

Ce travail est motivé par la construction d'un modèle de représentation et d'appariement de documents XML basé sur leur structure. Notre approche consiste d'abord à représenter chaque document XML par sa structure arborescente et ensuite à utiliser cette structure comme modèle de représentation pour la classification structurelle du document XML correspondant. La classification structurelle a pour rôle de regrouper des documents XML structurellement similaires dans des clusters (ou classes). Ceci aiderait, d'une part, à mieux organiser les documents XML et, d'autre part, à mieux répondre, en termes d'efficacité et d'efficacité, aux requêtes contenant des conditions structurelles. L'idée est que, si les documents partagent des structures similaires, ils sont plus à même de correspondre à la partie structurelle d'une même requête. Par ailleurs, une nouvelle mesure de similarité qui prend en compte certaines relations hiérarchiques entre les nœuds de chaque structure arborescente ainsi qu'un nouvel algorithme de clustering (classification non supervisée), ont été proposés. Pour valider notre approche, des expérimentations ont été menées sur deux collections XML. A cet effet, nous avons utilisé le corpus réel ACM SIGMOD Record et une collection XML synthétique, qui nous ont permis d'évaluer la performance de notre algorithme de clustering, la fiabilité de notre modèle de représentation ainsi que l'efficacité de notre mesure de similarité. Ces différents tests ont montré l'intérêt de notre approche