

Qualité des données *et* *défis de l'automatisation*

Le catalogage partagé, les catalogues collectifs, ainsi qu'un certain nombre de services associés, sont le résultat de la contribution de chacun à un enrichissement global. Alimenter un même silo de métadonnées a permis la création et le développement du Sudoc. Il est le fruit d'une collaboration entre les hommes avec l'aide de la machine. Mais, comme la plupart des catalogues, le Sudoc a une histoire. La fusion de catalogues préexistants – dont il est le résultat – a généré des doublons. Doublons bien trop nombreux pour être résorbés manuellement, d'où la tentation de l'automatisation.

Ainsi, on a confié à la machine la « décision » de fusionner les entrées. Certains rapprochements ont été heureux, réduisant le nombre de doublons pour un même auteur. Mais, dans d'autres cas, ces automatismes sont allés trop loin, fusionnant dans une même notice d'autorité des auteurs différents.

quement. Nous nous engageons donc dans un cercle vertueux de propagation de la qualité au sein d'un réseau et en direction d'autres réseaux interconnectés. On ne quitte pas tout à fait le travail en silos, mais ceux-ci ne fonctionnent plus isolément et s'insèrent graduellement dans des réseaux.

Dans cette démarche, la machine dépasse le rôle de simple « mise en relation ». Elle permet de traiter des quantités de données hors de portée de l'homme seul, elle cesse d'être un simple instrument pour devenir un collaborateur. Outre ces



quelques exemples strictement liés à nos catalogues, le dossier présenté dans ce numéro d'*Arabesques* nous montre à quel point les catalogues cessent d'être un but en soi, pour servir largement au-delà de leur objectif originel. Nous constaterons à quel point la qualité des données est primordiale, dans la mesure où, tout comme la qualité, les défauts aussi se propagent par ricochet.

Nous nous engageons dans un cercle vertueux de propagation de la qualité au sein d'un réseau et en direction d'autres réseaux interconnectés.

Ces premiers algorithmes, par trop imparfaits, nous ont conduits à mettre fin à l'expérience. Car, autant il est possible de fusionner automatiquement, autant scinder des rapprochements malheureux est difficile. De ces erreurs, il demeure encore aujourd'hui quelques scories difficilement rattrapables sans intervention humaine, chose que l'on cherchait justement à éviter.

Depuis ces expériences qui remontent à quelques années, la puissance de calcul des machines, les progrès en matière d'apprentissage, les travaux de recherche linguistique, auxquels l'Abes est parfois associée, ouvrent de nouvelles perspectives. La détection de candidats à la fusion devient plus fine et, pour éviter les erreurs passées, une contribution humaine est jugée indispensable. La machine propose, les catalogueurs et autres intervenants disposent.

Ces améliorations de la qualité de la base peuvent être utiles non seulement au Sudoc, mais aussi, grâce aux possibilités d'alignements entre divers référentiels, servir à d'autres et récipro-

Le travail du réseau sera primordial à l'heure du traitement des métadonnées des ressources électroniques. Il s'agira de récupérer les métadonnées fournies par l'éditeur, de leur faire subir des traitements automatiques avant de les soumettre aux réseaux pour leur amélioration, puis de disséminer ces améliorations. Ici encore nous pouvons parler de cercles vertueux, à condition que les circuits de traitement, de la répartition des rôles et des étapes des *workflows* soient clairement explicités : un chantier dans lequel l'Abes entend s'engager progressivement en commençant par des expérimentations.

Les bibliothèques deviennent laboratoires, lieux d'expérimentation et d'innovation. La machine contribue à la démultiplication de l'efficacité et du pouvoir de chacun au travers de réseaux de cellules interdépendantes. Nous constaterons, à la lecture de ce numéro, l'importance des perspectives qui s'ouvrent devant nous.

JÉRÔME KALFON
Directeur de l'Abes