

VERSUS : générateur de tableaux comparatifs à partir de bases de connaissances

Arnaud Giacometti, Béatrice Markhoff, Arnaud Soulet

Université de Tours, LIFAT, Blois
firstname.lastname@univ-tours.fr

Résumé. Les tableaux comparatifs sont utiles pour comparer des entités en dégageant leurs similarités et leurs différences non triviales. Le choix manuel des caractéristiques de comparaison reste une tâche complexe et fastidieuse. Cet article présente VERSUS qui est la première méthode automatique de génération de tableaux comparatifs à partir du Web sémantique. Pour cela, nous introduisons une mesure, nommée niveau de référence contextuel, pour évaluer si une propriété peut être une caractéristique intéressante pour comparer des entités. Cette mesure repose sur des contextes qui sont des ensembles d'entités similaires aux entités comparées. Nous montrons comment VERSUS sélectionne ces contextes et comment il évalue efficacement le niveau de référence contextuel à partir d'un point d'accès SPARQL public. Nous avons construit un benchmark à partir de Wikidata pour évaluer l'efficacité de VERSUS, avec une campagne d'évaluation manuelle des caractéristiques : la précision et le rappel sont élevés.

1 Introduction

Un tableau comparatif est un tableau à double entrée avec des entités à comparer en colonnes et les caractéristiques de comparaison en lignes. Le tableau comparatif est un outil particulièrement efficace pour la prise de décision en isolant les points communs et les différences significatives¹ entre les entités comparées. Par conséquent, cette technique analytique est populaire en science pour comparer des travaux, en culture pour comparer des oeuvres d'art ou dans le commerce pour comparer des produits ou des services. Dans ce contexte, cet article vise à automatiser entièrement le processus de génération d'un tableau comparatif d'un ensemble d'entités en interrogeant une base de connaissances disponible sur le Web sémantique telle que Wikidata (Vrandečić et Krötzsch, 2014). Par exemple, à partir d'Ada Lovelace et d'Alan Turing, nous voulons obtenir un tableau comparatif comme celui présenté par la table 1 construit automatiquement à partir de Wikidata. Au-delà des personnes, nous visons à comparer des entités telles que des lieux (pays, villes), des objets (tapisseries, statues), des institutions (universités, partis politiques), etc. Malheureusement, il n'y a pas de cadre théorique sur la conception des tableaux comparatifs pour déterminer si une caractéristique est intéressante pour comparer des entités. Cette tâche n'est pas anodine car dans 17 % des cas

1. Dans ce cadre, avoir des identifiants différents, ou des photos différentes, n'est pas considéré significatif parce que de tels traits ne sont partagés avec aucune autre entité.