Chaîne de traitement linguistique : du repérage d'expressions temporelles au peuplement d'une ontologie de tourisme

Stéphanie Weiser (1), Martin Coste (2) , Florence Amardeilh (1-2)

(1) MoDyCo – CNRS, Université Paris Ouest Nanterre La Défense –
200, avenue de la République, 92001 Nanterre.
sweiser@u-paris10.fr
(2) Mondeca – 3, cité Nollez, 75018 Paris.
martin.coste@mondeca.com
florence.amardeilh@mondeca.com

Résumé Cet article présente la chaîne de traitement linguistique réalisée pour la mise en place d'une plateforme touristique sur Internet. Les premières étapes de cette chaîne sont le repérage et l'annotation des expressions temporelles présentes dans des pages Web. Ces deux tâches sont effectuées à l'aide de patrons linguistiques. Elles soulèvent de nombreux questionnements auxquels nous tentons de répondre, notamment au sujet de la définition des informations à extraire, du format d'annotation et des contraintes. L'étape suivante consiste en l'exploitation des données annotées pour le peuplement d'une ontologie du tourisme. Nous présentons les règles d'acquisition nécessaires pour alimenter la base de connaissance du projet. Enfin, nous exposons une évaluation du système d'annotation. Cette évaluation permet de juger aussi bien le repérage des expressions temporelles que leur annotation.

Abstract This paper presents the linguistic data processing sequence built for a tourism web portal. The first steps of this sequence are the detection and the annotation of the temporal expressions found in the web pages. These tasks are performed using linguistic patterns. They lead to many questions which we try to answer, such as the definition of information to detect, annotation format and constraints. In the next step this annotated data is used to populate a tourism ontology. We present the acquisition rules which are necessary to enrich the portal knowledge base. Then we present an evaluation of our annotation system. This evaluation is able to judge the detection of the temporal expressions and their annotation.

Mots-clés : Annotation, expressions temporelles, ontologies, base de connaissance, tourisme

Keywords: Annotation, temporal expressions, ontologies, knowledge base, tourism

1 Introduction

Les travaux décrits dans cet article sont réalisés dans le cadre du projet RNTL EIFFEL. L'objectif global de ce projet est la mise en œuvre d'une plateforme logicielle permettant, autour d'une ontologie tourisme dédiée à un territoire, de sélectionner, classer et qualifier des contenus distribués sur le Web. La plateforme a pour but d'assister l'utilisateur dans la recherche et la construction d'un voyage à partir des ressources sélectionnées.

L'ontologie de tourisme que nous avons modélisée dans ce projet décrit les ressources touristiques, au sens large (activités, hébergements, logistique, patrimoine, territoires, voyages et itinéraires…) et organise la base de connaissance du territoire. Nous nous sommes particulièrement intéressés aux propriétés temporelles de ces ressources touristiques afin de pouvoir renseigner l'utilisateur sur les informations d'ouverture et de fermeture d'une offre touristique. Ces informations pouvant être obtenues à partir des contenus Web des ressources touristiques du territoire, c'est dans ce contexte que nous avons développé notre outil de repérage et d'annotation automatique des informations temporelles d'offres touristiques. Nous ne présenterons pas ici le repérage des objets touristiques ni leur localisation. Le lien entre ces différentes données sera fait ultérieurement.

Dans la suite de cet article, nous exposerons tout d'abord un bref état de l'art concernant le domaine de travail et les méthodes utilisées. Puis nous présenterons l'outil de repérage et d'annotation des informations temporelles suivi de la plateforme de peuplement d'ontologie. Nous finirons cet article par l'évaluation de notre outil d'annotation avant de conclure et d'envisager les travaux futurs.

2 L'apport du Web Sémantique et des outils de TAL

Les données actuelles des contenus Web sont souvent encore écrites en langage naturel, car elles sont destinées aux humains. L'approche du Web Sémantique consiste à décrire ces contenus en les annotant avec des informations non ambiguës, provenant le plus souvent d'une ontologie, afin de favoriser l'exploitation de ces contenus par des agents logiciels (Prié, Garlatti, 2004). Les ontologies, originaires des techniques de modélisation de la connaissance, fournissent les moyens d'exprimer les concepts d'un domaine en les organisant hiérarchiquement et en définissant leurs propriétés dans un langage de représentation des connaissances formel, favorisant le partage d'une vue consensuelle sur ce domaine entre les différentes applications informatiques qui en font usage.

Dans le cadre de ce projet, la tâche de peuplement d'ontologie repose sur la méthode des patrons lexico-syntaxiques issue du domaine du traitement automatique des langues et de la théorie des automates ; voir (Hearst, 1992) et (Morin, 1999) pour une présentation de cette méthodologie et des exemples d'utilisation. Pour effectuer le repérage et l'annotation des expressions temporelles, nous avons développé un ensemble de transducteurs à l'aide de l'outil Unitex[1]. Le graphe de la Figure 1 permet de repérer les horaires comme ceux de l'exemple suivant, et d'aboutir à la forme annotée ci-après.

Horaires D'ouverture Lundi et mercredi 9h-12h Jeudi 14h-17h Vendredi 14h-16h

[1] Unitex : http://www-igm.univ-mlv.fr/~unitex.

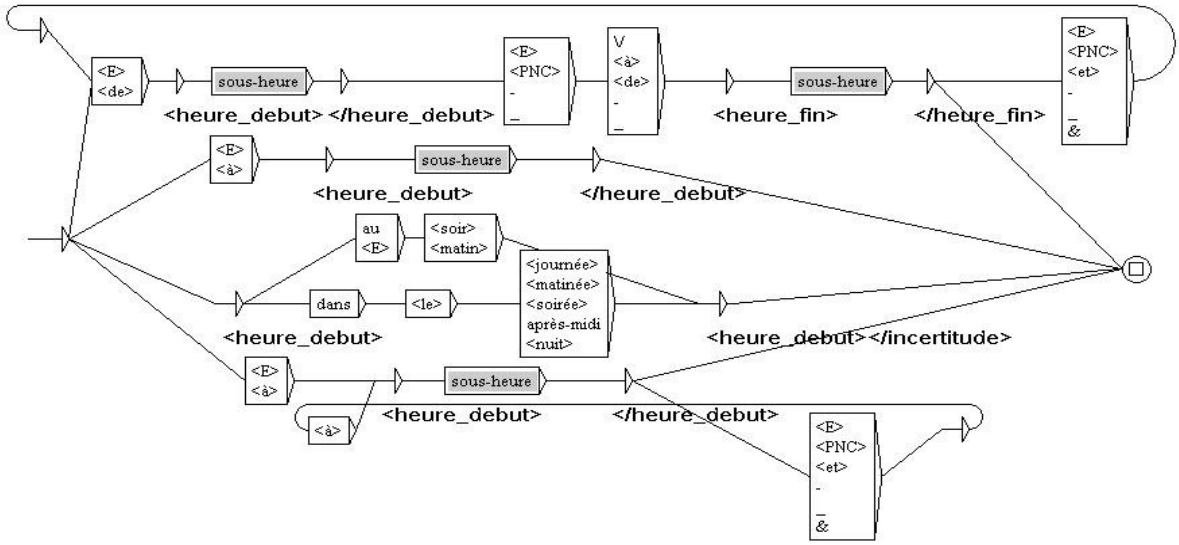

Figure 1 : Graphe Unitex repérant et annotant des horaires

<UT> <periode_Ouverture> Horaires d'ouverture <jour> Lundi</jour> et<jour> mercredi</jour> <heure_debut> 9h</heure_debut>-<heure_fin>12h</heure_fin><jour> Jeudi</jour><heure_debut> 14h </heure_debut>- <heure_fin>17h </heure_fin> <jour> Vendredi</jour><heure_debut> 14h</heure_ debut>-<heure_fin>16h </heure_fin> </periode_Ouverture> </UT> <description>"Horaires D'ouverture Lundi et mercredi 9h-12h Jeudi 14h-17h Vendredi 14h-16h"</description>

Les travaux d'extraction d'informations temporelles ou calendaires sont nombreux, voir par exemple (Battistelli et al., 2006) ou, pour un traitement plus large et en anglais, (Mani, Wilson, 2000). Nous nous distinguons de ces travaux car ils concernent des textes non structurés, prévus pour une lecture linéaire, tandis que nous travaillons sur des textes structurés conçus pour des pages Web. La forme et la structuration de la page Web influent en effet sur le contenu. De plus, des habitudes langagières particulières se sont développées en même temps que le Web et poussent à ne pas formuler les choses de la même manière que si elles étaient prévues pour un guide touristique papier par exemple (beaucoup d'abréviations, manque de ponctuation, retours à la ligne fréquents, etc.).

3 Repérage et annotation des informations temporelles dans les contenus Web

Les expressions temporelles que l'on souhaite repérer et annoter ont deux caractéristiques principales. D'une part, elles ont une visée informative et pratique. Il ne s'agit pas de dates historiques ou d'expressions descriptives du type *la nuit d'avant* mais d'informations touristiques : horaires d'ouverture et de fermeture, dates, périodes... D'autre part, elles sont formulées pour le Web, c'est-à-dire qu'elles s'insèrent rarement dans un texte réellement rédigé. La complexité des expressions temporelles varie donc énormément : certaines expressions sont très simples, d'autres peuvent devenir complexes, jusqu'au point d'être floues ou ambiguës. Le Tableau 1 présente les différents types d'expressions que l'on repère et annote, elles sont présentées plus en détail dans (Weiser et al., 2008).

Une fois ces expressions repérées, elles doivent être stockées dans la base de connaissance du projet, contrainte par l'ontologie de tourisme, intégrant une modélisation du « temps du tourisme ». L'ontologie constitue donc un pivot, une articulation dans ce projet et de nombreux ajustements ont été effectués pour pallier les contraintes engendrées par cette modélisation. Par exemple, chaque journée est découpée en « parties de journée » ; on a donc *lundi matin, lundi midi, lundi après-midi et lundi soir*. Cette modélisation est indispensable, entre autres pour représenter les plages horaires d'ouverture d'un restaurant.

Type d'expression	**Exemples**
Date	- Concert le 1er octobre.
Période (répétitive ou non)	- Festival de mai à juin. - Ouvert du lundi au samedi.
Horaire (répétitif ou ponctuel)	- Le concert commence à 20h. - Ouverture : de 8h à 12h et de 14h à 18h.
Exception	- Le camping est ouvert toute l'année sauf en janvier.
Combinaison	- De mai à juin, ouvert tous les jours sauf le mardi. De juin à septembre les après-midi, sauf lundi et mardi. Sur rendez-vous le reste du temps. - Fermé le 25 décembre et du 1er janvier au 23 février.

Tableau 1 : Expressions repérées et annotées

Par souci de simplicité et d'efficacité, le format XML a été choisi pour accomplir la tâche d'annotation. Un jeu de balises, décrit dans une DTD[2], elle-même présentée dans (Weiser, 2008), a donc été défini en fonction de ce qu'il était utile et possible d'annoter en rapport avec l'ontologie du projet. Si ce jeu de balises est spécifique à nos besoins, la plupart des expressions que l'on annote pourraient être couvertes par la norme TimeML (Pustejovsky et al, 2003) mais nous n'avons pas besoin d'un tel degré de détail. Ce qui nous intéresse c'est d'annoter les expressions à un niveau sémantique : savoir que l'expression correspond à une information d'ouverture et non de fermeture par exemple.

Voici quelques exemples d'expressions temporelles annotées par ce jeu de balises XML. L'expression textuelle se trouve dans la balise « description ».

1. <UT> <periode_Ouverture> <date_debut> de juin </date_debut> à <date_fin> septembre <date_fin> <incertitude/> <heure_debut> les après-midi </heure_debut> </periode_Ouverture> , <Exception> sauf <periode_Fermeture> <jour> lundi </jour> et <jour> mardi </jour> </periode_Fermeture> </Exception> .{S} <Exception> Sur rendez-vous le reste du temps </Exception> </UT> <description> "de juin à septembre les après-midi, sauf lundi et mardi.{S} Sur rendez-vous le reste du temps" </description>.{S}[3]

[2] Une DTD (Document Type Definition) permet de décrire un modèle de document XML.

[3] « {S} » est une marque insérée par Unitex pour marquer la fin des phrases.

2. <UT> Fermé <periode_Fermeture> le <date> 25 décembre </date> et </periode_Fermeture> <periode_Ouverture> <date_debut> du 1er janvier </date_debut> au <date_fin> 23 février </date_fin> </periode_Ouverture> .{S} </UT> <description> "Fermé le 25 décembre et du 1er janvier au 23 février.{S}"</description>

Ces deux expressions sont bien repérées et bien annotées. Elles reflètent le degré de complexité des expressions que l'on rencontre.

3. Ouverture <UT> <periode_Ouverture> Du <jour> Mardi </jour> au <jour> samedi </jour> de <heure_debut> 15h </heure_debut> à <heure_fin> 19h </heure_fin> </periode_Ouverture> </UT> <description> "Du Mardi au samedi de 15h à 19h" </description> et sur appel téléphonique

4. <UT> <periode_Ouverture> du <jour> lundi </jour> au <jour> vendredi </jour> </periode_Ouverture> </UT> <description> "du lundi au vendredi" </description> de 8h Ã 20h et le samedi de 8h Ã 12h

5. Ouverture : les week-ends et <UT> <periode_Ouverture> <jour> jours fériés </jour> </ periode_Ouverture> </UT> <description> "jours fériés" </description> de 14 à 19 heures tous les jours pendant les vacances scolaires de 14 à 19 heures

L'expression 3 est considérée comme incomplète car « et sur appel téléphonique » n'est pas repéré alors que cela devrait être annoté comme une exception. L'expression 4 est réellement incomplète : les horaires ne sont pas repérés à cause d'un problème d'encodage de la page Web. L'expression 5 n'est pas repérée dans son ensemble, une telle combinaison de jours et heures n'est pas prise en compte par notre système. Ces trois derniers exemples illustrent les principaux problèmes que l'on rencontre pour le repérage des expressions temporelles touristiques.

4 Le peuplement d'ontologie

La tâche de peuplement d'ontologie fait l'objet de nombreuses recherches et reste un véritable challenge, quel que soit le domaine étudié. L'annotation de grandes collections de documents nécessite des outils automatisés reposant sur les méthodes d'extraction d'information et d'apprentissage automatique. Parmi les approches les plus abouties, citons les outils KIM (Kiryakov et al., 2005) et OntoPop (Amardeilh, 2007). Pour plus de détails sur les outils existants, nous référons le lecteur à l'étude (Uren et al., 2006).

Dans le cadre de notre projet, nous avons décidé d'utiliser l'outil CA-Manager, nouvelle version d'OntoPop, développé dans le cadre du projet de recherche européen TAO[4]. La particularité du CA-Manager est qu'il conjugue une infrastructure UIMA avec les standards du Web Sémantique (web services et formats RDF/OWL). Il se veut un médiateur indépendant entre les outils d'analyse linguistique et les bases de connaissance. Il organise le flux d'annotation et de peuplement d'ontologie en une succession d'étapes (Amardeilh, Damljanovic, 2009), certaines étant optionnelles. Ces étapes peuvent être regroupées en trois principaux composants : l'extraction d'information ; la consolidation de l'information, notamment par l'exploitation d'algorithmes poussés de validation de contraintes reposant sur l'ontologie du domaine ; et enfin la sérialisation dans le format attendu par l'application finale et le stockage dans la base de connaissance.

[4] Projet européen IST TAO (Transitioning Application to Ontologies) : http://www.tao-project.eu

Dans notre projet, la phase d'extraction d'information repose sur un ensemble de transducteurs Unitex tel celui décrit ci-dessus. Le fichier XML produit est transformé par des règles d'acquisition de connaissance, décrites dans (Amardeilh, 2007) dans le schéma d'annotation interne exploité par le CA-Manager. Par exemple, dans la Figure 2, l'application de la règle d'acquisition présentée va permettre la création d'une instance de la classe « Période d'ouverture », elle même sous-classe de « Unité de Temps » (UT), dans l'ontologie de tourisme. Le schéma d'annotation dépend de la modélisation de l'ontologie du domaine. Les informations extraites et annotées sont automatiquement validées lors de la phase de consolidation en fonction des contraintes imposées par l'ontologie (restrictions, domaines et range) ainsi que l'élimination des doublons d'instances déjà contenues dans la base de connaissance attenante. Une fois ces instances validées ou invalidées, elles sont enregistrées dans la base de connaissance et peuvent ensuite si besoin être présentées à l'utilisateur final pour qu'il puisse valider manuellement et désambiguïser les connaissances jugées invalides par les algorithmes. La base de connaissance ainsi enrichie peut alors être exploitée via des requêtes sémantiques ou des actions de navigation effectuées par l'utilisateur final de l'application dans un portail dédié.

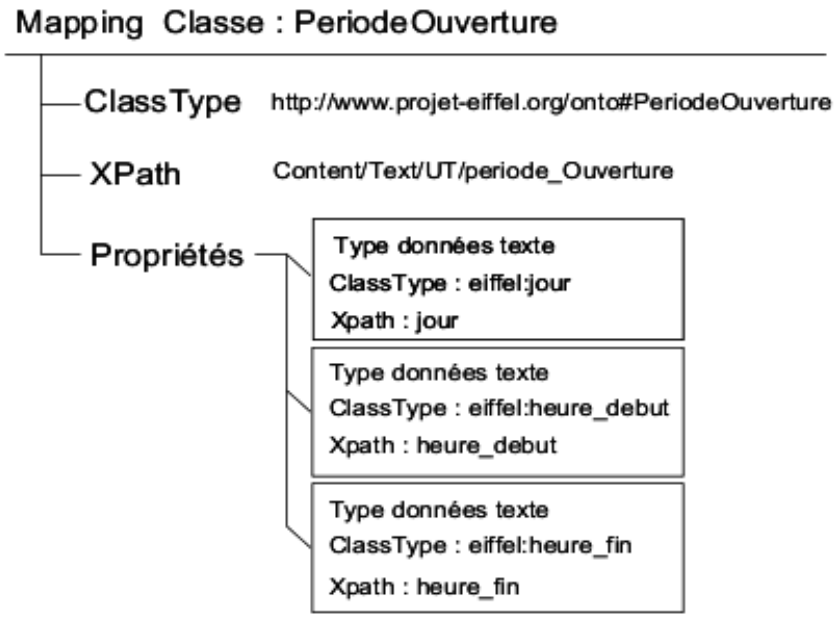

Figure 2: Règle d'acquisition de connaissance « Période Ouverture »

5 Évaluation du système

Dans cette partie, nous présentons une évaluation de notre système de repérage et d'annotation, en termes de rappel et précision. Pour plus de détails sur les questions d'évaluation, voir (Popescu-Belis, 2007). Si ces mesures de rappel et précision sont souvent utilisées en TAL, elles sont aussi parfois controversées, que ce soit dans le domaine de l'extraction d'information (Lavelli et al., 2004) ou dans celui de l'annotation sémantique (Maynard, 2005). En effet ces mesures sont trop rigides et ne permettent pas de rendre compte de résultats imparfaits ou partiels. Certaines solutions ont été envisagées mais elles n'ont pas donné lieu à des mesures standardisées tel le rappel et la précision que nous utilisons. Par exemple, (Freitag, 1998) avait proposé de classer les résultats dans trois catégories : résultats exacts ; résultats dans lesquels l'information attendue est contenue dans

le résultat obtenu ; résultats imbriqués où l'information obtenue dépasse l'information attendue.

5.1 Protocole d'évaluation

Pour procéder à une évaluation de notre système de repérage et d'annotation, nous lui avons donné à traiter un échantillon de 250 pages sélectionnées au hasard. Notre évaluation a pour but d'évaluer aussi bien le repérage que l'annotation. Au niveau des taux de rappel et précision, on considère comme correctes les expressions repérées à bon escient, même si elles le sont de façon incomplète ; en effet un repérage incomplet n'est la plupart pas nuisible pour le système (par exemple, pour *ouvert toute l'année le vendredi et le samedi*, si on ne repère pas la première partie, *ouvert toute l'année*, cela ne change rien au résultat final). On comptabilise les expressions manquées et les expressions repérées à tort. Au niveau de l'annotation, on considère comme bien annotées les expressions qui sont bien catégorisées (où on ne confond pas ouverture et fermeture par exemple).

Une fois les pages analysées par le système, nous distinguons celles dans lesquelles un repérage (et donc une annotation) a été effectué de celles dans lesquelles il n'y a pas de résultat. Pour celles qui ne donnent pas de résultat, nous vérifions si elles ne contiennent pas tout de même des expressions temporelles touristiques que l'on aurait dû repérer, afin de comptabiliser les expressions manquées. Les pages dans lesquelles un repérage et une annotation ont eu lieu sont également analysées manuellement. On cherche si elles contiennent des expressions manquées et, pour les expressions repérées et annotées, on vérifie si le repérage est effectué à bon escient, si l'expression est repérée dans son ensemble et si elle est bien annotée.

5.2 Résultats

Dans les 250 pages, 191 expressions sont dans un premier temps considérées comme à repérer[5]. Notre système repère 115 expressions : 67 expressions sont repérées à bon escient et 48 expressions sont repérées à tort. 124 expressions temporelles touristiques sont donc manquées par le système. Sur les 67 expressions repérées, 44 sont bien repérées et 23 sont repérées partiellement. Ces chiffres mènent aux taux de rappel et précision présentés dans la première colonne du Tableau 2. En qui concerne l'annotation, sur les 67 expressions repérées, 64 sont bien annotées et 3 comportent des erreurs d'annotation ; cela donne un pourcentage de 95,5 % d'annotations correctes. Certaines pages ayant des propriétés particulières, nous allons faire un bilan selon le nombre d'expressions par page et selon le nombre de pages. Sur les 250 pages analysées, 61 seulement contiennent des expressions temporelles touristiques. De nombreuses expressions temporelles non touristiques se trouvent aussi dans ces pages et mènent parfois à des repérages fautifs mais beaucoup ne sont pas repérées – à raison.

Parmi les 61 pages contenant des expressions à repérer, 18 pages ne sont pas annotées, ce qui mène à 51 expressions manquées. 23 pages sont traitées semi-correctement, c'est-à-dire qu'elles peuvent contenir des expressions bien repérées et bien annotées mais elles contiennent également des expressions manquées. 20 pages sont traitées correctement, pour un total de 24 expressions bien repérées et bien annotées. Nous remarquons que certaines pages sont à l'origine de beaucoup d'erreurs.

[5] Pour arriver à ce chiffre (191) on additionne les expressions repérées à bon escient et les expressions manquées et on soustrait les expressions repérées à tort.

Premièrement, en dehors de ces 61 pages, des expressions repérées à tort sont présentes dans 6 pages, comprenant donc 48 faux-positifs. Certaines de ces expressions apparaissent dans des pages non touristiques ; elles n'auraient donc pas dû se trouver à l'entrée de notre système et relèvent d'une erreur du module d'aspiration des pages. Nous décidons donc d'exclure ces pages (au total 3 pages contenant 41 expressions repérées sont exclues). Les taux ainsi obtenus sont présentés dans la deuxième colonne du Tableau 2. Nous avons conscience que le repérage fautif d'informations dans des pages non touristiques n'incombe pas à notre outil et qu'à l'avenir ces pages devront être éliminées avant de procéder à l'évaluation.

Deuxièmement, nous nous sommes intéressés au grand nombre d'expressions manquées par notre système. Une même page contient par exemple plus d'un tiers (49 expressions) des expressions manquées. Cette page présente un agenda. Elle ne peut pas être considérée comme non pertinente, car il s'agit tout de même d'un agenda touristique. En revanche, on peut avancer qu'à l'heure actuelle notre système n'est pas conçu pour analyser de telles pages. Il faudrait donc les identifier et en faire un traitement propre, prenant en compte le fait qu'il s'agit d'un agenda et que la page contient donc un grand nombre d'expressions temporelles. De plus les expressions temporelles contenues dans des pages-agenda ont souvent des formes très différentes de ce que l'on considère comme des « expressions temporelles touristiques ». Nous choisissons donc d'éliminer pour l'instant les pages-agenda de notre évaluation. Une fois ces pages exclues, on obtient les taux de rappel et précision de la troisième colonne du Tableau 2. Ces taux, bien meilleurs que les premiers taux bruts calculés, sont plus représentatifs des capacités de notre outil car nous avons éliminé quelques pages marginales qui avaient beaucoup d'incidence sur le rappel et la précision.

Taux bruts		Taux après exclusion des pages non pertinentes		Taux après élimination des pages-agenda	
Rappel	Précision	Rappel	Précision	Rappel	Précision
67 / 191 35 %	67 / 115 58,2 %	67 / 191 35 %	67 / 74 90,5 %	58 / 95 61 %	58 / 65 89,2 %

Tableau 2 : Taux de rappel et précision

Nous avons choisi de privilégier la précision, au détriment du rappel, afin d'assurer la fiabilité des informations stockées dans la base de connaissance.

6 Conclusion et perspectives

Nous avons présenté la chaîne de traitement linguistique développée pour un portail touristique sur Internet. En entrée de la chaîne sont fournies des pages Web touristiques converties en XML. Nous analysons automatiquement ces pages afin d'y repérer et d'y annoter les expressions temporelles liées aux objets touristiques. Le format d'annotation a été développé en fonction de l'ontologie de tourisme du projet. Nous avons ensuite présenté le mécanisme des règles d'acquisition de connaissances permettant de peupler l'ontologie à l'aide des informations annotées. Après avoir passé en revue les différents problèmes de modélisation et d'interaction entre nos modules, nous avons proposé une évaluation du module de repérage et d'annotation des expressions temporelles. Nous nous sommes appuyés sur les mesures de rappel et précision largement utilisées en TAL mais nous pourrions maintenant essayer de mettre au point d'autres mesures plus appropriées, permettant de juger plus finement des résultats partiels.

Outre le fait de continuer nos travaux sur le repérage proprement dit qui consiste à enrichir encore les données linguistiques contenues dans les transducteurs afin d'élargir le nombre d'expressions repérées, les règles d'acquisition des connaissances devront être complétées et évaluées. Nous sommes aussi en train de travailler, avec un autre partenaire du projet, à l'élaboration d'un moteur de raisonnement qui permettrait de calculer automatiquement les périodes d'ouverture en fonction des instances des périodes de fermeture créées dans la base de connaissance à partir des annotations.

Remerciements

Ce travail a été partiellement financé par le projet Eiffel ANR-05-RNTL-007.

Références

AMARDEILH F. (2007). *Web Sémantique et Informatique Linguistique : propositions méthodologiques et réalisation d'une plateforme logicielle*. Thèse de doctorat. Université Paris-Sorbonne.

AMARDEILH F., DAMLJANOVIC D. (2009). Du texte à la connaissance : annotation sémantique et peuplement d'ontologie appliqués à des artefacts logiciels. .In *Proceedings des 20èmes Journées Francophones d'Ingénierie des Connaissances (IC'2009)*.

BATTISTELLI D., MINEL J.-L., SCHWER S. (2006). Représentation des expressions calendaires dans les textes : une application à la lecture assistée de biographies. *Traitement Automatique des Langues* 47, 3, 1-26.

FREITAG D. (1998). *Machine Learning for Information Extraction in Informal Domains*. Thèse de doctorat, Université Carnegie Mellon.

HEARST M. A. (1992). AUTOMATIC ACQUISITION OF HYPONYMS FROM LARGE TEXT CORPORA. IN *INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS (COLING'92)*.

KIRYAKOV A., POPOV B., TERZIEV I., MANOV D., KIRILOV A., GORANOV M. (2005). Semantic annotation, indexing, and retrieval. In *J. Web Semantics, Science, Services and Agents on the WWW*, 2(1), Elsevier, 49-79.

LAVELLI A., CALIFF M. E., CIRAVEGNA F., FREITAG D., GIULIANO C., KUSHMERICK N., ROMANO L. (2004). IE evaluation: Criticisms and recommendations, In *Proceedings of the AAAI 2004 Workshop on Adaptive Text Extraction and Mining (ATEM 2004)*.

MANI, I. AND WILSON, G. (2000). Robust Temporal Processing of News. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL'2000)*, 69-76.

MAYNARD D. (2005). Benchmarking ontology-based annotation tools for the Semantic Web, in *Proceedings of the Workshop "Text Mining, e-Research and Grid-enabled Language Technology" in the UK e-Science Programme All Hands Meeting (AHM2005)*.

MORIN E. (1999). Acquisition de patrons lexicosyntaxiques caractéristiques d'une relation sémantique. *Traitement Automatique des Langues 40, 1,* 143-166.

POPESCU-BELIS A. (2007). Le rôle des métriques d'évaluation dans le processus de recherche en TAL . *Traitement Automatique des Langues 48, 1,* 67-91.

PRIÉ Y., GARLATTI S. (2004). Méta-données et annotations dans le Web sémantique, in *Le Web sémantique*, CHARLET J., LAUBLET P. et REYNAUD C. (Ed.), Hors série de la *Revue Information - Interaction - Intelligence* (I3), 4(1), Cépaduès, 45-68.

PUSTEJOVSKY J., CASTAÑO J., INGRIA R, SAURÍ R., GAIZAUSKAS R., SETZER, A., KATZ, G. (2003). TimeML: Robust Specification of Event and Temporal Expressions in Text. In *IWCS-5, Fifth International Workshop on Computational Semantics.*

UREN V., CIMIANO P., HANDSCHUH S., **VARGAS**-VERA M., MOTTA E., CIRAVEGNA F. (2006). Semantic annotation for knowledge management: requirements and a survey of the state of the art. In *J. Web Semantics, Science, Services and Agents on the WWW*, 4(1). 14-26.

WEISER S. (2008). Informations spatio-temporelles et objets touristiques dans des pages Web : repérage et annotation. Actes de *Recital 2008*, 131-140.

WEISER S., LAUBLET P., MINEL J.-L. (2008). Automatic identification of temporal information in tourism web pages. Actes de *LREC'08, the Sixth International Language Resources and Evaluation*, 127-131.