Application BASTRI
Fiches Equipes

ATOLL (SR0207CR)

Atelier d'outils logiciels pour le langage naturel
ATOLL ALPAGE (SR0098OR)

Statut: Terminée

Responsable : Eric Villemonte De La Clergeri

Mots-clés de "A - Thèmes de recherche en Sciences du numérique - 2024" : *Aucun mot-clé.*

Mots-clés de "B - Autres sciences et domaines d'application - 2024" : *Aucun mot-clé.*

Domaine : Systèmes symboliques
Thème : Organisation des contenus et de la langue

Période : 01/01/1997 -> 31/08/2007
Dates d'évaluation :

Etablissement(s) de rattachement : <sans>
Laboratoire(s) partenaire(s) : <sans UMR>

CRI : Centre Inria de Paris
Localisation : Rocquencourt
Code structure Inria :

Numéro RNSR : 199721345U
N° de structure Inria: SR0207CR

Présentation

Le traitement automatique des documents en langage naturel est devenu un enjeu essentiel, si l'on veut pouvoir utiliser efficacement les masses énormes d'information disponible sur la planète. Ce problème est encore renforcé par l'utilisation croissante d'Internet. Notre équipe-projet a pour objectif de développer des outils et techniques, théoriques ou appliqués, pour aider l'accès, le traitement et l'utilisation des documents en langage naturel.

Axes de recherche

- Analyse syntaxique : Exploration théorique et pratique de techniques d'analyse syntaxique pour différents formalismes grammaticaux servant au traitement de la langue naturelle. En particulier, des techniques de tabulation sont étudiées pour mieux gérer les ambiguïtés du langage et mise en oeuvre dans plusieurs systèmes :
    - SYNTAX : ce logiciel permet la compilation de grammaires non contextuelles (CFG), qu'elles soient déterministes ou non.
    - Grammaires à concaténation de chaîne (RCG) : cette hiérarchie de grammaires introduite par Pierre Boullier permet l'exploration efficace des formalismes grammaticaux faiblement dépendants du contexte (MCS). Une implantation des RCG existe qui est en particulier utilisée pour les grammaires d'arbres adjoints (TAG).
    - Automates à piles et programmation dynamique : cette approche repose sur l'utilisation d'automates à piles pour la description de diverses stratégies d'analyse et sur l'interprétation par programmation dynamique de ces automates pour construire des analyseurs syntaxiques tabulaires. Ces techniques sont validées dans le cadre du système DyALog pour des formalismes grammaticaux utilisant l'unification (DCG, Feature TAG, ...) ainsi que pour des programmes logiques.
- Infrastructure linguistique : développement d'un environnement de travail pour les TAG s'appuyant sur des représentations XML. Il comprend des analyseurs syntaxiques construits avec les RCG et DyALog mais également des serveurs pour accéder aux analyseurs, aux grammaires, et aux " forêts de dérivation " produites par les analyseurs.
- Acquisition de connaissances: Thème émergent concernant les

Contact

- **Responsable :** Eric Villemonte De La Clergeri
- **Tél :** 01.39.63.54.10
- **Secrétariat Tél :** 01.39.63.54.12

En savoir plus

- Site de l'équipe
- Site sur inria.fr
- Site du responsable
- Derniers Rapports d'Activité :

Documents sur la structure

- Intranet
- Privés

Décisions

- 3401 (25/06/2002) : modification
- 3620 (15/07/2002) : prolongation
- 5156 (10/10/2006) : prolongation
- 5700 (29/08/2007) : fermeture

Localisation

- **Adresse postale :** *Non renseignée*
- **Coordonnées GPS :** 48.83703, 2.103342

interactions entre " connaissance " et analyse syntaxique. Des connaissances "sémantiques" sur les mots aident l'analyse et inversement l'analyse syntaxique permet d'extraire ces connaissances à partir de corpora.

Relations industrielles et internationales

- Action Normalangue: cette action s'inscrit dans le cadre du programme national Technolangue et concerne les questions de normalisation de ressources linguistiques.
- ARC "Ressource lexicales pour les TAG" [RLT] : Action de Recherche Concertée en partenariat avec " Langue et Dialogue " (LORIA, Nancy) et TALaNa (Université Paris 7). L'objectif principal de cette ARC concerne l'acquisition semi-automatique de ressources lexicales pour une grammaire TAG du français, en s'appuyant sur l'analyse syntaxique de corpus.
- ARC Génération et Inférence [GENI] : Action de Recherche Concerté en partenariat avec Langue et Dialogue, Orpailleur (LORIA), Lattice et ILPL (IRIT, Toulouse). Dans le cadre de cette ARC, nous apportons une expertise sur les TAG et sommes intressés par des aspects de sémantique lexicale.
- Action FASTLING : Action INRIA-ICTII franco-portugaise entre ATOLL, le laboratoire CENTRIA de l'Université Nouvelle de Lisbonne (UNL) et le LIFO (université d'Orléans). Cette action prolonge une précédente ayant permis le développement d'un analyseur du portugais utilisant le système DyALog de l'équipe-projet ATOLL.
- Action Botanique : Ce projet, en cours de montage avec les services documentaires de l'IRD (ex ORSTOM), doit inclure une composante analyse de corpus botaniques par des techniques linguistiques.