



IBM Content Analytics

Libérer l'Information

Patrick HOFLEITNER

patrick_hofleitner@fr.ibm.com

Août 2011

TABLE DES MATIERES

| | |
|--|-----------|
| RESUME..... | 3 |
| INTRODUCTION | 4 |
| LA PROBLEMATIQUE..... | 5 |
| 1 L'EXPLOSION DU CONTENU NON-STRUCTURE | 5 |
| 2 LES SOCIETES DOIVENT RENDRE EXPLOITABLE LE CONTENU NON-STRUCTURE..... | 6 |
| LES RAISONS DE L'EVOLUTION | 7 |
| 1 L'EVOLUTION DES FONCTIONS D'ANALYSE..... | 7 |
| 2 DES TECHNOLOGIES EPROUVEES..... | 7 |
| 3 UN OUTIL INDISPENSABLE DANS TOUT DOMAINE D'ACTIVITE | 8 |
| IBM CONTENT ANALYTICS..... | 9 |
| POUR EN SAVOIR PLUS | 11 |
| 1 IBM CONTENT ANALYTICS | 11 |
| 1.1 Portail IBM Software..... | 11 |
| 1.2 IBM Redbooks | 11 |
| 2 IBM LANGUAGEWARE..... | 11 |
| 2.1 Portail IBM Software..... | 11 |
| 2.2 IBM alphaWorks..... | 11 |
| 3 IBM CLASSIFICATION MODULE..... | 11 |
| 3.1 Portail IBM Software..... | 11 |

Résumé

L'analyse de contenu est un nouveau domaine dans les fonctions d'analyse qui permet aux sociétés de bénéficier de nouvelles perspectives par l'exploitation des contenus non-structurés et semi-structurés. Ces contenus peuvent se présenter sous la forme de formulaires, documents, zones de commentaires dans des bases de données, pages Web, correspondances de clients et partenaires et toute autre information qui n'est pas stockée dans des zones structurées.

L'analyse de contenu rend possible l'accès, l'analyse et l'exploitation de contenus non-structurés et semi-structurés disponibles dans et au-dehors d'une organisation. C'est une extension naturelle de l'informatique décisionnelle (business intelligence) basée uniquement sur des données structurées.

IBM Content Analytics permet aux organisations d'exploiter le contenu non-structuré pour une meilleure appréciation de leurs domaines d'activité pour de meilleures prises de décision. Avec IBM Content Analytics les sociétés peuvent maintenant évoluer d'une stratégie de prise de décision basées uniquement sur les données vers une stratégie de prise de décision basée sur l'ensemble du contexte de leurs activités. Elles bénéficient alors d'un avantage compétitif dans un environnement en continuel mutation.

IBM Content Analytics



Introduction

Les décideurs sont en permanence à la recherche de nouveaux indices pour piloter efficacement leurs entreprises. Ils veulent identifier les sources premières des problèmes, repérer les opportunités plus rapidement et anticiper et exploiter les tendances d'évolution du marché. Pour ce faire ils s'orientent vers des solutions d'informatique décisionnelle pour suivre des critères de mesure de ventes, d'actions marketing et des opérations. Ces aperçus sont basés sur des données structurées internes à l'organisation.

Mais cela n'est qu'une vue partielle de la situation. A l'intérieur et à l'extérieur de l'entreprise les volumes d'information explosent. Clients, prospects, partenaires et influenceurs écrivent des commentaires, remplissent des formulaires, échangent avec les commerciaux, le marketing et le support. Cette masse d'information devrait être prise en compte dans la prise de décision. Cependant ce type d'information est difficile à suivre et à analyser objectivement. Elle est souvent disponible sous forme de contenu non-structuré ou semi-structuré comme des documents, des notes, des messages électroniques, des pages Web et autres.

De plus en plus d'entreprises prennent conscience de la valeur non exploitée du contenu non-structuré. Elles pourraient utiliser cette information pour répondre à des questions telles que :

- Est-ce qu'un appel au centre de support peut indiquer un risque de départ client ?
- Est-ce qu'une déclaration de sinistre présente des indices de tentative de fraude ?
- Est-ce que des retards de livraison provoquent des ventes manquées ?

Les sociétés ont besoin d'un moyen méthodique de trouver, extraire et analyser la masse de contenu non-structuré disponible aujourd'hui. Les fonctions d'analyse de contenu répondent à ce besoin.

Ce document traite du besoin d'analyse de contenu dans une organisation. Il aborde :

- Le besoin d'exploiter le contenu non-structuré pour une bonne conduite de l'activité.
- Les raisons justifiant une mise en œuvre rapide.
- Les principales caractéristiques de la solution IBM Content Analytics pour découvrir, affiner, visualiser et délivrer de nouvelles vues par l'analyse de contenus non-structurés.

La problématique

1 L'explosion du contenu non-structuré

Le contenu non-structuré est présent partout. Dans les messages électroniques, les documents, les blogs, les sites Web, les réseaux sociaux et les notes. Dans les systèmes de l'entreprise et aussi dans les réseaux sur des serveurs virtuels, à l'intérieur et à l'extérieur de l'organisation. Il est estimé que au moins 80 % de l'information générée journalièrement est non-structurée. Partie de cette information est pertinente pour l'activité de l'entreprise.

Cette information semi-structurée ou complètement en format libre échappe le plus souvent à une observation minutieuse. Il est difficile de consolider et exploiter parce que ce n'est pas organisé. Ce ne peut pas être recherché comme une base de données. Cependant il y a de l'information vitale qui pourrait être utilisée pour identifier les attentes clients, les faiblesses du centre de support, les incidents commerciaux, les tentatives de fraude, l'activité de garantie, les commentaires des clients et tout autre point sensible permettant de rendre l'entreprise plus efficace.



2 Les sociétés doivent rendre exploitable le contenu non-structuré

Les sociétés se handicapent elles-mêmes en ignorant la masse grandissante de contenu non-structuré et les informations utiles pour leurs activités qui y sont disponibles. Cependant le temps nécessaire et le coût d'examen, sélection et interprétation sont prohibitifs pour les sociétés désirant réaliser ces analyses par elles mêmes. Et même si elles le font, les résultats restent bien souvent subjectifs, anecdotiques et difficiles à suivre.

Les sociétés doivent disposer de moyens efficaces et économiquement viables d'analyse de contenu.

Comprendre la demande clients **en anticipant.**



Détecter les **déclarations frauduleuses** avant règlement.



Redéployer dynamiquement les ressources aux secteurs les plus tendus.



Sauver des vies en identifiant rapidement les risques.



Les raisons de l'évolution

1 L'évolution des fonctions d'analyse

L'analyse de contenu permet d'accéder, trier et analyser les contenus non-structurés et de les consolider avec les données structurées et tout autre type d'information disponible pour des comptes rendus et des analyses.

L'analyse de contenu est une extension naturelle de l'informatique décisionnelle (business intelligence). De nombreuses sociétés utilisent déjà l'informatique décisionnelle pour des prises de décision basées sur les données. C'est un processus de décision basé sur les informations des transactions passées et autres données structurées généralement stockées dans les entrepôts de données de l'entreprise.

Les sociétés peuvent compléter ces solutions d'informatique décisionnelle par des techniques d'analyse de contenu, qui peuvent alors être utilisées pour présenter des tendances basées sur du contenu non-structuré. L'analyse de contenu est particulièrement efficace pour suivre les évolutions du marché en quasi temps réel, comme une soudaine croissance de l'activité de garantie ou des réclamations clients. Cela donne aux sociétés la capacité de regrouper et analyser des informations critiques qui évoluent avec le marché et offre une puissante extension aux techniques traditionnelles d'informatique décisionnelle pour améliorer la prise de décision.

2 Des technologies éprouvées

Les capacités de l'analyse de contenu sont basées sur des technologies éprouvées qui ont été mises en œuvre depuis des années. Ces technologies peuvent absorber le volume croissant de contenu non-structuré provenant de la généralisation de l'utilisation d'Internet, des plates-formes de collaboration, des réseaux sociaux, des réseaux d'entreprise et des autres technologies d'information.

Plusieurs composants clés sont utilisés pour constituer une solution complète et efficace d'analyse de contenu :

- Technologies d'exploration, recherche et indexation pour identifier et marquer le contenu significatif dans une large variété de sources d'informations non-structurées.
- Technologies d'analyse sémantique pour trier, filtrer et catégoriser le contenu afin d'extraire les informations significatives et mettre en évidence automatiquement les tendances et anomalies.
- Technologies d'exploration et de présentation pour afficher des graphiques de corrélations, fréquences, déviations et tendances à des fins de compréhension rapide.

- Technologies ouvertes d'intégration pour mettre à disposition les extractions d'autres applications et systèmes.

Combinées, ces différentes technologies permettent d'identifier, trier, filtrer, organiser, visualiser et mettre à disposition les informations et vues critiques uniquement disponibles dans le contenu non structuré.

3 Un outil indispensable dans tout domaine d'activité

Avec l'explosion du volume du contenu non-structuré, il est devenu une nécessité de gestion d'investir dans un système d'analyse de contenu et d'exploiter l'information non-structurée afin de disposer d'un avantage compétitif. Par exemple imaginons qu'une compagnie d'assurance puisse utiliser l'analyse de contenu pour identifier les déclarations frauduleuses et les réduire de 10 % ? Un substantiel retour sur investissement peut être obtenu par les sociétés qui utilisent l'analyse de contenu pour suivre et intervenir dans des domaines ciblés comme les réclamations clients, les activités de garanties et les tentatives de fraude. Une solution d'analyse de contenu permet d'exploiter le contenu non-structuré omniprésent pour une amélioration significative des résultats de l'entreprise.

L'analyse de contenu peut être utilisée dans une grande variété d'industries et dans de nombreux contextes différents. Par exemple :

- Mesurer, suivre et améliorer le niveau de service d'un centre de support en analysant les rapports d'intervention.
- Adapter les plans de mise sur le marché selon les réclamations de garantie et les performances des produits.
- Développer une stratégie commerciale compétitive en analysant les informations sur la concurrence et les raisons d'actions gagnantes ou perdantes.
- Anticiper l'attrition du portefeuille clients ou les tentatives de fraude par l'analyse des déclarations.

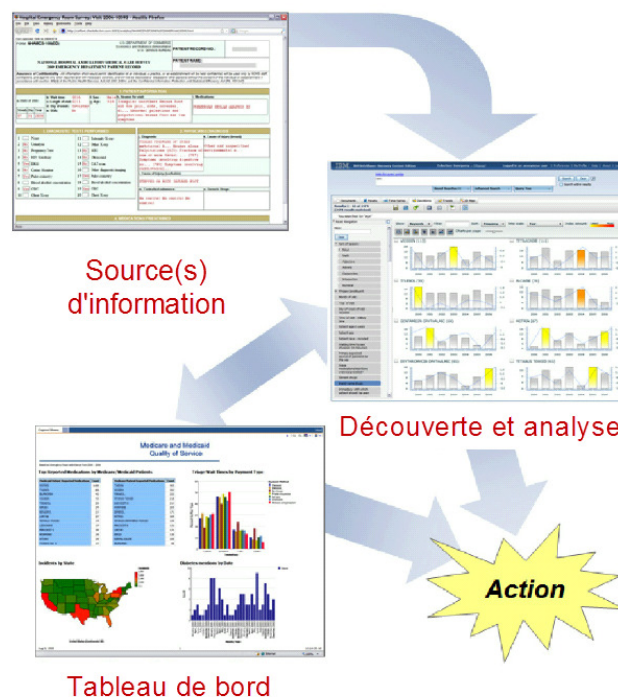
Tout type d'organisation peut bénéficier des aides apportées par l'analyse de contenu ; de l'industrie automobile à l'assurance en passant par le traitement de la délinquance.

IBM Content Analytics

IBM Content Analytics est une solution complète qui met à disposition des organisations tous les outils nécessaires pour libérer la valeur contenue dans l'information non-structurée. Une interface graphique et dynamique permet de découvrir des informations importantes en identifiant et faisant remonter des corrélations, mettant en évidence des tendances ignorées. Extension de l'informatique décisionnelle et amélioration du processus de décision avec l'information non-structurée.

IBM Content Analytics peut aider les sociétés à découvrir, affiner, visualiser et mettre à disposition de nouvelles vues sur l'activité. Il fournit des fonctions :

- Découvrir : Identifie et marque automatiquement les informations clés par l'analyse de toutes sources de contenu et l'identification des mots et expressions significatifs.
- Affiner : Permet des navigations et analyses en profondeur sélectives basées sur des critères clés et des catégories. L'information critique est extraite par l'application d'une série d'annoteurs qui identifient les concepts dans un texte et les relations entre ces concepts pour présenter le contenu pertinent aux analystes et décideurs.
- Visualiser : Met en évidence les écarts et les anomalies au travers de présentations graphiques et permet l'exploration pour des décisions et actions mieux circonstanciées.
- Mettre à disposition : Supporte l'intégration avec d'autres processus et applications par l'utilisation de standards ouverts. Cela inclut des entrepôts de documents ECM (Enterprise Content Management) comme Content Manager 8 et FileNet P8, des entrepôts de données comme IBM Data Warehouse ou Netezza, des outils complémentaires d'analyse et de présentation comme Cognos ou SPSS.



IBM Content Analytics utilise des technologies avancées basées sur des recherches réalisées par IBM et déployées depuis des dizaines d'années. Cette solution globale d'analyse de contenu utilise des algorithmes sophistiqués pour analyser les informations non-structurées dans presque tous les formats existants, incluant texte et données. Elle supporte aussi les contenus semi-structurés comme les documents XML, les pages HTML, les attributs de documents et les bases de données avec de longues zones de texte.

IBM Content Analytics peut être intégré avec des processus et des applications existants pour étendre leurs facultés d'analyse et de rapport. Il peut aussi être intégré avec des solutions de BI (Business Intelligence), ECM (Enterprise Content Management) et des entrepôts de données. Il peut également être utilisé comme ETL (Extract, Transform, Load) pour convertir du contenu non-structuré en données structurées.

IBM Content Analytics dispose avec IBM LanguageWare d'un environnement performant, basé sur Eclipse, de personnalisation et développement de fonctions avancées d'analyse de contenu pour une adaptation complète aux besoins spécifiques de chaque organisation.

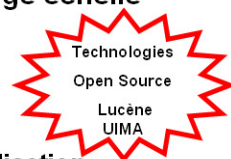
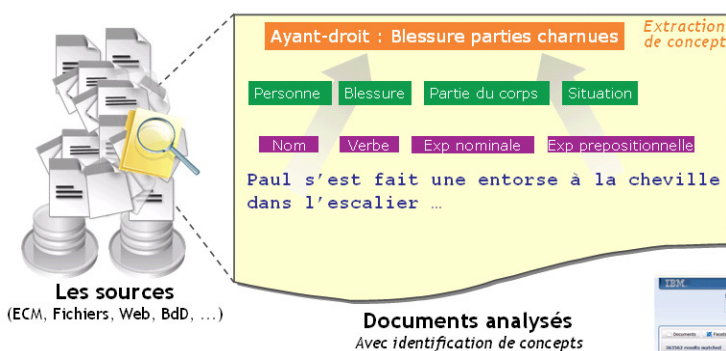
IBM Content Analytics intègre l'utilisation de la solution IBM Classification Module. Cet annotateur optionnel enrichit les fonctions d'IBM Content Analytics avec des bases de connaissances et des plans de décisions permettant de prendre en compte des plans de classement et des logiques spécifiques de traitement.

IBM Content Analytics est basé sur des standards ouverts (Open Source) comme Apache Lucene (Recherche texte) et UIMA (Unstructured Information Management Architecture).

IBM Content Analytics fonctionne dans les environnements AIX, Linux et Windows. La possibilité de répartir les différentes fonctions de traitement des contenus (Exploration, analyse-indexation et recherche) permet la prise en compte des volumes d'information et l'accompagnement des entreprises dans leur évolution.

Content Analytics permet des analyses inconcevables

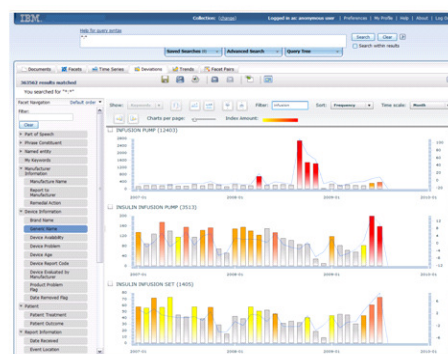
Consolide les conclusions, étend la compréhension sur une large échelle



Visualisation automatique

Concepts et sources sont présentés dans l'interface utilisateur

- Analyse document par document
- Consolidation des conclusions
- Evaluation sur de gros volumes d'informations
- Inconcevable par traitement manuel ou hors de coût et délai



Pour en savoir plus

1 IBM Content Analytics

1.1 Portail IBM Software

<http://www-01.ibm.com/software/data/content-management/products/search.html>

1.2 IBM Redbooks

IBM Content Analytics Version 2.2: Discovering Actionable Insight from Your Content

<http://www.redbooks.ibm.com/redpieces/abstracts/sg247877.html?Open>

2 IBM LanguageWare

2.1 Portail IBM Software

<http://www-01.ibm.com/software/ebusiness/jstart/textanalytics/infocenter.html>

2.2 IBM alphaWorks

<http://www.alphaworks.ibm.com/tech/lrw>

3 IBM Classification Module

3.1 Portail IBM Software

<http://www-01.ibm.com/software/data/content-management/classification/>