison de ces particularités, on considère généralement qu'un thésaurus ne décrit pas de hiérarchie conceptuelle et supporte assez mal une représentation des connaissances.
- Si l'on systématise l'aspect relationnel du thésaurus (en multipliant les descripteurs et les relations, notamment), on s'éloigne des outils de classification courants pour se rapprocher de ce qu'on appelle des ontologies, en obtenant ce qui correspond



peu ou prou à des réseaux sémantiques. Ceux-ci permettent tout d'abord l'élaboration de hiérarchies, un élément fondamental de toute ontologie. On considère que ce " chaînon manquant, " entre les thésaurus et les ontologies, est représenté par la catégorisation du célèbre moteur de recherche Yahoo !, peu formelle mais dessinant nettement une hiérarchie dans laquelle chacun des termes peut être situé. Ces hiérarchies sont également appelées des taxonomies. Une taxonomie est donc une hiérarchie de termes définis selon une relation plus ou moins précise. Cette relation est souvent la relation de spécialisation/généralisation (" is-a " en anglais). Cette relation peut être informelle ou formelle. Une hiérarchie formelle fondée sur la spécialisation/généralisation autorise l'héritage (certains auteurs parlent alors de subsomption), c'est-à-dire que si la baleine est un cétacé et la baleine à bosse une baleine, alors la baleine à bosse est un cétacé. Une hiérarchie informelle est moins regardante sur cet aspect (ainsi, dans le cas de Yahoo, les "Accessoires de mode" apparaissent dans la catégorie "Femmes").

- Dès lors que la subsomption est respectée (au moins, car une ontologie possède généralement plusieurs hiérarchies fondées sur autant de relations : fait-partie-de, s'utilise -concurrentement-avec, ...), on entre dans le domaine des ontologies, mais celles-ci possèdent plusieurs gradations internes, selon leurs caractéristiques. Ainsi, on peut distinguer les ontologies qui supportent la représentation des instances (c'est-à-dire des objets individuels, cette table précisément ou ce document et pas un autre). Ces ontologies permettent d'effectuer des opérations sur les instances et donc de modéliser plus précisément le monde.
- Enfin, les ontologies les plus formelles incluent des restrictions de valeur dans leurs relations (par exemple, un objet tombe dans la catégorie "licite" ou "illicite" selon le pourcentage de tel ou tel ingrédient qui le compose) et des contraintes logiques (un objet ne peut être humide et sec en même temps). Ces ontologies sont appelées "ontologies lourdes" ou "poids lourd" d'après l'appellation anglo-saxonne "heavyweight ontologies". Les autres ontologies sont dites, a contrario, "légères" ou "poids léger".

6.2 - Définition des ontologies

Si les ontologies trouvent leurs origines dans la philosophie (et notamment à Aristote), ce n'est pas avant 1991 que l'on trouve dans la littérature spécialisée la première définition des ontologies dans le cadre des systèmes d'information (en informatique). A partir de là, de nombreuses définitions ont fleuri, on n'évoquera ici que les trois plus importantes pour le projet :

- T. Gruber [Gruber, 1993] définit une ontologie comme " la spécification d'une conceptualisation ", c'est-à-dire la description formelle d'un état des choses.
- N. Guarino [Guarino, 1998] va plus loin en affirmant qu'une ontologie est une théorie logique rassemblant plusieurs conceptualisations, dans la mesure où un état des choses n'est pas suffisamment abstrait pour être autre chose qu'un modèle.
- J. Sowa [Sowa, 2001] est plutôt proche de la définition de Guarino, mais s'affranchit de la notion de théorie logique en décrivant (entres autres) des ontologies fondées sur des prototypes, où un concept n'est pas une catégorie, mais un prototype duquel les cas réels peuvent se rapprocher plus ou moins en fonction d'une distance sémantique.

Il existe bien d'autres définitions des ontologies, mais celles évoquées ci-dessus sont les plus représentatives des différentes approches des ontologies. Avec de nombreux auteurs, notons que dans bien des cas les ontologies sont des outils de référence pour une communauté. Celle de Ramcesh pourrait donc devenir une référence dans le domaine des tunnels.

6.3 - L'ontologie de Ramcesh

L'ontologie de Ramcesh est une ontologie de poids léger, pragmatique et basée sur deux relations, la relation de subsomption (est une sorte de) et la relation de voisinage. Pour représenter cette ontologie, la structure d'arbre a été retenue, car simple à mettre en œuvre et à réaliser. Elle est stockée dans une base de données relationnelle usuelle (MySQL) et représentée à l'aide de l'outil Mind Manager (voir www.mmdfrance.fr). A chaque syntagme, qui est un noeud de l'arbre, sont associés : une définition, des synonymes et des traductions dans d'autres langues que le français. La définition de voisinages sur le graphe support de l'ontologie, permet de mettre en évidence des régions sémantiques,

que l'on utilisera dans le glissement sémantique afin d'élargir certains questionnements.

6.4 - Le granule de connaissance

Le granule de connaissance est le composant informatique support de la connaissance créé pour le projet Ramcesh (Faure N., 2007) et que nous décrivons ci-après, similaire dans son origine comme dans sa forme à des approches existantes (Clark et Porter, 1997). Le granule est issu d'une analyse décrite dans (Faure N., 2006).

Le cœur du granule est composé de deux parties, les prémisses et les conclusions, qui rappellent les règles de production des systèmes experts. La relation d'implication entre ces deux composants est une des cinq suivantes :

- alors obligation : pour couvrir le caractère obligatoire des normes
- alors recommandation : pour couvrir le caractère recommandation des normes et autres textes.
- alors possibilité : qui est la relation par défaut et qui traduit un état de fait laissé au choix de l'ingénieur
- alors négation : pour traduire les interdictions
- alors risque : pour alerter l'utilisateur que certains désagréments sont alors possibles.

Prémisses et conclusions sont un ensemble de phrases reliées par des ET, mais permettant aussi la négation.

Une phrase est formée d'un thème (ce dont on parle), d'un prédicat (ce qu'on en dit) avec entre ces deux composants, une relation métier. La phrase peut ne comporter que le thème. Le thème, unique dans une phrase, peut être associé à plusieurs prédicats. Thème et prédicat sont des syntagmes de l'ontologie.

D'un granule, on peut faire référence à un autre granule, en s'assurant qu'une simple extension (réécriture) du premier granule ne suffit pas. La référence ne s'applique qu'aux prémisses.

Attention à ne pas confondre prédicat et valeur, car la relation fait partie du prédicat, d'où une difficulté d'écriture ; la relation correspond à une définition de la propriété et le prédicat à la valeur de cette propriété.

La relation métier peut être non significative, et n'introduit aucun prédicat, sinon elle est choisie parmi l'ensemble des relations métier qui sont les seuls mots de l'ontologie à ne pas être des substantifs. Nous

donnons ci-après quelques exemples de relations métier. Il y a deux grandes familles de relation métier : les relations conceptuelles ou absolues et les relations contextuelles ou relatives.

Relation conceptuelle ou absolue

- causalité, à cause, cause de, appliqué à, portant sur, contre
- matière, constitué de, contient, nuance, produit de
- caractérisation, de forme, de grandeur, de couleur, de volume, de niveau de risque, de propriété

Relation contextuelle ou relative

- relation avec l'espace, autour, proche, au dessous, au dessus, devant, derrière, à coté, sous
- partie, compris dans(contient), tout, modulo, comme, avec, du, sans
- relation avec le temps, pendant, avant, après, lors de, en même temps, tout le temps (permanent), à chaque
- comparaison, par rapport, très supérieur, supérieur, égal strict, a peu près égal, inférieur, très inférieur, grand, petit, faible (quand le comparé est implicite)
- qualité, après convenance, avec contrôle, sans contrôle, avec agrément, conforme à, suivant règle, d'une nomenclature, dans la tolérance, du domaine d'application, origine, maintenance
- méthode, avec prudence, dans l'urgence, suivant agrément
- en référence à, document, auteur, granule de connaissance, ouvrage, etc
- action, prendre en compte, faire, stopper, vérifier
- variation, augmenter, diminuer, statu quo

L'expérience de la mise en granule nous a fait découvrir les types de granule suivant :

- Granule qui commande une action : si....alors faire ceci
- Granule qui annonce un phénomène : si...alors il pleut
- Granule qui traduit un choix : si....alors choix dans une liste
- Granule check list : si....alors vérifier tout ça
- Granule norme : si.... alors utilisation norme N ou si norme N alors
- Granule cas : si....alors cas nommé.

Le granule comporte aussi la définition d'un univers, et d'un modèle de cet univers ce qui permet d'assigner des valeurs numériques à certaines grandeurs et de faire ainsi des comparaisons. L'ensemble des concepts présents dans un granule forme cet univers du granule (appelé aussi

signature), et les informations relatives aux individus afférents à ces concepts sont comprises dans ce modèle d'univers : informations numériques pour certaines grandeurs, cardinalité ou équivalence entre concepts.

D'autres éléments informatifs sont aussi attachés à chaque granule : les références des textes dont est issu le granule, le fragment de texte initial, la date de création, l'auteur, et d'autres éléments qui facilitent son maniement.

(voir figure 2)

6.5 - La continuité et le glissement sémantique

La continuité sémantique de Ramcesh est la possibilité de considérer des situations allant de l'évocation d'un simple concept en tant que tel au sein de l'ontologie, jusqu'à la description d'une situation compliquée comme un projet, à l'aide d'une association de granules.

Le glissement sémantique est contenu dans l'ontologie construite sous forme d'un arbre, ce qui permet de définir des voisinages sémantiques qui seront utilisés pour la construction du granule et l'élargissement automatique des questions.

Les principaux voisinages sémantiques sont : Fils et père, tous les syntagmes contenant un mot donné, tous les pères jusqu'à la racine, fils et petits fils, etc.

7 - Mise en œuvre du projet

La mise en œuvre du projet Ramcesh s'est faite par prototypage incrémental, c'est à dire que plusieurs écritures informatiques ont été réalisées afin de préciser interfaces et usages. Trois prototypes ont précédé la version actuelle.

7.1 - Développement des outils

7.1.1 - Options informatiques

Les options informatiques choisies correspondent aux possibilités de mise en ligne facile du système Ramcesh. Les bases de données sont supportées par MySQL, le langage de programmation est PHP et un navigateur est nécessaire pour afficher les écrans de Ramcesh.

Un site hébergeur permet l'accès via le Web pour un partage facile des tâches.

7.1.2 - Prototypage incrémental

Le prototypage incrémental est la technique de mise au point des interfaces et des outils du système Ramcesh. En 2004,

une première version permettait la mise en œuvre d'une première ontologie. En 2005 l'interface permettant la mise en forme des granules était réalisé, en 2006 des outils d'aide au développement de l'ontologie et des améliorations importantes conduisent à redéfinir les dialogues et les interfaces, ainsi que des ajouts dans les bases de données. La version 2007 intègre toutes ces améliorations dont l'efficacité a déjà été testée.

7.1.3 - Modes d'accès

Les droits d'accès aux possibilités de l'application définissent les profils des utilisateurs.

Nous trouvons ainsi :

- Le gestionnaire de l'application, c'est celui qui maintient l'application en ligne et fait qu'elle fonctionne du point de vue informatique. Il est en liaison avec le gestionnaire du site serveur.
- Le modérateur est l'homme de paix du domaine. C'est aussi le chef de projet qui doit trancher les discussions et les débats et apporter son savoir pour faire avancer le projet.
- L'expert ontologique est un expert du domaine qui connaît le maniement de Mind Manager, et qui grâce à cet outil met en forme, développe l'ontologie et transforme les fragments de texte en granules.
- Le chercheur de connaissance est un expert du domaine, qui connaît un traitement de texte, couper/coller, et qui trouve dans un texte les fragments de texte contenant un granule potentiel, une forme si...alors. Le texte travaillé devient une suite de fragments de texte. Tous les experts d'un domaine sont potentiellement des chercheurs de connaissance. Ce sont aussi les utilisateurs de la base de connaissances qu'ils ont contribué à créer.
- Le valideur est un expert du domaine chargé de relire les granules, et s'il les comprend, il les valide. Dans Ramcesh la notion de validation est importante et le valideur ne peut pas valider ses propres granules.
- Le visiteur peut naviguer dans l'ontologie et la base de granule. Il n'a accès aux éléments que un par un.

7.2 - Développement de la base de connaissance

Nous avons montré les raisons qui font que le monde de la géotechnique, au sens large, nécessite une approche originale avec un rôle important des experts.

Le fragment de texte permettant de construire le granule est identifié par la référence du document dont il provient.

Le titre, l'auteur, l'année, la source et la pagination doivent être saisis dans cet ordre et séparés par des virgules.

La pagination est de la forme pp:xx-xx (exemple : pp12-15).

Cette zone permet de saisir le nom du granule ainsi que l'interprétation du fragment de texte.

La première ligne contient le nom du granule suivant le caractère ";", la ligne suivante contient l'interprétation du fragment ou bien une traduction.

Cette liste contient les thèmes de granule. L'expert doit en choisir un parmi les thèmes de base auxquels tout le monde a accès et ses thèmes privés auxquels lui seul a accès.

La liste des concepts est générée automatiquement à partir du fragment de texte. L'expert peut rajouter des concepts en les séparant par des ";". C'est à partir de cette liste que le voisinage fonctionne.

Les phrases du granule sont identifiées par des identifiants de type Px. Ils permettront de saisir des contraintes booléennes entre les phrases.

Les thèmes et les prédicats des phrases de prémisse et de conclusion sont identifiés par des identifiants à deux lettres (AA -> ZZ). Ils permettront de saisir les contraintes de l'univers.

Les contraintes entre phrases s'écrivent en deux parties séparées par un point virgule. La première partie concerne les phrases de prémisse et la deuxième les phrases de conclusion. Les opérateurs booléens autorisés sont ET, OU, SAUF. On peut également utiliser des parenthèses pour la saisie des contraintes. Dans l'exemple ci-contre, la contrainte se lit : Si P1 OU P2 est réalisée, Alors il faut appliquer P3 ET P4.

Les contraintes de l'univers sont des contraintes mathématiques entre les identifiants de concepts.

Les opérateurs utilisés sont =, >, >=, <, <= et ont peut les utiliser en chaîne : AA=AB=AC>4; En fin de chaîne on peut finir par une contrainte numérique et dans tous les cas on doit finir par un ".".

Figure 2: Interface pour la création d'un granule

Important ne veut pas dire long et fastidieux. C'est d'ailleurs le point fort de Ramcesh de ne requérir des experts qu'une courte implication, qui de plus, peut être éparpillée dans le temps et l'espace.

Ce sont les experts du domaine qui contribuent au développement de la base de connaissance de Ramcesh qui comporte deux bases de données, celle qui contient l'ontologie et celle qui contient les granules.

Les paragraphes suivants décrivent ce développement qui comprend le choix des textes fait par un collège d'experts, la recherche des fragments de textes porteurs

de connaissance et leur traitement pour augmenter la base de connaissance de Ramcesh.

La base sera-t-elle complète un jour ? Sans doute jamais car le monde de la géotechnique est très productif et de nouvelles connaissances apparaissent tous les jours. Cependant, le processus cumulatif de Ramcesh permet d'approcher cette complétude grâce à une collaboration partagée du travail de traitement des textes. On peut imaginer, un jour, traiter les textes dès leur parution comme aujourd'hui les services documentations traitent l'existence des articles des revues.

7.2.1 - Le choix des textes

Le choix des textes à traiter est fonction de l'avancement de la mise en forme de la connaissance du domaine. Au début les textes choisis seront des textes riches en connaissances comme les cours, les recommandations et les règlements. Pour ces textes un travail en profondeur de formalisation a déjà été fait. Lorsque la base de connaissance sera importante on traitera des textes plus spécifiques comme les articles de congrès ou de revues. La collaboration avec un service documentaire est alors naturelle.

7.2.2 - La recherche des fragments de texte

Traiter un texte consiste à reconnaître les formes d'écriture contenant une relation de causalité. Si (ce contexte) alors (ces conséquences). La forme "si alors" n'est pas toujours explicite, on trouve fréquemment : dans ce cas... il se passe cela, qui est aussi une forme causale. A la fin du traitement on obtient une suite de fragments de texte contenant cette connaissance déductive.

Dans la pratique cette mise en évidence des fragments se fait avec un simple traitement de texte, l'expert élimine toutes les parties du texte ne correspondant pas à une forme causale. Le stylo scanner est aussi un bon outil de sélection pour les textes dont on ne possède qu'une copie papier.

7.2.3 - La mise en granule et l'augmentation de l'ontologie

La transformation de ces fragments de texte en granules, avec mise à jour de l'ontologie, est fortement assistée dans le processus Ramcesh. Ce processus comporte les étapes suivantes :

- Pour tous les fragments de texte

Vérification et complétude si nécessaire de l'ontologie vis à vis des syntagmes contenus dans les fragments de texte.

- Pour chaque fragment de texte

Mise en forme du granule correspondant, puis recherche des granules voisins (comparaison de signatures de granule), alors choix entre l'ajout du granule à la base de connaissance, son rejet avec ou non addition de sa référence dans le granule le plus voisin, ou modification d'un granule voisin pour compléter prémisses ou conclusion.

Cette partie peut être centralisée ou répartie et dans ce dernier cas des échanges de savoir faire entre "granuleurs" sont très souhaitables.

7.3 - Les applications

La première partie, correspondant à la création de la base de connaissance, étant faite ou tout au moins bien commencée, des applications spécifiques peuvent être envisagées. Nous en esquissons quelques unes.

7.3.1 - Encyclopédie

L'application encyclopédique de la base de connaissances correspond à la recherche dans la base de tous les granules

apportant une réponse totale ou partielle à une question.

Chaque granule possède une signature qui est la liste ordonnée par rang dans l'ontologie des syntagmes qui le compose.

A une question posée définie par un seul mot, l'application "Encyclopédie" propose par glissement sémantique (utilisation de la notion de voisinage sémantique) une liste de syntagmes dont l'utilisateur maîtrise l'extension par le choix du voisinage utilisé. Il choisit alors dans cette liste les syntagmes qui précisent sa question.

L'application classe alors les granules par degré de pertinence (ou nombre de syntagmes présents) et affiche les plus pertinents comme réponse à la question.

Le procédé est itératif, piloté par l'utilisateur, assisté par l'outil qui permet de préciser la demande, en indiquant à l'utilisateur des connexions sémantiques auxquelles il n'aurait pas forcément pensé.

Une application directe de cette approche est la recherche automatique de tous les éléments de règlement(s) à appliquer dans un contexte donné, le(s) règlement(s) ayant été mis sous forme de granules.

7.3.2 - Autres applications possibles

- Indexation automatique de documents

Les mots de l'ontologie contenus dans un document sont facilement indentifiables, y compris avec synonymie, et l'arborescence de l'ontologie permet de récupérer des niveaux plus élevés pour une indexation à la fois globale et détaillée. Les syntagmes de l'ontologie peuvent servir de mots clés, mais ordonnés de façon automatique.

- Fouille de texte et remplissage de formulaires

Pour un formulaire donné, l'ontologie supervisant la fouille de texte peut aider à compléter ce formulaire, même numérique. Ce formulaire renseigné peut contenir les données d'un programme de calcul. Cette fouille de texte peut aussi servir à rechercher des données pour des bases de données dont on connaît la difficulté à les faire remplir.

- Vérification de conformité et de cohérence

Un projet peut être décrit sous forme de granules. De même un texte réglementaire. Par simple comparaison des granules la machine peut donc vérifier si le projet est conforme à un règlement (ou à plusieurs règlements). Les incohérences du projet

seront facilement détectées du fait de granules contradictoires et l'utilisateur averti de ces incohérences devra les lever.

- Suivi de chantier et détection de risques

Les fiches de suivi d'un chantier peuvent être systématiquement analysées, dès leur réception, et des avertissements faits si certaines valeurs ne sont plus conformes aux règles, même numériques, explicitées dans les granules. Toute sortie d'une fourchette de valeurs est par exemple, automatiquement signalée.

- Aide à la conception

Aspect encyclopédique, détection d'incohérence, liaison avec codes de calcul permettent d'envisager une application type SIAD, centrée sur la construction d'un type d'ouvrage.

- Aide au dépouillement ou à la rédaction d'appel d'offres

La détection d'incohérence et le remplissage de formulaire à des fins de comparaison sont une première étape dans le dépouillement d'appels d'offres en vue du choix de l'un d'entre eux en fonction de critères qui peuvent être décrits sous forme de granules.

- Mémoire d'un bureau d'expertise ou de contrôle

Au fil des années les expertises s'accumulent et qui peut se souvenir de tous les détails des expertises afin d'assurer une cohérence entre les expertises faites dans des temps différents et par des collaborateurs différents ? L'outil Ramcesh permet le recueil et la recherche de ces détails en utilisant les granules de connaissance qui évitent la redondance et gardent le contexte. Pour un problème donné les nouveaux collaborateurs récupèrent rapidement le passé du bureau d'études. La confidentialité des expertises (judiciaires) est assurée car la référence inscrite dans le granule n'est dans ce cas, qu'un numéro de dossier.

- Base de cas pour raisonnement par analogie

Lors de la réflexion d'un expert, à la recherche d'une solution pour un problème donné, ce dernier utilise son expérience pour retrouver des cas similaires dont les solutions pourront être reprises. La structure du granule permet de stocker des cas, en donnant aux prémisses le rôle d'enregistrer la méta-connaissance du cas. (Faure et al, 2002) La rédaction du granule est alors : Si (éléments de description et intérêt



du cas) alors cas (référence). Le cas peut ainsi être retrouvé par son intérêt technique, mais aussi par sa référence (nom, lieu, etc.). Ramcesh permet donc, dans un même formalisme, de sauvegarder la connaissance et l'expérience.

- *Gestion d'un patrimoine*

Un patrimoine de tunnels, dans l'optique Ramcesh, est un ensemble de cas, augmenté de possibilités de mise à jour et de méthodes d'interrogation prédéfinies. Ramcesh fournit une façon de stocker les données de la famille d'objets grâce à son analyse de documents par un expert qui saura indiquer ce qu'il est important de conserver. La mise à jour des informations nécessitera un module spécifique.

7.3.3 - Intégration des quatre formes de connaissance

Dans le contexte informatique du Web dynamique, Ramcesh permet l'intégration des quatre aspects de la connaissance de l'ingénieur, cités auparavant. Concepts et relations sont définis dans l'ontologie et les granules, des outils de fouille de texte et/ou de données permettent, avec l'aide de l'utilisateur, de remplir des formulaires qui servent de données à des codes de calcul ré-écrits en PHP par exemple. De la connaissance aux calculs, la chaîne de conception est intégrée dans un seul contexte informatique.

7.3.4 - Mesures de l'innovation d'un texte

A partir de l'ontologie, soit complète, soit partielle, on peut analyser un texte de la façon suivante pour avoir une évaluation de son degré d'innovation, ce qu'il apporte de nouveau dans le contexte actuel. Ces mesures sont en effet dépendantes de l'ontologie, et comme cette dernière évolue dans le temps, ces mesures sont attachées à un état de l'ontologie et à recalculer en fonction des besoins. En cas de comparaisons, il est évident de travailler avec la même ontologie.

A un instant donné, on possède une ontologie définie par ses syntagmes, ses synonymes et son arborescence. On connaît le nombre de syntagmes de l'ontologie (nbsonto), le nombre de syntagmes par niveaux (nbsontor(i)), et les syntagmes finaux (feuilles de l'arbre) (nbsfonto).

Un outil (Mesuronto) permet d'obtenir pour un texte donné :

- le nombre de syntagmes de l'ontologie utilisés dans le texte, en différenciant

syntagmes finaux (nbsf) et non finaux (nbsnf). On connaît aussi le nombre de syntagmes de rang *i* utilisés dans le texte (nbsr(i)). Toutes ces valeurs sont normalisées pour un texte de 1000 mots.

- le texte mis en couleur comme avec l'outil Textonto permet d'évaluer par simple lecture le nombre de syntagmes nouveaux que produit ce texte (nbsnew). On a aussi compté le nombre de mots de plus de trois lettres du texte (nbsmot).

Ces valeurs permettent de définir un certain nombre d'indices à compléter par l'usage.

- Indice ontologique de détail du texte : soit le rapport $nbsf/nbsnf$ soit la formule : $(S(\text{poids}(i) * nbsr(i) / nbsonto(i)) / ? \text{Poids}(i))$ est une pondération du genre (20-i) pour donner aux premiers rangs plus de poids.
- Indice ontologique de pertinence du sujet : $(nbsf + nbsnf) / nbsonto$
- Indice ontologique simple de nouveauté : $nbsnew / nbsf$
- Indice cognitif simple de nouveauté : $nbgnew / nbg(\text{signature})$ (g pour granule de connaissance)

La signature ontologique du texte (ou son indexation) est la liste ordonnée par rang des mots de l'ontologie, avec (ou non) élimination des syntagmes feuilles.

Exemples pratiques d'application

- Une société savante locale doit sélectionner pour une conférence mondiale un certain nombre de papiers produits en nombre trop important par sa communauté. Les choix sont en général guidés par la pertinence avec les thèmes proposés et la nouveauté des idées du texte. Les différents indices ontologiques faciliteront le choix, et la recherche des granules dans chaque texte et la comparaison avec les granules de la base de connaissance garantira l'aspect innovant demandé.
- Un comité scientifique doit répartir un grand nombre de papiers dans différentes sessions. La signature ontologique du texte sera une aide précieuse.

7.4 - Mise en œuvre de Ramcesh

7.4.1 - Cas type de mise en œuvre

Pour orienter le lecteur, nous décrivons ici un scénario de mise en œuvre de Ramcesh au sein d'une entreprise ou d'un laboratoire.

Tout commence par une décision directrice. Il est en effet reconnu que tout projet de recueil de connaissance ne peut se développer sainement que si la décision "vient d'en haut". Une réunion de lancement du projet suit cette décision. Au cours de cette réunion, un domaine technique est défini et délimité. Cela revient à choisir un ensemble de textes à traiter. Ces textes sur le domaine choisi doivent contenir une grande partie de la connaissance du domaine (note de cours, normes, recommandations, documents de synthèse etc.). Cette liste de textes est confiée à un chef de projet.

Le chef de projet répartit alors les textes, voire parties de texte, entre gens du domaine, dont le rôle est de lire les textes pour y retrouver les formes déductives (si...alors, dans ce cas...on a, etc...) que l'on appellera fragments de texte. Quand tous les textes sont devenus des ensembles de fragments de texte, le chef de projet les confie à l'expert ontologique, qui avec l'aide des outils informatiques développés lors du projet Ramcesh, augmente l'ontologie et transforme en granule de connaissance les fragments de texte. Les participants au projet interrogent la base de connaissance (mode encyclopédique) pour évaluer la complétude de cette base. D'autres textes peuvent alors être traités, soit de façon épisodique soit de façon systématique suivant le choix de la stratégie de management des connaissances ou suivant les services rendus si une application a été développée.

7.4.2 - Fonctions de l'interface

L'interface accessible en ligne (avec mots de passe) comporte, pour l'instant, les fonctions suivantes :

- Gestion des comptes, Gestion des textes (indexation, suivi)
- Gestion individuelle de son compte (mot de passe), statistique
- Accès à l'information élémentaire (granule et syntagme), ajout de synonymes, modification définition, traduction
- Ajout d'un syntagme, ajout d'un granule, validation d'un granule
- Traitement global de l'ontologie (modifications branches, extraction de listes)
- Traitement de 'paquets' de fragments de texte
- Interrogation encyclopédique de la base de connaissance

8 - Conclusion et perspectives

8.1 - Conclusion

Un jeune stagiaire impliqué dans la réalisation d'un prototype de Ramcesh écrivait en guise de conclusion, que le travail sur Ramcesh était un exercice salutaire ; citons-le :

L'exercice de décrire l'univers, de cartographier la connaissance, d'écrire ce que probablement jamais personne n'avait encore écrit, de choisir et utiliser un mode de représentation plus porteur, ..., tous ces exercices, initiés avec RAMCESH, sont salutaires et à plus d'un titre :

- c'est l'occasion de rafraîchir les esprits, les mémoires, c'est le moment de la " remise à plat ", on lève des ambiguïtés qui végétaient, on se met d'accord, on décide une fois pour toute,
- cela donne le moyen de se repérer plus rapidement et aisément dans l'organisation (surtout pour les nouveaux collaborateurs),

- l'appropriation du métier est plus rapide, plus efficace, le transfert de connaissances et de compétences est favorisé, la communication améliorée,
- on valorise le domaine d'expertise de chacun et donc, de chaque individu. Présenter le périmètre de compétence de l'un éveille l'intérêt des autres et permet l'enrichissement mutuel,
- on pérennise le capital,
- on gagne en réactivité, en productivité (on se comprend plus rapidement), on gagne en qualité (on se comprend mieux, et on limite les conséquences d'une incompréhension),
- on crée de la connaissance (par juxtaposition de connaissances), on optimise des éléments d'organisation. (fin de citation).

Ce tableau fait avec l'enthousiasme de la jeunesse se précise chaque jour de plus en plus, les premières applications le confirment.

8.2 - Perspectives

Le système Ramcesh ouvre la voie à une réelle mise en forme de la connaissance, il peut s'adapter aux petites et grandes entreprises et toutes communautés. Il est facile à mettre en œuvre et procure très vite des retombées. La recherche de références est très simple et immédiate, le but original est atteint.

Le CETu et la société Solem mettent en œuvre cet outil pour le génie civil des tunnels et les problèmes de corrosion dans le bâtiment. L'ontologie a dépassé les 5000 syntagmes et bientôt 2000 granules garniront la base de donnée granules.

Pour de larges domaines une transposition avec des moyens informatiques plus puissants pour les bases de données sera sans doute nécessaire, mais le passage d'un système à l'autre sera transparent pour l'utilisateur.

BIBLIOGRAPHIE ●●●●●

Clark P., Porter B., Building concept representations from reusable components, Proceedings AAAI 1997

Ermine J.L., 2003, La gestion des connaissances, Hermès ed., 166p,

Faure N., 2006, Représentation des connaissances géotechniques, Actes Inforsid 2006, Hammamet

Faure N., 2007, Un système d'aide à la décision en géotechnique, Thèse Lyon 3, à paraître.

Faure N., 2006, The use of knowledge management in the management of tunnels and tunnel projects. Proc. Int. Conf. ITA/AITES, Seoul, Korea

Faure R.M., Mascarelli D., Vaunat J., Leroueil S., Tavenas F. ; 1992, Present state of development of XPENT, expert system for slopes stability problems.; 6th Int. Congress on Landslides Bell editor Christchurch. p1671-1678

Faure R.M., 1999, Data-bases and the management of landslides. Int. Symp. on Landslides. Shikoku (Japan). IS Shikoku'99, p1317-1330

Faure R.M., Locat J., Thimus J.F., Picarelli L. 2001, Base de données internationale sur les glissements de terrain, Revue Française de Géotechnique. Numéro 95 et 96 : Mouvements de terrain, pp 183-187.

Faure R.M., Robert A.A., Thimus J.F., W. Schubert, Bourgeois E., Bois A.P., 2002, KBT-Base de connaissances internationales les tunnels, TOS n°170, pp94-98

Grüber T., Toward Principles for the Design of Ontologies, In Formal Ontology in Conceptual Analysis and Knowledge Representation, edited by Nicola Guarino and Roberto Poli, Kluwer Academic Publishers. Substantial revision of paper presented at the International Workshop on Formal Ontology, March, 1993, Padova, Italy. Available as Technical Report KSL 93-04, Knowledge Systems Laboratory, Stanford University

Guarino N. (1998), Formal ontology and information systems. Amended version of a paper appeared in N. Guarino (ed.), Volume 46 Frontiers in Artificial Intelligence and Applications. IOS Press

Magnan J.P., (1992) CESSOL : Bilan du Développement d'un Système-expert, Geotechnique et Informatique, Proc. Int. Conf. on Geotechnics and Computers, Presses de l'École Nationale de Ponts et Chaussées, Paris.

Magnan J.P., Sève G., Pouget P. 1998 Quelques spécificités de l'analyse de risque pour les ouvrages de géotechnique ; The Geotechnics of Hard Soils - Soft Rocks. Balkem, Rotterdam, Evangelista & Picarelli pp 1109-1116 (in French)

Magnan J.P. (2002), L'organisation du travail en géotechnique : développement, normalisation et artisanat, Lettre de la Géotechnique 26 et 27, Société Internationale de la Mécanique des Sols et de la Géotechnique

Mascarelli D., Faure R.M., Kastner R., (1992) Anatomie d'un projet à base de connaissances, XPENT, système de travail en ingénierie des pentes, Actes du colloque Géotechnique et Informatique, ENPC, Paris.

Mascarelli D., (1994) Ingénierie des pentes instables : approche orientée modélisation de la connaissance, Thèse de doctorat. INSA Lyon.

Sowa, Knowledge Representation, Brooks/Cole, 2001



ANNEXE 1 : Exemple de granules

Granule 1 : relatif à l'étanchéité

Fragment de texte et référence

❑ Rec. AFTES GT9, collectif, 2001, TOS 168, pp134

❑ Le bituthène série 4000 peut être utilisé en étanchéité extrados si la température est supérieure à 5° C et si la protection est au moins équivalente à la classe 2 (voir article 7-4-2-3 du titre 3), sur dalle horizontale et sur paroi verticale sans charge d'eau.

Expression du granule, tous les syntagmes appartiennent à l'ontologie

Si extrados et dalle ou paroi et température et charge d'eau et protection-(caractéristique classe de protection, référence article 7-4-2-3 du titre 3)

Alors (possibilité) bituthène série 4000

Modèle d'univers : extrados=dalle, extrados=paroi, (température>5°, unité-celsius), charge d'eau=0, classe de protection=>2

Les conditions exprimées dans le texte se retrouvent dans le modèle d'univers.

Granule 2 : relatif aux règles parasismiques

Fragment de texte et référence

❑ Recommandations pour la conception et la protection parasismique des ouvrages souterrains, collectif, 2001, TOS 167, pp250

❑ si un sable, un sable vasard, un silt possède les caractéristiques suivantes, degré de saturation voisin de 100%, granulométrie assez uniforme de coefficient de Hazen inférieur à 15, de diamètre à 50% compris entre 0.05 et 1.5 mm, sous un état de contrainte effective verticale inférieur à 0.20 MPa en zone 1a et 1b, 0.25 MPa en zone 2, 0.30 MPa en zone 3, alors risque de liquéfaction.

Expression du granule, tous les syntagmes appartiennent à l'ontologie.

Si sable ou sable vasard ou silt et degré de saturation Sr et coefficient de Hazen Cu et diamètre à 50 % et contrainte effective verticale et zone sismique 1a et 1b

Alors risque liquéfaction du sol.

Modèle d'univers : Sr>90 %, Cu<15, D50>0.05 unité mm, D50<1.5 unité mm, Sigma'v<0.20 unité MPa

Ce granule sera triplé pour exprimer les conditions zone 2 et zone 3.

ANNEXE 2 : Recherche de fragments de textes pour faire des granules

Dans cette annexe nous présentons le travail de recherche systématique fait par un expert sur un document.

Nous avons pris le document : "Recommandations de l'AFTES sur l'étanchéité et le drainage des ouvrages souterrains", pour une mise en granule systématique des connaissances qu'il contient.

En lisant attentivement le document, nous avons référencé deux types d'informations utiles pour la Base de Connaissance RAMCESH.

1 - Les définitions

Ces définitions serviront à augmenter l'ontologie et seront introduites à côté de chaque syntagme de l'ontologie. L'utilisateur aura donc en ligne, par choix d'un syntagme et affichage de sa définition, la définition complète du syntagme.

Liste des 10 premières définitions trouvées dans le document :

- **Etanchéité** : Fonction qui fait qu'un produit ou un ensemble de produits s'oppose au franchissement par un liquide tel que l'eau.

- **Drainage** : Captage ponctuel ou surfacique d'arrivées d'eau dans un ouvrage souterrain. Cette eau est ensuite collectée et rejetée à l'extérieur par le réseau d'assainissement de l'ouvrage. Le drainage peut être provisoire, pour permettre par exemple la mise en place dans de bonnes conditions du complexe d'étanchéité, ou définitif et contribue de fait à la fonction étanchéité de l'ouvrage.

- **Dispositif d'Etanchéité Géomembrane (D.E.G.)** : Complexe étanchéité extrados indépendant composé de plusieurs matériaux qui remplissent chacun une fonction bien précise :

• écran de protection contre le poinçonnement placé contre le support et sous la géomembrane d'étanchéité,

• géomembrane d'étanchéité synthétique translucide,

• écran de protection mécanique placé sur la géomembrane lorsque le revêtement définitif est armé. Cet écran est également mis sous forme de bande, de 1,00 à 1,50 m de largeur au droit de chaque masque de bétonnage de la voûte.

- **Ecran de protection contre le poinçonnement statique** : Placé contre le support et sous la géomembrane d'étanchéité. Cet

écran de protection est parfois appelé protection "extrados" dans le cas d'un D.E.G. de tunnel ou de tranchée couverte réalisé avec emprises. Cet écran est essentiellement constitué par un géotextile non tissé. Celui-ci est parfois complexé avec un film de PVC.P ou de polyéthylène lui permettant, en cas de débit d'arrivée d'eau importante de constituer une première "barrière étanche" autorisant ainsi les opérations de thermosoudure des lès de géomembrane. Les caractéristiques physico mécaniques de cet écran de protection sont spécifiées à l'article 7.4.2.3. du fascicule 67 - titre III du CCTG.

- **Ecran de protection contre le poinçonnement dynamique** : Placé sur la géomembrane d'étanchéité, cet écran de protection est parfois appelé protection "intradados" dans le cas d'un D.E.G. de tunnel (avec armatures) ou de tranchée couverte réalisée avec emprises. Cet écran est obligatoirement constitué d'une membrane en matière synthétique (PVC.P ou polyéthylène). Les caractéristiques physico mécaniques de cet écran de protection sont spécifiées également à l'article 7.4.2.3, et par le n° 7 de l'annexe n°4 du fascicule 67 du titre III du CCTG.

- **Géoespaceurs** : Structure poly constituée de feuilles thermo formées, ou monofilament ou de toute autre structure, ayant pour fonction de créer un fort indice de vide facilitant l'écoulement de l'eau, soit en phase provisoire, soit en phase définitive.

- **Géocomposite de protection mécanique** : Association d'un géotextile non tissé et d'une feuille synthétique de faible épaisseur ; généralement en PVC.P ou en polyéthylène. La face constituée du géotextile est appliquée contre le support.

- **Ouvrage réalisé avec emprises** : Ouvrage réalisé à l'intérieur d'un soutènement (paroi moulée, paroi berlinoise, etc...) qui dans ce cas servira de support aux dispositifs d'étanchéité ou de drainage.

Ouvrage réalisé sans emprises : Ouvrage réalisé sans soutènement (fouille talutée, etc...) dans ce cas les dispositifs d'étanchéité et de drainage sont mis en œuvre après bétonnage de l'ouvrage, et directement sur celui-ci.

- **Géocomposite de drainage** : Association d'un géoespaceur avec un ou plusieurs géotextiles ayant une fonction filtration. Ces géocomposites de drainage sont généralement utilisés en drainage définitif



des piédroits de tranchées couvertes réalisés sans emprises (sans soutènement).

2 - La connaissance

Elle est représentée par les formes déductives, qui contiennent une implication de recommandation due au type de document. La même lecture attentive permet de mettre en évidence les fragments de texte (partie du texte du document) qui correspondent (dans notre cas) à des recommandations. Ces fragments de textes sont stockés dans un fichier qui les différencie par les mots-clés REFERENCE et FRAGMENT.

La REFERENCE comporte un titre, un auteur, une source ou type de document, une date et un nombre de pages. (tableau 1)

Le fichier obtenu est un fichier texte dont une partie est éditée ci-après (tableau 2) (17 premiers fragments sur plus de 200)

Jusqu'à ce stade l'outil de travail est un éditeur de texte. Dans notre cas un stylo-canner nous a aidés à ressaisir le texte des fragments de texte.

Ce fichier ainsi construit est alors appelé par l'interface de Ramcesh (voir figure 2) qui permet un traitement séquentiel, fragment de texte par fragment de texte. A chaque fragment de texte l'interface analyse le fragment et fournit la liste de syntagmes qui le compose. Avec cette liste l'utilisateur bâtit avec la souris le granule qui représente ce fragment de texte (voir mode d'emploi de Ramcesh).

REFERENCE

(titre) AFTES, recommandations relatives à : L'étanchéité et Le drainage des ouvrages souterrains Version 1 – approuvée par le Comité Technique du 3 mai 2000 (TOS 159) préparation et réception des supports de tranchées couvertes recevant un d.e.g.

(auteur) collectif

(source) recommandations de l'AFTES sur l'étanchéité et le drainage des ouvrages souterrains.

(année) 2000, nombre de pages : 19

Tableau 1 - Référence

Tableau 2 - Extrait de fichier texte

FRAGMENT

Si le support est de type paroi au coulis, alors il devra avoir la résistance mécanique requise par le CCTP avant la mise en œuvre du D.E.G.

FRAGMENT

Si les surfaces de coulis, dans un support de type paroi au coulis, sont superficiellement déshydratées et de faible cohésion, alors elles seront éliminées.

FRAGMENT

Si la résistance mécanique du coulis, dans un support de type paroi au coulis, change, alors il faut adapter la longueur des clous des systèmes de fixation intermédiaire du D.E.G.

FRAGMENT

Si il existe des désaffleurements entre les profilés métalliques et la paroi au coulis, dans un support de type paroi au coulis, alors ils ne devront pas excéder 5 cm.

FRAGMENT

Si les supports sont de type berlinoises et palplanches avec remplissage des ondes avec du polystyrène, alors le polystyrène doit épouser parfaitement l'onde de la palplanche.

FRAGMENT

Si les supports sont de type berlinoises et palplanches avec remplissage des ondes avec du polystyrène, alors le polystyrène doit être de classe EM selon la norme NFT 56 20I.

FRAGMENT

Si les supports sont de type berlinoises et palplanches avec remplissage des ondes avec du polystyrène, alors le polystyrène doit avoir une résistance minimale à la compression supérieure ou égale à 90 kPa.

FRAGMENT

Si il existe un support de tout type pour recevoir un D.E.G., alors devra être réalisé un solin d'angle à la jonction radier – support vertical.

FRAGMENT

Si le solin d'angle à la jonction radier – support vertical, est fait en mortier, alors il devra être arrondi avec un rayon supérieur à 5 cm.

FRAGMENT

Si il existe un support de type paroi préfabriqué et contre bajoyer béton, alors les désaffleurements entre parois ou panneaux ne devront pas excéder 5 cm.

FRAGMENT

Si les désaffleurements entre parois ou panneaux d'un support de type paroi préfabriquée et contre bajoyer béton, sont supérieures à 5 cm, alors les joints entre parois ou panneaux devront être chanfreinés à 45°.

FRAGMENT

Si les joints parois ou panneaux doivent être chanfreinés à 45°, alors ils devront être faits avec du mortier.

FRAGMENT

Si les joints entre parois ou panneaux sont chanfreinés à 45° avec autre type de matériaux que le mortier, alors ces matériaux ne devront pas être compressibles.

FRAGMENT

Si le support est de type paroi au coulis, alors la vérification des désaffleurements maximums des profilés métalliques en surface courante devra être réalisée.

FRAGMENT

Si le support est de type paroi au coulis, alors la vérification de la cohésion superficielle du coulis devra être réalisée.

FRAGMENT

Si le support est de type berlinoise, alors la vérification des désaffleurements entre profilés métalliques et plaques, panneaux, blocs de polystyrène en surface courante devra être réalisée.

FRAGMENT

Si le support est de type palplanches avec remplissage en polystyrène, alors la vérification des désaffleurements entre profilés métalliques et plaques, panneaux, blocs de polystyrène en surface courante devra être réalisée.