

TXM est un logiciel de textométrie développé actuellement dans le cadre de l'équipement d'Excellence Matrice, sous licence libre GNU General Public Licence version 3 et téléchargeable gratuitement depuis le site Sourceforge.

En tant que logiciel de textométrie, il permet de donner une vision linguistique globale et une vision linguistique locale des grands corpus de textes numériques qu'on lui soumet ; son développement s'inscrit dans le cadre général des Digital Humanities et de la demande croissante en analyse des textes et des discours dans les Sciences Humaines et Sociales (Histoire, Sociologie, Linguistique, etc.)

TXM permet à l'utilisateur d'articuler une lecture qualitative et une lecture quantitative des textes, avec deux types de fonctionnalités.

D'une part des fonctions documentaires qui permettent la navigation hypertextuelle dans le corpus, l'extraction de passage ou de concordances, la recherche plein texte de mots ou d'unité du discours. L'utilisateur peut ainsi construire une interprétation scientifique en la contrôlant à tout moment par une consultation précise et complète de ses données textuelles.

Et d'autre part des fonctions statistiques : calcul des mots spécifiques d'un texte ou d'une partie du corpus (représentant une période ou un type de locuteur par exemple), calcul des co-occurrences pour repérer les attractions particulières entre les mots, visualisation factorielle des textes ou des mots du corpus, etc.

C'est dans la bonne intelligence de ces deux types de lecture - lecture qualitative et lecture quantitative - que réside la richesse des parcours interprétatifs que l'analyste est invité à faire avec le logiciel.

En pratique, TXM prend en entrée des formats de texte classiques comme TXT ou XML mais prend aussi des fichiers issus de logiciels de lemmatisation et d'étiquetage morpho-syntaxique, tant et si bien que son traitement s'applique aussi bien au texte graphique, ou texte brut, qu'à des textes enrichis. Que cela soit d'un point de vue documentaire ou statistique, les requêtes deviennent ainsi complexes portant aussi bien sur des mots graphiques, que sur les lemmes ou les codes grammaticaux comme tel temps verbal ou tel enchaînement syntaxique. La structuration des textes peut par ailleurs être prise en compte dans les calculs : comme le découpage en paragraphes, le changement de locuteur dans un entretien ou la division en chapitres d'un ouvrage.

TXM bénéficie d'une liste de discussion fédérant son réseau d'utilisateurs dans l'esprit de son développement ouvert et modulaire, et pour faciliter son appropriation par une communauté en plein essor. Des fonctions d'import et d'export permettent la réutilisation des résultats et l'interopérabilité avec d'autres logiciels libres produits par d'autres communautés.

Pour répondre à la diversité des besoins et pratiques en SHS, TXM existe à la fois sous la forme d'un logiciel s'installant sur un ordinateur personnel (Windows, Mac ou Linux), et sous la forme d'un portail web permettant l'accès partagé et contrôlé à des corpus mis en ligne et interrogeables par le biais d'un simple navigateur Internet.