

# SERVEUR D'ONTOLOGIE POUR LA QUALITE DES SERVICES E-GOUVERNEMENT

Ludmila MANGELINCK\*, Cristelle ROUX\*

[ludmila.langlinck@synergiums.com](mailto:ludmila.langlinck@synergiums.com), [cristelle.roux@synergiums.com](mailto:cristelle.roux@synergiums.com)

\* Synergiums S.A. – 2, rue Pletzer – L-8088 Helfent-Bertrange (Luxembourg)

## Mots clés :

Ontologie, usage de l'information, citoyenneté, information géographique, e-gouvernement, mémoire d'entreprise, ingénierie des connaissances, modélisation des connaissances.

## Keywords:

Ontology, information usage, citizenship, geographical information, e-government, knowledge engineering, knowledge modeling.

## Palabras clave :

Ciudadanía, Sistema información geográfica, Gestión conocimiento

## Résumé

Afin de faciliter l'accès à l'information par le citoyen, dans le cadre d'e-gouvernement, et pour répondre aux besoins informationnels de tous les utilisateurs potentiels, nous proposons une approche basée sur la création d'une ontologie d'usage de l'information. Cette ontologie a été intégrée au sein d'un système de service d'e-gouvernement de manière à, d'une part, comprendre le besoin réel de l'utilisateur et, d'autre part, à déduire l'information requise pour satisfaire des requêtes citoyennes.

La méthodologie que nous avons développée, ainsi que les outils bibliométriques et sémantiques utilisés ont été examinés dans un cas pratique : satisfaire la plupart des besoins de l'utilisateur dans le domaine des infrastructures sportives en Région wallonne (Belgique) et plus spécifiquement ayant trait à l'information spatialisée (ou géographique).

Dans ce document, nous présentons l'approche détaillée, ensuite la méthodologie appliquée à la Région wallonne, ainsi que les complémentarités entre cette méthodologie et l'approche UML de la modélisation des besoins. Notre conclusion présente les modalités d'intégration possibles de cette méthodologie en temps que couche au sein d'un système d'e-gouvernement et précise les difficultés rencontrées pour appliquer cette approche.

# 1. Problématique – Contexte

Faciliter l'accès à l'information concernant les lignes de vie citoyenne élaborées dans un cadre de e-gouvernement est, à l'heure actuelle, une des priorités des gouvernements [2][13]. Ceci est particulièrement vrai dans le cadre de l'information géographique si l'on veut rentabiliser de manière non négligeable une information coûteuse à produire et à maintenir.

Fournir un accès à l'information à divers utilisateurs pose trois problèmes majeurs :

- ⇒ L'adéquation du besoin d'un utilisateur avec l'information disponible.
- ⇒ L'identification et la localisation des services qui pourraient répondre à son besoin.
- ⇒ La méconnaissance de la façon d'interroger le service adéquat.

La publication des métadonnées de découverte[4] ainsi que l'accès à des registres de référence pour le stockage des services, des données et des métadonnées solutionnent partiellement ces problèmes[2][5]. Ce qui existe actuellement permet à l'utilisateur de rechercher, de façon centralisée, des données et des services qui peuvent correspondre à son besoin.

Toutefois, pour exprimer un besoin, l'usager utilise ses propres termes qui ne sont pas forcément ceux présents au sein des listes de vocabulaires contrôlées et utilisées par les producteurs d'information tels que les administrations publiques. De plus, l'accès à ces listes ne lui permet pas de trouver, voire de déclencher ou de combiner les services utiles à la réalisation de son besoin.

Cet objectif peut être atteint en intégrant entre l'utilisateur et les services, des ontologies sur les usages de l'information.

Nous nous proposons de décrire comment ces ontologies sont utilisées au sein d'un serveur d'ontologies pour assurer la qualité des services de gouvernement puis nous décrirons comment nous avons construit ces ontologies. Pour ces descriptions, nous verrons tout d'abord l'aspect méthodologique. Nous présenterons (ou exposerons) ensuite la méthodologie appliquée dans le cadre du projet SAFIR, un projet du 6<sup>ème</sup> Programme Cadre de la Commission Européenne visant à donner aux utilisateurs de e-gouvernement d'autres moyens d'accéder aux services [19].

## 1 Méthodologie

### 1.1 Notre approche

Selon notre expérience et l'état de l'art que nous constituons régulièrement, nous avons défini une approche basée sur une ontologie des usages de l'information qui doit :

- ⇒ répertorier le vocabulaire des utilisateurs du domaine dans les usages de l'information. Par exemple, l'ontologie d'usage doit associer le terme « gymnase » au terme « salle de sport » ou encore plus généralement associe une « prise d'eau pour canadair » à un « lac ».
- ⇒ pour chaque usage, définir le contexte d'utilisation des données et l'information à associer aux informations géographiques. Dans l'exemple précédent, cela doit permettre d'associer à un « lac » une information de navigation aérienne dans le cadre de la « prise d'eau pour les pilotes de canadiens ».
- ⇒ pour chaque usage, établir les phrases employées par l'utilisateur pour formuler sa requête.
- ⇒ Enfin, intégrer les ontologies au sein d'un système de services de e-gouvernement afin, d'une part d'activer les services ou des combinaisons de services et, d'autre part, de déduire l'information requise et manquante au système pour répondre à une question posée. En effet, à la phrase « Je suis pompier, je cherche la piscine de Bastogne », comme un pompier en intervention a besoin d'un itinéraire pour accéder à la piscine, le système doit répondre « quelle est votre position », ce qui correspond à une information requise par le système pour calculer un itinéraire. Cela implique une connaissance des tâches de la localisation elle-même.

Nous dissocions alors (Fig. 1) :

- ⇒ **Ontologie du domaine** : qui reprend tous les concepts de domaine et les organise dans une taxonomie de concepts ;
- ⇒ **Ontologie des tâches** : qui reprend la façon d'accéder aux instances d'information. Par exemple, une ontologie des tâches de localisation représente la nécessité qu'une tâche d'itinéraire doit faire intervenir la référence sur la position de la personne et celle de l'objectif.
- ⇒ **Ontologie d'usage** : qui combine les deux ontologies précédentes dans le but de préparer les usages de l'information.
- ⇒ **Ontologie de service** : qui relie les usages et les services de e-gouvernement.

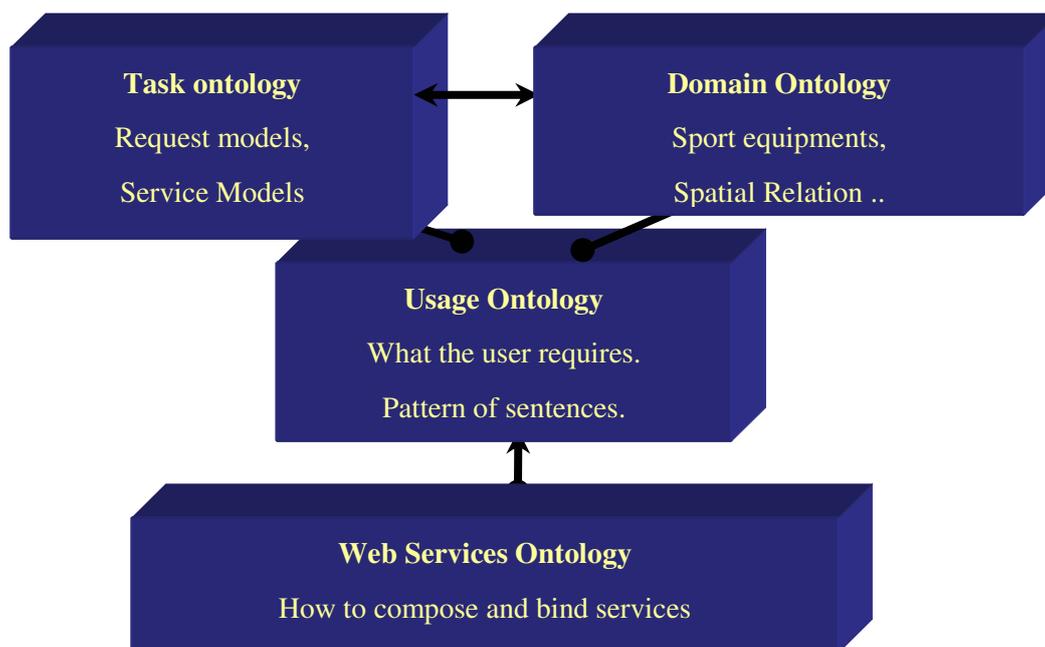


Figure 1 : Interaction des ontologies

Enfin pour prendre en compte le fait que chaque usager utilise ses propres termes pour qualifier un concept et la problématique du multilinguisme, nous centrons l'ontologie sur une notion de concept séparé des termes utilisés : le concept est une abstraction du monde réel partagée par un ensemble de personnes. Par exemple, le terme « club de sport » sera représenté dans deux concepts distincts : d'une part un concept représentant l'entité morale qui s'occupe d'une activité sportive et d'autre part un concept représentant le lieu où se rencontre ses membres.

## 1.2 L'analyse

Tous les usages de l'information géographique n'étant généralement pas connus de ses producteurs, nous avons choisi d'utiliser des méthodes de traitements sémantiques et bibliométriques de documents issus d'Internet afin d'identifier les usages potentiels[20]. Pour cela, nous avons amélioré une méthodologie que nous avons employée antérieurement [21]. La méthode a été initiée, en Région wallonne, dans le cadre du projet INFRASIG pour la définition de la politique des prix de l'information géographique.

Nous avons procédé de la manière suivante :

1. Constitution de différents corpus d'analyse.
2. Traitement des corpus par des analyses bibliométriques et sémantiques qui vont permettre d'une part la construction d'une ontologie de domaine et d'autre part l'extraction des usages de l'information.

3. Confrontation des usages et des concepts aux informations disponibles au sein des administrations et mise en évidence des liens avec les services d'e-gouvernement.
4. Représentation des différentes requêtes attenantes aux usages dans le système et des services pouvant répondre à ces requêtes.

Nous détaillerons dans cette partie les deux premiers aspects qui sont plus en lien avec les méthodologies de veille. Les deux dernières étapes constituent la validation et l'utilisation des résultats de la recherche des usages.

### 1.2.1 Constitution de corpus

Dans notre méthodologie, la constitution de corpus est une étape fondamentale : les corpus sont nécessaires pour :

- ⇒ la constitution d'une ontologie de domaine, que nous appellerons dans la suite « **corpus de référence** »,
- ⇒ la détection d'usages de l'information, que nous appellerons dans la suite « **corpus d'analyse** ».

Les deux utilisations différentes des corpus impliquent toutefois des contraintes très différentes sur leur contenu. La constitution d'ontologies implique une adéquation étroite entre les documents du corpus et le domaine concerné. En effet, les traitements sémantiques effectués sur ces documents révèlent les termes significatifs employés dans les documents. D'un côté, la détection des usages potentiels de l'information nécessite une grande quantité d'information afin de couvrir un maximum d'usages. De l'autre côté, si des documents impliqués ne sont pas relatifs au domaine, les ontologies créées seront biaisées. Les corpus ainsi créés sont extrêmement volumineux.

L'étape de constitution de corpus comprend l'identification des sources, la collecte des documents ainsi que des prétraitements.

Les sources concernées sont essentiellement issues d'Internet. En effet, aujourd'hui on postule qu'Internet peut être considéré comme représentatif de l'expression des citoyens. Mais plus encore, les publications officielles de l'administration et des associations liées aux administrations sur ce média nous permettent d'aborder également le vocabulaire des classes d'acteurs « administration ». En constituant ce corpus, nous accédons ainsi à deux des objectifs liés à notre approche : répertorier (a) le vocabulaire des utilisateurs et (b) les usages de l'information sur l'ensemble des classes d'acteurs « citoyens ». Concernant les corpus de référence, les sources sont identifiées par le biais d'entretiens entre les experts du domaine. En revanche, concernant les corpus d'analyse, les sources Internet sont identifiées uniquement sur base de moteurs de recherche. Les mots clés de la recherche sont définis en collaboration avec les experts.

Les documents sont ensuite collectés à l'aide d'outils d'aspiration. La profondeur de l'aspiration paramétrée est faible pour l'établissement des corpus de référence et forte concernant les corpus d'analyse.

Après l'étape de collecte des documents, les corpus doivent être prétraités afin d'être sûr que leur qualité soit adéquate pour les analyses en aval. Le prétraitement doit établir un équilibre entre :

- ⇒ le **silence** du corpus : l'information non présente dans le corpus mais indispensable au traitement,
- ⇒ et le **bruit** du corpus : l'information présente mais qui n'est pas liée à l'analyse.

Trois techniques peuvent être utilisées pour l'identification du bruit :

1. L'extraction et l'identification des métadonnées[23] des documents, plus particulièrement la source, la date, l'auteur, etc.
2. La sélection des documents contenant au moins un terme lié au domaine. Les documents qui ne contiennent aucun terme lié au domaine sont éliminés du corpus. Un pré-requis de cette technique est que la terminologie du domaine soit complète.
3. L'identification des documents peu relatifs au domaine par le calcul de statistiques d'identification des termes sur les documents.

Nous basons l'élimination du bruit sur les corpus de référence sur la première technique (métadonnées), et celle des corpus d'analyse sur la deuxième (l'absence de termes liés au domaine). La dernière technique ne permet pas réellement d'éliminer le bruit du corpus mais de vérifier l'adéquation du corpus à la terminologie du domaine et donc d'améliorer en parallèle corpus et terminologie. Tout document identifié comme peu relatif au domaine est évalué en deux catégories de façon supervisée :

- ⇒ soit effectivement peu intéressant et devant être éliminé du corpus ;
- ⇒ soit au contraire effectivement lié au domaine. C'est alors la terminologie qui doit être améliorée par une analyse plus profonde du document en processus itératif.

## 1.2.2 Traitements des corpus

Les traitements sur les corpus sont de deux types :

- ⇒ Une extraction terminologique[6][7] qui permet d'initialiser une ontologie de domaine.
- ⇒ Une extraction des usages de l'information, qui permet d'une part une structuration de l'ontologie par la détection des relations entre les concepts, et, d'autre part, de trouver les requêtes des utilisateurs.

Les ontologies de domaine sont initialisées en collectant les thésaurus, les dictionnaires et les classifications de termes disponibles. L'ensemble doit être converti en OWL[25], le langage utilisé pour notre projet. Simultanément, des corpus de référence sont collectés.

Une extraction terminologique est ensuite réalisée sur les corpus de référence. Cette extraction doit retrouver les termes spécifiques au domaine mais aussi et surtout des expressions du domaine telle que « art martial ». Dans la suite, une telle expression est nommée « mot composé ». (Une expression est un mot composé si la signification du mot composé est différente de celle des deux termes pris séparément).

L'extraction terminologique permet d'initialiser l'ontologie de domaine. Pour cela, il est nécessaire de vérifier les termes trouvés lors de l'extraction terminologique afin de créer un concept. Pour chaque terme, il convient de décider si le terme trouvé correspond à un concept déjà présent dans l'ontologie ou si un nouveau concept doit être créé. L'étape suivante de détection des usages va permettre de structurer cette ontologie et d'extraire les requêtes des utilisateurs.

Le problème initial consiste à: chercher des usages de l'information par des acteurs particuliers, c'est-à-dire que l'extraction des usages consiste à mettre en évidence des relations entre un acteur et plusieurs éléments du domaine, comme par exemple, des relations entre les termes « nageur », « adresse » et « piscine ». Pour cela, le calcul d'indice de corrélation entre les termes est adapté. Ce calcul d'indices peut-être basé :

- ⇒ soit sur le document : le calcul de l'indice entre deux termes se fera uniquement si les deux termes sont dans le même document ;
- ⇒ soit sur la phrase : les deux termes sont dans la même phrase.

L'analyse se base sur un calcul au niveau de la phrase, car notre expérience précédente nous a montré que les relations trouvées par une analyse sur le document sont trop bruitées.

Les traitements sur les corpus d'analyse sont définis :

- ⇒ Identification des grandes catégories de termes pour lesquelles des croisements potentiels existent. Ces catégories peuvent correspondre à des éléments de haut niveau de l'ontologie ou sont des catégories spécifiques à la recherche de l'usage de l'information géographique.
- ⇒ Identification des croisements potentiels entre ces catégories de termes.
- ⇒ Extraction des phrases qui contiennent au moins un terme d'une ou plusieurs catégories cibles.

- ⇒ Pour chaque phrase relevée, séparation des différents termes en catégories correspondant à des éléments de croisements à effectuer afin de trouver des relations potentielles. Les termes qui ne sont classés dans aucune catégorie sont néanmoins mémorisés: ils permettent de trouver un nouveau vocabulaire. Le lien avec la phrase et le document initial est conservé pour permettre à l'analyste d'étudier le contexte du croisement.
- ⇒ Calcul des indices de corrélation entre les différents termes des catégories pour les croisements identifiés.

Les croisements potentiels sont déduits de la constatation suivante : les usages de l'information géographique sont généralement déduits d'une liste de questions élémentaires [24] :

- ⇒ Où ? : La localisation. Exemple : où se trouve l'Atomium ou le Stade de France?
- ⇒ Qu'y a-t-il là ? : L'adressage. Exemple : quels sites remarquables y a-t-il dans ce pays ?

D'autres questions, classiquement identifiées dans la littérature, sont également utilisables dans le contexte de l'information géographique :

- ⇒ Et si ? : La simulation. Exemple : Si Bruxelles n'existait pas, quels seraient les flux entre les principales villes belges ?
- ⇒ Quand ? : Qu'est-ce qui a changé ? : L'analyse du changement. Exemple : quand a-t-on construit ce pont ?
- ⇒ Comment ? : L'analyse spatiale. Exemple : Comment puis-je me rendre là ? Comment est structurée (géographiquement) cette ville ?

Les trois dernières questions font intervenir des informations et des raisonnements qui ne sont pas disponibles à l'heure actuelle dans les systèmes d'informations géographiques. Nous ne les avons pas considérés dans un premier temps.

Selon le postulat précédent, la recherche de l'usage de l'information géographique nécessite d'ajouter aux termes du domaine:

- ⇒ Des termes spécifiques à la géomatique.
- ⇒ Des termes de localisation naïve : des termes utilisés par des non spécialistes de l'information géographique et qui révèlent des localisations d'objets tel que « où », ou des relations spatiales entre les objets tel que « dans ».
- ⇒ Des éléments géographiques et des éléments de repérage spatial, tels que les villes, les éléments de relief ou d'hydrgraphie, et plus généralement tout élément topographique pouvant apparaître sur des cartes.

Nous avons également besoin d'une classification exhaustive des acteurs.

## 2 Validation

La méthodologie et les outils ont été examinés dans un cas très pratique : atteindre la plupart des besoins de l'utilisateur dans le domaine des infrastructures sportives en Région wallonne (Belgique). Les besoins des usagers ont été divisés en deux groupes:

- ⇒ Les citoyens qui accèdent aux services de gouvernement de chez eux : les cas d'utilisations concernent les recherches d'information.
- ⇒ Les fournisseurs de services de e-gouvernement. Les cas d'utilisation concernent à la fois des besoins en information mais aussi et surtout la gestion des informations. Ce sont eux qui peuvent garantir la qualité des données.

Dans ce projet, nous avons pu mettre en place une double approche de définition des besoins :

- La méthode présentée ici pour couvrir les besoins des classes d'acteurs « citoyen ».

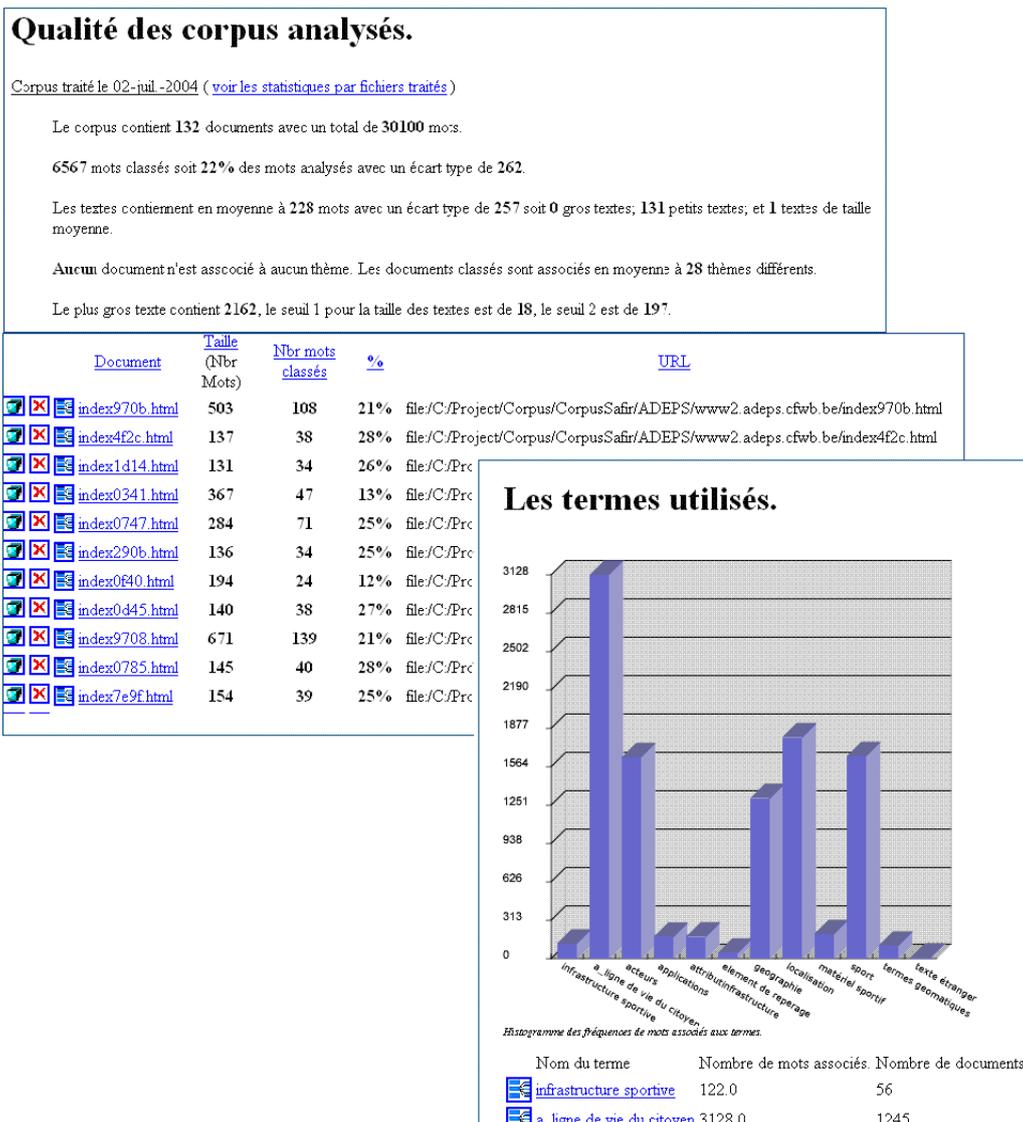
- Une modélisation des besoins par une approche classique UML. Cette modélisation des besoins couvre les besoins du second groupe d'acteur. La méthode employée est la définition de groupes d'utilisateurs. Plusieurs administrations et associations belges concernées par la gestion d'infrastructures sportives ont été invitées à formuler leurs besoins par rapport à l'établissement d'un « cadastre des infrastructures sportives en Région wallonne ».

Dans cette partie, nous précisons l'application de la méthode, les outils spécifiques que nous avons développés, les résultats obtenus ainsi que le système final.

## 2.1 Les outils utilisés

En complément des outils sémantiques et bibliométriques et d'un éditeur d'ontologie [18], nous avons dû développer des fonctionnalités additionnelles sur nos propres outils afin de couvrir l'ensemble des besoins spécifiques à la méthode. En effet, les outils à notre disposition ne nous permettaient pas ou peu de valider la terminologie ni de détecter le silence et le bruit dans les corpus. Pour cela, nous avons donc développé des fonctionnalités de calcul des statistiques d'indexation des corpus à partir de classification de termes.

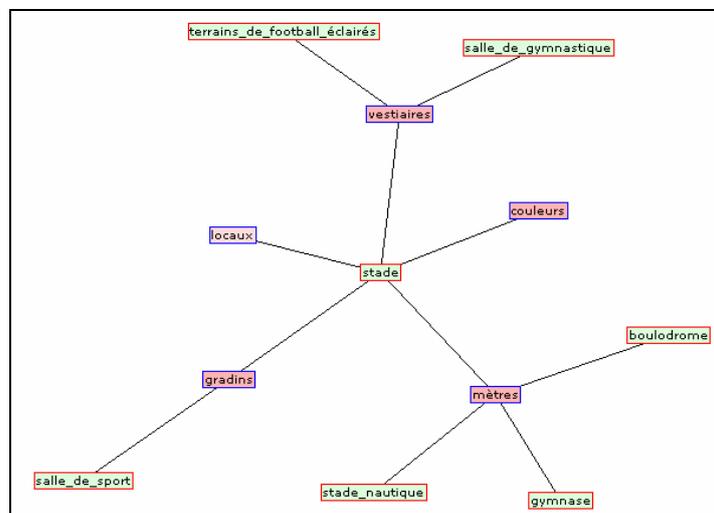
Basé sur un indexeur sémantique, nous analysons les résultats de l'indexation. Les statistiques sont visualisées par trois types d'interface: statistique sur le corpus, statistiques par documents traités et statistiques par groupes de termes.



Pour effectuer l'extraction des usages de l'information, un premier outil de prétraitement des corpus a été développé afin d'extraire les phrases en se basant sur les catégories de termes. Nous avons également développé des fonctionnalités en amont de l'outil XPlor[22]. Basé sur des calculs de cooccurrences de termes, le système propose un ensemble de termes qui semblent être en relation. Pour chaque relation potentielle, il vérifie les alternatives :

- si les termes ne sont pas déjà dans l'ontologie en temps que synonymes,
- s'ils font partie d'un même mot composé,
- s'ils sont en relation de subsomption (relation de spécialisation de concept nommée également « est-un »),
- si un autre type de relation existe entre les deux concepts.

L'opérateur a alors la possibilité d'associer les termes à des concepts ou de créer des relations entre les termes. Une visualisation graphique des relations potentielles est également possible.



Terme 1	Terme 2	Sont synonymes.	Forment un mot composé.	Se subsume.
		Concept	Concept	Concepts associé au termes
stade_français	stade			StadeFranAis>Stade
fronton	stade			
clubs_professionnels	clubs_amateurs			
clubs_omnisports	clubs_sportifs			
patinoire	piscine			
stade_de_france	stade			StadeDeFrance>Stade
terrain	stade			
stade	stade_français			Stade<StadeFranAis
piscine	centre_sportif			
terrains	terrain			
stade_municipal	stade			StadeMunicipal>Stade
gymnases	piscines			

Fréq.	Terme 1	Terme 2	Sont synonymes.	Forment un mot composé.	Se subsume.	Sont en Relation.
			Concept	Concept	Concepts associé au termes	Relation
142	hauteur	stade				KB 609068 Slot 41
128	longueur	stade				KB 609068 Slot 41
122	km	circuit				
116	km	parcours				
62	couleurs	stade				
60	équipement	piscines				KB 486945 Slot 20 KB 609068 Slot 2

## 2.2 Les résultats obtenus

L'ontologie de domaine comprend les concepts liés aux infrastructures sportives et installations sportives, au matériel sportif, aux activités sportives, aux acteurs, aux termes géographiques et à tous les besoins informationnels liés aux installations sportives: adresse, dimension, salles spécifiques telles que des salles de réunions, etc.. Elle comprend environ 10000 concepts.

Les besoins en information identifiés lors de cette phase d'extraction des usages de l'information peuvent être classés dans les catégories suivantes :

- ⇒ Accéder à une infrastructure sportive.
- ⇒ Assurer la sécurité des infrastructures sportives.
- ⇒ Pratiquer une activité sportive.
- ⇒ Pratiquer une activité libre.
- ⇒ Pratiquer via un club ou une association lié à une infrastructure sportive.
- ⇒ Gérer une infrastructure sportive.
- ⇒ Organiser des spectacles et événements sportifs.
- ⇒ Prévenir la violence dans les stades.
- ⇒ Promouvoir les infrastructures sportives.
- ⇒ Inaugurer une infrastructure sportive.
- ⇒ Contacter un club lié à une infrastructure sportive.

Comme nous pouvons le voir dans ces exemples, les usages retrouvés par cette méthode ne sont pas des usages réservés aux acteurs « citoyen » mais aussi aux acteurs « administration » et aux entreprises.

### 2.2.1 Confrontation des usages

Le but de cette étape est de vérifier les informations requises par les différentes classes d'utilisateurs. Voici quelques exemples d'usages détectés et confrontés avec les informations disponibles dans le système d'information géographique.

Cas d'usages	Information contenue dans la base
<b>Promotion d'infrastructures sportives</b>	non
<b>Traçage d'un itinéraire jusqu'à l'infrastructure sportive: Localisation d'une infrastructure sportive/équipement sportif</b>	Il est possible de localiser des infrastructures sportives grâce aux informations présentes dans la base de données alphanumérique; cette localisation peut être couplée à de l'information géographique (IG).
...Où se situe l'infrastructure sportive?	oui -> adresse + IG
<b>Facilité d'accès pour les handicapés</b>	oui -> attribut PMR pour les infrastructures
...Quels sont les différents aménagements sportifs que comporte telle ou telle infrastructure sportive?	oui -> installations présentes dans un site?
...Quelle est la capacité d'accueil de telle infrastructure sportive?	non -> seulement le taux de fréquentation mensuelle des installations
<b>Intervention sécurité (pompiers, police, agents de sécurité...)</b>	information présente (conformité normes incendie) mais très peu détaillée
...Quels sont les dispositifs de sécurité présents dans les infrastructures sportives?	non.
...Comment assurer la sécurité dans une infrastructures sportives?	non.

Cette confrontation permet :

- de préciser au fournisseur de services quelles sont les lacunes (services à développer) et les extensions possibles des bases de données ;
- de pouvoir établir le lien entre les requêtes et les services en précisant les liens entre le vocabulaire des utilisateurs et le vocabulaire des producteurs d'information.

## 2.3 Le système final

La solution finale est une couche ontologique qui s'intègre dans un contexte plus global de diffusion de l'information. Ce contexte est divisé en quatre couches :



- ⇒ Le registre de services publie tous les services d'e-gouvernement disponibles. Il sécurise également les accès.
- ⇒ Le portail contrôle le traitement des requêtes.
- ⇒ La couche ontologique joue le rôle d'un traducteur entre la requête et le service.
- ⇒ Une intégration de l'ontologie à un moteur de traitement textuel a été réalisée afin que le système puisse répondre à des requêtes textuelles. Une intégration avec un traitement vocal est en cours.

Un prototype a été réalisé. Les requêtes modélisées sont des requêtes d'accès aux infrastructures pour pratiquer des disciplines sportives. Voici quelques exemples de requêtes traitées par le système :

- « Je cherche des sports praticables à Namur »
- « Je veux le nom des endroits où pratiquer du football »
- « Je cherche le nom des endroits où faire du karaté »
- « Je veux le nom des endroits où pratiquer du tir au but »

Pour ces quatre usages, le système va avoir un comportement différent :

- Pour la première requête, le système va appeler un service qui recherche des aménagements sportifs pour une ville donnée avec comme paramètre « praticable » et « Namur ».
- Pour la deuxième, le système va appeler un service qui recherche des aménagements sportifs qui permettent de pratiquer une activité sportive avec comme paramètre « football ».
- Pour la troisième, le système va appeler un service qui recherche des aménagements sportifs qui permettent de pratiquer une activité sportive avec comme paramètre « art\_martial » dont le concept associé « ArtMartial » est une généralisation dans l'ontologie du concept « KarateDo ».

- Pour la quatrième, le système va appeler un service qui recherche des aménagements sportifs d'un type donné avec le paramètre `mur_de_frappe` relié au concept « MurDeFrappe ». Ce concept est obtenu à l'aide d'une relation dans l'ontologie nommée « sePratiqueDans » et qui est spécifié entre les concepts « TirAuBut » et « MurDeFrappe ».

### 3 Conclusion

La première phase du projet a permis de mettre au point une méthodologie à employer afin de garantir la qualité des services pour le citoyen. Les premiers tests ont été effectués avec succès. Ils ont prouvé la faisabilité de l'approche. Toutefois, l'analyse des usages doit être poursuivie et affinée. La modélisation d'interactions plus complexes est envisagée afin d'obtenir un système réellement utilisable par les citoyens. La phase de test par les usagers est prévue en 2006.

Cette expérimentation a également montré la complémentarité entre une approche UML et une approche d'extraction d'usages dans les corpus: l'approche UML est réservée à un public ciblé sur lequel on peut approfondir les besoins ; l'approche documentaire est réservée à un public générique, inaccessible via la voix classique de groupe d'utilisation. Les techniques d'extraction d'usages dans les textes vont permettre d'être plus précis quant aux requêtes formulées par les utilisateurs. En revanche les groupes d'utilisation vont permettre de définir les grandes classes d'utilisation de l'information et relier la requête au service de e-gouvernement.

Le principe de relever tous les termes du domaine et de diversifier les cas d'usages de l'information liés à un domaine est une aide conséquente dans l'approche de modélisation des besoins car elle permet d'envisager tous les cas de figure. Le fournisseur de l'information possède une meilleure visibilité sur l'utilité et la pertinence d'ajouter une information dans son système, de décider si cette information est aisément maintenable et ainsi de prendre la décision sur son intégration dans le système.

### 4 Bibliographie

- [1] F. Jacobiak, L'Intelligence Economique en pratique, Ed. d'Organisation, 2001.
- [2] Toward Citizen-Centred Service Delivery: a how-to guide for service improvement initiatives, published by institute for citizen-centred service, Canada, Marsh 2004.
- [3] L. Mac Kee, G. Pichler. GINIE : Geographic information Network in Europe, Final Report : Registries and e-Services. IST-2000-29493, August 2003.
- [4] M. Craglia : « MADAME : Méthodes d'accès aux données et aux métadonnées en Europe, Elargir l'accès aux informations du secteur public : guide des bonnes pratiques. Décembre 2000.
- [5] Paul SMITS (Editor for Architecture And Standards Working Group): "INSPIRE Architecture and Standards Position Paper", published by JRC-Institute for Environment and Sustainability, Ispra, under the identifier EUR 20518 EN, 2002
- [6] P. Séguéla and N. Aussenac-Gilles, " Extraction de relations sémantiques entre termes et enrichissement de modèles du domaine", Proceedings of the international conference on Knowledge capture, 2001
- [7] D. Bourrigault and C. Jacquemin, Construction de ressources terminologiques, In Ingenierie des langues, Jean-Marie Pierrel (Eds), pp 215-233, Hermès, 2000
- [8] W. Kuhn and M. Raubal, Implementing Semantic Reference Systems. M.F. Gould et al. (Eds.), 6th AGILE Conference on Geographic Information Science, Lyon, France. AGILE: in press. Online at <http://musil.uni-muenster.de>, 2003.
- [9] F.Fonseca and M. Egenhofer, "Ontology-Driven Geographic Information Systems", 7th ACM Symposium on Advances in Geographic Information Systems, Kansas City, MO, C. Bauzer Medeiros (ed.), pp. 14-19, November 1999.
- [10] N. Chomsky, Structures syntaxiques, Paris, Seuil, 1957.
- [11] N. Chomsky, Aspects de la théorie syntaxique, Paris, Seuil, 1971.

- [12] F. Bouille, "Structuring Cartographic Data and Spatial Processes with the Hypergraph-Based Data Structure", In Dutton, G. ed. Harvard Papers on Geographical Information Systems, Laboratory for Computer Graphics and Spatial Analysis, Harvard University, 1977.
- [13] The European E-Business Report 2004: 3rd Synthesis Report of the E-Business W@tch. Luxembourg: Office for Official Publications of the European Communities, 2004 (Enterprise Publications)
- [14] <http://framenet.icsi.berkeley.edu/>
- [15] <http://www.loa-cnr.it/DOLCE.html>
- [16] <http://ontology.teknowledge.com/>
- [17] <http://www.opencyc.org/>
- [18] <http://protege.stanford.edu/>
- [19] <http://www.safir-fp6.org>
- [20] C. Revelli, Intelligence stratégique sur Internet : Comment développer des activités de veille et d'intelligence économique sur le web, Dunod, 2000.
- [21] C. Roux and A. Lefèvre, Identification des besoins en information géographique, 7ème Journées internationales d'Analyse statistique des Données Textuelles, JADT'2004, Louvain-la-Neuve, 2004
- [22] C. Roux et D. Sosson, XPlor : un outil d'investigation en ligne sur des données relationnelles, VSST'2004, Toulouse, 2004.
- [23] <http://dublincore.org/>
- [24] B. Bucher, L'aide à l'accès à l'information géographique : un environnement de conception coopérative d'utilisations de données géographiques, thèse d'université, 2002.
- [25] <http://www.w3.org/2004/OWL/>