

Résumé

Mots-clés :

Abstract

Keywords :

Table des figures

Liste des tableaux

Liste des abréviations

Table des matières

Introduction Générale	2
1 Web et données sémantique	3
1.1 RDF	3
1.2 LINKED DATA (DONNÉES LIÉES)	4
1.3 Edition des données sémantiques	4
1.4 CRDT (type de donnée répliqué commutatif)	4
1.5 Les algorithmes d'édition de données sémantique	4
1.5.1 OR-Set	4
1.5.2 C-SET [4]	4
1.5.3 SU-Set	4
1.5.4 B-SET [5]	5
1.5.5 srCE	5
1.5.6 LD-SET	5
1.6 Etude comparative entre ces algorithmes	5
Conclusion	5

Introduction Générale

Dans le Web sémantique et dans le Web en général, un problème fondamental est la comparaison et l'appariement des données et la capacité de résoudre la multiplicité des références de données aux mêmes objets du monde réel, en définissant les correspondances entre les données sous forme de liens de données. La tâche de couplage de données devient de plus en plus importante à mesure que le nombre de données structurées et semi-structurées disponibles sur le Web augmente. La transformation du Web d'un « Web de documents » en un « Web de données », ainsi que la disponibilité d'importantes collections de données générées par des capteurs (« Internet des objets »), mènent à une nouvelle génération d'applications Web fondées sur l'intégration des données et des services. Parallèlement, de nouvelles données sont publiées chaque jour à partir de contenus générés par les utilisateurs et de sites Web publics. mène à une nouvelle génération d'applications Web En termes généraux, le couplage de données est la tâche de déterminer si deux descriptions d'objets peuvent être liées l'une à l'autre pour représenter le fait qu'elles se réfèrent au même objet du monde réel dans un cas décrivant les objets du monde réel à travers des sources de données hétérogènes, en supposant que la similarité entre deux descriptions de données est plus élevée, plus grande est la probabilité que la liaison inclut également la tâche de définir des méthodes, des techniques et des outils (semi-)automatisés pour Dans ce contexte, l'une des initiatives les plus importantes dans le domaine du Web sémantique est le lien entre les grandes collections de données déjà disponibles sur le Web [1]. exemple de données liées montre comment la tâche de liaison de données est cruciale sur le Web de nos jours. les méthodes et les techniques de liaison des données sur le Web sémantique. En outre, il existe d'importants travaux décrivant des domaines de recherche très proches des caractéristiques de liaison de données sémantiques qui nécessitent des solutions spécifiques tant en termes de nouvelles techniques que dans Par exemple, d'une part, le couplage de données nécessite de traiter la complexité sémantique typique du couplage ontologique, mais d'autre part, la grande quantité de données disponibles du champ de couplage de données sur le Web sémantique. Dans ce document, nous donnons une définition générale de ce domaine, afin de souligner les problèmes et de décrire les solutions. Nous allons mieux définir le problème de liaison de données, en discutant également de nombreux algorithmes. L'objectif de ce projet est d'étudier les approches existantes en matière d'édition collaborative des données sémantiques, afin de faire une comparaison collaborative entre ces approches, et de proposer un éditeur collaboratif distribué pour les données sémantiques capables de supporter des groupes dynamiques où les utilisateurs peuvent se joindre et partir à tout moment.[2]

Chapitre 1

Web et données sémantique

Le Web sémantique Pour bien comprendre , nous commençons par une définition. Le grand dictionnaire terminologique définit le terme sémantique comme « l'ensemble des relations entre les caractères, ou groupes de caractères, et leur significations, indépendamment de la façon de les employer ou de les interpréter. » Il précise par la suite que « si, en linguistique, la sémantique porte sur l'étude du sens à partir de la combinaison des mots, en intelligence artificielle, elle porte sur la capacité d'un réseau [le Web] à représenter de la manière la plus humaine possible des relations entre des objets, des idées ou des situations. » Le terme sémantique implique donc que la machine ne se contentera plus de présenter visuellement les données du Web, mais, en les reliant, elle pourra conserver les significations qui leur sont attribuables. Or, en transformant le contenu du Web pour qu'il soit « compréhensible » par la machine et non seulement présentable, nous permettons à cette même machine d'être plus efficace dans le traitement de l'information. Ainsi, le dialogue avec les moteurs de recherche devient possible. Nous sommes alors en mesure de nous exprimer dans des termes que nos ordinateurs peuvent aussi interpréter et échanger. Il est également possible d'automatiser, d'intégrer et de réutiliser l'information entre diverses applications. Le Web sémantique est décrit généralement comme un Web destiné aux machines. Disposer d'un Web dont le contenu est abordable par les machines peut apporter de grands bénéfices : L'automatisation de nombreuses tâches fondées sur le contenu comme la recherche de ressources ayant un contenu particulier, la comparaison du contenu de ressources (pages, bases de données, ontologies, etc. . .). Le Web sémantique permettrait de résoudre la relative difficulté de trouver de l'information sur le web.[3] Le Web sémantique a pour objectif de transformer le World Wide Web actuel, entièrement tourné vers la présentation des documents, vers un Web dont le contenu serait compréhensible par les machines. La vision s'appuie sur l'utilisation d'ontologies. Le Web sémantique n'est pas un Web séparé mais une extension du Web actuel, dans lequel l'information est bien définie, permettant ainsi aux ordinateurs et aux personnes de travailler en coopération. Le Web sémantique permettra aux machines de comprendre les documents et les données sémantiques, et non la parole humaine et les écrits.[3]

1.1 RDF

RDF est un acronyme de Resource Description Framework. RDF est une norme W3C des technologies Web sémantiques et la base de l'architecture standardisée des technologies Web Sémantique. Le RDF est un langage simple pour exprimer des modèles de données sous forme d'objets « ressources » et de leurs relations.

1.2 LINKED DATA (DONNÉES LIÉES)

Le Web est de plus en plus considéré comme un espace d'information global constitué non seulement de documents liés, mais aussi de données liées. Plus qu'une simple vision, le Web of Data qui en résulte est né de la maturation de la pile de technologies du Web sémantique et de la publication d'un nombre croissant d'ensembles de données selon les principes des Données Liées.[1]

1.3 Edition des données sémantiques

1.4 CRDT (type de donnée répliqué commutatif)

C'est un type de données répliqué pour lequel certaines propriétés mathématiques simples garantissent une cohérence éventuelle. Dans l'État style, les états successifs d'un objet devraient former un semi-réseau monotone, avec fusionner le calcul d'une limite inférieure. Dans le style op-based, concurrent En supposant seulement que le sous-système de communication assure la livraison éventuelle (en ordre causal pour les objets basés sur les opérations), les CRDT sont garantie de converger vers un état commun et correct, sans nécessiter synchronisation.

1.5 Les algorithmes d'édition de données sémantique

1.5.1 OR-Set

OR-Set est conçu pour la mise à jour basée sur les éléments, il n'est pas adapté pour les opérations basées sur les modèles comme les opérations de mise à jour SPARQ Afin de réduire la complexité de la communication, au lieu d'étiqueter chaque élément séparément Remarquez que dans le cas d'OR-Set, le nombre de tours serait égal à le nombre de triples exploités, car nous devons envoyer un message pour chaque un. Les triples en double n'affectent pas la complexité des tours, ils sont traités comme un insertion normale.

1.5.2 C-SET [4]

est un CRDT conçu pour le type de données de l'ensemble. compteur associé à chaque élément pour garder une trace du nombre de fois qu'il a été ajouté et supprimé. Un élément est considéré comme membre du C-Set si son compteur est supérieur à zéro. C-Set est convergent, laisse la causalité au sous-jacent réseau, mais malheureusement, quand une suppression suivie d'un insert sont exécutés l'intention n'est pas préservée. La figure 4 présente un exemple de violation de l'intention dans l'ensemble C. À partir d'un point où un élément x a 3 dans son compteur et donc, pas un membre de l'ensemble. Si deux nœuds décident simultanément pour insérer x et le supprimer immédiatement, ils convergeront vers un état où x est un membre de l'ensemble, ce qui n'était l'intention d'aucun d'eux.

1.5.3 SU-Set

SU-Set a été développée pour servir comme une base pour les entrepôts RDF qui pourrait s'implémenter dans un moteur d'inférence RDF. SU-Set modifie le comportement

des opérations à diffuser de telles sortes que l'ensemble des triplets seront affectés un par un.

1.5.4 B-SET [5]

repose sur l'entreposage de « pierres tombales », c.-à-d. les éléments supprimés sont simplement cachés à l'utilisateur via l'opération de recherche. L'utilisation de pierres tombales n'est pas appropriée pour les grands ensembles comme ceux que nous pouvons trouver dans le réseau de données, car leur complexité spatiale est élevée.

1.5.5 srCE

Une importante propriété du modèle srCE est que ses opérations ne s'exécutent pas directement sur l'ensemble final mais elles doivent d'abord s'exécuter sur l'un des multi-sets supplémentaires, selon le type de l'opération. Par la suite, une opération interne se génère en vue de produire un ensemble cohérent.

1.5.6 LD-SET

LD-Set est conforme à CRDT, il garantit la cohérence éventuelle dans tout les cas LD-Set concentre sur un mécanisme de réplication optimiste. Il vise à soutenir l'édition collaborative sociale de Données liées au sein d'une communauté virtuelle d'utilisateurs de différents sites, tout en maintenant une cohérence éventuelle afin de mise à jour simultanée

1.6 Etude comparative entre ces algorithmes

Conclusion

Bibliographie

- [1] C. Bizer, T. Heath, K. Idehen, and T. Berners-Lee, “Linked data on the web (ldow2008),” in *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, (New York, NY, USA), p. 1265–1266, Association for Computing Machinery, 2008.
- [2] T. Berners-Lee, J. Hollenbach, K. Lu, and J. Presbrey, “Tabulator redux : Browsing and writing linked data,” *CEUR Workshop Proceedings*, vol. 369, 01 2008.
- [3] T. Berners-Lee, J. Hendler, and O. Lassila, “The semantic web,” *Scientific American*, vol. 284, pp. 34–43, May 2001.
- [4] K. Aslan, P. Molli, H. Skaf-Molli, and S. Weiss, “C-Set : a Commutative Replicated Data Type for Semantic Stores,” in *RED : Fourth International Workshop on REsource Discovery*, (Heraklion, Greece), May 2011.
- [5] H. Zarzour and M. Sellami, “B-set : A synchronization method for distributed semantic stores,” in *2012 IEEE International Conference on Complex Systems (ICCS)*, pp. 1–6, 2012.