Avis d'expert par Benjamin de Boe, Product Manager Unstructured Data InterSystems

Un pêcheur à la ligne trouve ce qu'il cherche. En utilisant un sonar, le pêcheur trouve ce qu'il ne cherche pas, parfois même des trésors ! En matière de données, notamment médicales, cette approche est extrêmement séduisante par les perspectives cognitives qu'elle offre.

Une approche ascendante des données

Traditionnellement, utiliser un dictionnaire et une ontologie définis au préalable est la technique privilégiée pour obtenir des informations provenant d'une quantité importante de données, afin de trouver des liens entre elles. Cette méthode de recherche est coûteuse, car des experts capables de constituer un dictionnaire et une ontologie puis interpréter les résultats de la recherche de façon utile sont indispensables. Mais cette approche, dite descendante, souffre d'un inconvénient majeur : on ne peut trouver que ce que l'on cherche. Les liens pertinents qui sortent du cadre de la recherche n'étant en effet pas pris en compte.

A l'inverse, en matière de Big Data, l'approche ascendante peut se révéler un allié précieux. Le principe : une analyse en amont de toutes les données de façon à mettre à jour les liens qui les unissent. Pour bien comprendre, une analogie s'impose. Maritime en l'occurrence.

Comparons ainsi une base de données à un océan renfermant de nombreux trésors. Traditionnellement, le pêcheur lance sa ligne et s'il attrape quelque chose, c'est qu'un être vivant a mordu à l'hameçon. Un être vivant suffisamment grand que pour être pris au piège, mais trop faible pour s'en libérer.

Des animaux trop petits ou vivant trop profondément ne seront donc jamais attrapés. Le pêcheur peut naturellement utiliser d'autres techniques : différents types d'hameçons, un chalut, etc. Ce qui augmente de fait le prix du poisson tandis que le pêcheur ne pourra jamais attraper un trésor. En bref, le pêcheur ne peut que capturer ce qu'il cherche, jamais ce qu'il ne cherche pas. En matière de données, le principe est exactement le même...

De la donnée à la connaissance

Pour continuer l'analogie, un pêcheur qui souhaite utiliser l'approche ascendante installe un sondeur dans son bateau. Ce qui lui donne un aperçu de tout ce qui se trouve dans l'océan et sur le fond. Et ce sondeur est contrôlé par un puissant logiciel qui, sur la base de toutes ces images, découvre que certaines variétés de poissons ne vivent ensemble que par hasard, mais que d'autres espèces d'animaux marins par contre vivent en symbiose, etc. Et si le bateau vient à passer au-dessus du trésor légendaire, il le découvre également !

Mais revenons à la donnée au sens pratique. Imaginons un chercheur souhaitant trouver, dans un grand fichier de données, des patients ayant une pression artérielle élevée (hypertension). La méthode de recherche traditionnelle consiste à parcourir les données à l'aide d'un dictionnaire et d'une ontologie et à trouver de nombreux patients souffrant d'hypertension, mais aussi un dossier où ce terme « hypertension » est utilisé dans un autre contexte, comme « l'hypertension pulmonaire aiguë

» par exemple.

Chaque personne disposant d'une formation médicale sait que cette maladie est bien différente de l'hypertension, et c'est bien là le cœur du problème de l'approche descendante : des experts sont encore nécessaires pour interpréter les résultats de recherche. A fortiori dans le cas de très grandes bases de données et pareils résultats à filtrer, créer le bon fichier se révèle chronophage et coûteux.

De meilleurs résultats de recherches

Dans une approche ascendante, l'on repère également dans la base de données des patients les dossiers traitant de l'hypertension pulmonaire aiguë. Tout en allant beaucoup plus loin puisqu'on peut alors déceler que cette hypertension dans les poumons est apparue après que le patient a été victime d'une crise cardiaque (« infarctus du myocarde »). On ne recherche donc pas seulement des concepts spécifiques, on est également capable de corréler des informations entre elles pour déterminer les liens qui les relient potentiellement.

Pour y parvenir tout en évitant de déterminer des liens de causes à effets qui n'auraient pas lieu d'être, l'analyse du contexte dans lequel s'inscrivent ces concepts est primordiale. C'est ainsi que les outils d'analyse ascendante permettent de déterminer que l'hypertension et l'hypertension pulmonaire aiguë ne sont pas les mêmes pathologies. De façon à proposer au chercheur un dossier reprenant des patients souffrant d'hypertension, tout en mettant de côté les dossiers traitant d'hypertension pulmonaire aiguë. Le chercheur, qui n'a pas besoin de ces derniers dans le cadre de sa recherche spécifique, n'est pas noyé d'informations inutiles, sans qu'il ait besoin de faire le tri lui-même. En bref, l'approche ascendante offre donc de meilleurs résultats à moindre coût.

A propos d'InterSystems (www.InterSystems.fr)

InterSystems développe des technologies logicielles novatrices qui permettent la création d'applications de pointe. Avec sa passion de l'excellence et sa concentration sur le succès de ses clients, InterSystems offre des plates-formes de gestion de données et d'interopérabilité stratégiques et décisionnelles en temps réel, utilisées dans le domaine des soins de santé, des services financiers, gouvernementaux et de nombreux autres secteurs de l'industrie. Dans certains pays, comme la France, InterSystems propose également des applications de soins de santé unifiées basées sur ses technologies. Créée en 1978, InterSystems est une entreprise privée dont le siège est établi à Cambridge, Massachusetts (USA). Elle a des bureaux dans le monde entier et ses produits sont utilisés par des millions de personnes dans plus de 100 pays.