Méthodes formelles pour XML : logique

Pierre-Alain Reynier
http://www.lif.univ-mrs.fr/~preynier/XML/

Plan du cours

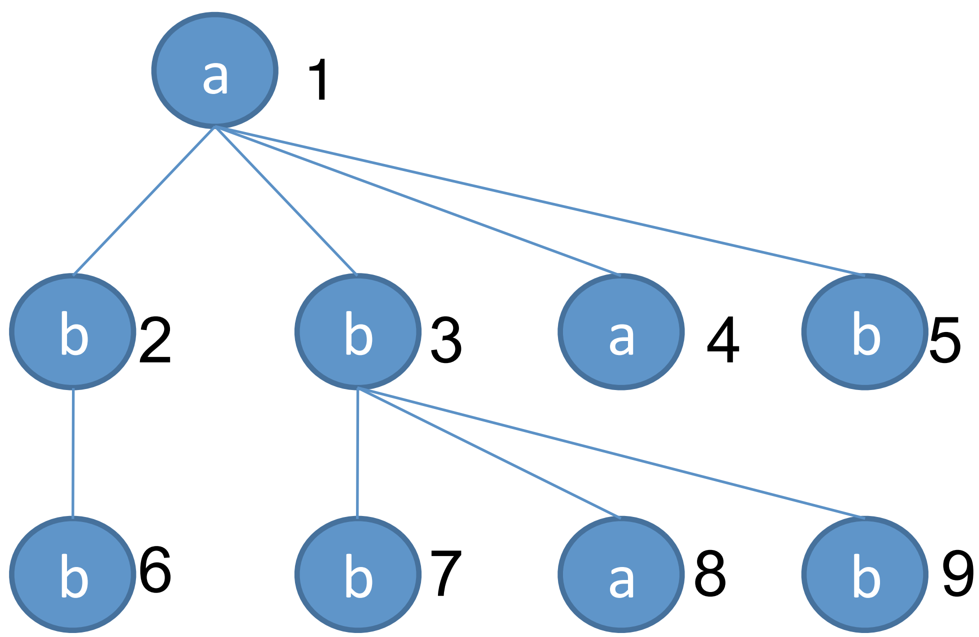
1 - Logique MSO : définition

2 - Utilisation pour XML

I - Logique Monadique du Second Ordre

Logique monadique du 2nd ordre

- Représentation de l'arbre comme une structure logique



E(1,2), E(1,3)… E(3,9)

S(2,3), S(3,4), S(4,5)…S(8,9)

a(1), a(4), a(8)

b(2), b(3), b(5), b(6), b(7), b(9)

Logique monadique du 2nd ordre

E(1,2), E(1,3)… E(3,9)

S(2,3), S(3,4), S(4,5)…S(8,9)

a(1), a(4), a(8)

b(2), b(3), b(5), b(6), b(7), b(9)

Syntaxe de MSO

$$\varphi ::= x = y \mid E(x,y) \mid S(x,y) \mid a(x) \mid \ldots \mid$$
$$\varphi \wedge \varphi \mid \neg \varphi \mid \exists x \varphi \mid$$
$$X \mid X(x) \mid \boxed{\exists X \varphi}$$

Set variable

Quantification over a set variable

Exemples de formule FO

Quelques formules du premier ordre = quantification sur des noeuds et pas des ensembles de noeuds.

- Le noeud x a un fils étiqueté par 'a' et un fils étiqueté par 'b' :

$$\varphi(x) = \exists\ y, z.\ E(x,y) \wedge E(x,z) \wedge a(y) \wedge b(z)$$

- Le père du noeud x (s'il existe) est étiqueté par un 'b' :

$$\forall y, E(y,x) \Rightarrow b(y)$$

Exercices : premier ordre

Ecrivez des formules pour les expressions suivantes :

- tous les fils du noeud x sont étiquetés par ‘b’
- le noeud x est une feuille
- le noeud x est le père du noeud y (formule à deux variables x et y)
- le noeud x est le grand-père du noeud y (formule à deux variables x et y)

Exemple de formule MSO

- Tout noeud ‘a’ a un descendant ‘b’
- Ceci correspond à la formule :

$$\varphi = \forall x\Big(a(x) \to \forall X\big((\alpha \wedge \beta) \to \delta\big)\Big) \quad \textit{where}$$

$$\alpha = X(x)$$

$$\beta = \forall y \forall z(E(y,z) \wedge X(y) \to X(z))$$

$$\delta = \exists y(X(y) \wedge b(y))$$

Pour tout noeud x étiqueté a: tout ensemble X qui
(α) contient x, et qui
(β) est clos par descendant,
contient un noeud étiqueté b

Lien MSO - automates d'arbres

Théorème : pour un ensemble L d'arbres, les propositions suivantes sont équivalentes :

- L = L(A) pour un certain automate ascendant A
  i.e. L est définissable par automate d'arbre ascendant

- L = {arbres T | T satisfait φ} pour une certaine formule φ de MSO
  i.e. L est définissable dans MSO à l'aide des opérateurs S(.,.) et E(.,.)

A propos des relations E* et S*

- * représente la clôture réflexive (xRx) et transitive (xRy et yRz implique xRz)
- E* représente donc la relation “descendant”
- S* représente donc la relation “même profondeur”
- Les relations E* et S* ne peuvent pas être exprimées dans FO, mais peuvent l’être dans MSO (cf TD)

2 - Utilisation pour XML

L’approche logique - automates

- Automates d’arbres = formalisme opérationnel
- Logique = formalisme descriptif, utilisé par exemple pour les langages de requêtes
- La théorie des langages formels unifie les problèmes liés aux automates et aux logiques : expressivité, décidabilité....

Quelques résultats

- Logique du premier ordre (FO) :
  - Satisfaisabilité indéc. en général
  - Décidable pour au plus deux variables : FO2
- Logique du second ordre (MSO) :
  - Décidable dans le cas des opérateurs S(.,.) et E(.,.) grâce aux automates d'arbres

Exemple d'application

Résultat : [Marx 2004]

Core XPath (fragment navigationnel de XPath) peut être étendu pour être complet par rapport à FO2(E*,*S**) et donc décidable