Apparition des balises dans la base de données :

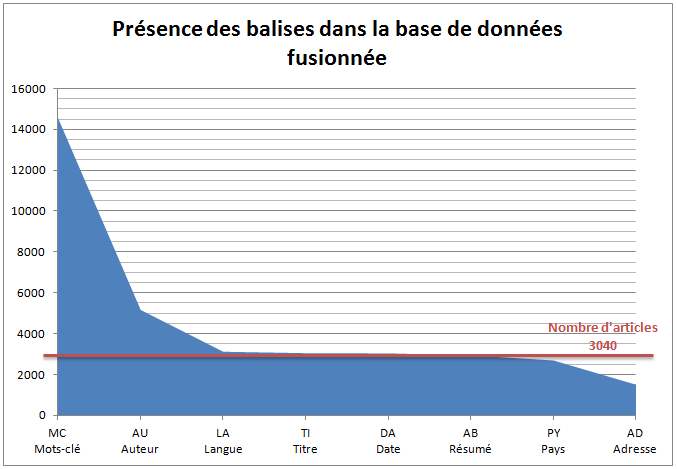


Figure : Fréquence d'apparition des balises dans la base de données

Nous avons effectué un comptage de chaque balise dans la base de données fusionnée. A partir de ce comptage, nous avons établi le graphique ci-dessus qui montre pour chaque balise, où elle se situe par rapport au nombre d’articles.

La base de données recense trois mille quarante (3040) articles. Le graphique indique donc que les mots-clé et les auteurs sont sur-représentés dans la base ; en effet, nous comptons plus de ces balises que d’articles. Ensuite, le nombre de balises langue, titre, résumé et dans une moindre mesure pays est à peu près égal au nombre d’articles. Ces balises sont donc présentes en moyenne une fois par article. Enfin, la balise adresse est sous-représentée ; elle n’est donc pas présente dans chaque article.

Nous avons choisi de ne pas représenter la balise MTM (multi-termes) sur le graphique car elle entraverait la bonne lisibilité de celui-ci. En effet, elle est présente 64968 fois.