

SAGE : préemption Web pour les services publics d'évaluation de requêtes SPARQL?

Thomas Minier, Hala Skaf-Molli, Pascal Molli

▶ **To cite this version:**

Thomas Minier, Hala Skaf-Molli, Pascal Molli. SAGE : préemption Web pour les services publics d'évaluation de requêtes SPARQL?. 30es Journées Francophones d'Ingénierie des Connaissances, IC 2019, AFIA, Jul 2019, Toulouse, France. pp.141. hal-02329807

HAL Id: hal-02329807
https://hal.science/hal-02329807

Submitted on 23 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SaGe : préemption Web pour les services publics d'évaluation de requêtes SPARQL⋆

Thomas Minier, Hala Skaf-Molli and Pascal Molli

LS2N – University of Nantes, Nantes, France
{thomas.minier,hala.skaf,pascal.molli}@univ-nantes.fr

Suivant les principes du Linked Open Data (LOD), les fournisseurs de données hébergent publiquement des millions de triples au format RDF. Cependant, fournir un service public qui permet à n'importe qui d'exécuter n'importe quelle requête SPARQL sur ces données est toujours un problème ouvert. Comme ces services sont soumis à une charge imprévisible de requêtes, le défi est d'assurer qu'ils demeurent *stables* malgré des variations en termes de taux d'arrivées des requêtes et des ressources nécessaires à leur évaluation.

Pour résoudre ce problème, la plupart des fournisseurs de données appliquent une politique d'utilisation équitable des serveurs basée sur des *quotas*. Ces derniers visent à empêcher les *effets convois* (Blasgen *et al.*, 1979), *c.a.d*, une requête longue à exécuter bloque l'évaluation des autres. Le principal défaut de cette politique est qu'elle empêche les requêtes interrompues de délivrer des résultats complets. Cela constitue une limite sérieuse pour les utilisateurs du LOD, qui peuvent vouloir exécuter des requêtes longues.

Nous pensons que le problème lié aux quotas ne réside pas dans l'interruption des requêtes, mais dans l'impossibilité pour les clients de *reprendre leur exécution* après interruption. Néanmoins, il n'existe pas de modèle de préemption pour le Web qui permet la suspension et la reprise de l'exécution de requêtes SPARQL. Dans (Minier *et al.*, 2019), nous proposons SaGe, un moteur d'évaluation de requête SPARQL basé sur la *préemption Web*. Il permet à un serveur Web de suspendre une requête en cours d'exécution après un certain temps, puis de reprendre son exécution ultérieurement. Une fois suspendue, l'état d'une requête est retourné au client, qui peut reprendre son exécution en renvoyant l'état au serveur.

La préemption Web engendre des coûts supplémentaires pour le serveur Web, qui doit suspendre la requête courante puis reprendre l'exécution de la suivante. En conséquences, le problème scientifique majeur est de maintenir ce surcoût marginal, quelle que soit la requête, afin d'assurer une exécution performante. Nos contributions sont les suivantes :

— Nous formalisons le modèle de *préemption Web* qui permet de suspendre et de reprendre l'exécution de requêtes SPARQL. Nous définissons aussi un ensemble d'opérateurs d'exécution préemptifs, dont la complexité d'arrêt et de reprise est bornée.

— Nous proposons SaGe, un moteur d'évaluation de requêtes SPARQL composé d'un serveur Web préemptif et d'un client Web intelligent qui permet l'évaluation de n'importe quelle requête SPARQL. Nous comparons ensuite les performances de ce nouveau moteur de requêtes avec les approches existantes. Nos résultats expérimentaux démontrent que SaGe surpasse de plusieurs ordres de grandeurs les approches existantes en termes de temps moyen d'exécution des requêtes et de temps d'obtention des premiers résultats.

Références

Blasgen M. W., Gray J., Mitoma M. F. & Price T. G. (1979). The convoy phenomenon. *Operating Systems Review*, **13**(2), 20–25.

Minier T., Skaf-Molli H. & Molli P. (2019). SaGe : Web Preemption for Public SPARQL Query Services. In *The World Wide Web Conference 2019 (WWW'19)*, San Francisco, United States.

⋆. Article complet publié sous le titre "SaGe: Web Preemption for Public SPARQL query services" dans les actes de *The World Wide Web Conference 2019 (WWW'19)*.