

L'enjeu des métadonnées dans les corpus textuels

Philippe Caron, Rodolphe Defiolle et Marie-Hélène Lay (dir.)

2019

Presses universitaires de Rennes
www.pur-editions.fr

IL DEVIENT superflu d'insister sur la révolution scientifique que représentent la numérisation et les capacités de mémoire dans la création de bases de données en sciences humaines. Le défi est désormais ailleurs : à la fois quantitatif et qualitatif. En effet la course est désormais à la constitution de très grosses ressources et les possibilités informatiques le permettent.

Ce changement d'échelle a une incidence automatique sur la vie académique : les recherches qui ne reposeront pas sur un fondement quantitatif appréciable risquent fort d'être frappées de caducité. Mais la constitution de ces gros ensembles nécessite un soin méticuleux, des moyens matériels, mais surtout humains qui ne sont pas à la portée de tous les chercheurs.

À supposer que ces conditions soient remplies, reste encore à s'entendre sur le contenu enregistré : aucun texte n'est numérisé et archivé de façon brute. Ce qui est à présent devant nous est donc autre : entre utilisation « maison » et normalisation des données, il s'agit de se frayer un chemin qui tout à la fois autorise les recherches les plus spécifiques sans pour autant priver la communauté d'une donnée échangeable et utilisable pour d'autres recherches.

Ce volume répond à cette préoccupation : comment faire migrer de vieux corpus vers d'autres standards, comment ménager des besoins particuliers et un désir légitime d'exporter ailleurs des documents, de quels outils pouvons-nous disposer pour réaliser au mieux cette quadrature du cercle ?