

Enumeration des réponses aux requêtes Xpath conditionnelles avec variables

Antoine VENANT

Equipe projet MOSTRARE Inria Lille Nord Europe

4 septembre 2009

- 1 Introduction
- 2 Preliminaires
- 3 Algorithme de Gottlob et Koch
- 4 Evaluation de Conditional Xpath avec variables
- 5 Énumération avec disjonctions
- 6 conclusion

XML

eXtensible Markup Language

- Langage de mise en forme de données sous forme arborescente.
- Applications nombreuses, page web, base de données(ex :bibliographie),...

```
<?xml version="1.0"?>
  <bib>
    <book>
      <author>
        D. Ervin Knuth
      </author>
      <title>
        The art of computer programming
      </title>
    </book>
  </bib>
```

Xml Path Langage

- Un langage pour la sélection de noeuds au sein d'un document XML.
- Un standard recommandé par le World Wide Web Consortium(W3C).
- Intégré au sein de technologies XML très utilisées : XQuery, XSLT, ...
- Plusieurs versions ont été définies.

Xpath 1.0

- Requêtes **monadiques**. Ex : Tous les livres d'un auteur donné. Dans une bibliographie.
- Une requête exprime un chemin dans le document XML.
- Evaluation efficace.

Xpath 2.0

Deux apports importants :

Xpath 2.0

Deux apports importants :

- Introduction de variables permettant des requêtes *n – aires*. Ex : tous les triplets (livre, auteur, titre) dans une bibliographie.

Xpath 2.0

Deux apports importants :

- Introduction de variables permettant des requêtes *n – aires*. Ex : tous les triplets (livre, auteur, titre) dans une bibliographie.
- Un nouvel opérateur permettant de sélectionner le complément d'une requête élève l'expressivité.

Xpath 2.0

Deux apports importants :

- Introduction de variables permettant des requêtes *n – aires*. Ex : tous les triplets (livre, auteur, titre) dans une bibliographie.
- Un nouvel opérateur permettant de sélectionner le complément d'une requête élève l'expressivité.

Mais une évaluation moins efficace, même sans les variables.

Une autre Extension

Enumeration
des réponses
aux requêtes
Xpath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
Xpath avec
variables

Enumération
avec
disjonctions

conclusion

Conditional Xpath

- Requêtes monadiques.
- Introduit des boucles « tant que » sur les chemins.
- aussi expressif que Xpath 2.0
- aussi facilement évaluable que Xpath 1.0.
- Possibilité d'introduire des variables pour obtenir un langage aussi puissant que Xpath 2.0.

Questions

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Enumération
avec
disjonctions

conclusion

Questions

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Enumération
avec
disjonctions

conclusion

- 1 Conditional XPath avec variables permet-il une meilleure évaluation que XPath 2.0 ?

Questions

Enumeration
des réponses
aux requêtes
Xpath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
Xpath avec
variables

Énumération
avec
disjonctions

conclusion

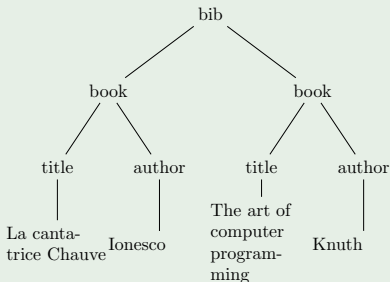
- 1 Conditional Xpath avec variables permet-il une meilleure évaluation que Xpath 2.0 ?
- 2 Peut on énumérer les réponses de façon à éviter de stocker en mémoire un ensemble de $n - upplets$ de noeuds ?

Arbre étiqueté d'arité non borné

Soit Δ un ensemble de symboles d'étiquette. L'ensemble T_Δ des arbres étiquetés dans Δ est le plus petit ensemble tel que

- 1 $a \in T_\Delta \forall a \in \Delta^*$.
- 2 $\forall a \in \Delta, \forall n \in \mathbb{N} \setminus \{0\}, \forall (t_1, \dots, t_n) \in T_\Delta^n, (a, t_1, t_2, \dots, t_n) \in T_\Delta$.

Exemple :



Nodes(*t*)

Enumeration
des réponses
aux requêtes
Xpath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
Xpath avec
variables

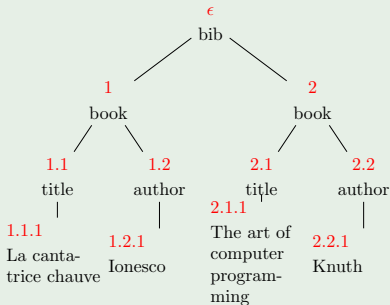
Énumération
avec
disjonctions

Conclusion

$$\text{Nodes}((a)) = \{\epsilon\}$$

$$\text{Nodes}(t = (a, t_1, t_2, \dots, t_n)) = \{\epsilon\} \bigcup_{k=0}^n k.\text{Nodes}(t_k)$$

Ex :



Predicats sur les arbres

Enumeration
des réponses
aux requêtes
Xpath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
Xpath avec
variables

Enumération
avec
disjonctions

conclusion

Prédicats

- 1 prédicats unaires $L = \{lab_a \mid a \in \Delta\}$ (tests d'étiquette).
- 2 prédicats binaires $R = \{\text{child}, \text{child}^*, \text{child}^{-1}, \text{child}^{-1*}, \text{next_sibling}, \text{next_sibling}^*, \text{next_sibling}^{-1*}\}$ (axes).

Sémantique de lab_a

$lab_a(t) = \{x \in Nodes(t) \mid x \text{ est étiqueté } a \text{ dans } t\}$.

sémantique des prédicats

Enumération
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

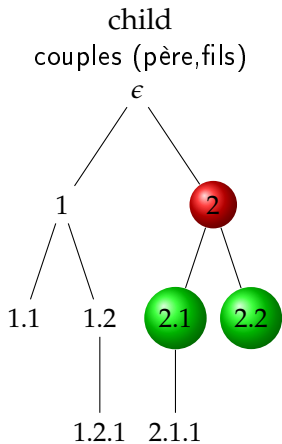
Preliminaires

Algorithme
de Gottlob et
Koch

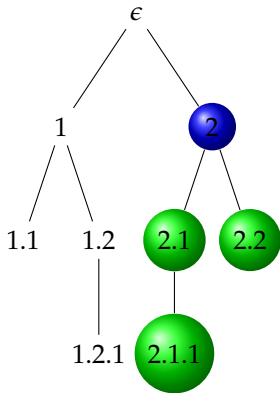
Evaluation
de
Conditional
XPath avec
variables

Enumération
avec
disjonctions

conclusion



child*
fermeture reflexive transitive



sémantique des prédicats

Énumération
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

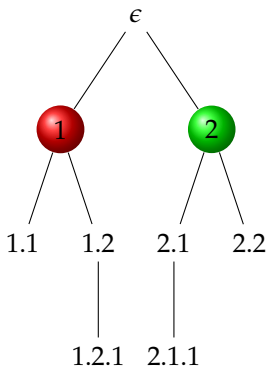
Algorithme
de Gottlob et
Koch

Évaluation
de
Conditional
XPath avec
variables

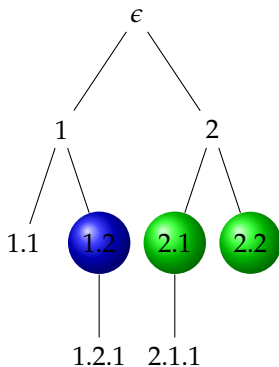
Énumération
avec
disjonctions

conclusion

next_sibling
couples (x, y)
 y premier voisin à droite de x



next_sibling*
fermeture réflexive transitive



Conditional Xpath avec variables

Enumeration
des réponses
aux requêtes
Xpath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
Xpath avec
variables

Énumération
avec
disjonctions

Conclusion

Syntaxe

$$\begin{aligned} \text{CondXPathExpr} &= /[F] \\ F &= r :: l \mid (r[F])^* \mid \text{not } F \mid F[F] \\ &\mid F \text{ or } F \mid \$x \end{aligned}$$

Restrictions sur les variables

- Nécessaires pour une évaluation efficace (comme pour Xpath 2.0).
- pas de variables dans une expression $(r[F])^*$.
- pas de partage de variables dans une expression $F_1[F_2]$ i.e. $\text{Vars}(F_1) \cap \text{Vars}(F_2) = \emptyset$.

Conditional Xpath avec variables

Enumeration
des réponses
aux requêtes
Xpath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
Xpath avec
variables

Enumération
avec
disjonctions

Conclusion

$(a^* b)$

$/[child^* :: *[$x[child[self :: a)]^*[child :: b[$y]]]]$. Selectionne tous les couples (x,y) tels que le chemin de x à y est dans $(a^* b)$.

paires (author,title) dans une bibliographie

```
/[  
  child^* :: book[child :: author[child :: *[$x]]][child ::  
  title[child :: *[$y]]  
]
```

Evaluation des axes

Enumeration
des réponses
aux requêtes
Xpath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
Xpath avec
variables

Enumération
avec
disjonctions

Conclusion

Π un ensemble de noeud de départ.

Evaluation de $child^* :: l$

Expression $child^* :: l$:

- Parcours en profondeur descendant de chaque branche du document.
- Si le noeud visité est dans $lab_l(t) \cap \Pi$ le collecter, passer un drapeau à 1
- collecter tous les noeuds suivants sur la branche.

Algorithme de Gottlob et Koch

Enumeration
des réponses
aux requêtes
Xpath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
Xpath avec
variables

Énumération
avec
disjonctions

Conclusion

Evaluation ensembliste :

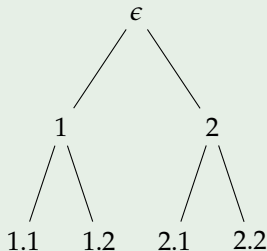
$$\begin{aligned}eval_{\downarrow}(\Pi, r :: l) &= \tilde{r}(\Pi) \cap Lab_l(t) \\eval_{\downarrow}(\Pi, (r[F])^*) &= \tilde{r}_{cond}(\Pi, eval_{\uparrow}(Nodes(t), F)) \\eval_{\downarrow}(\Pi, F_1[F_2]) &= eval_{\downarrow}(eval_{\downarrow}(\Pi, F_1), F) \\eval_{\downarrow}(\Pi, F_1 \cup F_2) &= eval_{\downarrow}(\Pi, F_1) \cup eval_{\downarrow}(\Pi, F_2)\end{aligned}$$

$$\begin{aligned}eval_{\uparrow}(\Pi, r :: l) &= \tilde{r}^{-1}(\Pi \cap Lab_l(t)) \\eval_{\uparrow}(\Pi; (r[F])^*) &= \tilde{r}_{cond}^{-1}(\Pi, eval_{\uparrow}(Nodes(t), F)) \\eval_{\uparrow}(\Pi, F_1[F_2]) &= eval_{\uparrow}(\Pi \cap eval_{\uparrow}(Nodes(t), F), P) \\eval_{\uparrow}(\Pi, F_1 \cup F_2) &= eval_{\uparrow}(\Pi, F_1) \cup eval_{\uparrow}(\Pi, F_2)\end{aligned}$$

- 1 Associer un ensemble d'affectations « sûres » à chaque position dans l'arbre syntaxique de la requête.
- 2 Double passage avec l'algorithme de Gottlob et Koch.

Ex :

`child* :: *[$x[child :: *[$y[child :: *[$z]]]]]` sur le document :



Exemple

Enumeration
des réponses
aux requêtes
Xpath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

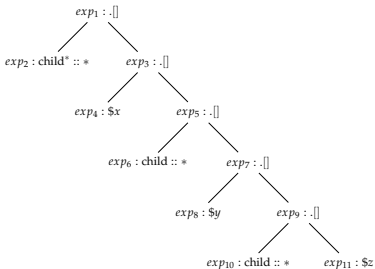
Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
Xpath avec
variables

Enumération
avec
disjonctions

conclusion



Exemple

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

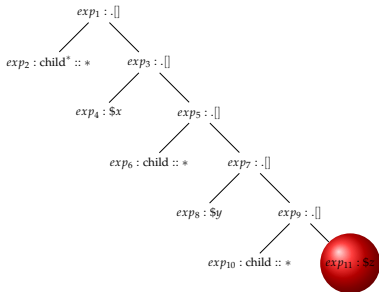
Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Enumération
avec
disjonctions

conclusion



- $E_{exp_{11}} = Nodes(t)$

Exemple

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

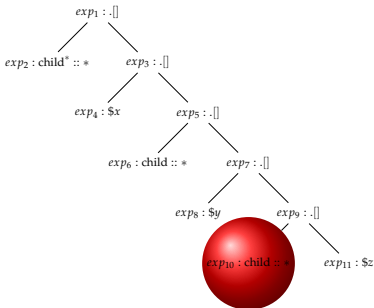
Introduction
Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Énumération
avec
disjonctions

conclusion



- $E_{exp_{11}} = Nodes(t)$
- $E_{exp_{10}} = \{\epsilon, 1, 2\}$

Exemple

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

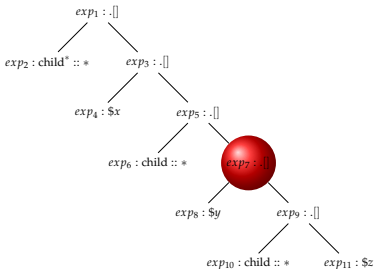
Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Enumération
avec
disjonctions

conclusion



- $E_{exp_{11}} = Nodes(t)$
- $E_{exp_{10}} = \{\epsilon, 1, 2\}$
- $E_{exp_7} = E_{exp_8} = E_{exp_9} = E_{exp_{10}}$

Exemple

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

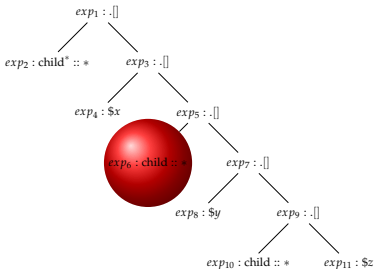
Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Énumération
avec
disjonctions

conclusion



- $E_{exp_{11}} = Nodes(t)$
- $E_{exp_{10}} = \{\epsilon, 1, 2\}$
- $E_{exp_7} = E_{exp_8} = E_{exp_9} = E_{exp_{10}}$
- $E_{exp_6} = \{\epsilon\}$

Exemple

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

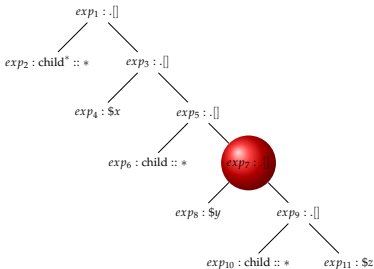
Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Enumération
avec
disjonctions

conclusion



- $E_{exp_{11}} = Nodes(t)$
- $E_{exp_{10}} = \{\epsilon, 1, 2\}$
- $E_{exp_7} \leftarrow E_{exp_7} \cap \{1, 2\} = \{1, 2\}$
- $E_{exp_6} = \{\epsilon\}$

Exemple

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

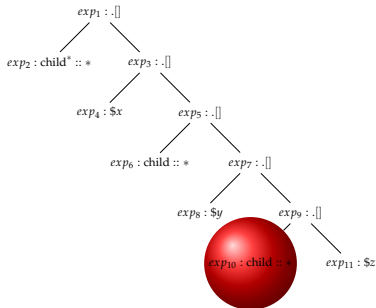
Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Énumération
avec
disjonctions

conclusion



- $E_{exp_{11}} = Nodes(t)$

- $E_{exp_{10}} \leftarrow E_{exp_{10}} \cap \{1,2\} = \{1,2\}$

- $E_{exp_7} \leftarrow E_{exp_7} \cap \{1,2\} = \{1,2\}$

- $E_{exp_6} = \{\epsilon\}$

Exemple

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

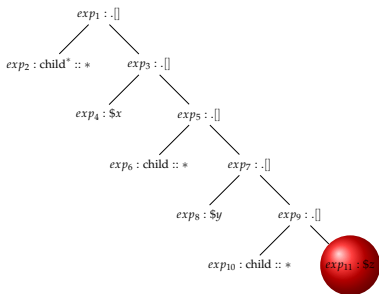
Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Enumération
avec
disjonctions

conclusion



- $E_{exp_{11}} \leftarrow E_{exp_{11}} \cap \{1.1, 1.2, 2.1, 2.2\} = \{1.1, 1.2, 2.1, 2.2\}$
- $E_{exp_{10}} \leftarrow E_{exp_{10}} \cap \{1, 2\} = \{1, 2\}$
- $E_{exp_7} \leftarrow E_{exp_7} \cap \{1, 2\} = \{1, 2\}$
- $E_{exp_6} = \{\epsilon\}$

Backtracking

Enumeration
des réponses
aux requêtes
Xpath condi-
tionnelles
avec
variables

Antoine
VENANT

introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
Xpath avec
variables

Enumération
avec
disjonctions

conclusion

- Toutes les positions auxquelles on trouve un symbole de variable \$x définit un point de backtrack.

Backtracking

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Énumération
avec
disjonctions

Conclusion

- Toutes les positions auxquelles on trouve un symbole de variable $\$x$ définit un point de backtrack.
- On part de b le point de backtrack courant. On retire la première valeur π_1 de E_b .

Backtracking

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Énumération
avec
disjonctions

Conclusion

- Toutes les positions auxquelles on trouve un symbole de variable $\$x$ définit un point de backtrack.
- On part de b le point de backtrack courant. On retire la première valeur π_1 de E_b .
- Passage descendant de b à partir de $\{\pi_1\}$.

Backtracking

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Énumération
avec
disjonctions

Conclusion

- Toutes les positions auxquelles on trouve un symbole de variable $\$x$ définit un point de backtrack.
- On part de b le point de backtrack courant. On retire la première valeur π_1 de E_b .
- Passage descendant de b à partir de $\{\pi_1\}$.
- On s'arrête à b' le point suivant de backtrack. On fait $E_{b'} \leftarrow E_b \cap \Pi$ (Π ensemble de noeuds courant).

Backtracking

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Énumération
avec
disjonctions

Conclusion

- Toutes les positions auxquelles on trouve un symbole de variable $\$x$ définit un point de backtrack.
- On part de b le point de backtrack courant. On retire la première valeur π_1 de E_b .
- Passage descendant de b à partir de $\{\pi_1\}$.
- On s'arrête à b' le point suivant de backtrack. On fait $E_{b'} \leftarrow E_b \cap \Pi$ (Π ensemble de noeuds courant).
- On recommence avec b' comme point courant.

conjonctions problématiques

Enumeration
des réponses
aux requêtes
Xpath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
Xpath avec
variables

Enumération
avec
disjonctions

conclusion

Exemple :

```
/(child* :: *[child-1 :: *[$x]])[next_sibling* :: *[$y]]
```

Problème

- il existe un noeud π tel que
 - π à un parent x .
 - π à un voisin à droite y .
- Deux affectations différentes pour π mène à la même solution :
 - 1 $(\pi \leftarrow 1.2, x \leftarrow 1, y \leftarrow 1.2)$
 - 2 $(\pi \leftarrow 1.1, x \leftarrow 1, y \leftarrow 1.2)$

conjonctions problématiques

Enumeration
des réponses
aux requêtes
Xpath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
Xpath avec
variables

Énumération
avec
disjonctions

Conclusion

Énumération récursive

Pour une position $p : exp_1[exp_2]$ problématique (par exemple) :

$$(child^* :: *[child^{-1} :: *[$X]])[ns^* :: *[$Y]]$$

- 1 Enumérer une solution pour exp_1 . Ex : $x \leftarrow 1$
- 2 Faire un passage descendant sur exp_1 en imposant cette valuation pour obtenir un ensemble Π de noeud compatibles avec cette solution. L'exemple donne $\{1.1, 1.2\}$.
- 3 Faire un passage depuis Π sur exp_2 pour obtenir des solutions pour exp_2 compatibles avec celle choisie pour exp_1 . L'exemple donne $y \leftarrow \{1, 1, 1.2, 2.1, 2.2\}$.
- 4 Recommencer avec une autre solution.

disjonctions $exp_1 \cup exp_2$

Enumeration
des réponses
aux requêtes
Xpath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
Xpath avec
variables

Énumération
avec
disjonctions

Conclusion

- 1 choisir exp_1
- 2 énumérer toutes les solutions
- 3 choisir exp_2
- 4 énumérer toutes les solutions.
- 5 faire l'union des deux ensembles

Complexité

Traitement d'une requête Q en $O(|Q||t||Q(t)||\log(|Q(t)|)|)$

Fragment du langage

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Enumération
avec
disjonctions

conclusion

Une seule disjonction, placée au sommet

$$\begin{aligned} \textit{CondXPathExpr} &= /[F] | \textit{CondXPathExpr} \cup \textit{CondXPathExpr} \\ F &= r :: l | (r[F])^* | F[F] | \textit{not } F | \$x \end{aligned}$$

Pour énumérer exp_1 or exp_2

- Calculer les tailles des ensembles de réponses pour les deux expressions.
- Si exp_1 a plus de solutions,
- Enumérer une réponse pour exp_1 , regarder si elle est solution dans exp_2 .
- Si oui la laisser pour plus tard.
- Enumérer une solution pour exp_2 .

Calcul de cardinalité

Enumeration
des réponses
aux requêtes
Xpath condi-
tionnelles
avec
variables

Antoine
VENANT

Introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
Xpath avec
variables

Enumération
avec
disjonctions

conclusion

- Associer à chaque noeud de E_{exp} le nombre de solutions
- Calculer ce nombre récursivement, au cours du backtrack.
- Utiliser la même astuce que précédemment pour traiter les conjonctions problématiques.

Enumeration
des réponses
aux requêtes
XPath condi-
tionnelles
avec
variables

Antoine
VENANT

introduction

Preliminaires

Algorithme
de Gottlob et
Koch

Evaluation
de
Conditional
XPath avec
variables

Enumération
avec
disjonctions

conclusion

- Resultat positif pour Conditional XPath avec variables.

- Resultat positif pour Conditional Xpath avec variables.
- Mais décevant vis à vis de l'énumération.

- Resultat positif pour Conditional Xpath avec variables.
- Mais décevant vis à vis de l'énumération.
- Un algorithme d'énumération cependant.

- Resultat positif pour Conditional Xpath avec variables.
- Mais décevant vis à vis de l'énumération.
- Un algorithme d'énumération cependant.
- Pour un fragment assez restrictif du langage.