Validation de graphes conceptuels

Juliette Dibie-Barthélemy*
Ollivier Haemmerlé*,** Eric Salvat***

*INA-PG, 16, rue Claude Bernard, F-75231 Paris Cedex 05
{Juliette.Dibie,Ollivier.Haemmerle}@inapg.inra.fr
**LRI (UMR CNRS 8623 - Université Paris-Sud) / INRIA (Futurs), Bâtiment 490, F-91405 Orsay Cedex
***IMERIR, avenue Pascot, B.P. 2013, F-66011 Perpignan
Salvat@imerir.com

Résumé. Les travaux menés en validation des connaissances visent à améliorer la qualité des bases de connaissances. Le modèle des graphes conceptuels est un modèle de représentation des connaissances de la famille des réseaux sémantiques, fondé sur la théorie des graphes et sur la logique du premier ordre. Nous proposons une solution pour valider sémantiquement une base de connaissances composée de graphes conceptuels. La validation sémantique d'une base de connaissances consiste à confronter ses connaissances à des contraintes certifiées fiables. Nous proposons d'utiliser des contraintes descriptives, exprimées sous forme de graphes conceptuels, qui permettent de poser des conditions sur la représentation de certaines connaissances dans la base. Ces contraintes introduisent une notion de cardinalité et sont soit minimales, soit maximales. Elles permettent respectivement d'exprimer "si A, alors –*au moins* ou *au plus*– **n fois** B". La satisfaction de ces contraintes par une base de connaissances repose sur l'utilisation de l'opération de base du modèle des graphes conceptuels : la projection.

1 Introduction

L'utilisation des Systèmes à Base de Connaissances (SBC) est encore limitée dans l'industrie, du fait de leur manque de fiabilité. Certifier leur qualité est donc un enjeu important pour leur diffusion. Les travaux menés dans le domaine de la validation des connaissances apportent des solutions pour garantir une meilleure qualité des SBC. La validation d'un SBC consiste à fournir des critères de qualité que le système doit vérifier et des outils permettant de vérifier ces critères. La plupart des travaux sur la validation des SBC ont concerné les systèmes à base de règles [Ayel et Rousset, 1990, Loiseau, 1992, Preece et Zlatareva, 1994, Bouali, 1996]. Les autres travaux ont essentiellement porté soit sur la validation de modèles de connaissances [Haouche et Charlet, 1996], soit sur la validation de réseaux sémantiques [Hors et Rousset, 1996]. Le travail présenté dans cet article s'inscrit dans le cadre général de la validation des connaissances. Il porte plus précisément sur la validation de bases de connaissances construites sur le modèle des graphes conceptuels.

Le modèle des graphes conceptuels est un modèle de représentation des connaissances de la famille des réseaux sémantiques. Il a été défini dans une volonté de proposer un modèle formel permettant d'allier l'expressivité de la langue naturelle à la puissance de la logique du premier ordre. Le modèle des graphes conceptuels présente, à nos yeux, deux intérêts essentiels. Le premier intérêt est que ce modèle, par essence graphique, permet à un utilisateur non initié de visualiser aisément les connaissances représentées. Le deuxième intérêt est algorithmique : les opérations de manipulation des connaissances [Mugnier et Chein, 1992] sont issues de la théorie des graphes et peuvent présenter des résultats expérimentaux intéressants en termes d'efficacité [Coulondre et Salvat, 1998].

Nos précédents travaux [Dibie *et al.*, 1998, Dibie-Barthélemy *et al.*, 1999] proposent une solution pour valider sémantiquement une base de connaissances construite sur le modèle des graphes conceptuels. La validation sémantique d'une base de connaissances consiste à confronter les connaissances assertionnelles de la base à des connaissances expertes supplémentaires données dans une optique de validation et certifiées fiables. Ces connaissances expertes sont exprimées sous la forme de *contraintes*, qui représentent une extension du modèle des graphes conceptuels. Nous proposons deux catégories de contraintes : des *contraintes existentielles* qui permettent de représenter de la connaissance qui doit ou ne doit pas se trouver dans la base, et, des *contraintes descriptives* qui permettent de poser des conditions sur la représentation de certaines connaissances dans la base. *Une base de connaissances est* dite *sémantiquement valide* relativement à un ensemble de contraintes existentielles et descriptives si elle satisfait chaque contrainte de l'ensemble.

Dans cet article, nous proposons une nouvelle définition des contraintes descriptives. Les contraintes descriptives permettent d'exprimer que si une connaissance A se trouve dans la base de connaissances, alors la connaissance B doit se trouver dans la base. La connaissance B, qui est de la forme A et C, est représentée par un *graphe contraint*, dont le sous-graphe partiel représentant A est appelé sa tête. L'étude de la satisfaction d'une contrainte descriptive par une base de connaissances repose sur la projection ou non d'un graphe particulier, appelé *graphe contraint n-étendu*, construit à partir du graphe contraint de la contrainte. Nous proposons deux types de contraintes descriptives : les contraintes descriptives minimales et les contraintes descriptives maximales.

L'article est organisé comme suit. La deuxième section fait une brève présentation du modèle des graphes conceptuels [Sowa, 1984] tel qu'il est formalisé dans [Mugnier et Chein, 1996]. La troisième section présente les règles [Salvat et Mugnier, 1996, Salvat, 1997], qui sont une extension du modèle des graphes conceptuels. La quatrième section donne la définition des contraintes descriptives. La cinquième section donne la définition des graphes contraints n-étendus. Les deux dernières sections présentent les contraintes descriptives minimales et maximales et étudient leur satisfaction par une base de connaissances.

2 Le modèle des graphes conceptuels

Une base de connaissances dans le modèle des graphes conceptuels est composée d'un support et de graphes conceptuels.

Le support est un ensemble de connaissances terminologiques permettant de déterminer le vocabulaire de base avec lequel sont représentées les connaissances. Il est composé d'un ensemble de types de concepts partiellement ordonné par la relation d'ordre *sorte_de*, notée $\leq$, d'un ensemble de types de relations, d'un ensemble de marqueurs individuels et d'un marqueur générique noté *, d'une application τ qui associe à tout marqueur individuel son type de concept minimum, et d'une application σ qui associe à tout type de relation le type de concept maximal de chacun de ses arguments.

Les graphes conceptuels permettent de représenter les connaissances factuelles construites sur le support. Un graphe conceptuel est un multigraphe non orienté, biparti, composé d'une classe de sommets concepts et d'une classe de sommets relations. Les arêtes adjacentes à un sommet relation sont totalement ordonnées. Chaque sommet a une étiquette définie par une fonction d'étiquetage : (1) un type de relation pour un sommet relation ; (2) un couple (type de concept, marqueur) pour un sommet concept (tout marqueur individuel doit satisfaire τ et les règles de voisinage de σ doivent être respectées).

Pour raisonner sur les graphes conceptuels, la notion fondamentale est la relation de spécialisation, notée $\leq$. La relation de spécialisation peut être définie de différentes façons [Sowa, 1984], et notamment, à l'aide de l'opération de projection, qui est un morphisme de graphes autorisant une restriction d'étiquettes des sommets.

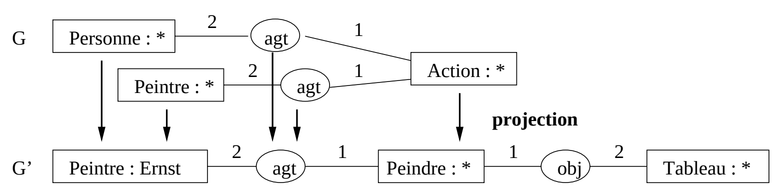

FIG. 1 – G se projette dans G', $G' \leq G$ (G' est plus spécifique que G).

Un graphe conceptuel est *irredondant* s'il ne se projette pas dans l'un de ses sous-graphes stricts. Un graphe conceptuel est *sous forme normale* si chacun de ses marqueurs individuels apparaît exactement une fois.

Dans l'interprétation logique des graphes conceptuels, une correspondance unique est définie entre les graphes conceptuels et la logique du premier ordre. Une formule bien formée Φ(G) de la logique du premier ordre est associée à chaque graphe conceptuel G. Elle est la fermeture existentielle de la conjonction des prédicats associés à chaque sommet de G. Un ensemble de formules logiques Φ(S) est associé au support S.

Exemple 1 *L'interprétation logique du graphe conceptuel G' de la figure 1 est :*
$\Phi(G') = \exists x,y\ Peintre(Ernst) \wedge Peindre(x) \wedge Tableau(y) \wedge agt(x, Ernst) \wedge obj(x, y)$.

La sémantique logique est consistante et complète : une correspondance existe entre la relation de spécialisation et la déduction logique (notée $\vdash$)

Théorème 1 (Consistance [Sowa, 1984] et complétude [Chein et Mugnier, 1992, Mugnier et Chein, 1996] de Φ). Soient $\Phi(S)$ l'ensemble des formules associées au support, G et H des graphes conceptuels sous forme normale. $G \leq H$ si et seulement si $\Phi(S), \Phi(G) \vdash \Phi(H)$.

Dans la suite nous considérons que les graphes conceptuels sont irredondants et sous forme normale. Nous notons $\mathcal{BC}$ une base de connaissances composée d'un support S et d'un graphe conceptuel irredondant sous forme normale Γ. En effet, un graphe conceptuel n'étant pas obligatoirement connexe, nous considérons un ensemble de graphes conceptuels comme un seul graphe conceptuel obtenu par juxtaposition des graphes conceptuels de l'ensemble.

3 Les règles dans le modèle des graphes conceptuels

Une règle est une règle d'inférence de la forme "si G_1 alors G_2", notée $G_1 \Rightarrow G_2$, G_1 et G_2 étant des graphes conceptuels simples avec possibilité de coréférence entre les sommets concepts de G_1 et G_2. Un lien de coréférence lie deux sommets concepts nécessairement de même étiquette. A ces sommets, Φ associe la même variable ou constante.

Définition 1 Une **règle** $R : G_1 \Rightarrow G_2$ est un couple de graphes conceptuels contenant chacun des sommets concepts repérés $(\lambda x_1, ..., x_n\ G_1,\ \lambda x_1, ..., x_n\ G_2)$, où $x_1, ..., x_n$ correspondent aux liens de coréférence entre G_1 et G_2 et sont appelés *points d'attache*. On note x_{i1} (resp. x_{i2}) l'occurence d'un x_i dans G_1 (resp. dans G_2). G_1 est appelé hypothèse de R et G_2 conclusion de R.

L'interprétation logique d'une règle $R : G_1 \Rightarrow G_2$ est définie comme suit : $\Phi(R) = \forall x_1 ... \forall x_n\ \Psi(R)$ avec $\Psi(R) = \Phi(\lambda x_1, ..., x_n\ G_1) \rightarrow \Phi(\lambda x_1, ..., x_n\ G_2)$, où $\rightarrow$ est l'implication logique et où les variables correspondant aux points d'attache entre G_1 et G_2 sont laissées libres dans les deux formules $\Phi(\lambda x_1, ..., x_n\ G_1)$ et $\Phi(\lambda x_1, ..., x_n\ G_2)$.

L'application d'une règle en chaînage avant sur un graphe conceptuel permet d'enrichir les informations contenues dans le graphe avec la connaissance contenue dans la règle.

Définition 2 Une **règle** $R : G_1 \Rightarrow G_2$ **s'applique à un graphe conceptuel** G s'il existe une projection π de G_1 dans G. Le graphe conceptuel résultant, noté $R[G, \pi]$, est obtenu à partir de G et G_2 en fusionnant chaque x_{i2} de G_2 avec $\pi(x_{i1})$.

Le chaînage avant est *adéquate* et *complet*, sa complétude nécessitant que le graphe conceptuel sur lequel est appliquée la règle soit sous forme normale.

Théorème 2 Un graphe conceptuel G est obtenu à partir de $\mathcal{BC}$ par une séquence d'applications de la règle R si et seulement si $\Phi(S)$, $\Phi(\Gamma)$, $\Phi(R) \models \Phi(G)$.

On remarquera que l'application d'une règle sur un graphe conceptuel sous forme normale ne produit pas nécessairement un graphe conceptuel sous forme normale.

E. Salvat donne une condition nécessaire et suffisante pour vérifier que l'application d'une règle sur un graphe conceptuel irredondant apporte de l'information supplémentaire, c'est-à-dire que le graphe résultant et le graphe d'origine ne sont pas équivalents.

Propriété 1 Soient G un graphe conceptuel irredondant et $R : G_1 \Rightarrow G_2$ une règle qui s'applique à G. Soit π_1 la projection de G_1 dans G selon laquelle R est appliquée à G pour construire $R[G, \pi_1]$. $R[G, \pi_1]$ **est équivalent à** G ssi il existe une projection π_2 de G_2 dans G telle que pour chaque point de connexion x_i, $\pi_1(x_{i1}) = \pi_2(x_{i2})$.

On dit qu'une **règle** $R : G_1 \Rightarrow G_2$ **n'est pas applicable à un graphe conceptuel irredondant** G si pour toute projection π_1 de G_1 dans G, $R[G, \pi_1]$ est équivalent à G.

4 Les contraintes descriptives

Les contraintes descriptives permettent d'exprimer que si une connaissance A se trouve dans la base de connaissances, alors la connaissance A et C doit se trouver dans la base.

Définition 3 Une *contrainte descriptive* est un triplet (GC, n, *type*) où GC est un graphe contraint, n un entier strictement positif et *type* le type de la contrainte.

Définition 4 Un *graphe contraint*, noté GC, est un graphe conceptuel irredondant sous forme normale dont un et un seul des sous-graphes partiels non isomorphe à GC est identifié comme étant sa *tête*. Ce sous-graphe partiel est noté head(GC).

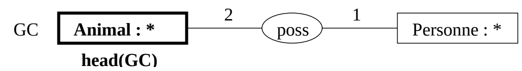

FIG. 2 – graphe contraint GC

Une contrainte descriptive peut être minimale (resp. maximale) et permet d'exprimer : "si on a head(GC), alors il faut avoir *au moins* (resp. *au plus*) n fois GC".

Exemple 2 *Intuitivement, la contrainte descriptive minimale (*GC*,* 2*, min) de la figure 2 signifie que tout animal doit avoir au moins deux maîtres. La contrainte descriptive maximale (*GC*,* 1*, max) signifie que tout animal doit avoir au plus un maître.*

5 Les graphes contraints n-étendus

Une contrainte descriptive (GC, n, *type*) permet d'exprimer que "si on a head(GC), alors il faut avoir –au moins ou au plus– n fois GC". Nous étudions la satisfaction d'une telle contrainte, qui se distingue par l'expression de la cardinalité n, à l'aide d'un graphe particulier, appelé graphe contraint n-étendu, construit à partir de GC. La construction d'un graphe contraint n-étendu nécessite de pouvoir distinguer deux sommets concepts. Nous introduisons pour cela une relation particulière, la relation binaire DIFF de signature ($\top$, $\top$), qui doit vérifier les deux propriétés suivantes :

- la symétrie : deux sommets concepts c_1 et c_2 sont, en effet, distincts si c_1 est différent de c_2 et si c_2 est différent de c_1. Nous proposons de vérifier cette propriété par construction : pour chaque couple de sommets concepts c_1 et c_2 que l'on veut distinguer, les sommets relations DIFF(c_1,c_2) et DIFF(c_2,c_1) sont ajoutés ;
- la non réfléxivité : un sommet concept ne peut pas être distinct de lui même. Pour vérifier cette propriété, nous proposons d'utiliser la contrainte existentielle négative GE^-_{DIFF} suivante, qui doit être satisfaite par tout graphe conceptuel G utilisant la relation DIFF (GE^-_{DIFF} ne doit pas se projeter dans G) :

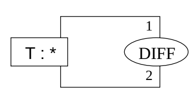

FIG. 3 – contrainte négative GE^-_{DIFF}

Définition 5 Soient un graphe contraint, noté GC_0, et GC_1, GC_2, ..., GC_{n-1} des graphes contraints isomorphes à GC_0. Chaque graphe contraint GC_i, $i \in [0, n-1]$, est composé de m sommets concepts génériques notés c_1^i, ..., c_p^i, c_{p+1}^i, ..., c_m^i où c_1^i, ..., c_p^i sont les sommets concepts génériques de head(GC_i). On appelle *graphe contraint n-étendu* ($n > 0$) le graphe conceptuel GCE_n obtenu comme suit :

1. somme disjointe de GC_0, GC_1, GC_2, ..., GC_{n-1} ;
2. $\forall j \in [1, p]$, fusion de c_j^0 et c_j^1 et ... et c_j^{n-1}, ce qui revient à fusionner tous les sommets concepts qui se correspondent de head(GC_0), ..., head(GC_{n-1}) ;
3. suppression des sommets relations jumeaux ;
4. mise sous forme normale ;
5. $\forall i \in [0, n-2]$, $\forall k \in [i+1, n-1]$, $\forall j \in [p+1, m]$, ajout d'un sommet relation DIFF entre les sommets concepts génériques c_j^i et c_j^k et d'un sommet relation DIFF entre les sommets concepts génériques c_j^k et c_j^i, ce qui revient à ajouter deux sommets relations binaires DIFF de signature ($\top$, $\top$) entre tous les sommets concepts génériques qui se correspondent de GC_0 n'appartenant pas à head(GC_0), ..., de GC_{n-1} n'appartenant pas à head(GC_{n-1}).

On appelle *tête* de GCE_n le sous-graphe partiel de GCE_n limité aux sommets concepts fusionnés en 2 et aux sommets relations de head(GCE_0) les reliant.

On remarquera que le graphe contraint 1-étendu GCE_1 est isomorphe à GC_0.

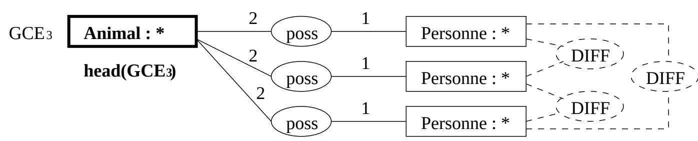

FIG. 4 – graphe contraint 3-étendu obtenu à partir du graphe contraint de la figure 2

Notation 1 *Un couple de sommets relations DIFF est représenté par un seul sommet relation DIFF en pointillé dont les arêtes en pointillé ne sont pas numérotées.*

Propriété 2 Un graphe contraint n-étendu est un graphe contraint.

Preuve 1 La preuve découle de la définition 5.

L'interprétation logique d'un graphe contraint n-étendu est la suivante :
$\Phi_{DIFF}(GCE_n) = \exists\, x_1 \ldots x_p\; x_{p+1}^0 \ldots x_m^0 \ldots x_{p+1}^{n-1} \ldots x_m^{n-1}\; \Phi_{libre}(GCE_n, x_1, \ldots, x_p, x_{p+1}^0, \ldots, x_m^{n-1}) \bigwedge_{i=0}^{n-2} (\bigwedge_{k=i+1}^{n-1} (\bigwedge_{j=p+1}^{m} \text{DIFF}(x_j^i, x_j^k) \wedge \text{DIFF}(x_j^k, x_j^i)))$

- $\Phi_{libre}(GCE_n, x_1, \ldots, x_p, x_{p+1}^0, \ldots, x_m^{n-1})$ représente la formule de variables libres $x_1, \ldots, x_p, x_{p+1}^0, \ldots, x_m^{n-1}$ telle que $\Phi(GCE_n) = \exists\, x_1 \ldots x_p\; x_{p+1}^0 \ldots x_m^{n-1}\; \Phi_{libre}(GCE_n, x_1, \ldots, x_p, x_{p+1}^0, \ldots, x_m^{n-1})$;
- $x_1 \ldots x_p$ sont les variables associées aux sommets concepts génériques de head(GCE_n) ;
- $x_{p+1}^i \ldots x_m^i$, $i \in [0, n-1]$, sont les variables associées aux sommets concepts génériques de GCE_n n'appartenant pas à head(GCE_n).

Exemple 3 *L'interprétation logique du graphe contraint* 3*-étendu* GCE_3 *de la figure 4 est :* $\Phi_{DIFF}(GCE_3)$ = $\exists$ *x, y, z, w Animal(x)* $\wedge$ *Personne(y)* $\wedge$ *poss(y,x)* $\wedge$ *Personne(z)* $\wedge$ *poss(z,x)* $\wedge$ *Personne(w)* $\wedge$ *poss(w,x)* $\wedge$ *DIFF(y,z)* $\wedge$ *DIFF(z,y)* $\wedge$ *DIFF(y,w)* $\wedge$ *DIFF(w,y)* $\wedge$ *DIFF(z,w)* $\wedge$ *DIFF(w,z)*

6 Les contraintes descriptives minimales

Une contrainte descriptive minimale (GC, n, min) permet d'exprimer que si la connaissance représentée par le graphe conceptuel head(GC) apparaît dans la base de connaissances, alors elle doit apparaître au moins n fois dans cette base sous la forme de la connaissance représentée par le graphe contraint GC.

L'étude de la satisfaction d'une contrainte descriptive minimale (GC, n, min) par $\mathcal{BC}$ repose sur la projection du graphe contraint n-étendu GCE_n dans Γ. Ce graphe contraint utilise la relation DIFF pour distinguer ses sommets concepts deux à deux. Pour pouvoir projeter GCE_n dans Γ, il faut donc introduire la relation DIFF dans Γ pour que la distinction implicite entre les sommets concepts de Γ soit comparable à la distinction explicite faite entre les sommets concepts de GCE_n par la relation DIFF.

En fait, comme la relation DIFF n'est, par construction, utilisée dans GCE_n qu'entre des sommets concepts de même type, il suffit d'introduire la relation DIFF dans Γ entre tous les sommets concepts ayant un sous-type de concept commun différent du type $\perp$. Pour cela, nous proposons, pour chaque couple de types de concepts t et t' de l'ensemble des types de concepts ayant un sous-type de concept commun différent de $\perp$, d'appliquer en chaînage avant la règle $RAjout_{DIFF}$ suivante à Γ :

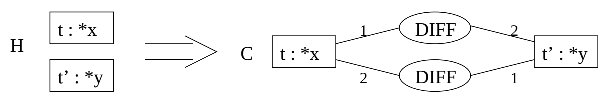

FIG. 5 – règle $RAjout_{DIFF}$ où les types de concepts t et t' ont un sous-type de concept commun différent de $\perp$

Définition 6 On note Γ_{DIFF} le graphe conceptuel obtenu par une séquence d'applications en chaînage avant de la règle $RAjout_{DIFF}$ à Γ, avec arrêt lorsque la règle n'est plus applicable, et, qui satisfait la contrainte existentielle négative GE^-_{DIFF}.

Propriété 3 Le graphe conceptuel obtenu par une séquence d'applications en chaînage avant de la règle $RAjout_{DIFF}$ sur Γ est irredondant et sous forme normale.

Preuve 2 Montrons par l'absurde que le graphe conceptuel, noté G_{res}, obtenu par une séquence d'applications en chaînage avant de la règle $RAjout_{DIFF}$ (cf. figure 5) sur Γ est irredontant. Soit π_1 une projection de H dans G_{res} selon laquelle $RAjout_{DIFF}$ peut être appliquée à G_{res} pour construire $RAjout_{DIFF}[G_{res}, \pi_1]$. Si $RAjout_{DIFF}[G_{res}, \pi_1]$ était un graphe conceptuel redondant, alors il serait équivalent à G_{res} et la règle ne serait pas applicable à G_{res} selon la projection π_1.

La preuve que G_{res} est sous forme normale est immédiate d'après la définition de la règle $RAjout_{DIFF}$, qui n'ajoute pas de nouveaux sommets concepts individuels.

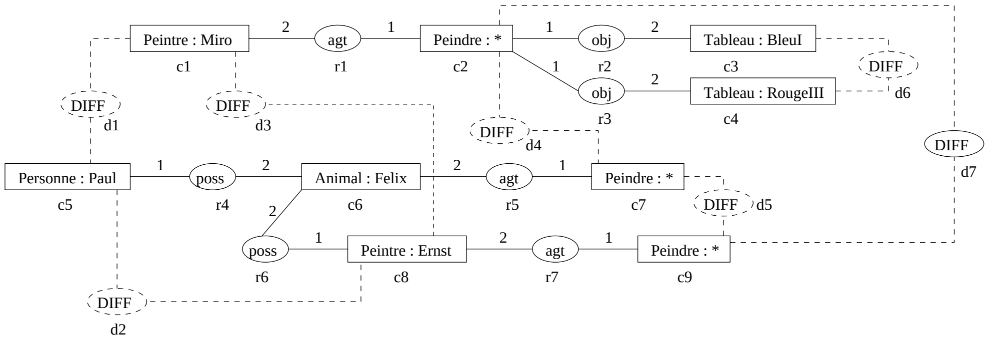

FIG. 6 – Graphe conceptuel Γ_{DIFF}

La définition de la satisfaction d'une contrainte descriptive minimale par une base de connaissances est la suivante :

Définition 7 Une *contrainte descriptive minimale* (GC, n, min) est *satisfaite* par $\mathcal{BC}$ si pour chaque projection Π de head(GCE_n) dans Γ_{DIFF}, il existe une projection Π' de GCE_n dans Γ_{DIFF} telle que $\Pi(\text{head}(GCE_n)) = \Pi'(\text{head}(GCE_n))$.

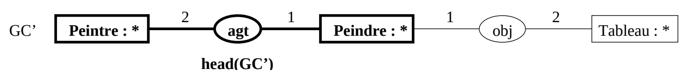

FIG. 7 – graphe contraint GC'

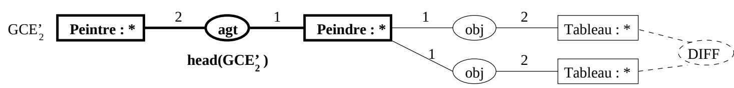

FIG. 8 – graphe contraint 2-étendu obtenu à partir du graphe contraint de la figure 7

Exemple 4 *La contrainte descriptive minimale* $(GC', 2, min)$ *(cf. figure 7), qui signifie que tout peintre qui peint doit peindre au moins 2 tableaux, n'est pas satisfaite par la base de connaissances* $\mathcal{BC}$*. Il existe, en effet, une projection* Π *de head*(GCE'_2) *(cf. figure 8) dans* Γ_{DIFF} *avec* f=Π*(head*(GCE'_2)*) le sous-graphe partiel de* Γ_{DIFF} *limité aux sommets* c_8*,* r_7 *et* c_9*, mais il n'existe pas de projection* Π' *de* GCE'_2 *dans* Γ_{DIFF} *telle que* Π'*(head*(GCE'_2)*)=*f *: le peintre Ernst peint, mais il ne peint rien.*

Pour étudier l'interprétation logique de la satisfaction d'une contrainte descriptive minimale par une base de connaissances, nous reprenons le théorème de J.F. Baget [Baget et Mugnier, 2002], qui propose de considérer une contrainte descriptive minimale comme une règle. Une contrainte descriptive minimale (GC, n, min) peut être considérée comme la règle R_{min} : head$(GCE_n) \rightarrow GCE_n$, dont l'interprétation logique est la suivante :
$\Phi(R_{min}) = \forall\ x_1, \ldots, x_p\ \Psi(R_{min})$
$\Psi(R_{min}) = \Phi_{libre}(\text{head}(GCE_n), x_1, \ldots, x_p) \rightarrow \exists\ x^0_{p+1}, \ldots, x^0_m, \ldots, x^{n-1}_{p+1}, \ldots, x^{n-1}_m$
$\Phi_{libre}\ (GCE_n, x_1, \ldots, x_p)$

Propriété 4 Une contrainte descriptive minimale (GC, n, min) est satisfaite par $\mathcal{BC}$ si et seulement s'il n'existe pas de graphe conceptuel Γ'_{DIFF} tel que $\vdash (\Phi(S) \wedge \Phi(\Gamma_{DIFF}) \wedge \Phi(R_{min}) \rightarrow \Phi(\Gamma'_{DIFF}))$ et $\vdash \neg(\ \Phi(S) \wedge \Phi(\Gamma_{DIFF}) \rightarrow \Phi(\Gamma'_{DIFF}))$.

Preuve 3 La preuve de cette propriété découle du théorème de [Baget et Mugnier, 2002].

7 Les contraintes descriptives maximales

Une contrainte descriptive maximale $(GC,\ n,\ max)$ permet d'exprimer que si la connaissance représentée par le graphe conceptuel head(GC) apparaît dans la base de connaissances, alors elle doit apparaître au plus n fois dans cette base sous la forme de la connaissance représentée par le graphe contraint GC.

Définition 8 Une *contrainte descriptive maximale* (GC, n, max) est *satisfaite* par $\mathcal{BC}$ s'il n'existe pas de projection de GCE_{n+1} dans Γ_{DIFF}.

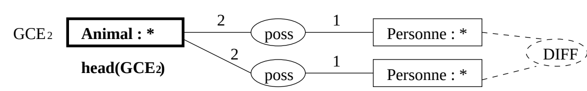

FIG. 9 – graphe contraint 2-étendu obtenu à partir du graphe contraint de la figure 2

Exemple 5 *La contrainte descriptive maximale* $(GC,\ 1,\ max)$ *(cf. figure 2), qui signifie que tout animal doit avoir au plus un maître, n'est pas satisfaite par la base de connaissances* $\mathcal{BC}$*. Il existe, en effet, une projection* Π *de* GCE_2 *(cf. figure 9) dans* Γ_{DIFF} *avec* $\Pi(GCE_2)$ *le sous-graphe partiel de* Γ_{DIFF} *limité aux sommets* c_5*,* r_4*,* c_6*,* r_6*,* c_8 *et* d_2 *: l'animal Felix a deux maîtres, la personne Paul et le peintre Ernst.*

L'interprétation logique de la satisfaction d'une contrainte descriptive maximale par une base de connaissances est la suivante.

Propriété 5 Une *contrainte descriptive maximale* (GC, n, max) est *satisfaite* par $\mathcal{BC}$ si et seulement si $\vdash \neg(\Phi(S) \wedge \Phi(\Gamma_{DIFF}) \rightarrow \Phi(GCE_{n+1}))$.

Preuve 4 La preuve repose sur la non projection de GCE_{n+1} dans Γ_{DIFF}.

8 Conclusion

Dans cet article, nous avons proposé de valider sémantiquement une base de connaissances construite sur le modèle des graphes conceptuels relativement à un ensemble de contraintes descriptives, qui permettent d'exprimer que si une connaissance A se trouve dans la base de connaissances, alors la connaissance A et C doit se trouver dans la base. Les contraintes descriptives peuvent être minimales ou maximales.

Nous avons introduit une nouvelle définition (par rapport à nos travaux antérieurs [Dibie *et al.*, 1998, Dibie-Barthélemy *et al.*, 1999]) des contraintes descriptives pour trois raisons essentielles. Tout d'abord, l'étude de leur satisfaction par une base de connaissances s'appuie sur une interprétation en logique du premier ordre. Ensuite, la notion de cardinalité qu'elles introduisent, "si A, alors –*au moins* ou *au plus*– **n fois** A et C", leur procure une bonne expressivité. Enfin, ces contraintes descriptives permettent de bien étudier la validité sémantique d'une base de connaissances. Elles permettent, en effet, d'étudier à la fois (1) la cohérence de la base : dans les contraintes descriptives maximales "si A, alors *au plus* A et C", l'expression "au plus" signifie qu'on rejette la connaissance supplémentaire qui ne respecte pas la contrainte et (2) la complétude de la base : dans les contraintes descriptives minimales "si A, alors *au moins* A et C", l'expression "au moins" signifie qu'on exige la présence de connaissances minimales dans la base.

Les algorithmes nécessaires à la validation sémantique des graphes conceptuels présentée dans cet article reposent sur l'opération de recherche et d'énumération de toutes les projections d'un graphe conceptuel dans un autre graphe conceptuel, qui est, dans le cas général, un problème NP-complet [Mugnier et Chein, 1992]. Ce résultat restreint l'utilisation de nos algorithmes à des applications de taille limitée. Toutefois, il est à noter qu'il existe des cas polynomiaux de certains des problèmes liés à la projection. Par exemple, la projection d'un arbre dans un graphe peut être réalisée de manière polynomiale [Mugnier et Chein, 1993]. De plus, les applications mettent généralement en œuvre des graphes conceptuels dont les étiquettes des sommets sont suffisamment discriminantes pour que des pretraitements réduisent significativement le risque d'explosion combinatoire [Salvat, 1993].

Nos algorithmes ont été implémentés sur la plate-forme CoGITo, qui est une boîte à outil pour la manipulation des graphes conceptuels [Haemmerlé, 1995]. On remarquera qu'une contrainte descriptive est représentée par un couple de graphes conceptuels : un graphe conceptuel pour représenter le graphe contraint et un graphe conceptuel pour représenter sa tête. La validation sémantique d'une base de connaissances peut se faire contrainte par contrainte ou en fonction d'un ensemble de contraintes combinées à l'aide d'opérateurs de conjonction et de disjonction.

Notre travail a été testé sur une application réelle en accidentologie, obtenue dans le cadre d'une collaboration avec le projet ACACIA de l'INRIA à Sophia Antipolis. Plus précisément, il a été testé sur une base de connaissances composée d'environ cent types de concepts, quatorze types de relations et sept graphes conceptuels de petite taille (4 à 5 sommets concepts et 3 à 5 sommets relations). Nous avons ajouté à cette base de connaissances cinq contraintes que nous avons obtenues à partir d'informations extraites du rapport de présentation de l'application d'accidentologie [Dieng, 1997]. Ce test relève plutôt du jeu d'essai, mais il montre néanmoins l'intérêt de l'approche sémantique dans la mesure où la donnée de contraintes simples permet d'étudier de manière pertinente le comportement d'une base de connaissances. Ce test montre également qu'une telle approche est limitée par l'existence de contraintes de validation qui doivent être données par un expert. Ce dernier point pose le problème de la validité des contraintes, que nous éludons pour le moment car nous supposons que nos contraintes sont fiables, mais qu'il serait intéressant d'étudier.

Une des prochaines étapes de notre travail sera d'utiliser nos contraintes pour valider sémantiquement une base de graphes conceptuels de plus grande ampleur (composée d'environ cent cinquante graphes conceptuels) développée dans le cadre du projet Sym'Previus, projet dont l'objectif est la réalisation d'un outil d'aide à l'expertise en microbiologie prévisionnelles pour la sécurité alimentaire.

Par ailleurs, nos précédents travaux portent sur la réparation d'une base de connaissances construite sur le modèle des graphes, qui n'est pas sémantiquement valide relativement à une ensemble de contraintes [Dibie-Barthélemy *et al.*, 2000]. Nous avons proposé des solutions pour restaurer d'une part la cohérence d'une base non valide relativement à un ensemble de contraintes existentielles négatives et descriptives maximales, et, d'autre part, la complétude d'une base non valide relativement à un ensemble de contraintes existentielles positives et descriptives minimales. Il serait intéressant d'adapter ce travail à nos nouvelles contraintes descriptives et de proposer un processus de réparation global à la fois des incohérences et des incomplétudes d'une base de connaissances [Djelouah *et al.*, 2002].

Références

[Ayel et Rousset, 1990] M. Ayel et M. C. Rousset. *La cohérence dans les bases de connaissances.* Cepadues, 1990.

[Baget et Mugnier, 2002] J. F. Baget et M. L. Mugnier. Extensions of simple conceptual graphs : the complexity of rules and constraints. *Journal of Artificial Intelligence Research*, 16 :425–465, 2002.

[Bouali, 1996] F. Bouali. *Diagnostic, validation et réparation de bases de connaissances : le système KBDR.* PhD thesis, Université PARIS-XI, centre d'Orsay, 1996.

[Chein et Mugnier, 1992] M. Chein et M.L. Mugnier. Conceptual graphs : fundamental notions. *Revue d'Intelligence Artificielle*, 6(4) :365–406, 1992.

[Coulondre et Salvat, 1998] S. Coulondre et E. Salvat. Piece resolution : Towards larger perspectives. In *ICCS'98, LNAI 1453*, pages 179–193, Montpellier, France, 1998. Springer Verlag.

[Dibie-Barthélemy *et al.*, 1999] J. Dibie-Barthélemy, O. Haemmerlé, et S. Loiseau. Constraints for validation of conceptual graphs. *EUROVAV'99*, pages 79–91, june 1999.

[Dibie-Barthélemy *et al.*, 2000] J. Dibie-Barthélemy, O. Haemmerlé, S. Loiseau, et E. Salvat. Validation et réparation sémantique des graphes conceptuels. *RFIA'2000*, pages 205–214, 2000.

[Dibie *et al.*, 1998] J. Dibie, O. Haemmerlé, et S. Loiseau. A semantic validation of conceptual graphs. In *ICCS'98, LNAI*, pages 80–93, France, 1998. Springer Verlag.

[Dieng, 1997] R. Dieng. Comparison of conceptual graphs for modelling knowledge of multiple experts : application to traffic accident analysis. Rapport de recherche 3161, INRIA, Sophia Antipolis, Avril 1997.

[Djelouah *et al.*, 2002] R. Djelouah, B. Duval, et S. Loiseau. Validation and reparation of knowledge bases. In *ISMIS 02, Lecture Notes in Artificial Intelligence 2366*, pages 312–320. Springer, 2002.

[Haemmerlé, 1995] O. Haemmerlé. *CoGITo : une plate-forme de développement de logiciels sur les graphes conceptuels.* PhD thesis, Université Montpellier II, Janvier 1995.

[Haouche et Charlet, 1996] C. Haouche et J. Charlet. KBS validation : a knowledge acquisition perspective. *ECAI'96*, pages 433–437, 1996.

[Hors et Rousset, 1996] P. Hors et M.C. Rousset. Modeling and verifying complex objects : A declarative approach based on description logics. *ECAI'96*, pages 328–332, 1996.

[Loiseau, 1992] S. Loiseau. Refinement of knowledge bases based on consistency. *ECAI'92*, pages 845–849, 1992.

[Mugnier et Chein, 1992] M.L. Mugnier et M. Chein. Polynomial algorithms for projection and matching. In H.D. PFEIFFER, editor, *Proceedings of 7th Annual Workshop on Conceptual Structures, Las Cruces, New Mexico*, pages 49–58, 1992.

[Mugnier et Chein, 1993] M.L. Mugnier et M. Chein. Characterization and algorithmic recognition of canonical conceptual graphs. In *ICCS'93, LNAI 699*, pages 294–311, Canada, 1993. Springer-Verlag.

[Mugnier et Chein, 1996] M.L. Mugnier et M. Chein. Représenter des connaissances et raisonner avec des graphes. *Revue d'Intelligence Artificielle*, 10(1) :7–56, 1996.

[Preece et Zlatareva, 1994] A. Preece et N. P. Zlatareva. A state of the art in automated validation of knowledge-based systems. *Expert Systems with Applications*, 7(2) :151–167, 1994.

[Salvat et Mugnier, 1996] E. Salvat et M.L. Mugnier. Sound and complete forward and backward chainings of graph rules. In *ICCS'96, LNAI 1115, Springer-Verlag*, pages 248–262, Australia, 1996.

[Salvat, 1993] E. Salvat. Algorithmes de projection de graphes conceptuels. Rapport de stage de DEA, Université Montpellier II, LIRMM, Juillet 1993.

[Salvat, 1997] E. Salvat. *Raisonner avec des opérations de graphes : graphes conceptuels et règles d'inférence*. PhD thesis, Université Montpellier II, LIRMM, Décembre 1997.

[Sowa, 1984] J.F. Sowa. *Conceptual structures : information processing in mind and machine*. Addison Wesley Publishing Company, 1984.

Summary

Research works done on knowledge validation aim at enhancing the quality of knowledge basis. Conceptual graph model is a knowledge representation model which belongs to the family of semantic networks, and is based upon the graph theory and first order logic. We give a solution to semantically validate a knowledge base expressed in terms of conceptual graphs. The semantic validation of a knowledge base consists in checking that the knowledge base respects some constraints that are given by an expert. We propose to use descriptive constraints, expressed in terms of conceptual graphs, which allow one to put conditions on the representation of some knowledge in the knowledge base. These constraints introduce a cardinality notion and are either minimal or maximal. They respectively allow one to express "if A, then –*at least* or *at most*– **n times** B". The checking of these constraints by a knowledge base is done by using the projection operation which is the ground operation of the conceptual graph model.