

Projet Data&Musée Représentations sémantiques et leur exploitation pour le traitement de données collectées dans des musées et monuments

Jean-Claude Moissinac, Nada Mimouni

► To cite this version:

Jean-Claude Moissinac, Nada Mimouni. Projet Data&Musée Représentations sémantiques et leur exploitation pour le traitement de données collectées dans des musées et monuments. DAHLIA 2019 :Informatique et Humanités numériques: quelles problématiques pour quels domaines?”, Jun 2019, Nantes, France. hal-02281249

HAL Id: hal-02281249

<https://hal.telecom-paris.fr/hal-02281249>

Submitted on 9 Sep 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.

Projet Data&Musée

Représentations sémantiques et leur exploitation pour le traitement de données collectées dans des musées et monuments

Jean-Claude Moissinac*, Nada Mimouni*
*Telecom Paris 46 rue Barrault 75013 Paris
Jean-claude.moissinac@telecom-paristech.fr
Nada.mimouni@telecom-paristech.fr
<https://www.telecom-paris.fr/>

Résumé. Data&Musée collecte de multiples données d'une centaine de partenaires, musées et monuments: billetterie, livres d'or, événements,... Nous avons fait le choix d'une représentation RDF de ces données pour faciliter les liens avec divers jeux de données: DBpedia, Wikidata, Joconde, DataTourisme ... Le but est d'entreprendre des explorations guidées par la sémantique des grands graphes que nous constituons.

Nous présentons ici le contexte de Data&Musée et les principaux choix que nous faisons liés aux représentations sémantiques.

1. Introduction

Data&Musée est un projet exploratoire visant à améliorer les systèmes d'information d'institutions culturelles, nous faisons l'hypothèse que la promotion du patrimoine culturel peut bénéficier des techniques récentes de représentation et d'exploration des connaissances. Pour cela, nous avons besoin d'enrichir des ensembles de données issues des partenaires du projet et d'en assurer l'interopérabilité pour parvenir à accroître la visibilité et l'accessibilité par un plus large public des institutions concernées.

Une approche reconnue pour assurer ce type d'exploitation de données est l'utilisation des technologies du web sémantique, qui ont montré leur puissance pour le développement des connaissances dans divers domaines comme le tourisme (Soualah Alila et al., 2016; Al-Ghossein et al., 2018), les villes intelligentes (Consoli et al., 2015; Gyrard et al., 2016) ou la valorisation du patrimoine culturel (Lodi et al., 2017). Ces techniques permettent d'assurer une représentation unifiée de données hétérogènes mais apparentées, ce qui facilite le liage et l'enrichissement de ces données.

Data&Musée reçoit, dans un premier temps, des données de 14 musées du groupement Paris Musées, et des 84 musées et monuments du Centre des Monuments Nationaux. En comptant les deux groupements, cela fait exactement 100 institutions qui fournissent des données. La fédération de ces institutions dans deux groupements facilite les interactions nécessaires à la collecte des données.

Data&Musée bénéficie des expertises de plusieurs entreprises spécialisées dans le traitement de certains types de données –billetterie, livre d'or- ou dans les services numériques pour institutions culturelles. Deux laboratoires de recherche contribuent à l'élaboration des solutions retenues : Telecom ParisTech et la chaire UNESCO ITEN.

Nous allons d'abord présenter ici la nature des données collectées et le volume de certaines d'entre elles. Nous évoquerons ensuite les choix de représentation et la méthode retenue pour leur intégration. Enfin, nous aborderons la nature des traitements que nous

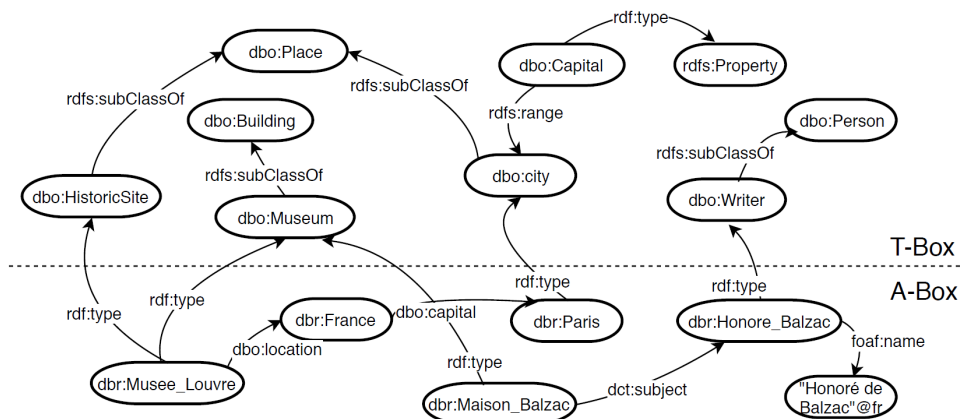
effectuons ou prévoyons sur ces données et des questions auxquelles nous espérons pouvoir répondre.

2. Les données

Plusieurs types de données sont intégrées ou utilisées, principalement des données de sources des partenaires et des données publiquement disponibles.

Les données des partenaires portent principalement sur les achats de billets, les contributions à des livres d'or électronique, les événements –expositions, ateliers, conférences...-, les consultations des sites web, des flux RSS, prochainement les collections. Ces données ont des représentations différentes selon leurs sources et leurs natures. A titre d'exemple, nous avons actuellement –au 15/6/2019- 997605 données de transactions commerciales, essentiellement des achats de billets et 128335 contributions à des livres d'or électroniques de musées ou monuments, comportant des notes sur l'accueil, la qualité des expositions,... et aussi des réponses libres.

Pour ce qui est des sources externes, les grandes bases de connaissances comme Yago, DBPedia, DBPedia-Fr et Wikidata sont des ressources très utiles car elles fournissent un stock de connaissances encyclopédiques semi-structurées suivant les principes du Linked Open Data. Ces bases nous fournissent des informations sur la position géographique des institutions, leur nature (château, site archéologique...), les artistes liés aux musées... Nous avons entrepris l'intégration de données du projet DataTourisme qui collecte et représente en RDF de nombreuses données issus des offices de tourisme de France. Nous exploitons aussi la partie de la base geonames concernant la France. Nous intégrons des données Nous exploitons aussi des données avec des représentations plus classiques –csv, json- comme la base Joconde, la liste des Musées de France, issues notamment de data.gouv.fr. Enfin, nous avons entrepris la collecte sur le Web de données sur des événements, principalement culturels (expositions, concerts...).



3. L'intégration des données

Le schéma ci-dessus donne une idée de notre approche. Chaque fois qu'une donnée concerne une institution, elle est reliée à l'URI tirée de DBpedia-Fr lui correspondant. Une méthode analogue est adoptée pour toutes les entités nommées facilement identifiables. Ainsi les données se trouvent aisément reliées par divers chemins au sein du graphe ainsi constitué.

NomPremierAuteur et al.

4. Les traitements

Outre des traitements assez classiques dans des tableaux de bord –courbes des ventes, ligne de temps des prochains événements...- , nous avons entrepris de tirer parti de la représentation en graphe en opérant des traitements comme la recherche de chemin reliant des nœuds et présentant certaines caractéristiques, mais aussi en travaillant avec des méthodes de construction de modèles vectoriels des entités de graphe. Ces travaux ont déjà débouché sur des résultats en matière de recherche de similarités entre éléments du graphe et devraient déboucher sur des applications comme la recherche de groupes (clusters) ou la recommandation.

5. Conclusion

Nous avons présenté les objectifs du projet Data&Musée. Ce projet a vocation à intégrer des données en provenance d'un grand nombre d'institutions culturelles ; de nouveaux partenaires devraient rejoindre le projet prochainement. Par ailleurs, nous établissons des passerelles avec des projets similaires ou complémentaires en France et à l'étranger.

Références

AL-GHOSSEIN M., ABDESSALEM T. & BARRÉ A. (2018). Open data in the hotel industry : leveraging forthcoming events for hotel recommendation. *J. of IT & Tourism*, 20(1-4), 191–216.

CONSOLI S., MONGIOVÌ M., NUZZOLESE A. G., PERONI S., PRESUTTI V., RECUPERO D. R. & SPAMPINATO D. (2015). A smart city data model based on semantics best practice and principles. In *WWW*.

GYRARD A., ATEMEZING G., BONNET C., BOUDAOU K. & SERRANO M. (2016). Reusing and unifying background knowledge for internet of things with lov4iot. In *2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud)*, p. 262–269.

LODI G., ASPRINO L., NUZZOLESE A. G., PRESUTTI V., GANGEMI A., RECUPERO D. R., VENINATA

C. & ORSINI A. (2017). *Semantic Web for Cultural Heritage Valorisation*, In *Data Analytics in Digital Humanities*, p. 3–37. Springer International Publishing : Cham.

SOUALAH ALILA F., COUSTATY M., REMPULSKI N. & DOUCET A. (2016). Datatourism: designing an architecture to process tourism data. In *IFITT and ENTER 2016 Conferences*.

Remerciements

Notre travail est soutenu par La Région Ile-de-France dans le cadre d'un projet FUI