

Découverte d'indicateurs de classement dans le Web des données

Cyril De Runz, Arnaud Giacometti, Béatrice Markhoff, Arnaud Soulet

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

Résumé. Analyser l'impact d'entités au sein de leur domaine est fondamental pour le comprendre. A cette fin, il est essentiel de disposer d'indicateurs numériques fins retranscrivant les spécificités du domaine. Cet article propose une approche pour découvrir automatiquement des indicateurs d'impact pour classer les entités. Bien que l'approche soit transdisciplinaire, les indicateurs de classement identifiés doivent néanmoins disposer d'une sémantique intradisciplinaire. Pour cela, notre approche s'appuie sur les bases de connaissances du Web des données, pas seulement pour faciliter le calcul opérationnel des indicateurs mais aussi pour profiter de leur transparence et de la formalisation explicite de leur sémantique. L'hypothèse simple mais centrale de ce travail est que chaque répartition inégale d'une quantité engendre un indicateur de classement pertinent. A cette fin, nous utilisons le coefficient de Gini pour identifier dans Wikidata les propriétés produisant des indicateurs de classement significatifs.

1 Introduction

Quel est le classement des peintres les plus importants ? Il n'est pas difficile de trouver une réponse à cette question, mais trouver la bonne s'avère plus compliqué. Par exemple, la table 1 présente les 10 peintres les plus importants selon trois sites. Ces classements s'accordent sur une partie des peintres (en gras dans les tableaux), mais certains artistes jugés comme primordiaux dans un seul classement sont occultés par les deux autres (en italique). Cela n'a évidemment rien de surprenant puisque ces classements découlent de l'avis subjectif d'une personne¹ ou d'un ensemble de personnes². Le troisième classement consiste à classer les peintres par popularité en mesurant le nombre de lectures de leur article dans une encyclopédie en ligne³ (voir l'indice en troisième colonne du classement artcyclopedia.com signifiant, par exemple, que la page consacrée à Salvador Dalí a reçu environ moitié moins de visites que celle de Pablo Picasso). L'utilisation d'un indicateur numérique est rassurant mais soulève aussi des questions de pertinence qui seront justement abordées dans cet article.

Notre question initiale concernant le classement des peintres pourrait être étendue à d'autres domaines : quels sont les processus biologiques les plus prépondérants ? Quelles sont les villes

1. www.theartwolf.com/articles/most-important-painters.htm

2. www.ranker.com/list/best-painters-of-all-time/ranker-art

3. www.artcyclopedia.com/mostpopular.html