

# GRAPHIK

## Représentation de connaissances et raisonnements

### Effectifs

au 30/06/2013 :

5 permanents (3 ETP)

7 doctorants (7 ETP)

4 autres personnels (4 ETP)

Nombre de thèses soutenues  
entre le 01/01/2008  
et le 30/06/2013 : 3

### Responsable :

Marie-Laure Mugnier

### Page Internet de l'équipe :

<http://www.lirmm.fr/graphik>

FORMALISMES LOGIQUES ET GRAPHIQUES, ONTOLOGY-BASED DATA ACCESS, INTÉGRATION SÉMANTIQUE DE DONNÉES, ARGUMENTATION, PRÉFÉRENCES, AIDE À LA DÉCISION BASÉE SUR DES CONNAISSANCES

## Présentation

Les travaux de GraphIK (Graphs for Inference and Knowledge representation) se situent dans le domaine de la représentation des connaissances et des raisonnements. Nous avons une approche fondée sur la logique mais utilisant également des graphes / hypergraphes étiquetés. Nous nous attaquons à quelques-uns des grands défis de ce domaine comme l'interrogation de grandes bases de connaissances, le raisonnement avec des bases hétérogènes et le traitement d'informations « imparfaites » (e.g. imprécises ou contradictoires). Notre objectif est d'étudier et développer des formalismes de représentation de connaissances et de raisonnement ayant des propriétés de généralité et d'efficacité algorithmique, et de les valider dans des systèmes à base de connaissances réels. Notre activité repose sur trois pieds : théorie (expressivité, propriétés structurelles, décidabilité et complexité des problèmes, algorithmes), développement logiciel (des outils qui implémentent nos résultats théoriques), et applications (qui nourrissent en retour les questions théoriques).

## Evolution de l'équipe

GraphIK a été créée en 2010 en tant qu'équipe commune avec INRIA. Elle s'appuie sur l'ex équipe RCR (Représentation de Connaissances et Raisonnements) du LIRMM et intègre des chercheurs informaticiens du laboratoire d'agronomie Montpellierain IATE. L'équipe RCR était reconnue internationalement comme l'une des équipes leaders sur le formalisme des graphes conceptuels [Chein Mugnier 2009]. Le nouveau projet d'équipe considère plus largement la logique comme un formalisme unificateur de nombreux langages de représentation de connaissances. Les résultats obtenus ont été très bien perçus dans la communauté plus compétitive de représentation des connaissances et raisonnements (dont la conférence emblématique est KR) et ont ouvert la voie à des collaborations fructueuses. La période de référence a vu deux recrutements, Madalina Croitoru (MCF, sept. 2008) et Souhila Kaci (PR, sept. 2011).

## Organisation et Vie de l'équipe

L'équipe se réunit sur une base hebdomadaire pour des exposés scientifiques internes ou invités. Des réunions par sous-groupes sur des thématiques plus précises rythment également la vie de l'équipe. Ces thématiques sont notamment associées aux projets financés (actuellement 3 projets ANR et 1 projet européen) mais pas seulement. Une journée annuelle de bilan et prospective est organisée.

Les doctorants sont co-encadrés, ce qui est un élément de synergie de l'équipe. La grande majorité des publications de l'équipe implique au moins deux permanents. L'équipe organise également un cours d'anglais hebdomadaire depuis l'automne 2010 auquel participent les doctorants et permanents qui le souhaitent (7 personnes actuellement).

D'un point de vue financier, un compte commun permet d'effectuer les dépenses qui ne relèvent pas d'un projet financé.

## Activités scientifiques

Nous présentons nos résultats selon les grandes problématiques scientifiques étudiées depuis 2008. Ces problématiques sont transverses aux axes qui structurent le projet d'équipe (voir site Internet). Les résultats sur la problématique Ontology-Based Data Access s'appuient sur les compétences historiques de l'équipe en logique du premier ordre, tandis que ceux obtenus sur les préférences et les systèmes argumentatifs pour l'aide à la décision correspondent à des compétences en développement dans l'équipe. Quant aux travaux sur l'intégration sémantique de données, ils ont consisté en plusieurs années de formalisation et expérimentations, qui ont permis de dégager des questions de recherche originales.

### Ontology-Based Data Access

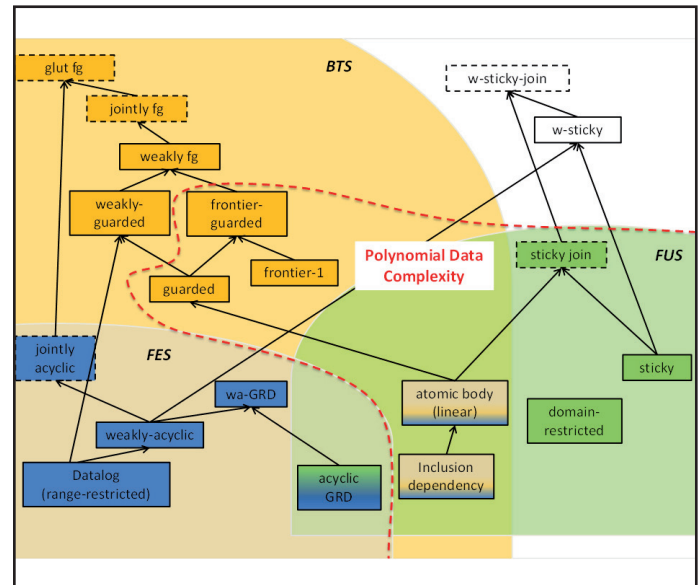
La question de la prise en compte dans l'interrogation de données d'inférences rendues possibles par une ontologie (représentant des connaissances générales sur le domaine d'application) suscite actuellement un énorme intérêt dans les domaines de recherche concernés (représentation de connaissances, bases de données, web sémantique). Cette problématique est connue sous le nom OBDA (Ontology-Based Data Access). Le problème de décision fondamental associé peut être exprimé ainsi : étant données une base de connaissances, composée de données  $D$  et d'une ontologie  $O$ , et une requête  $Q$ ,  $Q$  est-elle conséquence sémantique de  $D$  et de  $O$  ? Nous nous intéressons en particulier aux fragments de la logique du premier ordre permettant de décrire l'ontologie  $O$ , et au compromis entre l'expressivité de ces fragments et la complexité des raisonnements dans ces fragments. Notre approche utilise des règles existentielles pour décrire l'ontologie. Ces règles sont de la forme :

$$\forall x (Body[x] \rightarrow \exists y Head[x,y])$$

où  $x$  et  $y$  sont des ensembles de variables, et  $Body[x]$  et  $Head[x,y]$  sont des conjonctions d'atomes sans fonction. La puissance de ces règles provient du fait que les variables  $y$  qui sont propres à la tête de la règle ( $Head$ ) sont quantifiées existentiellement. L'application d'une règle permet ainsi de créer un nouvel individu, qui n'apparaît pas forcément dans la base de données. Cette faculté a été reconnue comme cruciale dans le cadre d'ontologies en monde ouvert. Les règles existentielles sont équivalentes aux règles de graphes conceptuels étudiées par l'ex équipe RCR, ainsi qu'à des contraintes très expressives étudiées de longue date en bases de données (TGDs) et au récent langage Datalog+/- proposé en bases de données dans le cadre d'OBDA. Enfin, elles généralisent les logiques de description actuellement proposées pour OBDA. La période 2009-2012 a vu l'équipe s'affirmer comme une des références internationales sur ces règles, comme l'ont souligné les experts lors de la première évaluation INRIA de GraphIK (fin 2011) : « *The group is also considered one of the few world-leading teams on decidable rule-based fragments of first-order logic [...] International leadership is further witnessed by many invitations to deliver plenary talks at conferences* ». L'expressivité des règles existentielles va de pair avec la

non-décidabilité des raisonnements associés. Nos travaux ont consisté en plusieurs étapes :

1. La définition de trois critères abstraits de décidabilité (saturation finie, saturation infinie mais à largeur arborescente bornée, réécriture finie) permettant de caractériser les classes décidables de règles connues.
2. La définition de contraintes sur les interactions entre règles permettant de combiner des classes décidables en conservant la décidabilité.
3. L'étude de complexité (classique et complexité de données) pour différentes classes décidables
4. L'étude d'algorithmes génériques (valables pour n'importe quel fragment d'une classe abstraite).



Ces résultats ont été publiés dans la revue Artificial Intelligence [Baget et al. 2011a], ainsi qu'à IJCAI 2009, 2011 [Baget et al. 2011b] & 2013, KR 2010 & 2012, et ils ont obtenu un best paper award en 2012 à la conférence plus spécialisée RR (Web Reasoning and Rule Systems). D'autre part, nous avons étudié les effets de l'introduction de négation atomique dans les requêtes dans un cadre plus simple où l'ontologie est réduite à un ensemble de prédicats préordonnés. Nous avons obtenu des résultats de complexité [Mugnier et al. 2012] et développé des algorithmes que nous avons testés expérimentalement sur des données aléatoires (DEXA 2010 et 2011, ainsi qu'un article à RFIA 2010 ayant reçu le prix AFIA). Enfin, nous avons proposé une extension non monotone des règles de graphes conceptuels basée sur la logique des défauts (article à ICCS 2010 ayant obtenu un best paper award). Nous étendons maintenant nos recherches à la prise en compte d'incohérences dans la base de connaissances, dans le cadre de deux projets ANR démarrés en 2013, ASPIQ\* et PAGODA\*\*.

### Préférences et systèmes argumentatifs pour l'aide à la décision

L'étude de préférences et de systèmes d'argumentation constitue un domaine de l'intelligence artificielle en plein essor. Historiquement, notre intérêt pour ces questions

\* ASPIQ : Techniques ASP pour l'interrogation d'informations web multi-sources hétérogènes à grande échelle

\*\* PAGODA : Practically robust Algorithms for OBDA

a été suscité par des besoins applicatifs en agronomie (collaboration avec INRA-IATE). La problématique applicative générale est celle de l'aide à la décision dans les filières agro-alimentaires, processus complexe allant des matières premières aux produits finis, voire à la gestion des déchets associés. Entrent en jeu de nombreux critères et acteurs aux objectifs généralement conflictuels. Une thèse co-encadrée avec l'IRIT (2007-2010) a proposé un cadre argumentatif appliqué à l'arbitrage entre acteurs des filières. Nous participons actuellement au projet européen **EcoBioCap** (2011-2015, porté par INRA-IATE) dont l'objectif est la conception d'emballages éco-efficaces et biodégradables, notre rôle étant de concevoir un système d'aide au choix d'un emballage pour un produit, prenant en compte des préférences associées à différents critères potentiellement conflictuels (premiers résultats présentés à COMMA 2012 et dans des congrès d'agro-alimentaire).

Sur le plan fondamental, la monographie [Kaci 2011] est consacrée aux formalismes de représentation des préférences et aux raisonnements associés. Les résultats récents ont consisté d'une part en une généralisation de la notion d'acceptabilité d'un ensemble d'arguments sous la forme d'une fonction d'acceptabilité conditionnelle qui a de meilleures propriétés explicatives (COMMA 2012), et en une analyse de différentes sémantiques de préférence relativement à des postulats usuels en logique (ECAI et FLAIRS 2012).

### Intégration sémantique de données

Le problème connu sous le nom de résolution d'entités consiste à déterminer si des références, provenant généralement de diverses sources de données, concernent la même entité du monde réel. C'est une problématique majeure du Web de données. En 2009, nous avons commencé une collaboration sur cette question avec l'ABES (Agence Bibliographique de l'Enseignement Supérieur). Dans le cadre du projet **SudocAd**, financé par le TGE Adonis, nous avons formalisé des métadonnées de l'ABES, de façon à pouvoir les représenter dans les langages du web sémantique, et implémenté un prototype d'identification d'entités au-dessus de CoGUI (cf. ci-dessous). Les résultats expérimentaux ont été jugés suffisamment prometteurs par l'ABES pour qu'elle participe à un financement de thèse et s'engage dans le projet ANR **Qualinca**. Ce projet, démarré au printemps 2011, est porté par GraphIK, et implique également l'INA (Institut National de l'Audiovisuel), le LIG et le LRI. Qualinca vise à développer des mécanismes combinant inférences logiques et méthodes numériques permettant de qualifier le niveau de qualité d'une base documentaire (relativement à la problématique d'identification

d'entités), de l'améliorer ou d'exploiter des bases en tenant compte de leur niveau de qualité.

### Logiciels (voir la liste sur la page Internet GRAPHIK)

- CoGUI est un outil pour construire et contrôler de façon graphique des bases de connaissances. C'est un logiciel libre écrit en Java et développé depuis 2005. Il supporte des bases de graphes conceptuels, tout en étant conçu pour être adaptable à d'autres formalismes graphiques, et possède des import/export en RDF5\*\*\*. Le site web associé fournit les différentes versions du logiciel, la documentation ainsi que des tutoriels.
- Alaska est une plate-forme générique dédiée OBDA. Les données sont stockées indifféremment dans différents systèmes (SGBD relationnels, triple stores RDF, bases de graphes), et les requêtes peuvent être traitées par ces systèmes ou par l'un de nos algorithmes génériques. Kiabora permet d'analyser une base de règles existentielles du point de vue de ses propriétés de décidabilité. Alaska et Kiabora devraient servir de briques logicielles dans notre future plate-forme dédiée OBDA basée sur les règles existentielles (Action de Développement Technologique INRIA 2013).

\*\*\* *Resource Description Framework (langage de base du web sémantique)*

### Visibilité internationale

Cette période s'est caractérisée par de nombreuses conférences invitées dans des colloques internationaux : ICCS 2009 & 2010, ICEIS 2010, RR 2011, Datalog 2.0 2012, RuleML 2012, Effost 2012 et 3 workshops associés à IJCAI 2011, ECAI 2012 et ESWC 2013. Nous avons eu des articles primés aux conférences ICCS 2010, RFIA 2010 et RR 2012 (ces 2 derniers articles étant co-écrits avec des doctorants). Nous sommes par ailleurs régulièrement membres des comités de programme des grandes conférences du domaine comme KR, IJCAI ou ECAI.

### Animation scientifique

#### Expertise

Nous participons aux conseils scientifiques de MIMOS (institut national de R&D de la Malaisie, depuis 2008), de l'ABES (depuis 2010) et du département CEPIA de l'INRA (depuis 2011). Nous avons également exercé une activité de formation pour MIMOS en 2009 et 2010. Nous participons au groupe de travail RDF du W3C depuis 2011.

#### Organisation de colloques

En 2009, nous avons initié une série de workshops associés à IJCAI : « Graph-based Knowledge Representation » (GKR@IJCAI). Nous avons assuré la présidence du comité de programme d'ICCS 2010 (et sommes membres du comité éditorial de cette conférence). En collaboration avec l'équipe Coconut, nous avons organisé ECAI 2012 à Montpellier (responsabilité de la trésorerie et de la publicité, recherche de sponsors, co-webmaster...). Nous participons également à l'animation de la communauté française, via le comité d'organisation des journées



---

d'Intelligence Artificielle Fondamentale (IAF) et divers comités de programme (IC, LFA, RFIA).

## Formation

Nous sommes fortement impliqués dans le master informatique de l'université Montpellier 2 : co-responsabilité du master (depuis 2011 ; env. 350 étudiants), responsabilité du master 1 (2005-2011), co-responsabilité de la spécialité de master Données Connaissances Langues (depuis 2011), porteur du projet de master LMD4 (ainsi que de licence LMD4).

## Faits marquants

---

- Création d'une équipe commune INRIA en 2010
- Résultats de décidabilité, complexité et algorithmes pour de larges fragments des règles existentielles (Datalog+/-)
- De nouvelles collaborations avec des équipes leaders dans les domaines concernés
- Organisation d'ECAI 2012 (20<sup>ème</sup> édition de European Conference on Artificial Intelligence) avec l'équipe Coconut

## Collaborations externes

---

- Karlsruhe Institute of Technology (Sebastian Rudolph) ; articles à IJCAI 2011, KR 2012 et AAMAS 2012
- Univ. Southampton (Talal Rahwan) ; articles à ECAI 2010 et IJCAI 2013 ;
- Univ. Luxembourg (Leon van der Torre); article à COMMA 2012 et co-encadrement de thèse
- Univ. Oxford, Data Management Group (Markus Krötsch et al.) : échanges scientifiques ; action INRIA commune à partir de l'automne 2013
- Collaborations avec des laboratoires de recherche français : CRIL, INRIA-OAK, LERIA, LIG, LRI et LSIS (dans le cadre des projets ANR) ; IRIT (publications, co-encadrement de thèse), INRA-IATE (cf. membres associés)

## Publications majeures

---

- M. Chein et M.-L. Mugnier, Graph-Based Knowledge Representation: Computational Foundations of Conceptual Graphs, Springer, (2009) 427p. [[Chein Mugnier 2009](#)]
  - S. Kaci, Working with Preferences: Less is More, Springer (2011) 202 p. [[Kaci 2011](#)]
  - J.-F. Baget, M. Leclère, M.-L. Mugnier et E. Salvat, On Rules with Existential Variables: Walking the Decidability Line, Artificial Intelligence 175, 9-10 (2011) 1620-1654 [[Baget et al. 2011a](#)]
  - J.-F. Baget, M.-L. Mugnier, S. Rudolph et M. Thomazo, Walking the Complexity Lines for Generalized Guarded Existential Rules, IJCAI'11: International Joint Conference on Artificial Intelligence (2011) [[Baget et al. 2011b](#)]
  - M.-L. Mugnier, G. Simonet et M. Thomazo, On the Complexity of Entailment in Existential Conjunctive First Order Logic with Atomic Negation, Information and Computation 215 (2012) 8-31 [[Mugnier et al. 2012](#)]
-