الجمهورية الجزائرية الديمقراطية الشعبية
République algérienne démocratique et populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement supérieur et de la Recherche scientifique
المدرسة العليا للإعلام الآلي 8 ماي 1945 - سيدي بلعباس
École supérieure d'informatique
8 Mai 1945 – Sidi Bel Abbes

Mémoire de fin d'études

Pour l'obtention du diplôme **d'ingénieur d'état**

Domaine : **Informatique**

Spécialité : **Systèmes d'information et Web – SIW**

Thème

Étude de l'efficacité énergétique des triple stores

Présenté par
Touahri Sara

Soutenu le : **Juillet, 2025**
Devant le jury composé de

Pr. AMAR BENSABER Djamel	Président du Jury
Pr. BELLATRECHE Ladjel	Encardant
Pr. BENSLIMANE Sidi Mohammed	Co-Encardant
Dr. LEHIRECHE Nessrine	Examinatrice.

Année universitaire : 2024/2025

Résumé

Les graphes de connaissances (KG) et le Web sémantique se sont imposés comme des piliers incontournables pour la gestion de grands volumes de données interconnectées. Au-delà de la performance brute, la consommation énergétique croissante des triple stores – moteurs dédiés au stockage et à l'interrogation de triplets RDF à l'aide de SPARQL – soulève aujourd'hui des questions économiques et environnementales pressantes. Cette prise de conscience motive une analyse fine de l'impact énergétique des traitements sémantiques.

Dans ce travail, nous avons d'abord exploré en profondeur la littérature existante sur l'optimisation de requêtes SPARQL et sur la mesure de l'énergie dans les bases de données relationnelles et non relationnelles. L'examen critique de ces contributions révèle deux lacunes majeures : (i) la plupart des études se concentrent encore sur la réduction du temps de réponse sans tenir compte de la facture énergétique ; (ii) les rares approches "éco-sensibles" restent souvent centrées sur les SGBD relationnels, laissant les triple stores en arrière-plan.

Partant de ce constat, nous proposons une méthodologie unifiée qui traite, d'un même élan, le temps d'exécution et l'énergie dépensée par les requêtes SPARQL : elle s'appuie d'abord sur un cadre d'audit énergétique combinant capteurs matériels et métriques logicielles, puis sur un modèle d'apprentissage automatique capable de prédire, pour chaque requête, le plan offrant le meilleur compromis entre rapidité et sobriété énergétique, enfin sur une stratégie d'optimisation intégrée à l'optimiseur de Virtuoso afin de sélectionner automatiquement ce plan « vert ». L'évaluation expérimentale menée dans ce PFE, ainsi que son implémentation, ouvrent des perspectives immédiates vers des triple stores véritablement durables.

Mots-clés— Web sémantique, Graphes de connaissances , Moteurs de stockage RDF , Requêtes SPARQL, Triplets RDF, Modèles d'apprentissage automatique, Apprentissage par renforcement, Optimisation des requêtes, Consommation énergétique, Réordonnancement des jointures

Abstract

Knowledge Graphs (KGs) and the Semantic Web have become indispensable pillars for managing vast amounts of interconnected data. Beyond raw performance, the growing energy consumption of triple stores—engines dedicated to storing and querying RDF triples with SPARQL—now raises pressing economic and environmental concerns. This awareness calls for a careful assessment of the energy footprint of semantic processing.

In this project, we first carry out an in-depth survey of the literature on SPARQL query optimisation and on energy measurement in both relational and non-relational database systems. A critical review of these contributions exposes two major gaps : (i) most existing studies still focus on reducing response time without considering the energy bill ; (ii) the few "energy-aware" approaches that do exist remain largely confined to relational DBMSs, leaving triple stores under-explored.

To bridge this gap, we propose a unified methodology that jointly addresses SPARQL execution time and energy consumption. It is built on (1) an energy-audit framework combining hardware sensors and software metrics, (2) a machine-learning model that predicts, for each query, the plan offering the best trade-off between speed and energy frugality, and (3) an optimisation strategy integrated into Virtuoso's optimiser to automatically select this "green" plan. The experimental evaluation and implementation carried out in this final-year project open up immediate prospects for truly sustainable triple stores.

Keywords— Semantic Web, Knowledge graphs, RDF storage engines, SPARQL queries, RDF triples, Machine-learning models, reinforcement learning, Query optimization, Energy consumption, Join reordering

الملخص

أصبحت رسوميات المعرفة والويب الدلالي ركائز أساسية لإدارة كميات هائلة من البيانات المترابطة. ومع ازدياد الاعتماد على مخازن الثلاثيات التي تُخزّن وتستعلم عن ثلاثيات RDF باستخدام SPARQL، برزت مسألة استهلاكها المتنامي للطاقة باعتبارها تحديًا اقتصاديًا وبيئيًا ملحًّا، الأمر الذي يستدعي تقييمًا دقيقًا لبصمتها الطاقية.

استعرض هذا العمل بعمق الأدبيات المتعلقة بتحسين استعلامات LQRAPS وقياس الطاقة في أنظمة قواعد البيانات العلائقية وغير العلائقية. وكشف الفحص النقدي لهذه الدراسات عن ثغرتين رئيسيتين: الأولى أنّ أغلب الأبحاث تركّز على تقليص زمن الاستجابة دون الالتفات إلى تكلفة الطاقة، والثانية أنّ المحاولات القليلة الحسّاسة للطاقة تظلّ في الغالب محصورة في قواعد البيانات العلائقية، مما يترك مخازن الثلاثيات خارج دائرة الاهتمام.

انطلاقًا من هذا الواقع، نقترح منهجية موحّدة تعالج زمن تنفيذ استعلامات SPARQL والطاقة المستهلكة لها في آنٍ معًا. تعتمد هذه المنهجية على إطار تدقيق طاقي يجمع بين مستشعرات مادية ومقاييس برمجية، ثم على نموذج تعلّم آلي يتنبأ لكل استعلام بخطة تنفيذ تحقّق أفضل توازن بين السرعة والاقتصاد في الطاقة، وأخيرًا على إستراتيجية تحسين مدمجة في مُحسّن virtuoso لاختيار هذه الخطة «الخضراء» آليًّا. إن التقييم التجريبي والتنفيذ اللذين أُجريا في هذا المشروع يفتحان آفاقًا فورية نحو مخازن ثلاثيات أكثر استدامة.

الكلمات المفتاحية—
الويب الدلالي، رسوميات المعرفة، محركات تخزين RDF، استعلامات SPARQL، ثلاثيات RDF، نماذج التعلّم الآلي، التعلّم المعزّز، تحسين الاستعلامات، استهلاك الطاقة، ترتيب الوصلات