
L'intelligence artificielle pour décrire, indexer et explorer des collections multimédias : comment faire sans Google ?

Guillaume Gravier*¹

¹IRISA (IRISA) – CNRS : UMR6074 – Rennes, France

Résumé

Décrire, indexer, explorer, analyser des collections de textes, d'images, de vidéos constitue une part importante du travail en humanités numériques, pour lequel les techniques d'intelligence artificielle ouvrent en théorie de nombreuses possibilités. En pratique, cependant, les limites des technologies grand public disponibles et immédiatement applicables à la description, l'indexation et la fouille de tels contenus sur une problématique donnée – les "technologies Google" par abus de langage – les rendent la plupart du temps inexploitable sans un considérable effort d'ingénierie pour les adapter. Nous argumenterons que faire "sans Google" requiert une co-construction de projets de recherche en humanités numériques et en intelligence artificielle pour, d'une part, faire évoluer la capacité des outils d'intelligence artificielle à appréhender de nouveaux problèmes rapidement et sans expertise particulière et, d'autre part, faire progresser la recherche en sciences humaines et sociales en l'outillant plus efficacement.

Après une brève présentation des principes sous-jacents de l'intelligence artificielle pour en comprendre l'intérêt et les limites, nous analyserons des projets de recherche à l'interface entre intelligence artificielle et humanités pour mettre en évidence les difficultés rencontrées et les questions de recherche qui en résultent en informatique. Nous prendrons notamment appui sur plusieurs projets : iCODA, qui concerne l'extraction d'information dans les archives du quotidien Ouest-France ; READ-IT, qui concerne la création d'une base de données outillée pour l'étude du patrimoine de la lecture ; ou encore ARCHIVAL, qui concerne l'exploration de fonds documentaires sur l'autogestion. Tous ces projets allient "concepteurs" et "consommateurs" de technologie pour répondre aux besoins des deux catégories.

Guillaume Gravier est directeur de recherche au CNRS et directeur de l'Institut de recherche en informatique et systèmes aléatoires (UMR 6074). Il est spécialiste de l'apprentissage statistique pour l'analyse de contenus multimédias. De 2014 à 2018, il a été responsable de l'équipe de recherche Linkmedia dédiée aux technologies pour l'exploration de collections multimédias. Il a été impliqué dans de nombreux projets en lien avec les humanités numériques, notamment aux côtés de l'Institut national de l'audiovisuel, de la Bibliothèque nationale de France ou encore de la Fondation des maisons des sciences de l'homme. Récemment, il a coordonné le projet iCODA sur l'exploration des archives du journal Ouest-France et, en binôme avec Brigitte Ouvry-Vial, le projet européen READ-IT sur le patrimoine de la lecture en Europe.

*Intervenant