

Le *text mining* clé de l'information stratégique



Alain Lot
Business intelligence
solutions director
IBM Europe

Solution d'intelligence décisionnelle, le *text mining* permet le traitement des données textuelles pour les transformer en informations stratégiques. Ces applications dans la banque touchent différents domaines : gain de temps considérable, fonctions d'analyse, de fidélisation, de gestion automatique d'informations non numériques...



Charles Huot
Business intelligence
Text mining manager
IBM Europe

Le fait de prendre des décisions importantes de manière quotidienne implique un accès rapide et efficace à l'information. Aujourd'hui, 80 % de l'information disponible se présente sous la forme textuelle, et cette tendance va en s'accroissant avec le développement des publications, des grandes bases de données, ainsi que des nouvelles formes de commercialisation. Chaque jour paraissent 1 000 livres nouveaux, 30 000 dépêches de presse, 30 rapports stratégiques. Les serveurs Internet, quant à eux, connaissent une croissance de 100 % par an et, pour qui recherche une information par mot clé, il n'est pas rare de se voir confronté à plusieurs milliers de documents à analyser.

Parmi les évolutions responsables de cette surabondance d'information, figurent la croissance du nombre des centres d'appels téléphoniques, l'ouverture de messageries électroniques auprès des clients, ainsi que la multiplication des sources d'informations financières (Reuters, Dow Jones, Prompt...).

Aux États-Unis, une enquête a été effectuée par Reuters au printemps 1997 auprès de 1 300 directeurs de sociétés. Ses conclusions, publiées sous le titre «Les assoiffés de l'information. Une enquête sur les effets de la surinformation», mettent en évidence qu'une surinformation étrangle le développement des affaires.

Parmi les résultats les plus significatifs, on note que :

- 43 % des directeurs interrogés pensent que les décisions importantes sont retardées pour cause de surinformation ;
- 48 % pensent qu'Internet est la cause principale de la surinformation ;
- 80 % jugent que la croissance rapide des volumes de communication au sein des entreprises, avec les clients aussi bien qu'avec les fournisseurs, est le principal facteur de la surinformation.

Comment maîtriser cette masse d'informations textuelles et prendre les bonnes décisions ?

De l'analyse linguistique à l'information décisionnelle

Le développement de l'analyse linguistique (qui permet la reconnaissance des mots significatifs et des mots identiques), de l'analyse de données (qui permet la détection de similarité de textes), ainsi que de la puissance des équipements informatiques a récemment entraîné des progrès spectaculaires permettant la segmentation, la classification, de même que l'accès aux textes.

Ces différentes fonctions, ainsi que le potentiel qu'elles induisent en termes d'informations stratégiques, sont intégrées pour former le *text mining*. Processus d'analyse, de classification et de découverte de l'information, dans tout type de document textuel, le *text mining* constitue ainsi

un outil d'aide à la décision pour les professionnels de l'économie et du marketing, les analystes, les conseillers financiers et les acheteurs.

Tout comme le *data mining*, il fait appel aux techniques les plus avancées dans le domaine de l'analyse des données, depuis la classification automatique jusqu'à la catégorisation. Sa particularité réside bien entendu dans le type d'information qu'il est amené à analyser, c'est-à-dire les textes.

Pour ce faire, le *text mining* utilise des modèles d'analyse linguistique, ce qui implique qu'un modèle différent soit nécessaire pour chaque langue analysée.

Le modèle va lire les documents phrase par phrase et les décomposer en formes grammaticales. L'analyse de texte permet également d'identifier automatiquement des mots composés, tels que «carte bancaire», «carte de crédit», «compte courant», «crédit immobilier», etc., et cette analyse peut contenir un modèle spécialisé. Ainsi, pour prendre l'exemple de la banque et des institutions financières, des éléments comme «chéquier», «carte bleue», «devise» peuvent être regroupés sous le terme générique «moyens de paiement».

Une fois que l'analyse linguistique a lu un ensemble de documents, le *text mining* utilise des méthodes de classification automatique pour regrouper ceux-ci par thèmes. Un calcul de similarité est réalisé entre chaque document pour déterminer une classification opti-

male. Les thèmes peuvent être calculés automatiquement (classification) ou être prédéfinis. S'ils sont prédéfinis, l'affectation automatique sera utilisée pour faire le routage de nouveaux textes dans des thèmes établis.

Améliorer la satisfaction du client

Plusieurs exemples montrent l'importance du *text mining* dans les banques ; en premier lieu, l'avantage compétitif que représente **une bonne prise en compte de l'information** que leur fournissent leurs clients.

Celle-ci provient essentiellement :

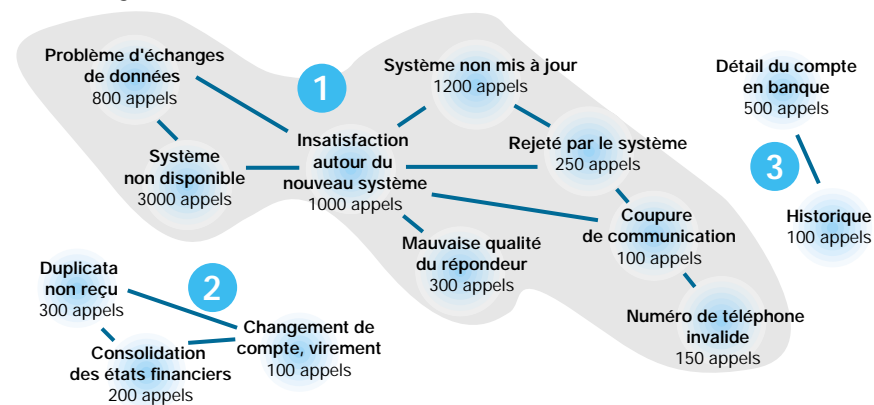
- des appels téléphoniques ;
- de la communication directe (face à face, au guichet) ;
- du courrier traditionnel ou de la messagerie électronique ;
- des enquêtes d'opinions.

Certaines banques américaines utilisent le *text mining* pour analyser et traiter les informations. Cette démarche permet de mieux comprendre les besoins des clients, leurs préférences, d'améliorer ainsi la description des portefeuilles clients et, par conséquent, d'être plus à même d'améliorer la satisfaction et la fidélisation de la clientèle.

L'information sur les clients a par ailleurs des implications importantes pour divers services de la banque ou de l'établissement financier, dans la mesure où le *text mining* permet de la caractériser et de la spécifier pour chacun d'eux. Cela va du routage automatique d'information, à la génération automatique de baromètres présentant la satisfaction dans le temps des détenteurs d'un produit spécifique. L'analyse rapide de l'information et sa prise en compte dans le processus de décision de la banque deviennent un élément clé dans l'évaluation des tendances et des demandes du marché. Par exemple, lors des récentes chutes spectaculaires des différentes places financières mondiales, le temps moyen pour qu'un client entre en relation par téléphone avec un agent était de 2 heures. On comprend dès lors toute l'importance de solutions de secours de type messagerie électronique avec intervention du *text mining*.

Une autre illustration de cette nécessité d'analyse se retrouve dans les résultats d'une étude réalisée par IBM et portant sur 10 000 appels téléphoniques reçus le 12 avril 1997. Ces appels ont été partiellement transcrits (passage de l'enregistrement vocal à une forme écrite interprétable par

Exemple d'analyse effectuée sur les appels téléphoniques d'une journée



l'ordinateur), puis analysés par le *text mining*. Le résultat est une **segmentation automatique** des appels clients, mettant en évidence l'existence de thèmes récurrents ①.

Ces thèmes sont automatiquement déterminés par le *text mining* qui va lire et analyser chaque appel. Une analyse quotidienne permet de voir graphiquement l'évolution des sujets traités dans les appels. Grâce au *text mining*, il devient donc possible, non seulement de détecter les thèmes d'insatisfaction, mais également de mesurer l'impact d'une action engagée pour redresser une situation particulière.

L'analyse de demandes (ou de plaintes) permet, par exemple, de déceler tout dysfonctionnement dans la mise en place d'un nouveau service, la solution *Customer relationship intelligence* (CRI) classant automatiquement les appels par catégorie. L'analyse fait ressortir trois réseaux correspondant à trois sujets :

- réseau 1 : nouveau système mis en place ;
- réseau 2 : appels liés à des transferts et des consolidations d'états ;
- réseau 3 : demandes liées au compte courant.

Des éléments spécifiques sont associés à chacun des réseaux. Par exemple, pour le premier, 3 000 appels pour indiquer que le système n'est pas disponible, 250 appels où le client indique qu'il a été rejeté par le système, etc.

La gestion de l'information financière non numérique

Autre exemple de l'importance du *text mining*, le développement sans cesse croissant de l'intelligence économique se traduit par une recherche d'information stratégique à travers la **compilation de l'ensemble des sources**

disponibles tels que journaux, magazines, rapports, télécopies, les messages électroniques, services de bases de données en ligne, enquêtes d'opinions.

Les supports et les formats de ces informations varient considérablement : dossiers traditionnels, livres, répertoires, documents électroniques, CD-Rom, microfilms, Internet.

Dans le cas de dépêches en provenance de Dow Jones, la solution *Text knowledge server* (TKS) analyse (littéralement, va lire les documents) et détermine le contenu de chacun. Il est ainsi possible de constituer automatiquement des rubriques nommées (par exemple, dividendes sur actions des sociétés).

Lorsque l'utilisateur accédera à cette rubrique, il trouvera un ensemble de dépêches parlant spécifiquement d'un sujet déterminé.

L'encadré ② indique que 250 dépêches (sur un total de 7 107, classées en 75 rubriques) parlent des dividendes des sociétés, 55 du marché du Nasdaq, 43 des investissements anglais aux États-Unis...

② Text mining sur les dépêches Dow Jones

	75 rubriques	7107 Documents
1. Dividendes sur actions des sociétés	250	(3,5 %)
2. Actions du marché informatique Nasdaq	55	(0,8 %)
3. Bonus sur les salaires des CEO	54	(0,8 %)
4. Fonds de la réserve du Trésor	48	(0,7 %)
5. Investissements anglais sur le marché américain	43	(0,6 %)
6. Annonces sur le budget de Washington	43	(0,6 %)
7. Etc.		