

Analyse incrémentale des usages pour le routage des requêtes dans les systèmes pairs à pairs

Taoufik YEFERNY*, Khedija AROUR**

*Faculté des Sciences de Tunis
Campus Universitaire, Tunis 1060, Tunisie
yeferny.taoufik@gmail.com

**Institut National des Sciences Appliquées
et de Technologie de Tunis, 1080 Tunisie
Khedija.arour@issatm.rnu.tn

1 Introduction

Les systèmes de recherche d'information en pair-à-pair (SRI P2P) constituent des grilles d'informations regroupant des millions des pairs connectés et tout pair participant peut partager un ensemble de ses documents comme il peut aussi accéder à n'importe quel document ou fichier partagé par les autres pairs. Par conséquent, la sélection des pairs qui peuvent répondre à ses requêtes devient un problème très important. Ce problème est appelé problème de routage des requêtes dans les SRI P2P. Nous présentons dans ce papier une amélioration de notre algorithme de routage des requêtes par apprentissage LearningPeerSelection (*LPS*) Yeferny et Arou (2009).

2 Présentation de l'algorithme *LPS*

La méthode de sélection des pairs de l'algorithme *LPS* supporte des nouvelles connaissances désignées par profil. Les profils représentent des corrélations sémantiques entre les requêtes passées et les pairs positifs. En effet, le but est de trouver des groupements des requêtes qui partagent les mêmes termes et des groupements des pairs qui ont agi positivement à un ensemble de requêtes. Pour ce faire, il a eu recours à une des techniques de l'extraction et gestion de connaissances (EGC), l'Analyse Formelle de Concepts (FCA), qui répond bien à ce problème. Deux contextes formels sont utilisés, le premier représente le lien entre les requêtes passées et les termes associés, appelé contexte $C1$. Le second, représente le lien entre les requêtes passées et les pairs positifs, appelé contexte $C2$. Un algorithme de génération de concepts formels est par la suite appliqué pour générer deux ensembles de concepts, notés par $E1$ et $E2$. Ces deux ensembles constituent une base $B(E_1, E_2)$ qui servira par la suite comme une base de connaissances pour l'algorithme de sélection des pairs.

3 Stratégie de mise à jour de la base de connaissances

Les bases des connaissances sont mises à jour après l'émission d'un certain nombre de requêtes. Pour ce faire, nous avons défini une nouvelle stratégie incrémentale de maintenance de ces bases. L'idée consiste à générer une base B_+ à partir de l'historique des nouvelles requêtes (les requêtes émises après la dernière opération de mise à jour). Par la suite, nous construisons la nouvelle base de connaissances en faisant l'union de B_+ et l'ancienne base. Avec cette stratégie, le nombre d'objets et d'attributs nécessaires pour la construction de la base B_+ est largement inférieur à celui de la stratégie statique Yeferny et Arour (2009).

4 Amélioration de l'algorithme *LPS*

```

1 Algorithme : LEARNINGPEERSELECTION( $B, Q, P_{max}$ )
2 Entrées :
3    $B$  : Base de connaissances ;
4    $Q$  : Requête ;
5    $P_{max}$  : Nombre maximum des pairs à sélectionner ; Sortie :
6    $selectedPeers$  : Liste des pairs sélectionnés
7 début
8    $selectedPeers = \emptyset$ 
9    $concept = getConcept(B.E1, Q)$ 
10  while  $concept$  et  $|selectedPeers| < P_{max}$  do
11     $SQPC := getSimilarConcept(B.E2, Extent(concept))$ 
12     $selectedPeers := selectedPeers \cup getSelectedPeers(SQPC)$ 
13     $B.E1 := B.E1 - \{concept\}$ 
14     $concept := getConcept(B.E1, Q)$ 
15  Return ( $selectedPeers$ )
16 fin

```

Algorithme 1 : NOUVELLE VERSION DE L'ALGORITHME DE SÉLECTION DE PAIRS

La complexité de l'algorithme *LPS* présenté dans Yeferny et Arour (2009) est de $O(|E_2| \times |E_1|)$, avec $|E_i|$ dénote la cardinalité de l'ensemble E_i . En effet, le problème majeur de cette solution est l'exploration importante d'un espace de recherche alors que la contrainte du seuil de nombre maximum des pairs à sélectionner n'a pas été exploitée. Pour réduire cet espace de recherche, nous proposons de considérer cette contrainte dans le processus de construction des pairs sélectionnés en parallèle avec la sélection des concepts similaires à Q (voir Algorithme1). La complexité est de $O(P_{max} \times (|E_1| + |E_2|))$ qui est inférieure à celle de l'algorithme *LPS*.

Références

Yeferny, T. et K. Arour (2009). Routage sémantique des requêtes dans les systèmes pair-à-pair. In *CORIA*, pp. 131–147.