

Approches d'analyse des données issues des médias sociaux appliquées au domaine de la
santé

par

Ulrick Serge Mba Engonga

Essai présenté au CeFTI

en vue de l'obtention du grade de maître en technologies de l'information

(Maîtrise en génie logiciel incluant un cheminement de type cours en technologies de
l'information)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Longueuil, Québec, Canada, octobre 2015

Sommaire

En 2014, une étude, *NETendances 2014*, réalisée par le CEFRIO et intitulée « Les médias sociaux, au cœur du quotidien des québécois – Usage du Web, médias sociaux et mobilité¹ » montre qu’au cours des dernières années, les médias sociaux se sont démocratisés pour prendre aujourd’hui une place préminente sur le Web; cette étude note que plus de huit internautes québécois sur dix (85,0 %) utilisent les médias sociaux pour consulter du contenu, se connecter à leur compte, relayer ou partager du contenu, interagir avec d’autres ou créer du contenu. Cette proportion représente 71,7 % des adultes québécois. Cela montre un intérêt majeur des médias sociaux dans notre société. En parallèle, plusieurs enquêtes réalisées aux États-Unis et au Canada signalent depuis quelques années que les médias sociaux, notamment les blogs, forums, tels que *Twitter* et les sites de réseautage en ligne comme *Facebook*, sont de plus en plus utilisés par les individus pour obtenir toute sorte d’information, notamment sur la santé.

Plusieurs organismes ont un site web et des comptes sur les principaux médias sociaux. Ils démontrent par là un intérêt pour les nouvelles technologies de l’information en général et les médias sociaux en particulier. Leur objectif principal est de contrôler leur image, d’améliorer leur communication en s’adressant directement à leur public cible. Toutefois, peu ont véritablement compris et saisi l’énorme avantage que peut représenter l’analyse et le traitement des données issus des conversations sur les médias sociaux pour atteindre, voire améliorer leurs objectifs. Cela est d’autant plus vrai dans le domaine de la santé publique; plusieurs grands hôpitaux et organismes en santé et soins médicaux ont une forte présence sur les médias sociaux et sont suivis par beaucoup d’utilisateurs qui s’intéressent et commentent l’actualité et autres sujets qui se réfèrent à ces grands hôpitaux. Il n’en demeure pas moins que les données générées par ces personnes et l’opinion qu’elles expriment via les médias sociaux ne sont pas encore très

¹ <http://www.cefrio.qc.ca/netendances/medias-sociaux-coeur-quebecois/>

prises en compte, comme cela se fait déjà dans les domaines comme le commerce électronique ou la politique. Or, avec les médias sociaux, les patients sont de plus en plus informés et engagés dans leur choix sur les organismes qui prendront en charge de leurs soins de santé. Étant donné que les patients sont des acteurs au cœur même de la prestation des soins de santé, leurs conversations et les données qu'ils produisent sont un indicateur important pour mesurer l'opinion sur la qualité des soins; les données générées sont aussi une mine d'or pouvant aider à améliorer leur image et obtenir des précieuses informations décisionnelles.

Jusqu'à présent, les systèmes d'information développés pour les organismes répondaient au besoin de traitement des données internes. En ouvrant leurs portes aux plates-formes de médias sociaux, des questions se lèvent sur comment les organismes en santé peuvent traiter ces données externes pour en tirer de la valeur. Cette essai a donc pour objectif de voir quelles sont les « approches d'analyse et de traitement des de données issues des médias sociaux dans le domaine de la santé ».

Cet essai se concentre sur les méthodes et techniques pour extraire, analyser et classifier les données issues médias sociaux. Ces données sont générées par des personnes qui s'expriment au sujet des principaux établissements en santé au Québec. La méthodologie de recherche appliquée cherche à collecter les données issues du réseau social Twitter, pour trois grands hôpitaux : Le CHUM, le CHU Ste-Justine et l'Hôpital Montréal pour enfants. Une fois collectées, ces données sont analysées à l'aide trois types d'analyses : l'analyse de sentiments, l'analyse de l'influence et l'analyse du volume. Les résultats ont permis de démontrer que les non seulement les réseaux sociaux sont utilisés comme source d'information, mais aussi que les données issues de ces réseaux peuvent être exploitées pour obtenir des indicateurs clés pouvant être utiles à une meilleure prise de décision de la part des exécutifs au sein des organismes en soins de santé au Québec.

Remerciements

Dans le cadre de cet essai, je tiens à remercier mon directeur académique M. Pierre-Martin Tardif et Mme Lynn Legault de m'avoir guidé et conseillé dans les choix à faire pour mener à bien ce travail.

Je veux aussi remercier M. Claude Cardinal, Directeur adjoint au Centre de formation en technologies de l'information (CeFTI). Merci pour votre dévouement et votre disponibilité tout au long de ce cursus de Maîtrise en génie logiciel à l'Université Sherbrooke.

Enfin, je veux remercier spécialement, ma femme Rita que j'aime tant, mon fils David ainsi que tous mes proches. Merci pour votre soutien et vos prières durant ces moments d'absences et de dur labeur.

Table des matières

Sommaire.....	iii
Remerciements	v
Table des matières	vi
Liste des tableaux	ix
Liste des figures.....	x
Glossaire	xi
Liste des sigles, des symboles et des acronymes.....	xii
Introduction	1
Chapitre 1 Mise en contexte et problématique	3
1.1 Mise en contexte	3
1.2 Problématique	7
1.3 Sujet, objectif et hypothèse.....	9
Chapitre 2 Revue de la littérature	11
2.1.1 Les médias sociaux	11
2.1.2 Définition des médias sociaux.....	11
2.1.3 Catégories de médias sociaux.....	12
2.1.4 Twitter.....	13
2.2 L'environnement des soins de santé.....	14
2.2.1 Perception du public sur le système de santé canadien	14
2.2.2 Opinion sur la qualité des soins de santé au Québec	17
2.2.3 Défis de l'industrie des soins de santé	18
2.2.4 Le patient, un acteur au cœur du système de santé.....	19
2.3 Les médias sociaux au sein de l'industrie des soins de santé.....	21
2.3.1 Utilisation des médias sociaux dans la santé à l'échelle mondiale.....	21
2.3.2 Utilisation des médias sociaux dans le système de santé au Canada.....	24
2.3.3 Mesures de médias sociaux dans l'environnement de santé.....	25
2.4 L'analyse des médias sociaux.....	26

2.4.2	Approches d'analyse de sentiments.....	31
2.4.3	Analyse du volume	37
2.4.4	Analyse de la portée potentielle.....	38
2.4.5	Analyse du niveau d'engagement.....	38
2.5	Conclusion de la revue de littérature	38
Chapitre 3 Méthodologie		40
3.1	Type de recherche.....	40
3.2	Milieu, population et échantillon.....	40
3.3	Déroulement de l'enquête.....	41
3.4	Démarche méthodologique.....	41
3.4.1	Collecte de données	42
3.4.2	Formatage et insertion des données.....	43
3.4.3	Mesure et évaluation des mesures	43
3.4.4	Présentation des résultats.....	43
3.5	Outils de mesure	44
3.5.1	Application ASTAS de collecte et traitement de données :	44
3.5.2	Architecture de l'application ASTAS.....	45
3.5.3	Diagrammes UML de l'application ASTAS	46
3.5.4	Base de données.....	47
3.5.5	Open Authentication (OAuth)	47
3.5.6	L'API de Twitter.....	48
3.5.7	Classe principale main de l'application ASTAS	49
3.5.8	Semantria pour Excel de LEXALYTICS	50
Chapitre 4 Expérimentation et analyse des résultats		52
4.1	Ensemble de données collectées	52
4.2	Résultats et discussions.....	53
4.2.1	Analyse du sentiment.....	53
	<i>Discussion de l'analyse des sentiments</i>	55
4.2.2	Analyse de l'influence	56
	<i>Discussion sur le niveau d'engagement</i>	57
	<i>Discussion sur la portée potentielle</i>	59
4.2.3	Analyse du volume	59
	<i>Discussion sur les utilisateurs les plus actifs</i>	61

<i>Discussion sur les utilisateurs les plus influents</i>	64
<i>Discussion sur les autres mesures</i>	65
<i>Discussion sur le volume dans le temps</i>	68
Conclusion	69
Liste des références	74
Bibliographie	79
Annexe I API Twitter	80
Annexe II API Datumbox.....	85
Annexe III Code source de l'application ASTAS et données collectées avec Excel	91
Annexe IV Résultats de l'analyse des sentiments	92

Liste des tableaux

Tableau 1.Catégories de gazouillis	14
Tableau 2.Mesures d'analyse d'influence	37
Tableau 3.Champs de la table Gazouillis	47
Tableau 4.Champs d'un gazouillis collecté	52
Tableau 5.Termes de recherche et nombre de gazouillis collectés.....	53
Tableau 8.Niveau d'engagement.....	57
Tableau 9.Portée potentielle	58
Tableau 10.Chumontreal: Utilisateurs les plus actifs	60
Tableau 11.ChuSteJustine: Utilisateurs les plus actifs	60
Tableau 12.HopitalChildren: Utilisateurs les plus actifs	61
Tableau 13.Chumontreal: Utilisateurs les plus influents.....	62
Tableau 14.ChuSteJustine: Utilisateurs les plus influents.....	63
Tableau 15.HopitalChildren: Utilisateurs les plus influents.....	64
Tableau 16.Autres mesures de volume.....	65
Tableau 6.Sentiments approche basée sur une machine d'apprentissage.....	92
Tableau 7.Sentiments avec approche basée sur un lexique	94

Liste des figures

Figure 1.Séquencement de la revue de littérature.....	11
Figure 2.Perception des canadiens sur la qualité des services de santé.....	16
Figure 3.Opinion sur le système de santé au Québec et dans d'autres provinces et pays.....	17
Figure 4.Définition du sentiment.....	27
Figure 5.L'analyse des sentiments	28
Figure 6.Approches et techniques d'analyse de sentiments.....	31
Figure 7.Approche basée sur un lexique	32
Figure 8.Démarche méthodologique	42
Figure 10.ASTAS App: Interface des résultats pour le CHUM.....	44
Figure 11.Architecture du système.....	45
Figure 12.Diagramme UML de cas d'utilisation	46
Figure 13.Diagramme de classes	46
Figure 15.REST Twitter API 1.1	49
Figure 16.Semantria Excel de LEXALYTICS	51
Figure 19.Chumontreal: Utilisateurs les plus influents	62
Figure 20.ChuSteJustine: Utilisateurs les plus influents	63
Figure 21. HopitalChildren: Utilisateurs les plus influents	64
Figure 22.Chumontreal: Volume dans le temps	66
Figure 23.ChuSteJustine: Volume dans le temps	67
Figure 24.HopitalChildren: Volume dans le temps	67
Figure 17.Sentiments avec l'approche basée sur une machine d'apprentissage	93
Figure 18.Sentiments avec l'approche basée sur un lexique	94

Glossaire

Twitter	Service gratuit de messagerie sociale pour envoyer et recevoir des messages courts en temps réel.
Gazouillis	Message texte d'environ 140 caractères posté sur le média social Twitter.
Username	Pseudonyme précédé d'un @, identifiant un utilisateur; ex: « @username ».
Retweet	Gazouillis d'un autre utilisateur, qui est republié tel quel. Précédé de RT; Exemple : « RT @username ... Retweet ... »
Mention	Gazouillis qui inclut au moins l'User Name d'un autre utilisateur. Exemple : « ... @username de mention ... »
Replies	Gazouillis débutant par l'User Name d'un autre utilisateur pour lui répondre. Exemple : « @username replies a un gazouillis ... »
Framework	Cadre de travail ou ensemble de composants structurant un logiciel.
Abonné	Abonné d'un compte qui peut suivre les gazouillis de ce compte.
Hashtag	Terme précédé d'un #, peut être associé à une conversation. Exemple: « #hashtag »
Fréquence	Nombre d'occurrences d'un terme dans le texte.

Liste des sigles, des symboles et des acronymes

CEFRIO	Centre facilitant la recherche et l'innovation dans les organisations, à l'aide des technologies de l'information et de la communication (TIC)
CHUM	Centre hospitalier de l'Université de Montréal
CSUM	Centre de santé de l'Université McGill
API	<i>Application Programming Interface</i> : est un ensemble normalisé de classes, de méthodes ou de fonctions qui sert de façade par laquelle un logiciel offre des services à d'autres logiciels
ASTAS	Analyse de Sentiment Twitter Appliquée à la Santé.
SVM	<i>Support Vector Machines</i> : sont une classe d'algorithmes d'apprentissage.
PHP	<i>Hypertext Processor</i> : Langage de programmation utilisé pour développer des pages Web.

Introduction

Depuis la création d'Internet et ce jusqu'à un passé récent, les sites web étaient des plateformes fournissant de l'information en lecture seule pour la plupart des utilisateurs. Suite à l'avènement du Web 2.0 et surtout l'utilisation massive des médias sociaux, tels que Twitter ou Facebook, le nombre d'utilisateurs a considérablement augmenté et leurs rôles a évolué en celui de pourvoyeur d'information sur le web. Les médias sociaux ont donc fourni un canal d'information bidirectionnel, le premier crée un pont entre les utilisateurs et les propriétaires des sites, et le deuxième entre les utilisateurs eux-mêmes.

Dans la même tendance, l'industrie des soins de santé publique a abondamment adopté les médias sociaux comme moyen de communication pour rester en contact avec leurs patients et diffuser de l'information. Traditionnellement, les patients n'avaient accès à de l'information que lorsqu'ils étaient en face d'un professionnel en soins de santé médicaux, où encore dans des revues spécialisées écrites par des experts. Actuellement, avec l'utilisation de nouveaux moyens technologiques tels que les téléphones intelligents directement connectés sur les médias sociaux, les patients peuvent rechercher de l'information sur la santé, participer au débat sur un sujet en santé, apprécier un service ou un produit provenant d'un organisme de santé; ils peuvent même entrer en contact avec des experts en santé.

Ces interactions créent un volume important d'informations potentiellement utilisables. Le patient est beaucoup plus informé, a de plus en plus de pouvoir dans la prise de décision quant aux services qui lui seront octroyés et peut dès lors, choisir quel organisme a, selon lui, le meilleur potentiel pour lui répondre à ses besoins et lui administrer les soins médicaux souhaités. Il est donc important pour ces organismes en soin de santé de savoir comment collecter les données provenant des médias sociaux et surtout comment analyser ces données afin de mieux comprendre leur écosystème social, et déceler les opinions et tendances qui leur permettraient

de maîtriser leur communication et d'améliorer l'image qu'ont leurs patients par rapport aux soins et services offerts.

Le sujet de cette étude porte sur les approches d'analyse des données issues des médias sociaux appliquées au domaine de la santé. Pour explorer ce sujet, l'essai s'organise comme suit :

Le premier chapitre effectue une mise en situation permettant de placer le lecteur dans le contexte dans lequel se déroule l'étude. Ce chapitre formule aussi la problématique de recherche et énonce les questions, les objectifs ainsi que l'hypothèse de recherche. Le deuxième chapitre présente la revue de littérature qui fera un survol des articles ou ouvrages qui traitent de l'analyse de données issues des médias sociaux, ainsi que les défis des organismes en soins de santé au Québec et leur utilisation des médias sociaux. Le troisième chapitre présente la méthodologie de recherche choisie ainsi que sur les instruments et techniques de collectes de données. Cette partie définit la population et l'échantillon d'étude. Elle explique aussi le déroulement de la collecte et présente l'analyse des données recueillies. Enfin, le quatrième chapitre présente les résultats obtenus. Ce chapitre permet de voir si les résultats obtenus vérifient l'hypothèse élaborée au départ de cet essai. Enfin des recommandations seront aussi proposées.

Chapitre 1

Mise en contexte et problématique

1.1 Mise en contexte

Un des enjeux les plus importants au Québec aussi bien pour les gouvernants, les décideurs, que pour la population est la santé². En effet, à la question de savoir quels sont les trois sujets les plus importants, un sondage CROP-Radio-Canada révèle que 35 % des gens citent les soins de santé et la qualité des services reçus dans les hôpitaux. Un des éléments essentiels dans ces hôpitaux est le département des urgences. Il est un des points d'accès critiques car on y offre les premiers services à la population; les urgences sont aussi un important carrefour entre les professionnels en santé et les patients qui y sont traités. Ces départements des urgences sont responsables en moyenne de 50 % des admissions des établissements de soins de santé au Québec. C'est donc, une porte d'entrée majeure pour les patients et un excellent contexte permettant mesurer l'opinion sur la qualité des soins.

Le système de santé en général et les grands hôpitaux n'ont pas bonne presse au Québec, c'est un problème qui date même depuis les années 70. Très souvent, dans les médias, on voit des reportages de salles d'attentes engorgées, de patients sur des civières longeant les couloirs des hôpitaux. Le temps d'attente au service d'urgence dans les hôpitaux est un sujet qui préoccupe beaucoup de québécois³. Malgré les promesses et les efforts, plus d'un million de patients ont dû attendre en moyenne 17 heures 30 minutes sur une civière aux urgences au cours de l'année 2013. Cette situation est encore pire à Montréal où l'attente est encore plus longue. Les malades

² <http://ici.radio-canada.ca/sujet/elections-quebec-2014/2014/03/10/007-sondage-crop-election-enjeux.shtml>

³ <http://www.hebdosregionaux.ca/monteregie/2014/04/02/en-cette-campagne-electorale-quels-sont-les-enjeux-qui-vous-preoccupent>

doivent attendre en moyenne 21 heures sur une civière avant d'obtenir une chambre, une place dans une autre ressource du système de santé ou même de retourner à la maison.

Le Québec n'est pas le seul à devoir faire face à cette problématique. De nombreuses provinces sont aussi aux prises avec l'engorgement de leurs urgences. En Ontario, le surpeuplement des urgences est également un problème et le temps moyen en 2006 était de trois-heures et dix-huit minutes. Dans la région d'Edmonton, le temps d'attente moyen était de quatre heures et trente-six minutes alors qu'aux États-Unis, il était de trois heures et douze minutes. Bien sûr, il existe des différences quant aux définitions et à la qualité des données susceptibles d'expliquer une partie de la différence mais il semble manifeste que le Québec soit dans une classe à part au chapitre des temps d'attente dans les urgences.

Un effort de transparence a été fait de sorte que les temps d'attente dans toutes les urgences de la province sont accessibles. Chaque année, le journal La Presse, comme d'autres sites Internet, publient des sondages et des palmarès sur les urgences⁴. Si une note moyenne de C est souvent octroyée, la situation est en réalité plus sombre car le temps qu'un patient passe à attendre en moyenne sur une civière ne cesse d'augmenter. En 2014, le palmarès des urgences révèle que ce temps d'attente est en moyenne de seize heures quarante-deux minutes. À Montréal, où l'on retrouve les plus grands hôpitaux en termes de nombre de lits, l'attente est encore plus longue et n'a pas beaucoup diminué au fil des ans.

Outre le fonctionnement des urgences, un des sujets qui font encore plus parler dans la santé est la gouvernance des projets de grands hôpitaux tels que le Centre universitaire de l'Université de Montréal (CHUM) et le Centre universitaire de santé McGill (CUSM). En effet, la rémunération des directeurs généraux et médecins, les dérapages budgétaires, le gaspillage de fonds publics sont de nombreux sujets qui se retrouvent dans la presse et dans les discussions au sein de la population québécoise.

⁴ <http://www.lapresse.ca/multimedias/201405/13/01-4766374-neuvieme-palmares-des-urgences-tableau-et-carte-interactive.php>

Si l'image ou la perception de la plupart des organismes en santé publique et en soins médicaux au Québec stagne, une poignée d'hôpitaux parviennent à améliorer quelque peu leur fonctionnement, leur communication et leur performance dans le temps d'attente aux urgences; la rénovation et l'amélioration des services y contribuent grandement. On peut citer par exemple, l'exemple du nouveau centre d'urgences inauguré à l'Hôpital juif de Montréal⁵. L'ambition clairement affichée de ce nouveau département d'urgences est de réduire considérablement le temps d'attente des patients et d'en faire un des plus bas de la province. Ce service de l'urgence de l'hôpital juif, qui est un des plus achalandés en ce qui concerne le nombre de patients sur civières, veut s'assurer du maintien et de l'amélioration de la qualité des soins. Bien que le nombre d'employés et de civières demeure sensiblement le même, les installations élargies et les infrastructures perfectionnées visent à mettre en œuvre les meilleures pratiques en matière de soins d'urgence et d'introduire des innovations conformes aux tendances actuelles. L'objectif consiste à donner aux patients des soins d'urgence de qualité supérieure tout en favorisant leur retour dans la communauté le plus rapidement.

D'autres exemples d'effort pour améliorer les prestations de soins et l'image du système de santé au Québec est la mise en place des grands hôpitaux comme le CHUM ainsi que CUSM et des hôpitaux complètement dédiés aux soins des enfants comme l'Hôpital Ste-Justine ou l'Hôpital Montréal pour enfants. Parmi ces grands hôpitaux, voyons l'exemple du CHUM.

Histoire du CHUM

Le Centre hospitalier de l'Université de Montréal (CHUM) est une initiative récente, il est né de la fusion de trois hôpitaux, soit l'Hôtel-Dieu, l'Hôpital Notre-Dame et l'Hôpital Saint-Luc. Le 1er octobre 1996, à la suite de la signature d'un protocole d'entente entre les représentants du Gouvernement, de la Régie régionale de la santé et des services sociaux de Montréal-Centre, de l'Université de Montréal et des trois établissements concernés, le CHUM a été officiellement constitué.

⁵ <http://www.ledevoir.com/societe/sante/399088/20-minutes-promet-l-hopital-general-juif>

Mission du CHUM

Le CHUM est le centre hospitalier universitaire de l'Université de Montréal. Il offre prioritairement des soins et des services surspécialisés à une clientèle adulte régionale et suprarégionale. Innovateur et exemplaire, doté d'un centre de recherche qui le distingue, le CHUM se démarque comme pôle unique de développement, de mise en pratique et de transfert des connaissances à travers ses activités intégrées de soins, de recherche, d'enseignement, d'évaluation des technologies et modes d'intervention en santé, et de promotion de la santé.

Le patient est au cœur des valeurs du CHUM

Le CHUM est un établissement universitaire, un lieu de connaissances et de transfert du savoir, d'innovation, de technologies et de pratiques de pointe, mais il est aussi un milieu humain. En mettant le patient au cœur de son action, le CHUM agit en étroite collaboration avec lui et ses proches. Il fournit les meilleurs soins et joue un rôle déterminant dans l'amélioration de la santé et du bien-être de la population québécoise.

Chaque jour, son savoir-faire se traduit par des milliers d'interactions. Le CHUM accorde beaucoup d'importance à la qualité des comportements. Un sourire, de l'attention, une présence chaleureuse et la courtoisie sont des priorités au CHUM.

Les valeurs de l'institution reflètent l'ambition d'offrir la meilleure qualité et la meilleure expérience de soins et de services aux patients et à leurs proches. Le respect de la dignité et de l'autonomie des patients humanise leurs relations avec les différents professionnels de la santé. Pour qu'elles aient du sens et orientent les actions, tous s'appliquent à traduire les valeurs en actions concrètes.

Vision du CHUM

Porteur d'un grand projet pour le Québec, le CHUM se distingue parmi les grands centres hospitaliers universitaires à l'échelle internationale par :

- la qualité des soins et des services offerts à ses patients, considérés comme des partenaires des épisodes de soins et de sa mission;
- le leadership suprarégional et le rayonnement international de ses secteurs de pointe en soins, en recherche et en enseignement;
- ses pratiques novatrices, performantes et intégrées, au service du bien-être et de la santé du patient et de la population;

- la qualité de l'enseignement dispensé pour l'ensemble des professions de la santé, l'innovation de ses pratiques de formation et l'efficacité du transfert de connaissances;
- l'excellence dans ses pratiques de gestion et sa capacité à créer un milieu de travail et de formation sain et stimulant, permettant à son personnel et à ses étudiants d'offrir leur plein potentiel;
- l'engagement constant du personnel, des médecins, des étudiants et des bénévoles;
- la qualité de ses partenariats au sein d'un réseau intégré en santé et en éducation, ainsi qu'avec ses partenaires issus des milieux scientifique, économique et communautaire;
- son implication citoyenne et sa contribution au projet de société du Québec.

Nouvelle stratégie de communication

Le CHUM a entrepris de nombreux travaux majeurs pour offrir une expérience de qualité à ses patients. Aussi, on remarque que cet organisme a adopté une approche de communication moderne basée sur les médias sociaux. En plus de son site web, le CHUM a créé plusieurs comptes actifs sur les médias sociaux les plus connus tels que Twitter⁶ et Facebook. Ce nouveau mode de communication est plus direct et va de pair avec les autres moyens que sont la presse écrites et les revues scientifiques.

1.2 Problématique

L'utilisation des médias sociaux par les utilisateurs-patients est fondamentale, car elle leur permet de partager des opinions et rechercher des informations sur divers sujets. Sur les différentes plateformes de médias sociaux, les utilisateurs-patients peuvent émettre des sentiments et participer à des discussions. Certains utilisateurs partagent des informations sur des sujets relatifs à la santé, comme leurs expériences avec des centres de santé publiques et des établissements de soins médicaux. Ils peuvent ainsi apprécier un service, recommander un médecin ou des cliniques. Ils peuvent aussi se plaindre de certains aspects particuliers dudit

⁶ <https://twitter.com/chumontreal> <http://facebook.com/chum.montreal>

établissement, tels que la bonne gouvernance, la qualité du service, le fonctionnement dans le département des urgences (heure d'attente, compétence et attitude des soignants). Ces données peuvent être d'une grande valeur si elles sont extraites, traitées et analysées. Cela a intéressés plusieurs concepteurs de logiciels. Il existe plusieurs applications informatiques qui permettent de faire l'analyse des données issues des médias sociaux et la détection de sentiments exprimés dans les conversations.

Le problème est que le volume de ces données produites et partagées entre les différents utilisateurs de médias sociaux connaît une croissance importante au cours de ces dernières années, sans pour autant qu'on sache très bien exploiter ses données. Or, l'analyse des médias sociaux peut être utilisée pour générer des informations stratégiques ou pour détecter les opinions qui se dégagent des conversations issues de ces médias sociaux. Elle peut aussi être utilisée pour comprendre le comportement des utilisateurs et leurs recommandations sur de l'image d'une institution ou la qualité de services reçus.

La revue de littérature fait état des recherches et études de cas menées dans plusieurs domaines - les domaines les plus en vogue étant l'industrie musicale et cinématographique ou encore la politique - Ces recherches portent notamment sur comment collecter et analyser les données issues des médias sociaux. Dans le domaine des affaires, on utilise l'analyse des médias sociaux pour cibler la publicité et pour faire la surveillance de marque. Quant au domaine de la politique, l'analyse des sentiments est utilisée pour connaître les tendances de votes. Par contre, il y a moins d'attention sur ce sujet dans le domaine de la santé et des soins médicaux, encore moins au Québec. Si la plupart des organismes de ce domaine ont une présence sur les principaux médias sociaux tel que Twitter - c'est le cas du CHUM de CHU Sainte-Justine, de l'Hôpital Montréal pour enfants - il n'en demeure pas moins que peu savent et entreprennent une réelle analyse des messages émis sur eux sur les médias sociaux. Peu regardent avec attentions les commentaires écrits sur eux par leurs patients ou par les journalistes. Or, l'analyse de ces conversations peut leur offrir la possibilité de convertir cette masse de données en informations stratégiques permettant de prendre des décisions éclairées. Cela peut aussi servir à mesurer l'impact de ces décisions ciblées au fil du temps.

1.3 Sujet, objectif et hypothèse

La perception du public et la couverture médiatique du système de santé sont une préoccupation majeure au Québec. Les sujets tels que la gouvernance des grands hôpitaux et l'expérience vécue dans les services d'urgence font l'objet de plusieurs conversations sur des médias sociaux comme Twitter. L'analyse des conversations issues des médias sociaux peut s'avérer particulièrement intéressante dans le domaine de la santé publique. En effet, il serait intéressant de pouvoir trouver des voies et moyens permettant de collecter et analyser les données issues des médias sociaux, l'objectif étant d'engendrer des informations de valeur comme par exemple l'opinion des patients et l'appréciation qu'ils ont de la qualité des services qu'ils reçoivent dans les organismes en soins de santé. C'est pourquoi le sujet de cet essai porte sur les approches d'analyse des médias sociaux appliquées aux conversations sur les établissements de santé au Québec. L'étude mesurera plus précisément les données issues des commentaires postés sur le média social Twitter et qui concernent trois grands hôpitaux : le CHUM, le CUSM et Centre hospitalier Ste-Justine. Afin de mieux explorer ce sujet, la question de recherche suivante est posée :

Est-ce que les approches d'analyse des conversations issues des médias sociaux peuvent aider à identifier l'appréciation de l'image et de la qualité de service dans les organismes en soins de santé et leur fournir des indicateurs clés pour une meilleure prise de décision?

Cette question comporte plusieurs autres interrogations qui aideront à faire le tour du sujet :

- Qu'est-ce qu'un média social?
- Qu'est-ce que l'analyse des médias sociaux?
- Quelles sont les données pouvant être collectées et comment les extraire?
- Quelles mesures pour l'analyse de médias sociaux en santé?
- Quelles sont les approches d'analyse des médias sociaux et comment transformer les données analysées en information stratégique?

L'hypothèse de cet essai s'énonce comme suit :

Les approches d'analyse des conversations issues des médias sociaux peuvent aider un organisme en soin de santé à identifier l'appréciation de leur image, de la qualité de leurs services et leur fournir des informations stratégiques pour une meilleure prise de décision.

Chapitre 2

Revue de la littérature

Les principaux sujets sont appuyés par des références de la littérature. Le séquençement de la revue de la littérature est schématisé dans la figure suivante :

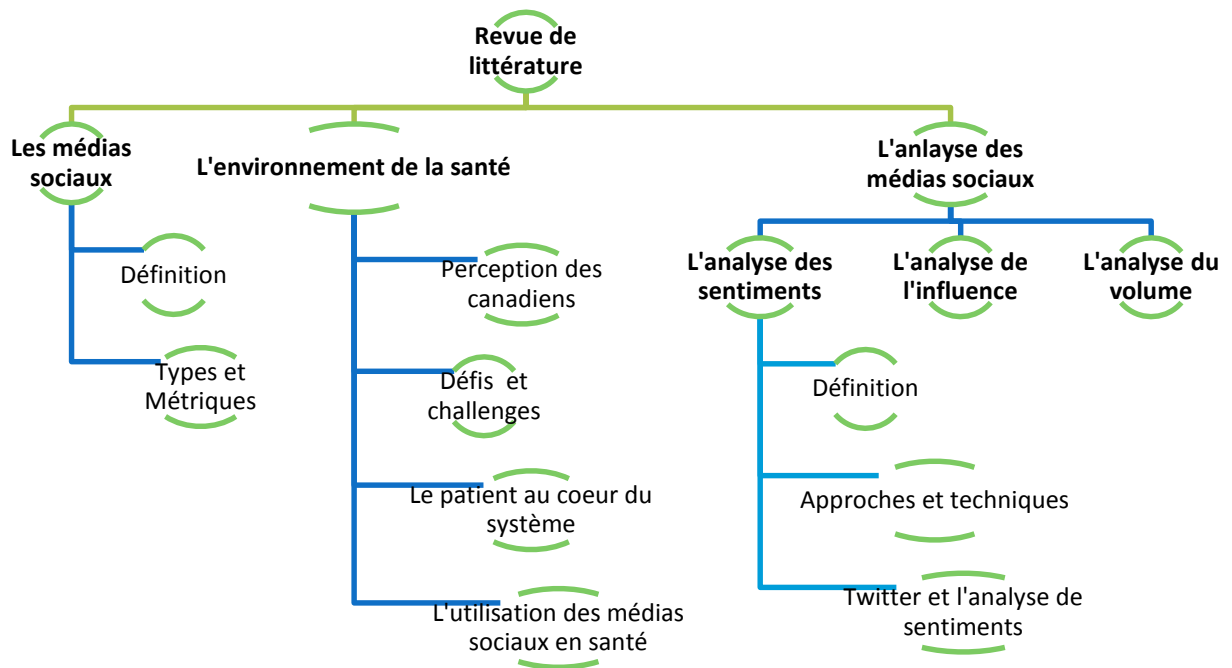


Figure 1. Séquençement de la revue de littérature

2.1.1 Les médias sociaux

2.1.2 Définition des médias sociaux

Sur Internet, plusieurs écrits confondent les termes médias sociaux et Web 2.0. Ce n'est pas le cas de Kaplan et Haenlein (2009) [1]. Ces derniers affirment que pour définir les médias sociaux,

une distinction claire doit être établie entre les médias sociaux et d'autres termes associés comme le Web 2.0 ou le contenu généré par les utilisateurs.

Le terme Web 2.0 a été pour la première fois utilisé par Tim O'Reilly (2005) [2]. Par Web 2.0, l'auteur fait référence au développement d'Internet comme un média interactif. La forme et le contenu initialement disponible sous formes de sites Web qui ne pouvaient être consultés passivement que par les utilisateurs; cela est qualifié de Web 1.0. En permettant aux utilisateurs d'interagir et de créer du contenu eux-mêmes, Internet a évolué vers ce qui est communément appelé Web 2.0. Cependant, Kaplan et Haenlein (2009) considèrent le Web 2.0 « comme la plate-forme pour l'évolution des médias sociaux ». Après avoir clairement distingué les concepts associés, Kaplan et Heanlein (2009) définissent les médias sociaux comme « un groupe d'applications basées sur Internet qui s'appuient sur les fondements idéologiques et technologiques du Web 2.0, et qui permettent la création et l'échange de contenu généré par l'utilisateur» [1].

2.1.3 Catégories de médias sociaux

Il existe plusieurs applications de médias sociaux sur Internet. Ces applications peuvent être réparties dans les catégories suivantes :

2.1.3.1 *Les réseaux sociaux*

Une définition des réseaux sociaux a été élaborée par Boyd et Ellison (2007) [3]. Selon les auteurs, un service web peut être défini comme étant un site de réseau social s'il permet aux utilisateurs :

- 1) De construire un profil public ou semi-public dans un système;
- 2) D'articuler une liste d'autres utilisateurs avec lesquels ils partagent une connexion;
- 3) Voir et parcourir la liste de connexions et contenus effectués par d'autres au sein du même système.

Kaplan & Heanlein (2009) définissent les sites de réseaux sociaux comme des fonctions permettant de s'interconnecter avec d'autres utilisateurs en les invitant à avoir accès à leurs pages personnelles [1]. Il existe plusieurs réseaux sociaux, les plus connus étant : Facebook (www.facebook.com) ou encore LinkedIn (www.linkedin.com).

2.1.3.2 Les blogues et micro-blogues

Une définition d'un blog est proposée par Vickery et Wunsch-Vincent (2007) [4], un blog est «un type de site Web qui affichent généralement les entrées horodatées dans l'ordre chronologique inverse». La plateforme de blog la plus connue est WordPress (www.wordpress.com).

Les blogs permettent aux gens de partager leur histoire ou une opinion sur plusieurs sujets. Bien que généralement gérés par une personne physique, les blogs sont souvent interactifs puisque les lecteurs ont la possibilité de laisser des commentaires. Une variante sur les blogs normaux est l'application de type « *micro-blogging* » qui permet aux utilisateurs d'envoyer des messages courts d'un nombre maximum de caractères. L'application de « *micro-blogging* » la plus célèbre est *Twitter* (www.twitter.com).

2.1.4 Twitter

Avec plus de 271 millions d'utilisateurs actifs et 500 millions de « *gazouillis* » postés par jour, *Twitter* est devenu l'un des médias sociaux les plus utilisés⁷. C'est un micro-blogue qui permet à un utilisateur d'écrire et lire des messages sur Internet. Ces courts messages publiés sont appelés « *gazouillis* » ou gazouillis en français et sont limités à 140 caractères. Après s'être enregistré, il est possible non seulement de poster « *gazouilliser* » un message, mais aussi de « Retweeter » un message pour l'afficher à ses abonnés tel qu'un utilisateur l'a soi-même vu. Pour accéder aux « gazouillis » postés d'autres utilisateurs, il faut s'abonner à leur compte. Un utilisateur qui choisit de suivre un autre devient un abonné ou « *abonné* ». Il est possible pour

⁷ <https://business.twitter.com/fr/basics/learn-twitter>

un utilisateur de restreindre la lecture de ses messages en gardant privé l'accès à son compte, en évitant donc de le rendre public. Les messages sont visibles par l'abonné après validation d'une requête d'ajout.

Les gazouillis postés peuvent concerner toute sorte de sujet, toutefois le contenu de *Twitter* peut être classé en différentes catégories. Selon une étude réalisée sur 2000 gazouillis par Pear Analytics en 2009⁸, les gazouillis peuvent être classés en six catégories :

Catégorie	Pourcentage
Pointless babble	40%
Conversational	38%
Pass along value	9%
Self-promotion	6%
Spam	4%
News	4%

Tableau 1. Catégories de gazouillis

2.2 L'environnement des soins de santé

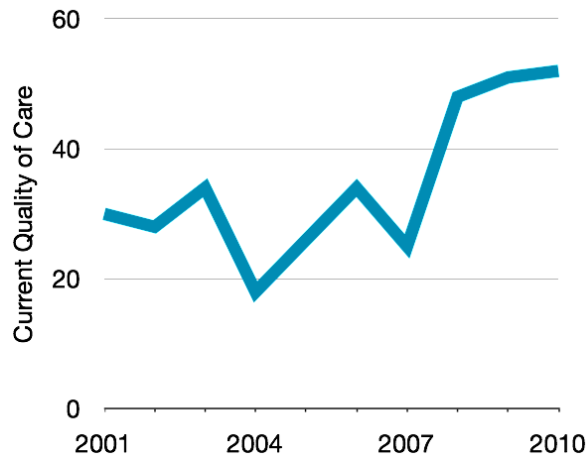
2.2.1 Perception du public sur le système de santé canadien

Les services de soins de santé font partie des préoccupations majeures des Canadiens. La perception de ces derniers sur les services en soin de santé a d'ailleurs fait l'objet d'un rapport [5] rédigé par le Dr Sokora, professeur à l'Université McGill, et présenté à la fondation canadienne de la recherche sur les services de santé. Voici quelques conclusions:

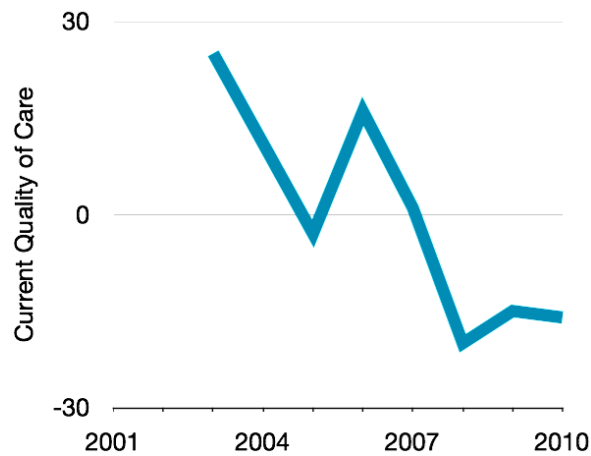
- Les attitudes des Canadiens à l'égard du système de santé semblent être assez partagées.
- Alors que les perceptions des niveaux de soins passés et présents sont relativement positives, les perceptions de l'avenir du système sont plutôt négatives.
- Au Québec, on tend à avoir une opinion moins favorable du système de santé qu'ailleurs au Canada.

⁸ <http://www.pearanalytics.com/blog/2009/twitter-study-reveals-interesting-results-40-percent-pointless-babble/>

- La vaste majorité des Canadiens déclarent avoir accès à un médecin de famille et jugent les services reçus « bons » ou « excellents ». En revanche, l'accès en temps opportun à des médecins et à des spécialistes obtient une mauvaise note. Les Canadiens estiment que l'accès aux services d'urgence est déplorable.
- La tendance générale de la couverture médiatique et de l'opinion publique au cours des dernières années montre que les enjeux liés aux temps d'attente et aux pénuries de médecins ont été quelque peu éclipsés par les épidémies (la grippe porcine, par exemple), l'activité physique et la nutrition.



Données provenant du Bulletin de santé d'Ipsos : « Quelle note attribuez-vous à la qualité globale des services de santé disponibles pour vous et votre famille? » [% des répondants donnant la note A ou B moins % des répondants donnant la note C ou F.]



Données provenant du Bulletin de santé d'Ipsos : « En général, croyez-vous que les services de santé de votre collectivité s'amélioreront grandement, s'amélioreront un peu, se détérioreront un peu ou se détérioreront grandement au cours des deux ou trois prochaines années? » [% des répondants donnant la note A ou B moins % des répondants donnant la note C ou F.]

Figure 2. Perception des canadiens sur la qualité des services de santé

Source: Perceptions du public et couverture médiatique du système de santé canadien © FCRSS

2.2.2 Opinion sur la qualité des soins de santé au Québec

En 2013, une enquête internationale est réalisée par le Commonwealth Fund montre qu'au Québec, seulement 23 % des répondants jugent que le système de santé est assez efficace. C'est l'un des résultats les plus défavorables parmi l'ensemble des pays. Au Canada, ce pourcentage est de 42 % et, dans certains pays, il dépasse les 50 % (Pays-Bas, Suisse et Royaume-Uni).

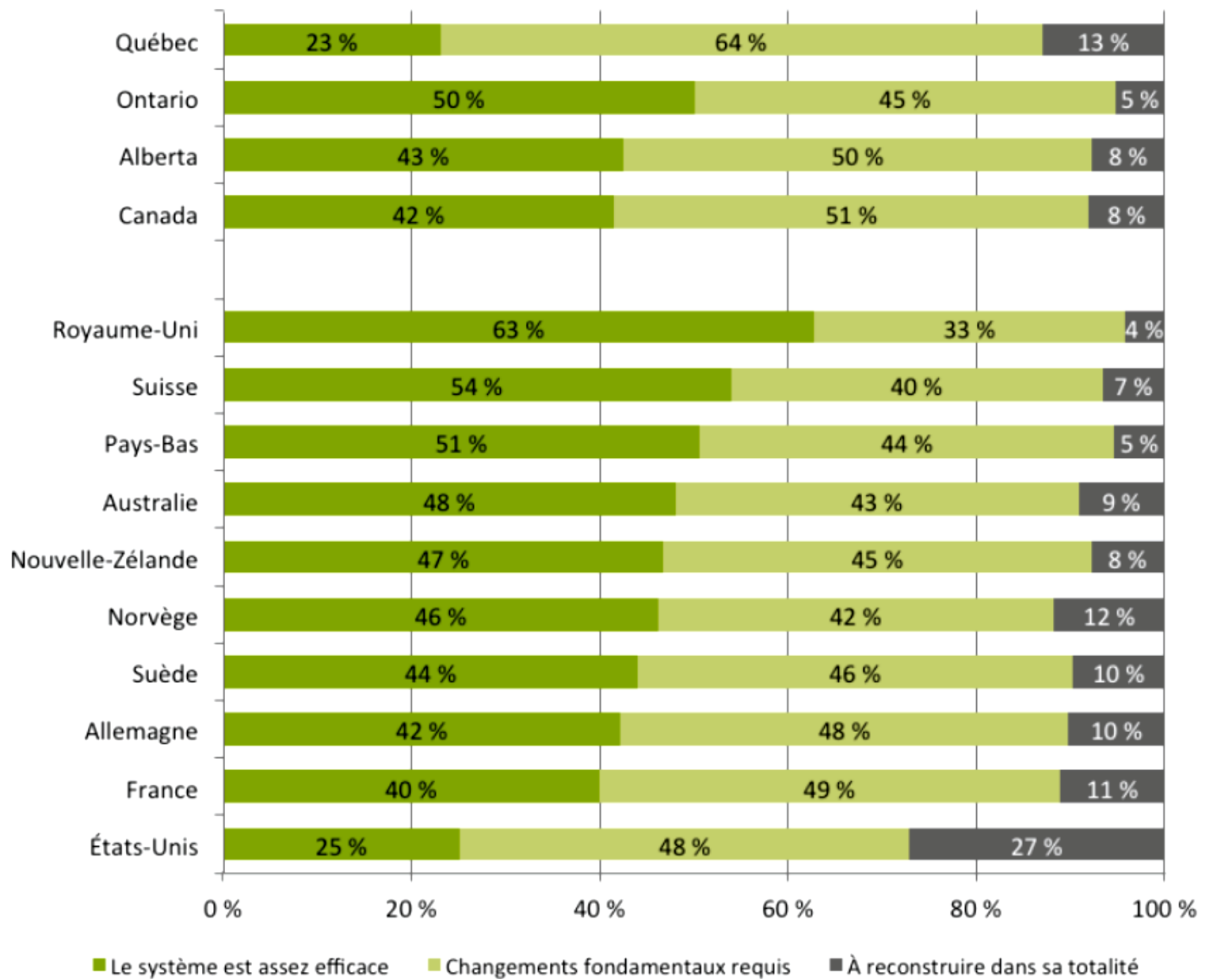


Figure 3. Opinion sur le système de santé au Québec et dans d'autres provinces et pays

Source: http://www.csbe.gouv.qc.ca/fileadmin/www/2013/CWF/CSBE_Rapport_Commonwealth_Fund_2013.pdf

2.2.3 Défis de l'industrie des soins de santé

Les organismes de soins de santé publique tels que les hôpitaux ont des défis qu'ils souhaitent relever. Ces défis sont relatés dans une étude d'Akinsi et al. [6].

2.2.3.1 *Augmentation des parts de marché*

Dans un monde où les règles du marché libre prévalent, les institutions œuvrant dans le monde de la santé publique sont de plus en plus soumises à une forte concurrence. Les hôpitaux et autres organismes en santé offrent des types de services similaires, ils doivent donc travailler plus fort pour rivaliser avec les concurrents pour gagner de nouveaux clients. Comprendre les facteurs qui influencent les choix des consommateurs est devenu primordial pour les organismes en soins médicaux, car cela leur permettra de personnaliser leurs soins, de maîtriser les coûts et d'améliorer l'expérience des patients. Il leur faut désormais développer des stratégies marketing de rétention et d'expansion.

2.2.3.2 *Améliorer la qualité des soins*

Les organismes en soins de santé ont besoin de moyens plus efficaces et plus rentables pour partager les meilleures pratiques permettant d'éliminer les erreurs médicales, de trouver des formes plus efficaces de traitement, ou d'améliorer la communication médicale. L'amélioration de la qualité des soins est à ce sens un des défis majeurs.

2.2.3.3 *Réduction des coûts*

Avec les médias sociaux, les organismes en soins de santé ont trouvé un moyen efficace de communiquer avec les patients. À moindre coût, ils peuvent encourager la prévention et promouvoir le bien-être.

2.2.4 Le patient, un acteur au cœur du système de santé

2.2.4.1 Déclaration d'engagement du CHUM envers les patients

La déclaration d'engagement⁹ du CHUM envers les usagers et leurs proches, est une preuve que le patient est au cœur du système de santé. Voici un extrait :

« Nous, du Centre hospitalier de l'Université de Montréal, proclamons notre volonté de tout mettre en œuvre pour vous satisfaire en améliorant de façon constante la qualité de nos interventions. Conformément à notre mission et à nos valeurs, nous nous engageons à :

- assurer l'accessibilité aux soins et aux services;
- offrir des soins et des services efficaces et pertinents;
- assurer la continuité des soins et des services dispensés;
- intervenir avec compétence, empathie et respect;
- pourvoir un environnement sécuritaire et confortable. »

2.2.4.2 Le patient a d'ores et déjà le choix

Les organismes du milieu de la santé publique se rendent de plus en plus compte du pouvoir que détiennent les patients. Ce pouvoir s'est considérablement accru. Les choix étant devenus de plus en plus réfléchis, les patients disposent de plus en plus d'information accessible sur Internet. Selon Javalgi et al (1991) « un aspect clé du succès dans le marketing de soins de santé est la capacité à élaborer des stratégies axées sur le marché en fonction du comportement de choix des consommateurs » [7]. Dans une autre étude, l'auteur révèle qu'il est important pour les gestionnaires des organismes de soins de santé de discerner quelle perception les patients potentiels ont la réputation de leur institution et de leurs produits. Afin de créer une perception positive dans l'esprit des patients qui deviennent des consommateurs, un agent de

⁹ http://www.chumontreal.qc.ca/sites/default/files//documents/A_propos/PDF/declarationengagement.pdf

commercialisation de soins de santé peut utiliser plusieurs éléments provenant des données des médias sociaux à des fins de marketing [8].

2.2.4.3 *La responsabilisation des patients*

Un concept apparaît de plus en plus dans la littérature. Il est souvent appelé «la responsabilisation du patient». Wolinsky (1988) [9], note que la distribution du pouvoir entre le médecin et le patient est devenue plus équilibrée. Quant à Akinci et al (2004) [6], ils indiquent que les patients sont devenus des consommateurs mieux informés et plus responsables quant aux soins de santé qui leur sont promulgués. Aussi, les patients sont devenus des participants plus actifs dans les décisions concernant les processus de traitement et dans le choix de leurs fournisseurs de soins de santé [6]. Les patients prennent le contrôle sur leur propre santé (gestion de la maladie en raison de leur meilleure connaissance acquise sur Internet). Ils aimeraient être considérés comme des partenaires à part entière dans les discussions et la communication avec les professionnels qui leur offrent des soins de santé.

2.2.4.4 *Les patients définissent les critères de choix des hôpitaux*

Plusieurs études ont été réalisées sur le sujet du choix des hôpitaux par les patients. Dans ces études, les auteurs reconnaissent que les patients magasinent entre les hôpitaux et font des comparaisons entre les fournisseurs de soins de santé. Dans l'étude menée par Javalgi et al (1991) [7], les critères prédominants étaient la proximité et le confort. Pour confirmer la robustesse des indicateurs trouvés, Akinci et al (2004) a réalisé une étude en Turquie afin de tester la robustesse des indicateurs trouvés antérieurement pour le choix de l'hôpital. Cette étude conclut que l'accessibilité des services hospitaliers est de loin le facteur le plus important dans le choix de l'hôpital [6]. Outre la proximité, l'image de l'hôpital et sa réputation sont d'une importance dans le choix de l'hôpital par les patients turcs. Plusieurs autres critères émanent de ces études, entre autres l'équipement disponible, la présence de médecins spécialistes et de la disponibilité de technologies modernes. Si Javalgi et al (1992) montrent que le critère des conseils fournis par les médecins généralistes a une moindre influence sur le choix des patients, cela est contredit par Leister et al, (2007), Berendsen et al, (2010), qui eux confirment

l'importance de l'avis du médecin dans le choix du patient pour un fournisseur de soins de santé [10], [11].

2.3 Les médias sociaux au sein de l'industrie des soins de santé

Les médias sociaux sont de plus en plus présents dans toutes les sphères de la société. Les nombreux utilisateurs des médias sociaux s'avèrent être aussi des patients ou des experts en soins de santé publique.

2.3.1 Utilisation des médias sociaux dans la santé à l'échelle mondiale

Les médias sociaux sont utilisés par les organismes de santé dans le monde entier comme étant un moyen efficace d'étendre l'expansion, de favoriser l'engagement et d'accroître l'accès aux informations d'ordre médical [12]. Une des valeurs intrinsèques des médias sociaux est que les utilisateurs peuvent s'engager activement et se connecter à l'information, aux organisations et aux autres utilisateurs, parmi eux des experts du monde de la santé. La valeur ajoutée est que cela crée une occasion de dialoguer et un partage de connaissances.

Dans le monde de la santé, les médias sociaux contribuent à améliorer l'accès à l'information de santé [13]. Malgré une utilisation répandue, il y a peu d'étude portant sur l'utilisation des médias sociaux médicaux ainsi que l'impact de l'analyse des données issues de ces médias sociaux médicaux.

Au niveau mondial, les médias sociaux sont souvent utilisés pour jouer le rôle d'informer les populations quant aux risques de catastrophes naturelles. Ainsi, une architecture de surveillance de maladies en temps réel a été développée par Sofean and Smith. Ils traquent les messages relatifs aux maladies postés sur les médias sociaux à l'aide de Twitter, avec une performance de 88 % [14]. Ainsi une détection plus rapide permet d'avoir plus de temps pour préparer les ressources et mieux répondre [15].

En 2011, un agent des communications de l'Organisation Mondiale de la Santé (OMS) a utilisé Twitter pendant le tremblement de terre qui a frappé le Japon et a été capable de déjouer les

idées fausses sur la santé [16]. Durant ce même évènement, des médecins ont utilisé Twitter pour faire face à l'inaccessibilité des autres méthodes de communication et pour informer leurs patients souffrant de maladies chroniques et qui souhaitaient des prescriptions [17].

En Turquie, une campagne a été organisée, visant à accroître l'utilisation du préservatif via une page Facebook. Ainsi, les publicités Facebook, les sites de partage de vidéos et les blogues ont permis aux ventes de préservatifs en ligne d'augmenter de 8 % [18].

Aux États-Unis, 60 % des départements de services de santé utilisent les médias sociaux, même si leur utilisation est encore à un stage précoce. Une des organisations faisant partie des leaders dans l'utilisation des médias sociaux médicaux est le Centers for Disease Control and Prevention (CDC). Les conseils provenant de sa plateforme ont été un guide pour plusieurs organisations et son succès a été un élément déclencheur pour l'utilisation accrue des médias sociaux dans le domaine de la santé en Ontario [12] [14]. Ainsi, en 2009, la CDC a mené une vaste campagne de communication sur la grippe H1N1 grâce à l'utilisation de Twitter, MySpace et YouTube. En adaptant les messages à chaque plateforme, le CDC a pu s'engager dans un dialogue avec le public permettant de renforcer la confiance. Il y avait une augmentation des notes de satisfaction du public à l'époque de cette campagne.

2.3.1.1 *Point de contact entre les patients et les médecins*

En dehors du contact téléphonique ou visuel, Bennett (2012) [20] explique que les médias sociaux sont un point de contact supplémentaire pour les clients. Dans une autre étude, Eckler et al. (2010) [21] confirment ce point. Cette étude dit que les médias sociaux peuvent faciliter, renforcer et même améliorer la communication médecin-patient. En partageant leurs histoires, ou en ayant les mêmes diagnostics, les patients ne sont plus isolés. Ils sont en mesure de se trouver en ligne. Les médias sociaux deviennent un outil pour suivre la progression de leurs maladies et de partager des expériences avec d'autres patients dans un état médical similaire (Keckley & Hoffman, 2010) [38].

2.3.1.2 Outil d'éducation des patients

Selon Bennett (2012) [20], les médias sociaux constituent un prolongement naturel des efforts pour rejoindre et enseigner. Fournir de l'information sur des sujets de santé peut informer et même éduquer les patients. Pour Eckler et al. (2010) [21] l'éducation à travers les médias sociaux est un prolongement de la visite d'un patient, ce qui permet d'améliorer la circulation de l'information et de la cohérence des soins, en particulier pour les maladies qui nécessitent des soins à long terme. De plus, les patients peuvent obtenir plus aisément une procédure chirurgicale ou médicale à l'avance lorsqu'ils reçoivent de l'information via les médias sociaux (Bennett, 2009) [20]. Eckler (2010) [21] indique que par le biais de l'éducation en ligne, les patients deviennent plus actifs dans l'autogestion ou la prévention des maladies. Cela est communément appelé la « gestion de la maladie » ou encore le « maintien de la santé et du bien-être ».

2.3.1.3 Communication en période de crise

Le terme crise peut référer à un désastre naturel ou une quelconque situation externe à l'hôpital. Ainsi, les institutions de soins de santé peuvent utiliser les médias sociaux pour informer la communauté lors des situations de crise comme une maladie pandémique et contagieuse. Selon Keckley (2011) [19], le Centre de Deloitte pour les solutions de santé souligne qu'un réseau social comme Twitter peut également être utilisé pour diffuser des instructions et pour agir, par exemple en cas de la grippe H1N1.

La situation de crise peut par ailleurs être interne. Il peut s'agir du cas d'un hôpital où il s'est produit un incident fatal ou un autre scandale pouvant potentiellement nuire à la réputation ou l'image de l'institution de santé. Si les médias sociaux sont hors du contrôle de l'institution, ils peuvent être un carburant pour propager une telle crise avec de fausses rumeurs, des informations erronées. Il est crucial pour un établissement de soins de santé de savoir maîtriser cet outil de communication à titre préventif pour éviter une atteinte à la réputation. Selon (Bennett, 2009) [20], une réponse rapide via les médias sociaux en ligne tels que Twitter permet

à un hôpital de prendre le contrôle du message et de tenir la communauté à jour avec des informations en temps réel.

2.3.2 Utilisation des médias sociaux dans le système de santé au Canada

L'agence Ipsos Reid a réalisé une étude et publié un rapport en 2012. Ce rapport montrait que 95 % des Canadiens âgés de moins de 55 ans et 68 % des Canadiens de plus de 55 ans ont accès à l'Internet [22]. Parmi ceux qui ont accès, 80 % l'ont à la maison et 62 % sont inscrits dans les réseaux sociaux et 30 % les visitent au moins une fois par jour. De plus, 64 % de ceux qui sont souvent connectés ont utilisé Internet et les médias sociaux pour rechercher des informations médicales ou liées à la santé [23].

2.3.2.1 Utilisation accrue de technologies mobiles au Canada

Le développement des téléphones intelligents – aussi connus sous le nom de *Smartphones* – s'est encore accrue et offre de plus en plus la possibilité d'utiliser les médias sociaux [24]. En 2012, 62 % des Canadiens étaient propriétaires d'un téléphone intelligent et parmi ceux-ci, 60,7 % avait accès à des médias sociaux de toutes formes au moins une fois par mois sur leurs téléphones intelligents [24]. Les changements qui en résultent pour les relations entre les personnes et le niveau de connaissances sont importants. Quarante et un pourcent des Canadiens qui utilisent les réseaux sociaux avouent communiquer plus avec les gens en ligne que ceux hors ligne [24].

2.3.2.2 Quand est-il du de l'utilisation des médias sociaux dans le domaine de la santé au Canada?

Il y a eu un changement fondamental dans ce que les citoyens attendent des deux services de livraison de l'information et des rencontres avec des professionnels de la santé (26,31). Hesse et al (2009) font valoir que, tout comme les consommateurs s'attendent maintenant à vingt-quatre heures d'accès à l'information grâce à des applications en ligne, ces applications sont utilisées pour rechercher des informations de partager leurs expériences en matière de santé (y compris

les notes et les commentaires sur les rencontres cliniques et professionnels de la santé individuels). La lecture de ces messages en ligne peut influencer les attitudes des patients vers les questions de santé et peut affecter la décision de se conformer aux traitements prescrits ou des conseils médicaux (6, 13, 20, 22). En outre, on observe une tendance dans les comportements des patients. Ces derniers arrivent souvent à leurs rendez-vous médicaux en ayant pris le soin de rechercher l'information en santé et de poster des questions à d'autres réseaux sociaux composées d'autres patients ou d'experts. Puis, une fois, le rendez-vous terminé, ils accourent sur les réseaux sociaux pour s'informer, commenter et partager leurs opinions.

Il existe quelques examens complets sur l'utilisation des médias sociaux aux bureaux de santé publique qui ont été réalisés, notamment en Ontario. En Juillet 2011, la plupart (33 sur 36) des unités de santé publique en Ontario utilisaient les médias sociaux. *Twitter* et *Facebook* étant des plateformes les plus souvent utilisées [19].

En 2010, l'étude de Schein et al. a compilé un rapport détaillé sur l'utilisation des médias sociaux par le Bureau de santé publique de l'Ontario. Les auteurs ont conclu qu'il y a une utilité potentielle pour les médias sociaux dans la santé publique (ainsi que les pièges potentiels), en particulier dans l'amélioration de la portée et la promotion de messages. L'engagement des utilisateurs à créer du contenu et fournir des commentaires améliore la confiance du public, la loyauté et la confiance dans l'information. En plus de recommandations de gestion, ils recommandent la coordination de matériaux et de collaboration aux niveaux municipal, provincial et fédéral [18].

2.3.3 Mesures de médias sociaux dans l'environnement de santé

Les organismes en soins de santé s'aventurent de plus en plus sur les médias sociaux souhaitant utiliser ces nouveaux médias pour mieux communiquer avec la population. Ils sont donc à la recherche de mesures efficaces pour justifier leurs investissements (sorte de retour sur investissement ou ROI). Schmidt (2001) [36] en définit quelques-unes :

- Les leaders d'opinion sur les réseaux sociaux des organismes : cette mesure permet aux organismes d'identifier quels sont les utilisateurs les plus actifs ou ayant le plus d'influence. Ils peuvent être utilisés pour relayer un message en temps de crise.
- Le taux d'engagement : cette mesure est très importante, car plusieurs organismes se limitent à compter le nombre de « j'aime » (*likes*) sur leur fils d'actualité. Une telle mesure permet de savoir jusqu'à quel point vos sujets suscitent l'engouement au sein de la communauté.
- Les sujets viraux : pour faire une campagne efficace, il est important pour un organisme de savoir quel genre de contenu captive le plus d'attention. Cela peut s'agir d'un mot dièse (*hashtag*), d'une vidéo, etc.

2.4 L'analyse des médias sociaux

La popularité des médias sociaux devient de plus en plus grandissante. En effet, les réseaux sociaux tels que *Facebook*, *LinkedIn*, *Instagram* ou *Twitter* comptent chacun des millions d'utilisateurs et en enregistrent toujours plus. *Twitter* par exemple a plus 200 millions d'utilisateurs actifs, 500 millions de gazouillis par jour et plus de deux millions de recherches.

En tant que décideur d'une organisation en soin de santé, les données issues des médias sociaux *Twitter* offrent une mine d'informations à exploiter. Ainsi, posséder uniquement un compte sur un média social n'est plus suffisant. Il faut aussi savoir quelle est la bonne façon d'influencer efficacement l'audience. Les questions qui se posent sont : comment gérer son compte pour avoir la meilleure audience, comment exploiter les données produites par les conversations? Pour répondre à ces questions, on utilise l'analyse des médias sociaux.

L'analyse des médias sociaux peut être défini la pratique de la collecte des données issues des plateformes de réseaux sociaux et l'analyse de ces données pour générer des informations utiles pour prendre des décisions stratégiques.

Pour un organisme en soins de santé, il faut connaître quelles sont les personnes qui s'intéressent à l'actualité qui vous concernent. Comment, où et quand interagissent-ils, quels sont les

contenus les plus attrayants et quelle est l'opinion de ces utilisateurs à propos de votre institution.

Lorsqu'on désire effectuer l'analyse des médias sociaux, il y a différents types ou catégories d'analyse possible. Il est par exemple possible de faire une analyse d'influence ou encore du volume. Toutefois, l'utilisation la plus commune de l'analyse des médias sociaux est l'analyse des sentiments pour connaître quelle est l'opinion générale des utilisateurs sur votre image et la qualité de vos services.

2.4.1.1 L'analyse des sentiments

2.4.1.2 Définition du mot sentiment

Qu'est-ce qu'un sentiment? C'est une question à laquelle il faut répondre avant de savoir qu'est-ce que l'analyse des sentiments. Le mot sentiment signifie une opinion, un avis que l'on porte sur quelque chose¹⁰ [25]

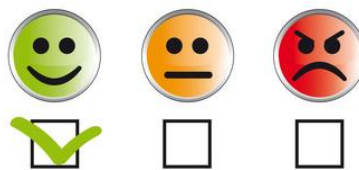


Figure 4. Définition du sentiment

Source: <http://www.clarabridge.com/sentiment-analysis-and-business-sense/>

2.4.1.3 Définition de l'analyse des sentiments

Plusieurs termes existent dans la littérature pour désigner l'analyse des sentiments, on peut citer les termes analyse d'opinions, exploration d'opinions (*opinion mining*), ou analyse de sentiment (*sentiment analysis*). En réalité, il y a très peu de différence entre les dénominations et toutes convergent vers l'exploration et l'analyse des sentiments exprimés dans des textes subjectifs.

¹⁰ <http://www.larousse.fr/dictionnaires/francais/sentiment/72138>

Pour Pang et Lee (2008) [26], les termes d'analyse d'opinion et analyse de sentiments sont interchangeables et expriment un sens commun. Ils s'utilisent par conséquent très fréquemment. Par définition, l'analyse des sentiments est le traitement informatique d'un texte qu'on explore afin d'identifier le sentiment ou l'opinion générale exprimée. On parle souvent de classifier la polarité (positive, négative ou neutre) d'un texte. L'exploration se fait concernant une entité et cette entité peut représenter des individus, des produits, des évènements ou encore des sujets variés qui sont susceptibles d'être couverts par des commentaires, critiques ou appréciations.

Le processus d'analyse des sentiments se fait en plusieurs étapes selon Pang et Lee (2008) [26]. En effet, un système informatique autonome ayant pour fonction l'analyse des sentiments d'un texte doit être capable d'extraire les textes, de distinguer les textes neutres, puis il doit être capable de détecter les paragraphes ayant des jugements évaluatifs. Enfin, il importe que ce système soit capable de présenter globalement le sentiment exprimé dans l'ensemble du texte.

Pour identifier la polarité d'un texte, l'analyse de sentiment se base sur plusieurs caractéristiques prises en compte, entre autres : la fréquence d'un terme, l'importance d'un terme dans le texte, les négations dans les phrases.

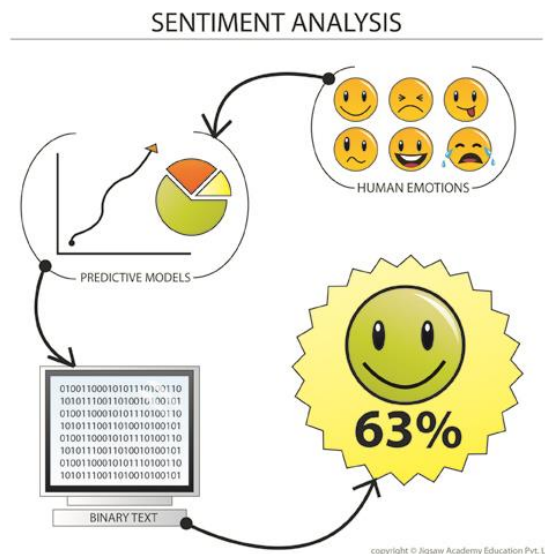


Figure 5.L'analyse des sentiments

Source: <http://analyticstraining.com/2011/sentiment-analysis/>

2.4.1.4 *Intérêt pour l'analyse des sentiments issus des médias sociaux*

L'analyse des sentiments issus des réseaux sociaux est devenue un sujet de recherche en vogue. Selon Pang et Lee (2008) [26], les tous premiers systèmes d'analyse automatique des sentiments ont surtout été appliqués aux textes traditionnels tels que les critiques en ligne et les articles de la presse. Avec la masse d'information générée par les utilisateurs sur les réseaux sociaux, on observe de plus en plus d'engouement sur l'analyse des sentiments qui y sont exprimés.

Il existe plusieurs entreprises et organismes qui souhaitent connaître les opinions échangées à leur égard sur les médias sociaux. L'intérêt est en premier lieu commercial. Dans l'étude de Pang et Lee (2008) [26], on note que « ce que pensent les gens » ou l'opinion des autres a toujours été un élément d'information important pour la plupart des personnes ou entreprises, au moment de la prise de décision. Bien avant l'utilisation des réseaux sociaux, les gens manifestaient leur désir de connaître les sentiments des autres, de leurs familles, de leurs amis, avant de se décider. Cela se faisait à travers de recommandations via le bouche-à-oreille. Avec l'arrivée du Web 2.0, la possibilité de connaître les opinions des gens est beaucoup plus accessible à travers les réseaux sociaux. Ces gens ne font pas partie de l'entourage proche, ni même des critiques dont le but masqué est de promouvoir l'image d'un organisme ou de son produit. Pang et Lee (2008) [26] affirment donc que l'opinion issue des personnes sur les réseaux sociaux est plus objective et valable.

En dehors de l'intérêt des personnes connectées à connaître les sentiments des autres pour se décider, il y a aussi plusieurs autres organismes et entreprises qui désirent connaître les sentiments et opinions les concernant. Un exemple d'intérêt commercial que l'on trouve dans l'étude de Pang et Lee (2008) [26] est celui d'un commerçant désireux de connaître pourquoi son nouveau produit ne se vend pas bien. Il s'est donc intéressé aux opinions de ses consommateurs vis-à-vis de son produit. Le résultat fut précieux pour lui car en connaissant ces opinions il a pu adapter son offre aux impératifs du marché.

Dans un autre registre, on voit l'intérêt que porte le monde politique sur les opinions des populations sur les médias sociaux. Pang et Lee (2008) [26] notent qu'au lieu d'utiliser les sondages ou publicités qui coutent très chers, certaines entreprises ou organismes ont là un

moyen peu coûteux de savoir l'intérêt ou le sentiment que leur porte l'ensemble de la communauté qui est de plus en plus connectée sur les médias sociaux.

Ainsi il y a un intérêt pour l'analyse des sentiments exprimés dans les réseaux sociaux. C'est un des sujets les plus intéressants de nos jours. Les entreprises l'utilisent en effet pour promouvoir leurs produits et surveiller leurs images. Il est devenu nécessaire de savoir construire des outils de surveillance des médias sociaux, de concevoir des systèmes informatiques capables d'extraire et traiter automatiquement les informations publiées sur des organismes afin d'identifier le sentiment général qui se dégage de ces commentaires.

Un tel système nécessite au moins deux modules : l'un doit être responsable de capter des informations issues des médias sociaux, selon des critères de recherches comme des mots clés, l'autre module quant à lui doit réaliser l'analyse des sentiments, c'est-à-dire être capable d'identifier la polarité de chaque texte et d'évaluer le sentiment général pour déduire ce que les gens pensent.

Dans la littérature, R. Van't Ende (2013) [35] utilise cette formule pour calculer le sentiment général :

$$\text{Sentiment} = (\text{Nb gazouillis positifs} - \text{Nb gazouillis négatifs} + \text{Nb gazouillis neutres}) / (\text{Nb total de gazouillis})$$

2.4.1.5 Exemple d'analyse de sentiments spécifique à Twitter

Dans leur étude, Tumasjan et al. (2010) [30] effectuent une analyse des gazouillis mentionnant les noms des six partis politiques qui sont représentés au Parlement allemand. Cette étude conclut que *Twitter* est utilisé comme un forum de discussion politique en ligne : plus d'un gazouillis sur trois faisait partie d'une conversation et près de 20 % de tous les messages étaient des Retweet, à savoir des gazouillis qui répondent à un autre gazouillis. Ces résultats ont démontré que *Twitter* ne s'utilise pas que pour exprimer ses opinions, mais également pour discuter avec d'autres internautes. L'étude a également révélé que 0,6 % des utilisateurs dans le corpus avaient publié 21,2 % de tous les gazouillis tandis que 50,3 % des utilisateurs avaient posté 10,2 % de tous les gazouillis. Ces chiffres ont indiqué que Twitter peut être considéré

comme une très bonne source d'information parce que les gens partagent et discutent de leurs opinions sur un sujet librement. Il peut être également utilisé en tant que sources de données pour une analyse d'opinions visant à savoir le sentiment général.

2.4.2 Approches d'analyse de sentiments

Les approches automatiques d'analyse de sentiments peuvent être divisées en deux catégories. Dans la revue de littérature, on retrouve en général deux approches d'analyse de sentiments :

- l'approche basée sur un lexique (*Lexicon based approach*) utilise un lexique prédéfini;
- l'approche basée une machine d'apprentissage (*Machine Learning based approach*) utilise la classification subjective des textes à partir d'un large ensemble de données.

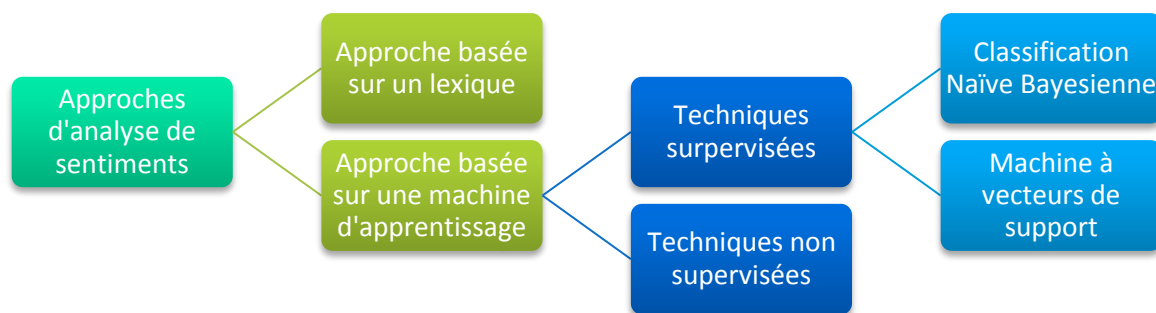


Figure 6. Approches et techniques d'analyse de sentiments

2.4.2.1 L'approche d'analyse des sentiments basée sur un lexique

L'approche d'analyse des sentiments basée sur un lexique est celle qu'on retrouve le plus dans la revue de littérature. Elle aide à identifier la polarité d'un texte en se servant de deux catégories de mots prédéfinis et pondérés, appelé lexique ou dictionnaire. Elle identifie tous les mots positifs ou négatifs au sein d'un texte. Le dictionnaire est constitué d'un petit ensemble de mots d'opinion subdivisés en deux catégories. L'une des catégories comporte des mots dont la terminologie est plus positive, tandis que l'autre catégorie regroupe les mots associés à un sentiment plus négatif. Pour établir un dictionnaire, ils commencent généralement par les mots les plus intuitifs et qui expriment un sentiment positif, le nombre de ces mots est amplifié par leurs synonymes constitue la catégorie positive. Une catégorie de mots négatifs peut être

automatiquement constituée à partir des antonymes de mots de la catégorie positive, puis elle est amplifiée par d'autres mots jugés négatifs. À chaque mot est associée une pondération positive ou négative. La somme des mots est soit positive ou négative et représente une évaluation globale du sentiment dans le texte.

L'algorithme utilisé dans l'approche basée sur un lexique peut être exprimé comme suit :

1. Initialiser la note totale du sentiment $s \leftarrow 0$
2. Pour chaque jeton du texte, vérifier la présence dans le lexique :
 - a. Si le jeton est présent :
 - i. Si le jeton est positif, alors : $s \leftarrow s + w$
 - ii. Si le jeton est négatif, alors : $s \leftarrow s - w$
3. Regarder le sentiment total du texte
 - a. Si $s >$ au seuil, alors le sentiment du texte est positif
 - b. Si $s <$ au seuil, alors le sentiment du texte est négatif

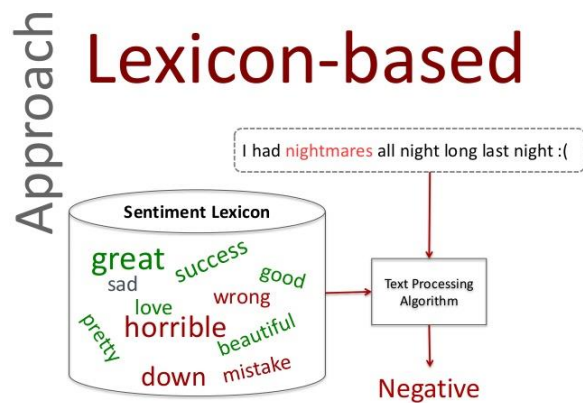


Figure 7. Approche basée sur un lexique

Source: <http://www.slideshare.net/Staano/senticircles-for-contextual-and-conceptual-semantic-sentiment-analysis-of-twitter>

Voici des exemples trouvés dans la littérature qui utilisent l'approche basée sur un lexique :

- O'Connor et al. (2010) ont utilisé le lexique de sentiment MPQA de Weibe et al. (2005) afin d'identifier le sentiment général dans les gazouillis mentionnant Barack Obama. Afin de classer chaque gazouillis, ils ont simplement compté le nombre de mots positifs et négatifs selon le lexique de sentiment. Les résultats de leur étude montraient une corrélation entre le sentiment général dans les gazouillis et l'opinion publique exprimé dans le sondage Gallup.
- Thelwall et al. (2010) ont développé un algorithme appelé *SentiStrength*, basé sur un lexique, et qui assignait à un texte donné une polarité (positive/négative) correspondant à une valeur comprise entre 1 à 5. Ils ont trouvé une liste de 298 termes positifs et 465 négatifs.

2.4.2.2 Difficultés de l'approche basée sur un lexique

Dans leur étude, Pang et Lee (2008) [26] font état des difficultés liées à l'approche basée sur un lexique. Selon eux, il y a des difficultés qui peuvent surgir lors de l'analyse et qui entravent dans une certaine mesure la fiabilité des résultats.

- La taille du texte : Des micros blogues comme Twitter ont des messages dont la taille est limitée. Si dans certains cas cela peut s'avérer être un avantage car les auteurs vont droit au but pour exprimer leurs opinions, cela pose le problème que l'opinion exprimée peut être dépendant d'un seul mot et ce mot peut ne pas être présent dans le lexique ou le dictionnaire de données et peut causer une perte d'opinion. Pang et Lee (2008) citent le fait qu'une phrase qui exprime un sentiment, mais qui pourtant ne contient pas de mots explicitement subjectifs.
- La variation des abréviations : À cause de la spontanéité, du contexte et de la limite des caractères, les messages postés peuvent contenir plusieurs types d'abréviations, des émoticônes ou même de nouveaux mots incompréhensibles mais bien utilisés dans le vocabulaire en vogue sur les réseaux sociaux, par exemple : *lol*, *gr8 pour great*, *mr6* pour merci.

- Le contexte : Un autre exemple est souvent lié au contexte, en effet Pang et Lee (2008) notent qu'une phrase peut être interprétée positivement ou négativement selon le contexte dans lequel elle est prononcée. Ce qui est difficile à déterminer.
- La structure d'un texte : Pang et Lee (2008) signalent une difficulté de cette approche, liée à la structure d'un texte qui malgré le fait qu'il comporte plusieurs mots positifs, est en réalité négatif à cause de la dernière phrase qui comporte un mot négatif particulier.

En prenant en compte les difficultés qu'ils ont pu observer dans leur étude, Pang et Lee (2008) stipulent que l'approche basée sur un lexique n'est pas viable pour déterminer le sentiment général d'un texte, si elle se fonde uniquement sur la somme des mots à connotations positive ou négative. Cette approche, ni la subjectivité des mots, ni le contexte, encore moins la position des mots d'une phrase, ne sont pris en considération.

Les difficultés de cette approche s'intensifient lorsqu'on applique cette méthode basée sur un lexique, aux textes plus courts comme ceux issues du réseau social *Twitter*. Cette plateforme ne permet en effet que des commentaires de 140 caractères maximum. Des chercheurs comme Moreno-Ortiz et Pérez Hernández (2012) [27] ont appliqué la méthode basée sur un lexique sur des textes provenant de *Twitter*. Leur étude avait pour objectif de promouvoir l'utilisation des systèmes automatiques pour l'analyse de sentiments de textes courts. Les investigateurs avaient été demandés d'analyser automatiquement un corpus de 60 000 gazouillis et de déterminer pour tout gazouillis la polarité et l'intensité du sentiment exprimé. Ils ont utilisé le logiciel *Sentitext*¹¹, basé sur un lexique espagnol semi automatiquement collecté et manuellement révisé. À partir des mots négatifs et positifs dans un texte, le logiciel sait calculer la valeur sentimentale globale du texte sur une échelle de zéro à dix. Cette étude a révélé une importance quant à l'instrument de mesure, de même que le contexte dans lequel le sentiment peut être exprimé lors d'une analyse de sentiment.

En somme, les études de Pang et Lee (2008) [26] ainsi que celles de Moreno-Ortiz & Pérez-Hernández (2012) [27] font état de de difficultés qui peuvent restreindre la fiabilité du résultat

¹¹ *Sentitext* est un logiciel d'analyse autonome des sentiments basé sur un lexique (<http://www.sentitext.com/>)

d'un système bâti sur l'approche basée sur un lexique. Il n'en demeure pas moins que c'est une technique populaire pouvant être utilisée pour déterminer l'opinion du contenu d'un texte lorsque le contexte et l'instrument de mesure sont maîtrisés.

2.4.2.3 L'approche basée sur une machine d'apprentissage

L'approche basée sur un lexique présente un inconvénient majeur qui est l'incapacité de trouver des mots d'opinion avec des orientations spécifiques au domaine et au contexte.

L'approche d'analyse de sentiments basée sur une machine d'apprentissage (*Machine Learning based approach*) utilise des algorithmes statistiques et des fonctionnalités linguistiques de classifications. À partir d'un ensemble (ou modèle) de données appelé « *data set* », les algorithmes sont entraînés et peuvent, à partir de ces modèles, faire des prédictions sur d'autres données qu'ils peuvent rencontrer à l'avenir. Ces algorithmes varient dans leur capacité à performer sur de grandes séries de modèles, et un modèle qui ne ressemble à aucun de ceux précédemment connus par l'algorithme peut être susceptible d'être mal interprété. L'approche basée sur une machine d'apprentissage est subdivisée en méthodes d'apprentissages supervisées et non-supervisées.

a) Apprentissage supervisé

Les méthodes dites supervisées sont celles qui utilisent de grands ensembles de données labellisées; elles nécessitent la formation de deux ensembles de données: un ensemble d'apprentissage et un autre de test. L'ensemble de données d'apprentissage (*ou training set*) est utilisé par un classificateur automatique pour apprendre les caractéristiques de différenciation des textes afin d'entraîner le système. L'ensemble de données de test est quant à lui utilisé pour vérifier la façon dont ce classificateur (ou l'algorithme de classification) performe.

Il existe plusieurs algorithmes de classification dans les techniques dites supervisées : la classification naïve bayésienne (*NB: Naïve Bayes classification*), l'entropie maximale (*Maximum Entropy*), les arbres de décision ou encore le (*Support Vector Machines, SVM*).

Dans la revue de littérature, il y a plusieurs comparaisons de différents algorithmes de classification de texte.

Go et al. (2009) ont été les premiers à faire des travaux sur l'analyse des sentiments avec les données issues de *Twitter*. À cause de l'insuffisance d'un lexique en bonne et due forme, ils ont employé une approche permettant de d'entraîner le classificateur de leur machine d'apprentissage. Ils ont commencé par télécharger un grand nombre de gazouillis via l'API *Twitter* et ont utilisé les émoticônes présents dans les gazouillis comme indicateurs labellisés. Les gazouillis qui contenaient les émoticônes positifs étaient positifs et ceux qui contenaient des négatifs étaient négatifs. Ils ont par contre éliminés les gazouillis contenant des émoticônes à la fois positifs et négatifs. À la fin, ils eurent un large ensemble de données (data set) de 1600000 gazouillis avec 800000 de chaque classe (positive ou négative). Les algorithmes de classification utilisés furent le classificateur naïf de Bayes (*Naïve Bayes Classifier*), *Maximum Entropy* et *Support Vector Machine* (SVM). Ce dernier avait leur préférence car il obtenait un taux de fiabilité de 82.9 %.

Classification naïve bayésienne (Naïve Bayes classifier)

L'algorithme naïf de Bayes est une technique connue et utilisée pour des tâches de classification. Cette technique est appelée naïve car elle suppose que les probabilités combinées sont indépendantes les unes des autres. Selon Segaran (2007) [33], les classificateurs bayésiens sont souvent utilisés pour la classification des textes car ils nécessitent beaucoup moins de puissance de calcul que d'autres méthodes. Jurafsky et Martin (2009) [34] confirment cette affirmation. Pour eux, le classificateur naïf de Bayes est souvent utilisé comme une bonne méthode de référence avec des résultats qui sont suffisamment bonnes pour une utilisation pratique.

Pang et Lee (2008) font une comparaison de la performance de trois des classificateurs les plus connus à savoir : le classificateur naïf de Bayes (*Naïve Bayes classifier*), le classificateur d'entropie maximum (*Maximum Entropy classifier*) et le classificateur SVM (*Support Vector Machines*). Le résultat dévoilait que lorsque l'ensemble de données (d'apprentissage et de test) est petit, l'algorithme de classification naïf de Bayes est le plus performant. Tandis que SVM l'est pour des ensembles plus larges.

b) Apprentissage non-supervisé

Les méthodes supervisées ne peuvent souvent pas être utilisées car les grands ensembles de données ne sont pas toujours disponibles. Dans ce cas, les techniques dites non-supervisées sont une option pour une analyse de sentiments qui ne nécessite pas des grands ensembles de données labellisées (pré-marquées).

Les machines d'apprentissages utilisant des techniques non-supervisées reçoivent seulement, en entrées, un ensemble de données non labellisées. Ces méthodes non supervisées peuvent aussi être utilisées pour étiqueter un ensemble de données qui peut ensuite servir d'ensemble d'apprentissage pour les techniques supervisées.

Connaître le sentiment ou l'opinion générale des utilisateurs à propos de votre marque ou de la qualité de vos services est très bénéfiques. Toutefois, cela peut être insuffisant pour justifier le retour sur investissement consenti sur une analyse des médias sociaux.

2.4.3 Analyse du volume

Une bonne analyse du volume permet d'en connaître les caractéristiques de l'audience et du contenu publié dans les messages ayant un rapport avec votre compte. Plusieurs mesures peuvent être considérées lorsqu'on mesure le volume :

Mesure	Description
Utilisateurs les plus actifs	Permet de savoir les utilisateurs qui participent le plus aux conversations sur votre marque.
Utilisateurs les plus influents	Permet de savoir les utilisateurs ayant les plus grand nombre d'abonnés
Nombre mentions	Permet de savoir combien de fois votre marque a été mentionnée par un utilisateur durant une période spécifique. En ayant cette mesure, nous pouvons avoir une bonne idée de l'intérêt que portent les (ou chaque) utilisateur(s).
Volume au cours du temps	Mesure permettant de voir l'évolution et l'impact de votre marque à travers le temps.
Mots principaux	Permet de comprendre le contenu ayant le plus d'impact.
Hashtags principaux	Permet de connaître les termes de recherche ayant le plus d'impact.
Langue	Permet de savoir la langue la plus utilisée

Tableau 2. Mesures d'analyse d'influence

2.4.4 Analyse de la portée potentielle

Elle permet non seulement de connaître le nombre de personnes qui vous suivent, mais également l'audience potentielle que l'on peut atteindre si l'on veut encore connaître une plus grande expansion. Voici la formule pour mesurer la portée potentielle :

Portée potentielle = # d'utilisateurs mentionnant votre nom + # de leur *abonnés*.

2.4.5 Analyse du niveau d'engagement

L'engagement est une bonne manière de justifier le retour sur investissement ou le succès d'une stratégie d'expansion sur les réseaux sociaux. Cette mesure permet de savoir comment les utilisateurs, qui suivent votre compte ou conversent sur vous, interagissent et jusqu'à quel point ils sont engagés. Les utilisateurs qui s'engagent présentent plusieurs opportunités : ils peuvent agir comme avocat défenseur des causes de votre institution. Ils peuvent fournir des feedbacks sur la qualité des services et peuvent vous permettre de mieux comprendre d'autres utilisateurs. La formule suivante permet d'évaluer le niveau d'engagement d'un compte :

Engagement = @Replies + Retweet + Mentions / (Nb gazouillis + @Replies + Retweet + Mentions)

2.5 Conclusion de la revue de littérature

L'analyse des sentiments est un domaine qui montre un intérêt grandissant dans nos sociétés car plusieurs organismes dans différents domaines portent un intérêt croissant à savoir quelles sont les opinions des gens qui s'expriment sur les réseaux sociaux tels que *Twitter*. Les études démontrent que *Twitter*, comme bon nombre de médias sociaux, peut être un indicateur fiable pour identifier clairement l'opinion générale des utilisateurs de ce média social. Sur base des difficultés de la méthode basée sur un lexique, reportée dans la revue de littérature, il convient d'utiliser une technique d'apprentissage supervisée basée sur une stratégie de classification

naïve bayésienne parce que c'est une méthode simple et intuitive dont la performance est similaire à d'autres approches. Elle allie efficacité (performance optimale de temps) avec une précision raisonnable. En plus, elle requiert relativement peu de données d'entraînement pour estimer les paramètres nécessaires à la classification.

En se basant sur cette revue de littérature, nous pouvons procéder à l'expérimentation en appliquant une méthodologie de recherche. La section qui suit présentera l'approche méthodologique préconisée pour mener à bien cette étude.

Chapitre 3

Méthodologie

3.1 Type de recherche

L'étude préconisée dans cet essai est descriptive et corrélacionnelle. En effet, cette étude consiste à déterminer les caractéristiques des données et la nature des opinions dans le but d'établir les associations entre elles. Cette description constitue l'objectif de la recherche. Elle doit faire ressortir les aspects tels que le niveau d'engagement des utilisateurs et leur appréciation de l'image d'un (ou des) organisme(s) en soins de santé au Québec.

Nous procédons par une approche quantitative qui nous permettra de recueillir des données observables et quantifiables. Ces données mesurées seront analysées selon différentes méthodes de la revue de littérature.

Cette recherche s'appuie aussi sur des instruments et technique de recherche. Elle aboutit à des données quantifiables et observables qui permettront de faire des analyses descriptives et des présentations sous forme de tableaux et de graphiques.

3.2 Milieu, population et échantillon

Le principal milieu de mesure de cette étude est le réseau social *Twitter*, plus précisément l'ensemble des messages postés ayant un lien avec les mots clés préalablement définis.

La population cible est l'ensemble des utilisateurs ayant un compte *Twitter*. La méthode d'échantillonnage est probabiliste et le principal critère de sélection des utilisateurs et des gazouillis retenus pour l'étude est: les messages publiés doivent avoir un lien avec le sujet ou les termes ayant un rapport avec le compte d'un hôpital ou un organisme en soins de santé du Québec. Internet étant un vaste réseau, sans frontières, la population ne sera pas limitée à une

région ou une ville donnée. De même, le critère de sexe et d'âge ne devrait pas être tenu en compte. Cependant ces critères peuvent faire l'objet d'un intérêt particulier si besoin se fait sentir.

L'*API Twitter* nécessite qu'on respecte des règles préétablies. Non seulement la durée de connexion à leurs serveurs est limitée, mais les gazouillis ne demeurent pas dans l'historique. Ainsi, nous comptons collecter au sein d'une base de données un nombre consistant de messages (environs 500) en lien avec le(s) compte(s) de notre étude. Il est à noter, que tout dépend de l'intérêt des utilisateurs. La fréquence des connexions varie et on peut bien se avoir une quantité importante de gazouillis à analyser. Toutefois, cela peut aussi avoir l'effet contraire.

3.3 Déroulement de l'enquête

En premier lieu, nous allons déterminer les caractéristiques des gazouillis que nous allons suivre. Le contenu de ces gazouillis doit être en rapport avec le compte d'un organisme en soin de santé. Pour cela, des termes sont définis et utilisés pour identifier potentiellement les gazouillis connexes à notre sujet d'étude. En outre, nous allons bâtir une application logicielle permettant de se connecter au compte créé et de collecter les gazouillis et Retweet reliés aux mots clés.

Nous utiliserons les techniques d'analyse d'influence et de sentiments pour déterminer l'engagement et l'opinion qui se dégagent des gazouillis recueillis et traités. Un sentiment positif ou négatif (ou neutre) est affecté à chaque gazouillis subjectif. Une catégorisation des résultats nous permettra aussi de déduire plus d'informations pertinentes pour d'éventuels décideurs et leurs plans d'actions.

3.4 Démarche méthodologique

La démarche méthodologique consiste en quatre grandes étapes: la collecte de données, le formatage des données, la détection des sentiments, la classification des sentiments et la présentation des résultats.

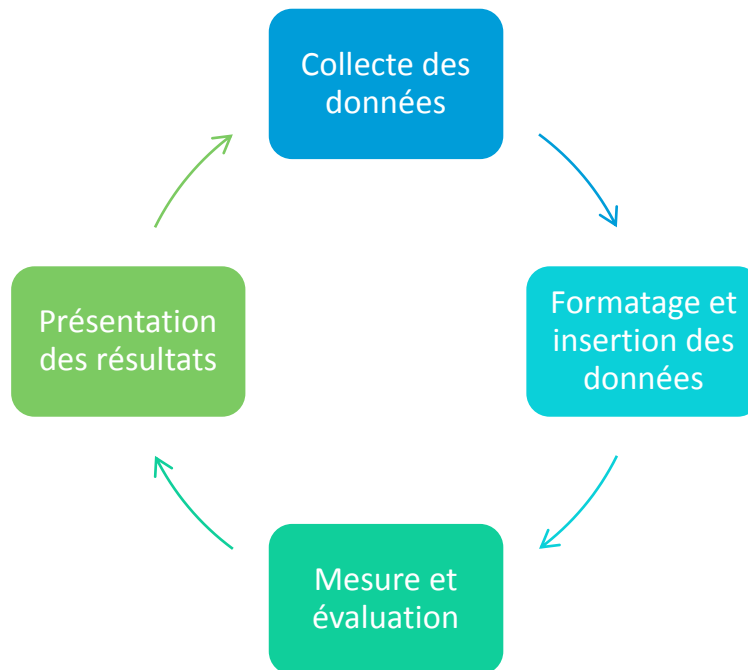


Figure 8. Démarche méthodologique

3.4.1 Collecte de données

Les données à recueillir proviennent du réseau social Twitter. Ce sont les commentaires postés sur les comptes des trois grands hôpitaux choisis: @CHUMontreal pour le Centre hospitalier de l'Université de Montréal, @CHUSteJustine pour le l'Hôpital Ste-Justine et @HopitalChildren pour l'Hôpital Montréal pour enfants. Ainsi la collecte se fera sur les gazouillis des utilisateurs qui font mentions de l'un de ces trois noms d'utilisateurs. Les abonnés expriment généralement leurs opinions sur les micros blogues ou autre réseau social. Ces messages sont pour la plupart accessibles via des services Web. Étant donné le volume énorme de ces données et la vitesse à laquelle elles évoluent, il est quasi impossible de les collecter à la main. C'est pourquoi ces réseaux sociaux, comme *Twitter* ou *Facebook* fournissent des services web et des *APIs* qui permettent à des applications tiers de se connecter à leurs serveurs et d'extraire les données en fonction des termes de recherches utilisés pour filtrer le contenu auquel on est intéressé.

3.4.2 Formatage et insertion des données

Les données issues des médias sociaux sont collectées à partir de plusieurs plateformes qui adhèrent à leur propre format de données, le plus souvent non structurées. Dans cette étape, il convient de formater les données dans une seule base uniforme de sorte que dans les étapes ultérieures, on puisse procéder à l'analyse de ces données. En outre, cette étape permet de préparer le texte pour permettre de sélectionner les attributs nécessaires à l'analyse, car chaque plateforme n'offre pas le même catalogue d'attributs. Une table est donc spécialement créée dans la base de données de l'application et les données stockées suivront le format de cette table.

3.4.3 Mesure et évaluation des mesures

À ce stade, on examine tout d'abord la subjectivité de chaque texte collecté et formaté au sein d'une base de données. Seuls les textes subjectifs sont pris en compte. Cette sous-étape peut aussi être combinée à l'étape précédente de préparation du texte. Une fois que nous avons l'ensemble des phrases subjectives, il convient de déterminer les mesures retenues, puis de les évaluer. Pour cette étude, en premier lieu il s'agit du sentiment de chacune de ses phrases en base de données. L'analyse de sentiments se fait en utilisant des techniques de calcul et de classification statistiques. En outre, après avoir collecté les messages postés dans la base de données, il est possible d'évaluer plusieurs autres mesures nous aidant à estimer l'influence. L'engagement et la portée potentielle sont évalués, ainsi que d'autres caractéristiques du contenu pour savoir quels utilisateurs, *hashtags*, langue a le plus d'influence.

3.4.4 Présentation des résultats

Les mesures sont présentées sous une forme compréhensive et descriptive. Pour cela, les résultats sont présentés en nous servant des tableaux et de graphiques. Le but est de donner une vision des tendances afin de soutenir les discussions et la conclusion.

3.5 Outils de mesure

3.5.1 Application ASTAS de collecte et traitement de données :

Pour réaliser l'analyse des sentiments sur des postes *Twitter* concernant un établissement de santé, il convient d'avoir un système autonome permettant non seulement d'extraire convenablement les données, mais aussi de pouvoir les traiter, formater en base de données et les analyser pour exprimer le sentiment qui se dégage. Pour cela, une application a été développée. L'application est nommée ASTAS pour (Analyse de Sentiments Twitter Appliqué à la Santé). Aussi, l'outil *Semantria* de *LEXALYTICS* permettra de mieux analyser les données et les présenter les résultats sous forme graphique.

Voici une interface graphique montrant un aperçu de l'application ASTAS :



Figure 9.ASTAS App: Interface des résultats pour le CHUM

3.5.2 Architecture de l'application ASTAS

L'application nommée ASTAS (Analyse de Sentiments Twitter Appliquée à la Santé) est un programme développé à l'aide du Framework *PHP CodeIgniter*, ainsi que la bibliothèque *Datumbox* qui nous aide à implémenter l'*API Twitter* pour la connexion et la collectes des données à *Twitter*. La librairie *Datumbox* nous permet également d'implémenter l'analyse des sentiments des gazouillis extraits. Le logiciel *Semantria* permet de faire une analyse de sentiment plus approfondie et une meilleure schématisation du sentiment général et des termes les plus utilisés.

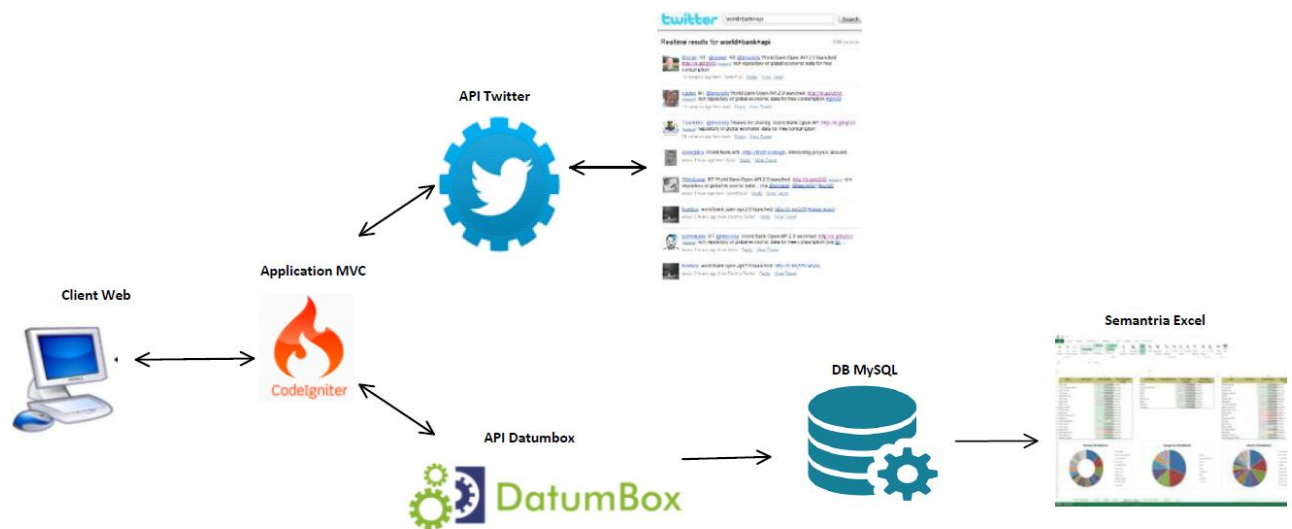


Figure 10. Architecture du système

3.5.3 Diagrammes UML de l'application ASTAS

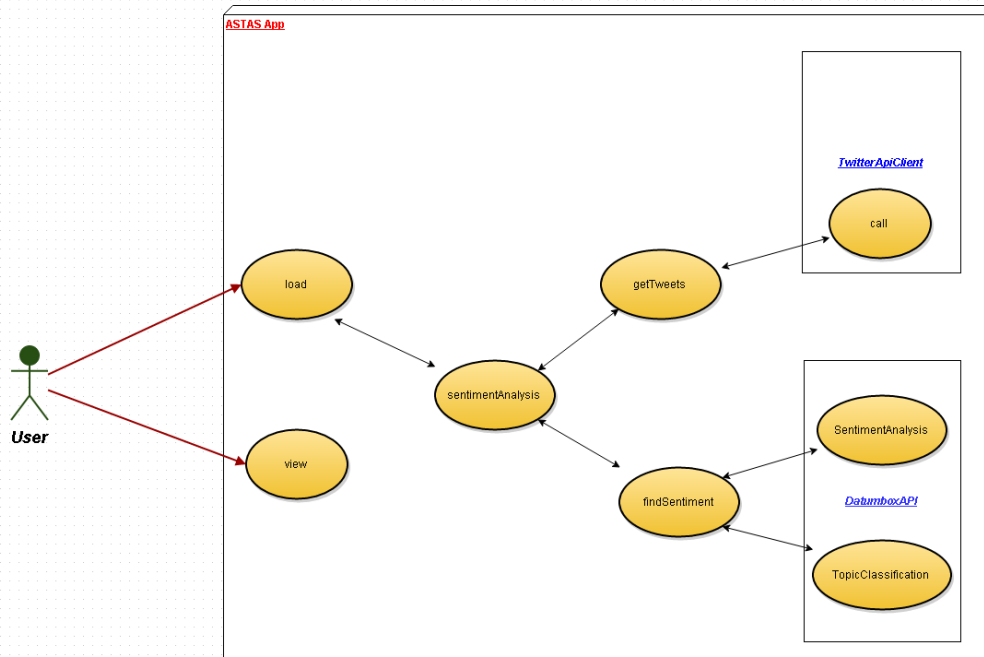


Figure 11. Diagramme UML de cas d'utilisation

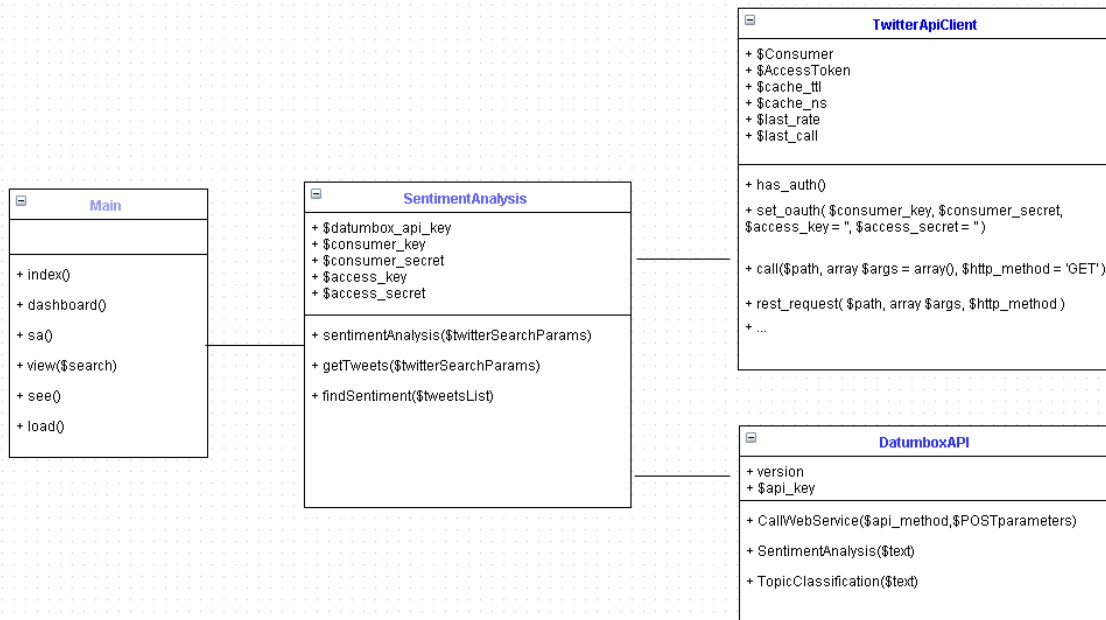


Figure 12. Diagramme de classes

3.5.4 Base de données

Il est important de décider quelles données sont les plus importantes pour notre application, afin de stocker les bonnes informations provenant de *Twitter* dans la base de données. Les champs les plus importants sont considérés. Bien qu'il existe d'autres tables nécessaires pour le fonctionnement de l'application, nous présentons principalement la table qui nous intéresse pour faire l'extraction des gazouillis et l'analyse des données.

Champs	Description	Type
ID	Identifiant du Gazouillis	INT(20)
UserName	Nom d'utilisateur	VARCHAR(100)
LocalTimeStamp	Date du message	Timestamp
Text	Corps du message	TEXT
Language	Langue du message	VARCHAR(100)
Source	Plateforme d'émission du gazouillis	VARCHAR(300)
Location	Localisation géographique du gazouillis	VARCHAR(100)
Hashtags	Termes associé aux gazouillis	VARCHAR(100)
Urls	Liens du gazouillis	VARCHAR(300)
UserMentions	Mentions utilisateurs	VARCHAR(300)
AbonnéCount	Nombre d'abonnés	INT(11)
Name	Nom complet de l'utilisateur	VARCHAR(100)
Sentiment	Sentiment du gazouillis	VARCHAR(100)
Topics	Sujets dont parle le gazouillis	VARCHAR(100)

Tableau 3. Champs de la table Gazouillis

3.5.5 Open Authentication (OAuth)

En mars 2013, *Twitter* a commencé à exiger l'authentification *OAuth* pour toutes les sessions de l'*API*. Cela permet de mieux protéger la vie privée des utilisateurs, tout en améliorant le suivi. Depuis la version 1.1, l'*API Twitter* requiert maintenant l'authentification *OAuth*. En effet, que ce soit pour une application seule, ou même pour une demande utilisateur, *Twitter API 1.1* exige l'authentification pour les nombreux appels d'*API* spécifiques à l'utilisateur.

OAuth est un protocole libre, créé par Blaine Cook et Chris Messina. Il permet d'autoriser un site web à utiliser l'API sécurisée d'un autre site web pour le compte d'un utilisateur. OAuth permet aux utilisateurs de donner, à un site «consommateur», l'accès à des informations personnelles provenant d'un site « fournisseur » de service ou de données, ceci tout en protégeant le pseudonyme et le mot de passe des utilisateurs.

3.5.6 L'API de Twitter

Twitter fournit des interfaces de programmations qui permettent, à l'aide de fonctions, d'interagir avec son service et de se connecter via des applications tierces pour obtenir des informations issues de Twitter (gazouillis, *abonnés*, etc.). Ainsi, L'API *Twitter* 1.1 permet d'accéder à la base de données Twitter et de récupérer/publier plusieurs informations. Quatre classes composent l'API Twitter:

- *SEARCH*: Permet d'interroger *Twitter* pour récupérer des données simples, essentiellement des gazouillis.
- *REST*: Extension avancée de *REST* qui permet d'accéder à des fonctionnalités avancées de *Twitter* : chercher des utilisateurs, des « *abonnés* », voir les « *status* », éditer des informations sur son compte, etc.
- *STREAMING*: permet de communiquer avec *Twitter* en mode *streaming*. La particularité de cette API avancée est de permettre l'accès à de gros volume de données Twitter et d'être moins contrainte par les limites d'accès et d'interrogations de Twitter.
- *WEBSITES*: permet d'intégrer des fonctions de base Twitter dans des sites Web.

3.5.6.1 L'API REST

REST demeure la façon la plus courante pour accéder aux données Twitter. En utilisant les jetons sécurisés obtenus via l'authentification *OAuth*, l'utilisateur fait des demandes à *Twitter* pour des données spécifiques, par exemple pour poster un gazouillis à un utilisateur spécifique.

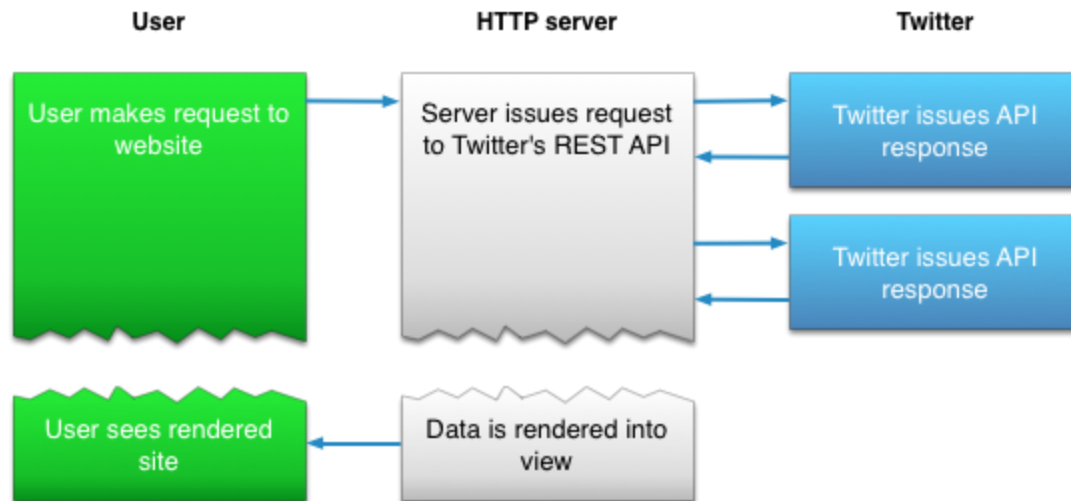


Figure 13.REST Twitter API 1.1

Source: <http://dev.twitter.com>

3.5.6.2 L'API Streaming

L'API Streaming permet de recevoir des gazouillis et notifications en temps réel de Twitter. Cependant, il nécessite une haute performance, une forte persistance de données et une connexion permanente entre le serveur local et *Twitter*.

Différence entre l'API REST et l'API Streaming

Twitter a des contraintes qui limitent les activités d'une application tierce qui utilise son API. L'utilisation de l'API REST est assujettie à plusieurs limites promulguées par *Twitter* comme le temps de connexions pour collecter les données. L'API Streaming permet par contre une connexion de longue durée aux serveurs de *Twitter* et de collecter un volume large de données.

3.5.7 Classe principale main de l'application ASTAS

Le but de la classe est d'utiliser les fonctionnalités offertes par les API *Twitter* et *Datumbox*, permettant de collecter les gazouillis et d'évaluer leur polarité.

Attributs

- \$datumbox_api_key: Votre Datumbox clé API
- \$consumer_key: Votre Consumer key de Twitter
- \$consumer_secret : Votre Consumer secret de Twitter
- \$ACCESS_KEY: Votre Access key de Twitter
- \$access_secret: Votre Access secret de Twitter

Fonctions

- Fonction load(\$twitterSearchParams)
 - *Description* : fonction Search/gazouillis de l'API Twitter pour faire la recherche en fonction des paramètres de recherche.
 - *Paramètre(s)*: \$twitterSearchParams, critères de recherche.
 - *Valeur de retour*: \$gazouillis, liste des gazouillis qui correspondent aux critères.
- Fonction findSentiment(\$gazouillisList)
 - *Description* : Cette fonction récupère la liste de gazouillis trouvés et évalue leurs sentiments.
 - *Paramètre(s)* : \$gazouillisList, liste des gazouillis retrouvés
 - *Valeur de retour* : Retourne le tableau contenant le sentiment, le topic et l'ID.
- Fonction CallWebService (\$api_method,\$POSTparameters)
 - *Description* : fonction appelle l'API Datumbox
 - *Paramètre(s)* : \$api_method,\$POSTparameters
 - *Valeur de retour* : Retourne le résultat en JSON contenant le sentiment.

3.5.8 Semantria pour Excel de LEXALYTICS

Semantria Excel est une solution d'analyse de sentiments bâtit comme un *Add-in* Excel. Il le transforme en un outil puissant et facile à utiliser pour la surveillance et la visualisation des données de Twitter, Facebook, les enquêtes et d'autres données non structurées. Il offre de puissants outils de visualisation Excel.

L'analyse des sentiments

À l'aide d'un moteur de catégorisation, *Semantria Excel* identifie ce qui est positif, négatif ou neutre dans le contenu des gazouillis, et découvre ce que vos clients disent de la marque et l'expérience client. Il est possible d'utiliser pour cela un lexique préétabli.

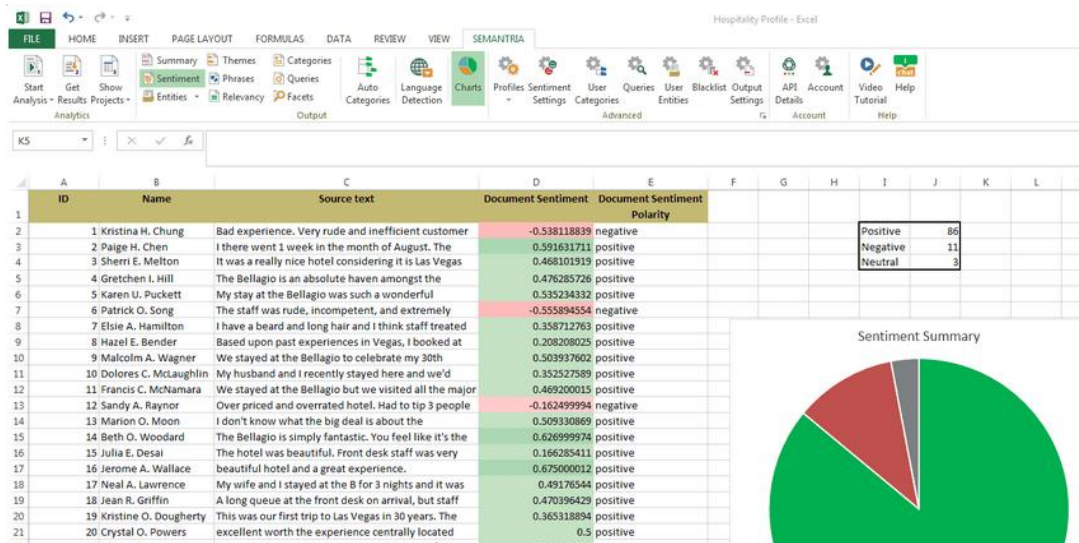


Figure 14. Semantria Excel de LEXALYTICS

Source: <http://predictiveanalyticstoday.com>

Chapitre 4

Expérimentation et analyse des résultats

Ce chapitre décrit l'expérimentation au cours de laquelle les données provenant des conversations sur Twitter ont été collectées et enregistrées au sein d'une base de données MySQL. Ces données forment notre ensemble de données ou « *dataset* ».

Le contenu de ce chapitre est organisé de telle sorte que nous commençons par décrire les mesures utilisés, puis pour chacune des mesures nous présentons les mesures observées et les résultats. Ces résultats seront commentés.

4.1 Ensemble de données collectées

À l'aide des termes de recherches ayant un rapport avec les trois grands centres de soins de santé choisis, les fonctions, appelant l'API Twitter pour s'authentifier et collecter les gazouillis, ont été utilisées. La structure d'un gazouillis collecté est:

ID	User Name	Local Time Stamp	Text	Language	Source	Location	Hashtags	Urls	User Mentions	Abonné Count	Name
...

Tableau 4. Champs d'un gazouillis collecté

Les trois grands organismes en soins de santé que nous avons choisis sont :

- Le Centre hospitalier de l'Université de Montréal, pour son importance au Québec;
- Le Centre hospitalier Ste-Justine, pour son rôle majeur dans les soins aux enfants;

- L'Hôpital Montréal pour enfants, pour sa position quant aux soins données aux enfants et pour le fait que c'est un des hôpitaux du Centre de santé de l'Université McGill (CSUM).

Dans la revue de littérature, les termes de recherches sont en général des mots dièses ou *hashtags*. Cependant pour cet essai, le nom du compte de chaque organisme est privilégié comme terme de recherche. En effet, lors d'une première phase de collecte, en prenant le terme de recherche #chum pour collecter des données du CHUM, une majeure partie des gazouillis recueillis n'avait rien à voir avec l'hôpital ou le compte du CHUM; idem pour l'hôpital Montréal pour enfants qui utilise un acronyme plutôt qu'un terme #hopitalmontrealpourenfant. Les termes de recherches pour chaque hôpital sont: @chumontreal, @ChuStjeJustine et @HopitalChildren. Avec ces termes, on a pu collecter un ensemble de données de plus de 500 gazouillis ; ce qui est remarquable compte tenu des restrictions imposées par Twitter. La collecte des données a été effectuée durant une période débutant le 29 juin jusqu'au 28 juillet 2015, soit environ un mois de mesures.

	Terme de recherche	Nombre de gazouillis trouvés
CHUM	@chumontreal	71
CHU Ste-Justine	@ChuStjeJustine	80
Hôpital Montréal pour enfant	@HopitalChildren	383

Tableau 5. Termes de recherche et nombre de gazouillis collectés

4.2 Résultats et discussions

4.2.1 Analyse du sentiment

Il est très important de connaître le sentiment général se dégageant des conversations ayant un rapport avec votre marque. En analysant quel contenu dont les réactions sont positives versus celles qui sont négatives, les décideurs d'un organisme en soins de santé peuvent avoir en temps

réel, l'appréciation que les gens se font d'un service, d'une nouvelle campagne ou de toute initiative susceptible de soulever des opinions diverses.

Pour mesurer le sentiment, nous avons utilisé les deux approches d'analyse de sentiment vues dans la revue de littérature, à savoir :

- L'analyse de sentiment basée sur un lexique et l'analyse de sentiment basée sur une machine d'apprentissage. Pour l'approche basée sur un lexique, nous avons utilisé Excel 2013 comme outil de mesure avec le plugin Semantria Excel. Un lexique de mots positifs et négatifs a été élaboré et le contenu de chaque gazouillis a été évalué en fonction de ce lexique pour déterminer si ce gazouillis exprime un sentiment positif ou plutôt négatif.
- L'analyse basée sur une machine d'apprentissage. Des API de *Datumbox* ont été utilisées dans l'application ASTAS. La première fonction est de déterminer la subjectivité d'un gazouillis, une fois le gazouillis caractérisé subjectif, on utilise l'évaluation du sentiment en utilisant le classificateur naïf de Bayes (*Naives Bayes Classifier*). Une fois le sentiment trouvé, le gazouillis est mis à jour dans la base de données MySQL.

Il est à noter qu'il existe des gazouillis dont le sentiment peut être qualifié de neutre, en ce sens qu'il n'exprime ni une opinion positive ni négative. Enfin, il y a aussi des gazouillis non subjectifs, c'est-à-dire dont les approches utilisées n'ont pu déterminer s'ils expriment un sentiment.

Une fois le sentiment de chaque gazouillis trouvé, que ce soit avec l'approche basée sur un lexique, que l'approche utilisant une machine d'apprentissage, nous utilisons une formule inspirée de la revue de littérature, celle de Van't Ende (2003). Dans ses recherches, cette dernière a pu établir une corrélation entre le score du sentiment et l'image ou l'appréciation de la marque.

- % de gazouillis positifs = Nb gazouillis positifs / Nb total de gazouillis;
- % de gazouillis négatifs = Nb gazouillis négatifs / Nb total de gazouillis;
- % de gazouillis neutres = Nb gazouillis neutres / Nb total de gazouillis.

Et en définitive :

Sentiment = (Nombre de gazouillis positifs - Nombre de gazouillis négatifs + Nombre de gazouillis neutres) / (Nombre total de gazouillis).

Les résultats, pour chaque grand hôpital, sont présentés dans les tableaux et graphiques de l'Annexe IV.

Discussion de l'analyse des sentiments

La première analyse, celle du sentiment consistait en deux approches, comme vu dans la revue de littérature: une basée sur une machine d'apprentissage et l'autre basée sur un lexique. Les tableaux contiennent le nombre de gazouillis mesurés avec un sentiment positif, négative ou neutre, et ce, pour les deux approches. Il y a quatre possibilités de valeur du champ Sentiment dans la base de données : « *positive* », « *negative* », « *neutral* » et « *none* ». Ainsi pour quatre gazouillis, le sentiment général est positif s'il est supérieur ou égal à 50 %.

En comparant les deux approches, nous pouvons constater que l'approche d'analyse de sentiments basée sur une machine d'apprentissage est beaucoup plus efficace et précise dans la détermination des sentiments. En effet, pour les trois grands hôpitaux, le sentiment général dépasse toujours la moyenne 50 %, mais surtout le nombre de gazouillis neutres ou n'ayant aucun sentiment est beaucoup plus élevé dans l'approche basée sur un lexique que sur celle avec une machine d'apprentissage. Cela confirme ce qui a été mentionné dans la revue de littérature par Pang et Lee (2008). Ces derniers ont insistés sur le manque de précision et les difficultés en rapport avec l'approche basée sur un lexique. Cette différence s'explique par le fait qu'avec l'approche basée sur un lexique les sentiments peuvent être exprimés d'une manière ambiguë au sein d'un message posté. De plus, le message posté peut exprimer un sentiment, mais si les mots figurants dans le lexique ne sont pas retrouvés, le gazouillis sera considéré neutre ou n'ayant aucun sentiment. La différence peut aussi être expliquée par le fait que l'approche basée sur une machine d'apprentissage utilise des algorithmes de classification mieux entraînés sur de plus gros ensemble de données.

Enfin, nous pouvons observer que dans l'ensemble de données, un grand nombre de gazouillis postés provenaient d'un compte appartenant à l'un des grands hôpitaux. Le sentiment est alors

biaisé pour ces gazouillis car les gazouillis postés sont beaucoup plus promotionnels ou informatifs. Seuls ceux qui sont postés par des utilisateurs différents peuvent réellement être pris en compte. Toutefois, le but premier était de pouvoir déduire un sentiment à chaque gazouillis et d'en déduire l'opinion générale. Cette opinion est plutôt positive dans la plupart des hôpitaux. Ce qui par corrélation peut signifier que l'image ou la perception générale des utilisateurs est dans l'ensemble plus positive que négative.

4.2.2 Analyse de l'influence

Il est très important de savoir quelle influence votre organisation a sur les réseaux sociaux. En effet, tout effort consenti en termes de temps ou encore d'argent doit être justifié par un retour sur investissement qui peut s'exprimer en terme financier ou d'une autre manière. Dans le cadre de l'analyse des médias sociaux, il existe deux mesures aidant à mesurer l'influence d'une marque auprès des utilisateurs qui la suivent ou mentionnent un terme associé au compte de cette marque:

- **Le niveau d'engagement**

L'engagement est une mesure permettant aux décideurs de votre organisation de savoir à quel niveau les utilisateurs sont engagés et participent activement aux conversations autour de votre marque. Idéalement pour un gazouillis posté par le compte utilisateur de l'organisation, il devrait avoir au moins une réaction qui peut être soit un « *Replies* », un « *Retweet* », une « *Mention* ».

La formule pour mesurer l'engagement est :

$$\text{Engagement} = (\text{Nombre de Replies} + \text{Nombre de Retweet} + \text{Nombre de Mentions}) / (\text{Nombre de Gazouillis} + \text{Nombre de Replies} + \text{Nombre de Retweet} + \text{Nombre de Mentions})$$

- *@Replies*: Gazouillis commençant par l'identifiant du compte de l'organisation. Il symbolise une réponse à un gazouillis. Exemple: «@chumontreal ...bravo».
- *Retweet*: C'est lorsqu'un utilisateur partage directement un gazouillis mentionnant votre terme, avec leur public. Il est précédé d'un RT. Exemple: "RT @chumontreal: bravo... ».

- *Mentions*: Gazouillis utilisant l'identifiant du compte de l'organisation, mais qui n'est pas un Replies ou une Mention. Exemple: «bravo ... @chumontreal ...».

Le tableau suivant résume l'engagement mesuré et évalué.

	chumontreal	ChuSteJustine	HopitalChildren
Nb Replies	3	4	11
Nb Retweet	27	18	202
Nb Mentions	44	62	167
Engagement =	51,03 %	51,22 %	49,8 %

Tableau 6. Niveau d'engagement

Discussion sur le niveau d'engagement

Pour savoir comment fonctionne une campagne de présence sur les médias sociaux, le niveau d'engagement peut être utilisé. En effet, pour un gazouillis posté par le compte utilisateur d'un grand hôpital, il faudrait au moins une réponse (*Replies*), une mention (*Mentions*) ou un transfert ou retranscription tel quel de ce message (*Retweet*). Ainsi, le niveau minimal d'engagement qu'un compte d'un organisme en soin de santé doit avoir est de deux quarts, soit 50 %.

En observant le tableau ci-dessus, nous constatons bien que le niveau d'engagement est supérieur pour les comptes du CHUM (51,03 %) et du CHU Ste-Justine (51,22 %). Par contre pour l'Hôpital Montréal pour enfants il est de 49,8 %. Cela est d'autant plus étonnant car le compte de dernier organisme était celui avec le plus de gazouillis publiés (soit 383), alors que les deux premiers n'atteignaient pas 100 gazouillis. Ce résultat démontre la pertinence et l'importance de cette mesure, car il ne suffit pas d'être soi-même actif sur les réseaux sociaux en publiant des messages sur son propre compte, encore faudrait-il que ces messages soient repris ou qu'ils suscitent une réaction de la part des utilisateurs. C'est ainsi qu'on peut savoir si oui ou non le message véhiculé a de l'impact sur l'audience.

Enfin, nous observons que dans les trois hôpitaux, le nombre de réponses est très faible par rapport aux autres catégories de gazouillis (*Mentions* ou *Retweet*). Cela indique que les messages postés ne suscitent pas de débat au sein de la population; ce qui peut de prime abord être vue comme un assentiment général, mais qui en réalité peut traduire par un désintérêt par rapport aux messages postés et la communication véhiculée. Cette dernière doit donc être plus ciblée, mieux réfléchie afin d’avoir le plus de rétroactions de la part des utilisateurs, permettant ainsi d’avoir leur avis positifs ou non sur les actions et campagnes menées.

- **La portée potentielle**

Cette mesure permet de savoir le nombre de personnes potentiels dans le réseau social de l’organisme. En détenant une telle information, il est possible de prendre une meilleure décision pour savoir si cela vaut la peine ou non de prendre le risque de faire une campagne en fonction du nombre de personnes que cette campagne peut atteindre. En ayant une grande portée potentielle, on dispose de plus de personne qu’on peut influencer.

Par exemple, si un utilisateur a un abonné et publie un message avec un mot dièse spécifique, la portée potentielle est de un. Mais si ce même utilisateur a mille abonnés, en postant un message, la portée potentielle est de mille, et ce nombre croît de façon exponentielle, surtout si ces abonnés disposent d’un grand nombre d’abonnés.

$$\text{Portée potentielle} = \sum \text{Utilisateurs} + \sum \text{Nb Abonnés de chaque Utilisateur}$$

	chumontreal	ChuSteJustine	HopitalChildren
Nb d’utilisateurs	56	57	183
Nb d’abonnés de ces utilisateurs	176 611	152 648	269 450
Portée potentielle =	176 667	152 705	269 633

Tableau 7.Portée potentielle

Discussion sur la portée potentielle

La portée potentielle permet à une campagne de rejoindre un plus grand nombre de personnes que la réalité. En effet, dans le tableau ci-dessus, le nombre d'utilisateurs qui sont abonnés ou qui sont intéressés par les comptes des trois grands hôpitaux est dérisoire (en moyenne un peu plus de 50 pour le CHUM et le CHU Ste-Justine et 183 pour l'hôpital Montréal pour enfants). Cependant lorsqu'on s'intéresse au réseau des utilisateurs, on se rend compte que le nombre de personnes pouvant être exposées aux messages postés par ces hôpitaux est beaucoup plus grand. Une campagne de recrutement est sans doute nécessaire.

4.2.3 Analyse du volume

L'analyse du volume donne l'habileté de comprendre l'impact et l'intérêt de votre compte. Certaines organisations se limitent souvent à compter le nombre de «*abonnés*» ou le nombre de «*like*»; toutefois, il existe plusieurs autres mesures permettant de connaître les caractéristiques de votre audience.

- **Les utilisateurs les plus actifs**

Cette mesure permet de savoir quels sont les utilisateurs les plus actifs parmi ceux qui engagent une ou des conversations avec les termes en rapport avec votre compte. On peut évaluer cette mesure en évaluant le nombre de gazouillis postés par utilisateurs :

a) Centre Hospitalier de l'Université de Montréal

Nom de l'utilisateur	Nb de gazouillis
Fondation du CHUM	7
Éric Soulard	3
Elisa Pucella	2
André Picard	2
Sandra Aube	2
Thierry Alquier	2
CNWNouvelles	2

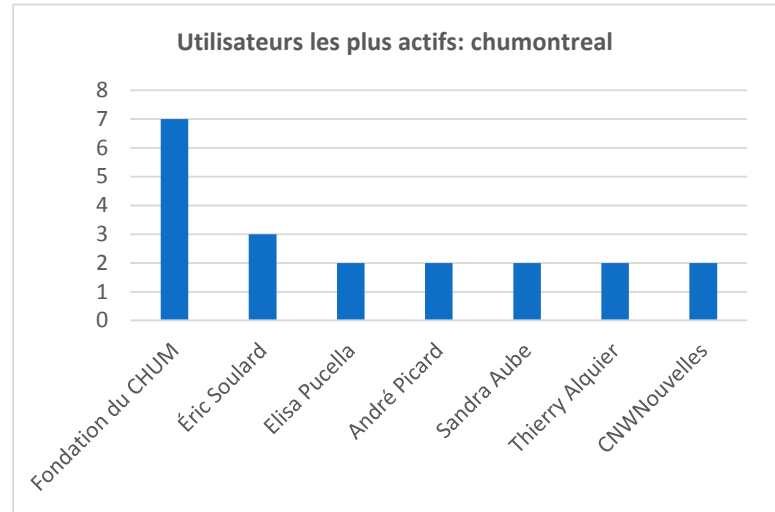


Tableau 8.Chumontreal: Utilisateurs les plus actifs

b) Centre Hospitalier de Ste-Justine

Nom	Nb de gazouillis
Thedalekbieber	5
Doctor Who Society	5
HotTammami	4
BIO PLS (AIMH Niall)	3
Plongée Total Diving	3
Manuel Doré	3
Matin Toronto	2
Patrick Mercier	2

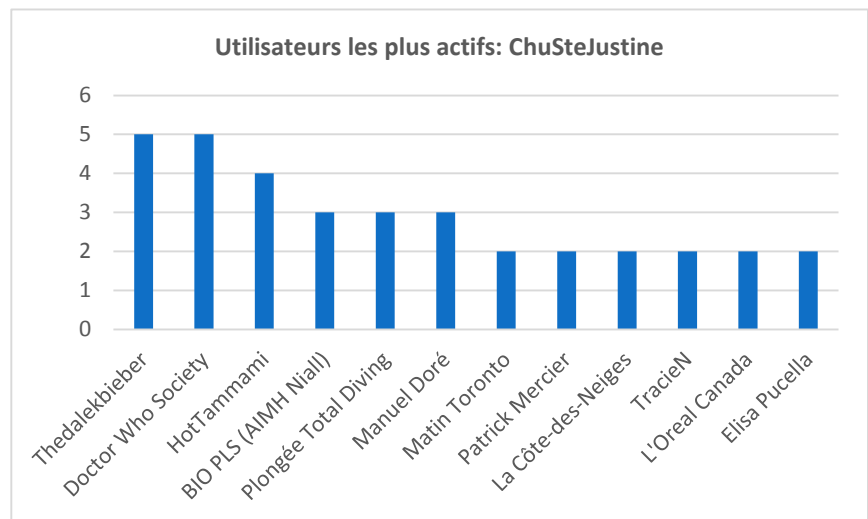


Tableau 9.ChuSteJustine: Utilisateurs les plus actifs

c) Hôpital Montréal pour enfants

Nom	Nb de gazouillis
MUHC	26
Mamanpourlavie.com	16
Montreal Children's	14
AndyCollins for Kids	12
Terri	8
Dept. of Pediatrics	8
MedExtra.com	7
Claude M	7
Keiylikko	6
aline levi	6
Pfizer Canada	6

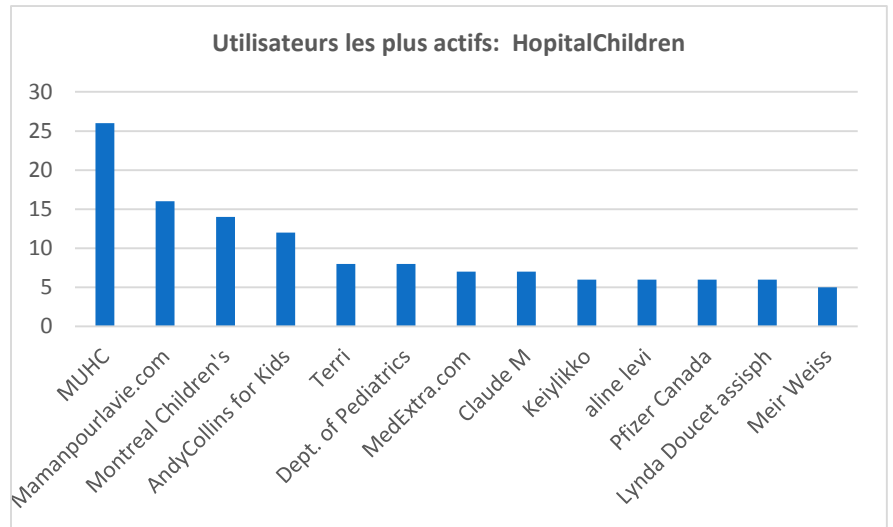


Tableau 10. HopitalChildren: Utilisateurs les plus actifs

Discussion sur les utilisateurs les plus actifs

Un des faits observés par les tableaux ci-dessus est que les utilisateurs les plus actifs sont de caractéristiques diverses. On retrouve en effet des fondations ou ONG, mais aussi des personnes particulières. Cette mesure peut aider les organismes en soins de santé à savoir quels sont les utilisateurs les plus actifs. Toutefois cette mesure ne signifie pas grand-chose si ces utilisateurs n'ont pas une grande audience. Il convient de regarder une autre mesure qui sont les utilisateurs les plus influents.

- **Les utilisateurs les plus influents**

Un utilisateur est influent lorsqu'il possède un grand nombre d'abonnés. En effet, même si un certain utilisateur X est plus actif (en terme de nombre de gazouillis postés) sur le compte de l'organisme qu'un autre utilisateur Y, ce dernier utilisateur Y peut être le plus influent parce qu'un seul de ses gazouillis sera vu par un nombre plus important d'abonnés. On évalue donc

cette mesure en mesurant les utilisateurs ayant le plus grand nombre d'abonnés, sur le compte de votre organisation.

a) *Centre Hospitalier de l'Université de Montréal*

UserName	Nom	Nb d'abonnés
CanoeNouvelles	Canoe.ca Actualités	41428
picardonhealth	André Picard	37308
UMontreal	Université Montréal	15350
ProstateCancerC	ProstateCancerCanada	9679
CNWNouvelles	CNWNouvelles	9437
CNWNews	CNWNews	7642
ArianeKrol	Ariane Krol	7282
Danielelorain	Danièle Lorain	7068
hgagneTVA	Harold Gagné	6105
PublicServiceCa	PublicService Canada	5149
Pilar_Institute	Pilar Institute	5030
IRSC_CIHR	IRSC	3630
CNWHealthcare	CNWHealthcare	2339

Tableau 11.Chumontreal: Utilisateurs les plus influents

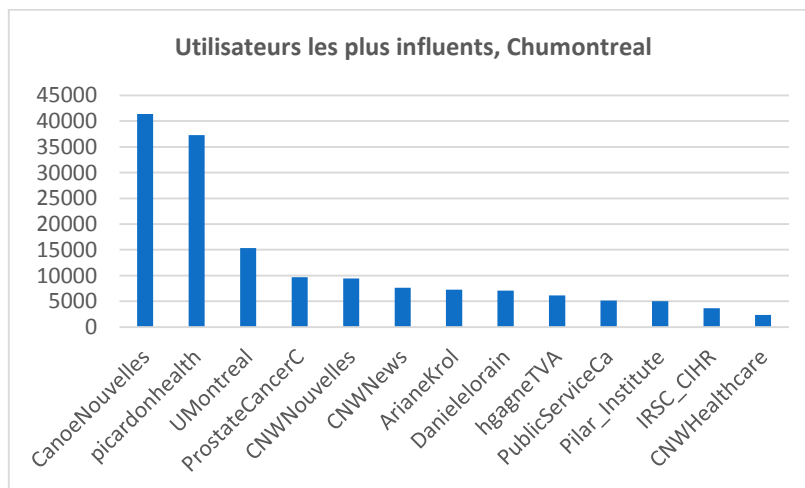


Figure 15.Chumontreal: Utilisateurs les plus influents

b) *Centre hospitalier de Ste-Justine*

UserName	Nom	Nb d'abonnés
CanoeNouvelles	Canoe.ca Actualités	41566

tracycityline	Tracy Moore	40346
UMontreal	Université Montréal	15342
plongee	Plongée Total Diving	6925
Naitreetgrandir	Naître et grandir	6551
PublicServiceCa	PublicService Canada	5148
L'OrealPress_Can	L'Oreal Canada	3754
cbcHomerun	CBC Homerun	3678
BossDivaShop	Unplugged	2967
uMontreal_news	uMontreal News	2622
CMNCanada	ChildrnsMiracleNetwrk	2537

Tableau 12. ChuSteJustine: Utilisateurs les plus influents

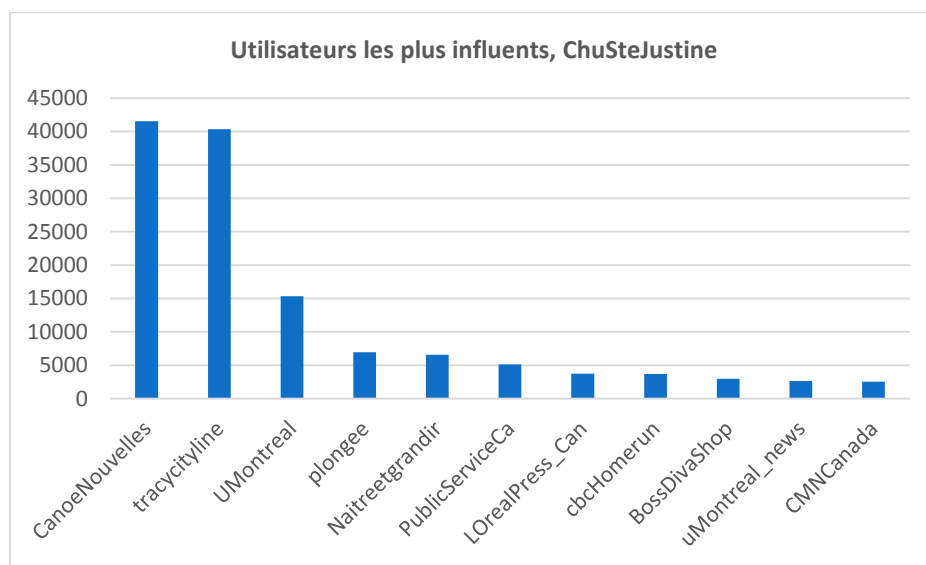


Figure 16. ChuSteJustine: Utilisateurs les plus influents

c) Hôpital Montréal pour enfants

UserName	Nom	Nb d'abonnés
SidLee	SID LEE	43306
CIHR_IRSC	CIHR	27467
Akillorn19	Alex Killorn	15111
StolleryKids	Stollery Kids	9879
CharityNewsUK	CharityNewsUk	6878
Naitreetgrandir	Naître et grandir	6474
hgagneTVA	Harold Gagné	6120
mamanpourelavie	Mamanpourelavie.com	5743
webcoupdoeil	coup d'oeil	5466

Le_Book	LE BOOK	5440
PublicServiceCa	PublicService Canada	5151
IGAQC	Supermarchés IGA	4956
AboutKidsHealth	AboutKidsHealth.ca	4776

Tableau 13. HopitalChildren: Utilisateurs les plus influents

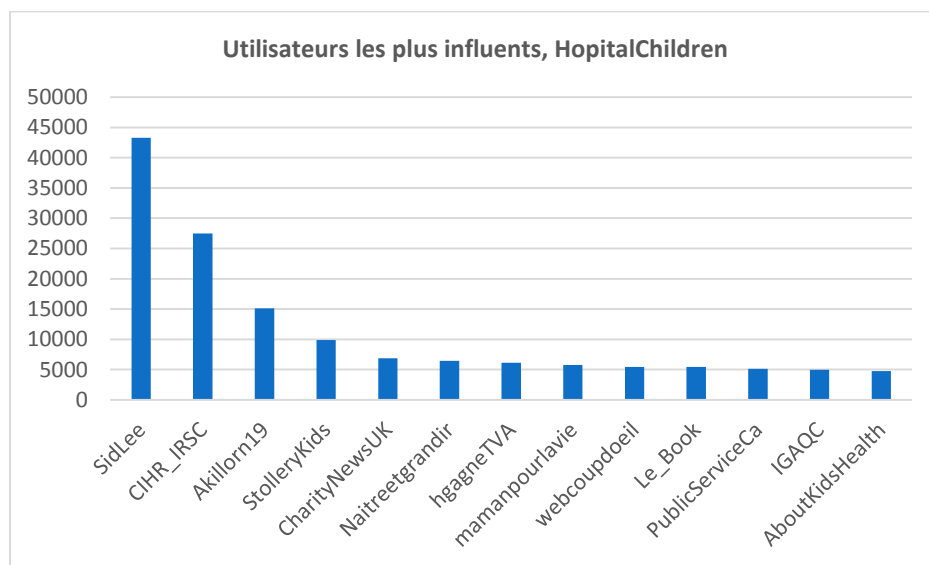


Figure 17. HopitalChildren: Utilisateurs les plus influents

Discussion sur les utilisateurs les plus influents

La mesure consistant à connaître les utilisateurs les plus influents est beaucoup plus intéressante. Un utilisateur, bien qu'il soit moins actif, peut être plus influent car un seul de ses avis sur l'image ou l'appréciation des services reçus, peut être suivi par des milliers d'utilisateurs qui sont ses abonnés. Dans le cadre d'une campagne ciblée, il convient de recruter de tels utilisateurs pour avoir un plus grand impact.

Aussi, on observe que parmi les utilisateurs les plus influents dans les messages postés sur les 3 grands hôpitaux, on retrouve les comptes utilisateurs de groupes de presse tels que Canoe Nouvelles et CNW Nouvelles. Cela rejoint le point soulevé dans la revue de littérature. En effet, les journaux et la presse peuvent jouer un rôle de relais auprès des utilisateurs pour une plus grande pédagogie ou une meilleure communication en temps de crise.

- **Autres mesures :**

	Chumontreal	ChuSteJustine	HopitalChildren
Language	Fr (64); En(7)	Fr (50); En(25)	Fr (245); En(135)
Hashtags	- CHUM (6); - marijuana (5); - TVA fibrosekystique dondorganes (4);	- montrealcomiccon (9); - DuracellAlimenteLesSourires (3); - Montreal (3); - RCES (3);	- heatsafety (24); - baignadesécuritaire (14); - JoinUs (11); - chaleuretsécurité (11); - swimsafety (10);
Topics	- Business & Economy (16); - Science (12); - Society (7); - Sports (6); - Health (5);	- Recreation & Activities (21); - Business & Economy (12); - Sports (7); - Home & Domestic Life (6); - Health (5);	- Home & Domestic Life (97); - Sports (47); - Health (47); - Recreation & Activities (40); - Arts (28);
Locations	- Canada (10); - Montréal (6); - Québec (4);	- Canada (6); - Montréal, Québec (4); - Toronto (4); - Canada/Québec (3);	- Montreal, Quebec (41); - Québec, Canada (19);
Sources	- Twitter Web Client (29); - Twitter for iPhone (12); - SocialOomph (7); - GazouillisDeck (6); - Twitter for iPad (5);	- Twitter for Android (21); - Twitter for iPhone (20); - Twitter Web Client (14); - GazouillisDeck (9); - Hootsuite (5);	- Twitter Web Client (119); - Twitter for iPhone (98); - Hootsuite (57); - Twitter for Android (49); - Twitter for iPad(12);

Tableau 14. Autres mesures de volume

Discussion sur les autres mesures

En regardant le tableau ci-dessus, on observe que pour le cas des trois grands hôpitaux, la langue utilisée en majorité sur les réseaux sociaux est le français. En effet le pourcentage du français est à plus de 90 % pour le CHUM et à plus de 60 % dans le cas du CHU de Ste-Justine et de l'Hôpital Montréal pour enfants. C'est assez étonnant car ce dernier hôpital est à majorité anglophone, et fait partie du Centre universitaire de santé McGill (CSUM).

En outre, on observe aussi que les mots dièses ou hashtags ainsi que les topics les plus utilisés sont variés. Toutefois, on note que les hashtags ainsi que les topics qui concernent la santé occupent une place importante. Cela confirme ce qui a été évoqué dans la revue de littérature, à savoir que la santé est un sujet majeur dans les conversations sur les médias sociaux.

Enfin, nous observons que les sources ou outils utilisés sont hétérogènes: pour le CHUM et l'hôpital Montréal pour enfants, la majorité des gazouillis postés le sont à partir du site web de Twitter. Toutefois, les appareils et applications mobiles sont aussi très utilisés. L'application mobile Twitter for iPhone est l'une des plus utilisés, ce qui dénote que les iPhones et les technologies de Apple sont en vogue au sein de la population canadienne tel que mentionné dans la revue de littérature.

- **Le volume dans le temps**

Le volume de gazouillis a été mesuré dans le temps pour chaque grand hôpital. Cette mesure détermine le nombre de gazouillis publiés par jour. Cela donne les graphiques suivants:

a) Centre hospitalier de l'Université de Montréal

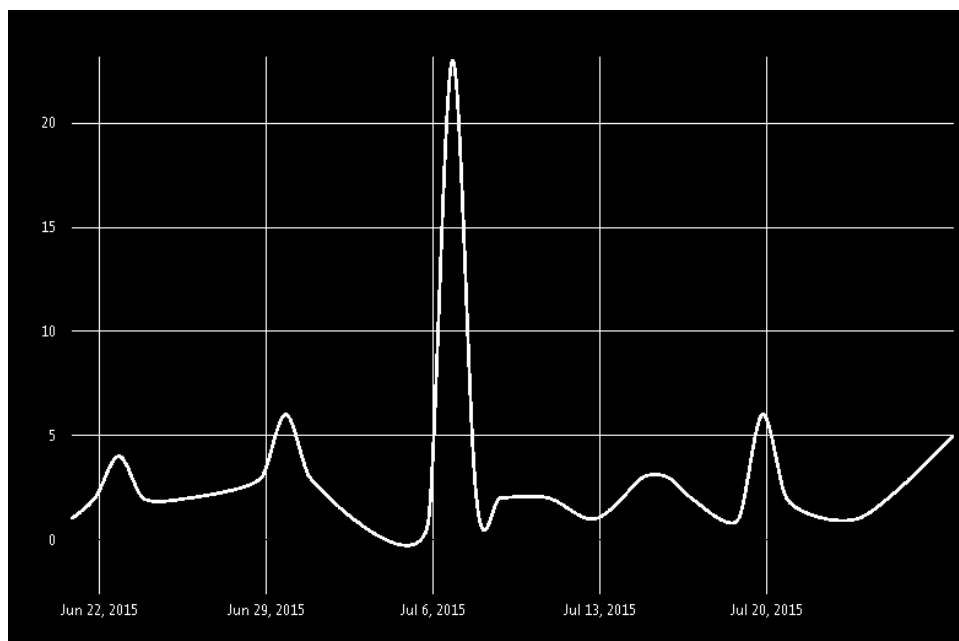


Figure 18.Chumontreal: Volume dans le temps

b) *Centre hospitalier de Ste-Justine*

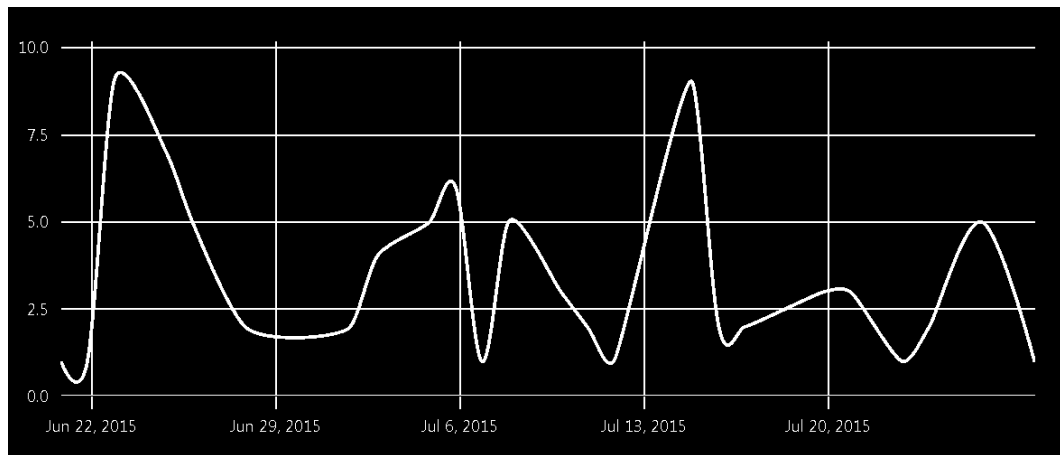


Figure 19.ChuSteJustine: Volume dans le temps

c) *Hôpital Montréal pour enfants*

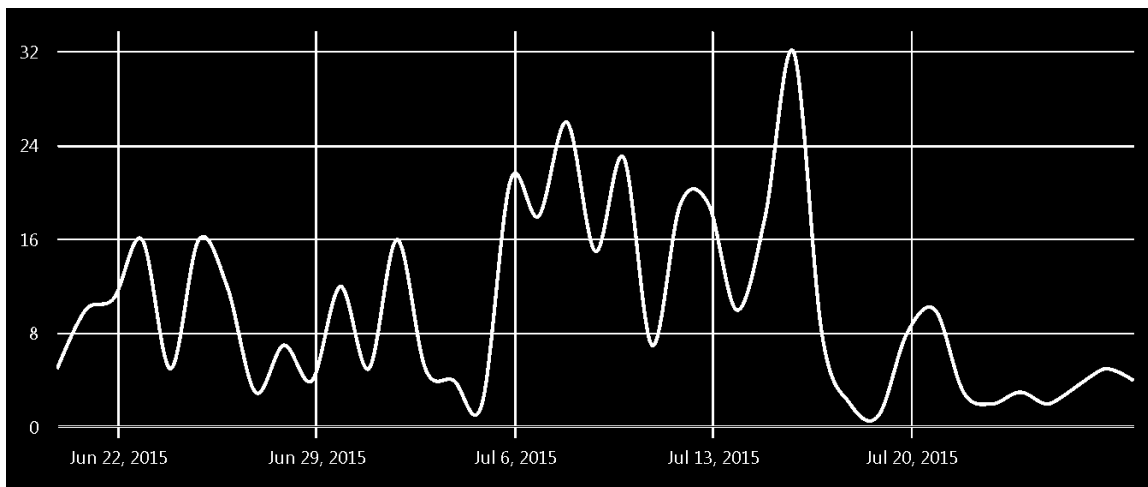


Figure 20.HopitalChildren: Volume dans le temps

Discussion sur le volume dans le temps

Les figures 22, 23 et 24 présentent l'évolution du nombre de gazouillis publiés dans le temps pour les comptes *Twitter* des trois grands hôpitaux sur lesquels se base cet essai. En effet, on observe que le volume de gazouillis publiés au quotidien varie peu pour les trois grands hôpitaux : Pour le CHUM, le volume dans le temps est de l'ordre de 5 gazouillis par jour. Pour le CHU Ste-Justine il est entre 5 et 10 gazouillis par jour ; et enfin pour l'Hôpital Montréal pour enfants, ce nombre varie entre 16 et 21 gazouillis postés par jour.

On observe un cas particulier : durant toute la période de mesure, la variation a été très forte sur le compte du CHUM. En effet, durant la période entourant le 6 juillet 2015, le volume de gazouillis est passé de 5 à plus de 25. Cette différence s'explique par l'évènement de la démission du directeur général du CHUM. Ce qui a suscité de nombreuses réactions dans le débat.

Conclusion

Le but visé par cet essai est d'explorer les solutions d'analyse des données issues des médias sociaux appliqués au domaine des soins de santé. Pour ce faire, le cheminement de cette étude a commencé par une problématique qui soulignait le fait que bien que la plupart des organismes en soins de santé sont présents sur les médias sociaux, très peu utilisent efficacement l'analyse de l'énorme volume de données engendrées par les utilisateurs.

La revue de littérature a permis de voir les différentes recherches effectuées ces dernières années. Après avoir défini ce que sont les médias sociaux, on a vu qu'ils sont largement utilisés au sein de la société canadienne. Cette utilisation s'étend de plus en plus dans le domaine de la santé où les utilisateurs prennent les médias sociaux comme sources d'informations ou plateforme de discussion pour donner leurs opinions sur l'image et les sujets en rapport avec un ou plusieurs organismes en soins de santé.

Twitter, grâce à son API, permet de collecter des données à partir de termes de recherches. Après avoir défini les mesures utilisées, nous avons procédé à l'implémentation d'un outil de mesure, l'application ASTAS qui utilisait les API de *Twitter* et de *Datumbox* pour la collecte et l'analyse de données issues des médias sociaux. Les trois grands hôpitaux sur lesquels s'est basée l'expérimentation, qui a duré un mois, sont: Le centre hospitalier de l'université de Montréal (CHUM) pour son importance politico-économique au Québec, le CHU Ste-Justine et l'Hôpital Montréal pour enfants, pour leurs rôles majeurs dans les soins octroyés aux enfants. La collecte a permis d'enregistrer un ensemble de données de plus de 500 gazouillis publiés.

À partir de cet ensemble de données, l'information a été analysée en utilisant les techniques et approches issues de la revue de littérature. En premier lieu, on a effectué une analyse de sentiments en utilisant deux approches: l'approche basée sur un lexique et l'autre qui est basée sur une machine d'apprentissage. Ensuite, on a effectué l'analyse d'influence, permettant de connaître le niveau d'engagement et la portée potentielle d'utilisateurs. Enfin, une analyse de

volume a permet d'aller voir en profondeur quels sont les utilisateurs les plus actifs et influents, quels sont les mots dièses, les langues ou les sources les plus utilisés.

Ce chapitre présente donc les conclusions qui peuvent découler des analyses effectuées sur l'ensemble de données collectées. L'objectif étant de confirmer que les données issues des médias sociaux peuvent servir d'indicateurs fiables pour des décisions stratégiques.

En premier lieu, l'analyse de sentiment a été la principale analyse effectuée. L'analyse de sentiments exprimés dans les gazouillis pour les trois grands hôpitaux choisis a révélé qu'une majorité des messages postés sont subjectifs et expriment au moins un sentiment positif, négatif ou neutre. Toutefois, on observe une différence dans l'approche d'analyse de sentiments basée sur un lexique qui présentait un taux de sentiment neutre assez élevé, par rapport à celui obtenu par l'approche basée sur une machine d'apprentissage. Ce constat est aussi le même pour le nombre de messages postés qui n'exprime aucun sentiment (*none*). Ce nombre s'avérait assez élevé dans l'approche basée sur un lexique alors qu'il était assez bas dans celle basée sur une machine d'apprentissage. Cette conclusion est une confirmation des études précédemment réalisées qui stipulaient que si la méthode basée sur un lexique est un peu plus facile d'utilisation, l'approche basée sur une machine d'apprentissage (avec le classificateur naïf de Bayes) est beaucoup plus efficace pour déterminer le sentiment général au sein d'un corpus de gazouillis.

L'opinion ou sentiment général qui se dégage est plus positif que négatif. Cela s'explique tout d'abord que dans la formule, les messages neutres sont regroupés avec les positifs. Mais cela s'explique surtout par le fait que la majorité des messages ont été postés par les propriétaires des comptes ou des utilisateurs qui sont apparentés à ces organismes en soins de santé (comme les fondations issues de ces grands hôpitaux). Or, il est évident que le propriétaire du compte ou un utilisateur apparenté ne peut écrire de messages de connotation négative, pouvant nuire à sa propre image. En outre, les messages postés suscitaient très peu de débats. Tout cela fait que l'analyse de sentiment semble quelque peu biaisé. Toutefois, l'objectif premier était de démontrer qu'il est possible d'extraire le sentiment général des gazouillis postés sur le compte de médias sociaux d'un organisme en soin de santé. Des études plus poussées et une stratégie

permettant de susciter plus de débats permettront de mesurer avec beaucoup plus de précision le sentiment général qui se dégage dans les gazouillis publiés et une corrélation avec l'appréciation de l'image ou de la qualité des services d'un organisme en soins de santé.

En second lieu, on a procédé à l'analyse d'influence. Pour déterminer l'influence, il a fallu définir les types de messages postés, parmi lesquels les *mentions*, les *retweet*, les *replies*. À l'aide d'une formule, on a pu déterminer le niveau ou taux d'engagement. Étant donné que pour un gazouillis publié sur le compte, il faudrait au moins une réponse sous la forme d'une *mention*, *retweet* ou *replie*, nous avons déterminé que le niveau d'engagement moyen est de 50 %. Les trois grands hôpitaux atteignent en moyenne ce score, avec un score légèrement au-dessous de la moyenne pour l'Hôpital Montréal pour enfants, alors que cet organisme en soins de santé présentait le plus grand ensemble de gazouillis collectés pendant la durée (soit 383). On remarque qu'il y a très peu d'interactions et la majorité des gazouillis sont créés par les comptes utilisateurs appartenant à ces organismes. Ce niveau d'engagement moyen est confirmé par l'autre mesure qui est la portée potentielle. La portée potentielle exprime les utilisateurs potentiels qu'on peut rejoindre pour d'éventuelles campagnes. Cette mesure évalue non seulement les utilisateurs réels qui suivent le compte d'un organisme en soins de santé, mais il prend également en compte les abonnés de ces utilisateurs. Le contraste est saisissant et l'écart entre le nombre d'utilisateurs réels et les nombres d'utilisateurs potentiels, est énorme. Par exemple, pour le CHUM, le nombre d'utilisateurs est de 56, alors que la portée potentielle (nombre d'utilisateurs potentiels) est de 176 667. Si ce dernier nombre peut être utilisé pour justifier une nouvelle campagne, il n'en demeure pas moins que l'écart fait conclure que les organismes en soins de santé devraient adopter de stratégies de recrutement plus efficaces.

La dernière analyse est celle du volume. Elle a permis d'identifier, pour chacun des comptes des grands hôpitaux, les utilisateurs les plus actifs et les plus influents. Les utilisateurs les plus actifs sont ceux qui publient le plus de messages sur les comptes appartenant aux organismes en soins de santé. Ces utilisateurs sont très importants car ils peuvent être utilisés comme public cible pour obtenir des opinions, des commentaires ou autre point de vue, de manière assez rapide. Aussi,

dans le cas où l'on veut susciter un ou plusieurs débats, les organismes en soins connaîtront quels sont les utilisateurs clés pour amplifier leur message.

L'analyse du volume a aussi permis d'identifier les utilisateurs les plus actifs et les influents. Les utilisateurs influents sont ceux qui ont un très grand nombre d'abonnés. En effet, même si ces utilisateurs ne sont pas les plus actifs, un message publié ou retransmis peut-être vu par des milliers d'autres utilisateurs. Parmi ces utilisateurs, il y avait des comptes appartenant à des grands groupes de presse. En temps de crise, ces utilisateurs joueraient un rôle très important de relais d'information. Par exemple, pendant les crises du virus H1N1 ou du virus Ébola, la population en général et la communauté présente sur les réseaux sociaux ont été très inquiètes et ont exprimés des commentaires plus ou moins vrais sur ces épidémies. En ciblant le message vers cet audience d'utilisateurs les plus influents, les organismes en soins de santé s'assurent l'utilisation d'une stratégie efficace de communication en temps de crise.

Enfin, en regardant les graphiques du volume de gazouillis publiés dans le temps, la tendance confirme le faible taux de participations vu dans les autres analyses et mesures. Ceci est dû au manque de débats suscités sur les comptes de réseaux sociaux de ces hôpitaux. Toutefois, lorsqu'un sujet suscite un débat, on a pu constater une augmentation significative du nombre de gazouillis publiés par jour. Le cas du Centre hospitalier de l'Université de Montréal l'illustre bien. En effet, le volume moyen de gazouillis publiés par jour durant la période d'expérimentation était d'environ cinq. Ce nombre a augmenté de manière significative, allant de 5 à 26 gazouillis publiés par jour. En examinant de plus près les messages postés, on a identifié qu'ils ont été émis lors de la démission du directeur général du CHUM, le Dr Turgeon. Ceci fait dire qu'en définitive, il y a un grand besoin d'adapter les messages postés afin qu'ils suscitent plus de réactions et d'engouement de la part de la communauté présente sur les médias sociaux.

En somme, les mesures effectuées et les résultats obtenus démontrent que les médias sociaux en général et *Twitter* en particulier peuvent servir de source de données permettant d'obtenir des indicateurs clés pour une meilleure prise de décision. Un conseil d'administration ou l'équipe marketing d'un organisme en soins de santé disposant de telles informations ayant de la valeur

peut prendre des décisions ou orienter leurs stratégies en se basant sur des résultats graphiques compréhensibles par la haute direction. Ces résultats peuvent servir d'explicatifs pour le retour sur l'investissement consentie pour être présent sur les médias sociaux et tirer profits des avantages qu'offrent ces plateformes.

Il est à noter le faible nombre d'étude qui porte sur l'analyse des médias sociaux en santé. Si la plupart des études ont lieu aux États-Unis, elles portent essentiellement sur d'autres domaines d'industries telles que la politique, le commerce électronique ou l'industrie cinématographique. Cette absence de littérature n'a pas permis d'effectuer une bonne comparaison avec nos résultats. Aussi, il serait intéressant de comparer les résultats d'une analyse de sentiments avec un sondage effectué, en temps de crise (lorsque le taux de participation aux débats est au haut). Ce serait un bon comparatif de l'efficacité des approches préconisées.

Enfin, il convient de définir un *framework* ou cadre de travail permettant de doter à tout organisme en soins de santé, les outils, la documentation et la formation nécessaire pour pouvoir adopter une stratégie de communication par les médias sociaux. Ce cadre de travail permettra aussi une analyse efficiente des leurs données issues des médias sociaux afin d'en tirer plusieurs informations utiles pour la prise de décision.

Liste des références

- [1] Kaplan, A. M. et Haenlein, M., *Users of the world unite! The challenges and opportunities of social media*, Business Horizons, Vol. 53 (No.1), 2009, p. 59-68.
- [2] O'Reilly, T., *What is Web 2.0: design patterns and business models for the next generation of software*, 2007. <<http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>>. Consulté le 3 octobre 2014.
- [3] Boyd, D. et Ellison, N., *Social Network Sites: Definition, History, and Scholarship*, Journal of Computer-Mediated Communication, 13 (1), 2007, article 11.
- [4] Vickery, G. et Wunsch-Vincent, S., *Participative Web And User-Created Content: Web 2.0 Wikis and Social Networking*, Organization for Economic Cooperation and Development (OECD) Paris, 2007.
- [5] Sokora, S., *Perceptions du public et couverture médiatique du système de santé canadien*. Fondation canadienne de la recherche sur les services de santé, Ottawa, Canada, 2001.
- [6] Akinci, F., Ezel Esatoglu, A. et Tengilimoglu, D. et Parsons A., *Hospital Choice Factors: A Case study in Turke*, Health Marketing Quarterly, Vol. 22 (No. 1), 2004, p. 3-19.
- [7] Javalgi, R., Rao, S. R. et Thomas, E. G., *Choosing a hospital: Analysis of Consumer Tradeoffs*, Journal of Health Care Marketing, Vol. 11 (No. 1), 1991, p. 12-22.
- [8] Javalgi, R., Whipple, T., McManamon, M. et Edick V., « *Hospital image: a correspondence analysis approach* », Journal of Health Care Marketing, Vol. 12 (No. 4), 1992, p. 34-41.

- [9] Wolinsky, F. D., *The Sociology of Health: Principles, Practitioners, and Issues*. Belmont (CA): Wadsworth Publishing Company, 1998.
- [10] Leister, J. et Strausberg, J., *Why do patients select a hospital?* Journal of Hospital Marketing & Public Relations, Vol. 17 (No. 2), 2007, p. 13-31.
- [11] Berendsen, A. J., de Jong, G. M., Schuling, J., Bosveld, H.E.P., de Waal M.W.M. et Mitchell G. K., *Patient's need for choice and information across the interface between primary and secondary care: a survey*. Patient Education and Counseling, Vol. 79 (No. 1), 2010, p. 100-105.
- [12] Centers for Disease Control and Prevention (CDC). *CDC's Guide to Writing for Social*. <<http://www.cdc.gov/socialmedia/tools/guidelines/pdf/guidetowritingforsocialmedia.pdf>>. Consulté le 05 mars 2014.
- [13] Schein, R., Wilson, K. et Keelen, J., *Literature review on effectiveness of the use of social media: A report for Peel Public Health*, 2010. <<http://www.peelregion.ca/health/resources/pdf/socialmedia.pdf>>. Consulté le 05 avril 2014.
- [14] Chunara, R., Freifeld, C. et Brownstein, J., *New technologies for reporting realtime emergent infections*. Parasitology. 139(4), 2012, p 1843-1851.
- [15] Friedrich, MJ., *Cholera and social media*. Journal of American Medical Association. 307(7), 2012, p 656.
- [16] Merchant, R., Elmer, S. et Lurie, N., *Integrating social media into emergency preparedness efforts*. New England Journal of Medicine. 365(4), 2011, p 289-91.
- [17] World Health Organization, *Mixed uptake of social media among public health specialists*. Bulletin of the World Health Organization. 89(11), 2011, p 784-5

- [18] Purdy, C. H., *Using the internet and social media to promote condom use in Turkey*. *Reprod Health Matters*.19 (37), 2011, p 157-65.
- [19] Keckley, P., *Social Networks in Health Care: Communication, collaboration and insights*. Deloitte Center for Health Solutions, 2011.
<https://www.ucsf.edu/sites/default/files/legacy_files/US_CHS_2010SocialNetworks_070710.pdf>. Consulté le 05 mars 2014
- [20] Bennett, W., *The Personalization of Politics Political Identity, Social Media, and Changing Patterns of Participation*. *The ANNALS of the American Academy of Political and Social Science*. 644, 2012, p 20-39.
- [21] Eckler, P., Worsowicz, G. et Rayburn, W., *Social Media and Health Care: An Overview*. *Practice Management*, 1046-1049, 2010, p 265.
- [22] Lura Consulting, *Social Media Plan: Best Practices Overview*. Simcoe Muskoka District Health Unit, 2010 .
<http://dvqlxo2m2q99q.cloudfront.net/000_clients/53739/file/smdhu-socialmedia-plan-best-practices-overview.pdf>. Consulté le 05 03 2014.
- [23] Duong, K. et Flosi S., *Canada digital future in focus 2013*, 2013
<http://www.comscore.com/Insights/Presentations_and_Whitepapers/2013/2013_Canada_Digital_Future_in_Focus>. Consulté le 05 février 2015.
- [24] Thackeray, R., Neiger, BL. et Keller, H., *Integrating social media and social marketing: a four-step process*. *Health Promotion Practice*;13(2), 2012, p165-168.
- [25] Dictionnaires de français LAROUSSE.
<<http://www.larousse.fr/dictionnaires/francais/sentiment/72138?q=sentiment#71335>>
Consulté le 03 mars 2015.

- [26] Pang, B. et Lee, L., *Opinion mining and sentiment analysis*. Foundations and Trends in Information Retrieval, 2(1-2), 2008, p1-135.
- [27] Moreno-Ortiz, A. et Pérez Hernández, C., *Lexicon-Based Sentiment Analysis of Twitter Messages in Spanish*, 2012, <<http://www.daedalus.es/TASS>>. Consulté le [05-03-2014].
- [28] Wiebe, J., *Instructions for annotating opinions in newspaper articles*. Departement of Computer Science Technical Report TR-02-101, University of Pittsburgh, Pittsburgh, 2002.
- [29] Wiebe, J., Wilson, T. et Cardie, C., *Annotating expressions of opinions and emotions in language*. Language Resources and Evaluation (formerly Computers and the Humanities), 39, 2005, p164-210.
- [30] Tumasjan, A., *Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment*, Proceedings of the Fourth AAAI conference on Weblogs and Social Media, 2010.
- [31] Qiu, G., He, X., Zhang, F., Shi, Y., Bu J. et Chen C., *DASA: dissatisfaction-oriented advertising based on sentiment analysis*. Expert Syst Appl, 37, 2010, p. 6182–6191.
- [32] Hatzivassiloglou, V., McKeown, K., *Predicting the semantic orientation of adjectives*. In: *Proceedings of annual meeting of the Association for Computational Linguistics*. (ACL'97); 1997.
- [33] Segaran, T., *Programming Collective Intelligence*. Sebastopol: O'Reilly Media, Inc, 2007.
- [34] Jurafsky D., Martin, J.H., *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. New Jersey: Pearson Education, Inc, 2009.
- [35] Van't Ende, R., *Sentiment analyse van gazouillis: de correlatie tussen de merkbeoordeling op Twitter en de werkelijke merbeoordeling van consumenten*. Master thesis, Tilburg University, 2013.

- [36] Schmidt, J., *Social media metrics for health care*. Social Media and Health. 2011.
- [37] Pak, A., et Paroubek P., *Twitter as corpus for sentiment analysis and opinion mining*. Proceeding of LREC, 2010.
- [38] Keckley, P. H., Hoffman, M., *Social Networks in Health Care: Communication, Collaboration and Insights*. Deloitte, 2010.

Bibliographie

Fortin, M. *Fondements et étapes du processus de recherche*. Chenelière
Éducation, Montréal, 2010

Annexe I

API Twitter

Twitter est un réseau social *Open source* qui a plusieurs *API* ou interfaces de programmation qui permettent l'utilisation d'applications tierces. Une application tierce est un programme non écrit par le programme principal qui ajoute des fonctionnalités. *Twitter* impose des limites sur les *API* pour protéger ses utilisateurs, mais, il est très clair que les *API* peuvent être utilisées. Plusieurs applications disponibles qui utilisent des *API Twitter*, mais la plupart des applications sont développées afin que les utilisateurs aient directement accès aux gazouillis. Une application peut être installée sur l'ordinateur ou appareil mobile. Une fois que l'application est définie par un utilisateur, son compte *Twitter* est lié à l'application et l'utilisateur n'a plus à se connecter à *Twitter* de lire ou de répondre aux gazouillis. La manipulation des gazouillis et Retweet se fait via une application tierce. En utilisant les *API Twitter*, une application peut être développée pour enregistrer les gazouillis sur un sujet donné.

Procédure d'enregistrement de l'application pour obtenir les clés

Il faut créer une application sur le site des développeurs de twitter, générer quatre valeurs-clés appelées *Consumer key*, *Consumer secret*, *Access token*, et *Access token secret*. On va ensuite utiliser ces quatre valeurs pour authentifier et envoyer nos requêtes en *PHP* grâce à une classe.

S'enregistrer auprès de Twitter

Nos requêtes doivent être authentifiées via *OAuth*, cela implique que l'on crée un client identifiable auprès de *Twitter*.

C'est une action qu'on effectuera une seule fois par site ou application sur lesquels on souhaite faire des requêtes Twitter. Cela se passe sur le site de Twitter dédié aux développeurs, à l'adresse <https://dev.twitter.com>. Connectez-vous avec votre compte Twitter:

Sign in with your Twitter account

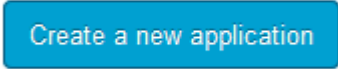
Please log in to access that page.



A login form with two input fields. The first field is labeled 'Username:' and contains the text 'NOE_interactive'. Below it is a link that says 'New to Twitter? Sign up!'. The second field is labeled 'Password:' and contains a series of dots. Below the fields is a blue button labeled 'Log in'.

Une fois connecté, rendez-vous sur la page "*My applications*", soit en cliquant sur dev.twitter.com/apps, ou en y accédant via le menu déroulant qui s'affiche au survol de votre image de profil.

Cette page liste les applications que vous avez déjà créées. Nous allons en créer une nouvelle, en cliquant sur le bouton "*Create a new application*":



Un formulaire vous est présenté. Précisez le nom de votre application et sa description. Dans le cadre d'un site personnel par exemple, vous pouvez ici donner simplement votre nom dans le champ '*Application Name*'. Pour les champs '*Website*' et '*OAuth redirect_url*', mettez simplement l'*url* de votre site web dans les deux cas en incluant `http://`

Create an application

Application Details

Name: *

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description: *

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website: *

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens.
(If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL:

Where should we return after successfully authenticating? For [@Anywhere applications](#), only the domain specified in the callback will be used. [OAuth 1.0a](#) applications should explicitly specify their `oauth_callback` URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

Developer Rules Of The Road

Last Update: September 5, 2012.

Rules of the Road

Twitter maintains an open platform that supports the millions of people around the world who are sharing and discovering what's happening now. We want to empower our ecosystem partners to build valuable businesses around the information flowing through Twitter. At the same time, we aim to

Une fois le formulaire envoyé, vous serez sur la page qui détaille les informations de votre application.

Ce qui nous intéresse particulièrement, c'est la section *OAuth settings*, car nous avons besoin des valeurs *Consumer key* et *Consumer secret* pour la suite. Gardez donc ces informations à portée de main.

Your access token

Use the access token string as your "oauth_token" and the access token secret as your "oauth_token_secret" to sign requests with your own Twitter account. Do not share your oauth_token_secret with anyone.

Access token	ddddddd-xxxXXxxxxXXxxXXXXXXXXxxxxXXXxxxXXxxxxXX
Access token secret	XXxxXXxxXXXXxxxxxxXXxxXXxxXXxxXXxxXXxxXXxxXXxx
Access level	Read-only

[Recreate my access token](#)

Nous allons avoir besoin de l'*access token* et de l'*access token secret* pour la suite, alors gardez cette page à portée de main.

Intégration

Pour autoriser notre application à faire des requêtes à l'API, nous avons besoin de quatre valeurs. Deux pour l'application que nous utilisons: *Consumer key* et *Consumer secret*, et deux pour l'utilisateur qui utilise l'application *access token* et *access token secret*.

Annexe II

API Datumbox

How to build your own Twitter Sentiment Analysis Tool (posté sur le blog)

September 2, 2013 Vasilis Vryniotis . 23 Comments

Machine Learning & Statistics Online Marketing Programming

Twitter Sentiment AnalysisIn this article we will show how you can build a simple Sentiment Analysis tool which classifies gazouillis as positive, negative or neutral by using the Twitter REST API 1.1v and the Datumbox API 1.0v. Even though the examples will be given in PHP, you can very easily build your own tools in the computer language of your choice.

Update: The Datumbox Machine Learning Framework is now open-source and free to download. If you want to build a Sentiment Analysis classifier without hitting the API limitations, use the `com.datumbox.applications.nlp.TextClassifier` class.

You can find the complete PHP code of the Twitter Sentiment Analysis tool on Github.

Social Media Monitoring & Sentiment Analysis

Social Media Monitoring is one of the hottest topics nowadays. As more and more companies use Social Media Marketing to promote their brands, it became necessary for them to be able to evaluate the effectiveness of their campaigns.

Building a Social Media Monitoring tool requires at least 2 modules: one that evaluates how many people are influenced by the campaign and one that finds out what people think about the brand.

Evaluating the generated buzz is usually performed by using various KPIs such as the number of abonnés/friends, the number of likes/shares/RTs per post and more complex ones such as the

engagement rate, the response rate and other composite metrics. Quantifying the buzz is usually straightforward and can be performed by using basic statistics.

On the other hand, being able to evaluate the opinion of the users is not a trivial matter. Evaluating their opinions requires performing Sentiment Analysis, which is the task of identifying automatically the polarity, the subjectivity and the emotional states of particular document or sentence. It requires using Machine Learning and Natural Language Processing techniques and this is where most of the developers hit the wall when they try to build their own tools.

Thankfully Datumbox simplifies the process of using Machine Learning since it offers several API functions which allow you to build custom Social Media Monitoring tools in no time. Some of the services that are available to the API users are the Sentiment Analysis, the Twitter Sentiment Analysis and the Subjectivity Analysis API functions. In this article we will focus only on the Twitter Sentiment Analysis method, nevertheless as you can easily find out the rest of the functions work similarly.

Performing Sentiment Analysis on Twitter

Performing Sentiment Analysis on Twitter is trickier than doing it for large reviews. This is because the gazouillis are very short (only about 140 characters) and usually contain slangs, emoticons, hash tags and other twitter specific jargon. This is the reason why Datumbox offers a completely different classifier for performing Sentiment Analysis on Twitter.

Building the Sentiment Analysis tool

In order to build the Sentiment Analysis tool we will need 2 things: First of all be able to connect on Twitter and search for gazouillis that contain a particular keyword. Second evaluate the polarity (positive, negative or neutral) of the gazouillis based on their words. For the first task we will use the Twitter REST API 1.1v and for the second the Datumbox API 1.0v.

To speed up the development we will use 2 classes: The great PHP-Twitter-API client written by Tim Whitlock and the Datumbox PHP-API-Client offered by our service. As you will soon find out getting the gazouillis via the Twitter API is the most complicated task of this tutorial.

Create your own Twitter Application

Unfortunately Twitter made it more complicated for developers to use their API. In order to be able to search for particular gazouillis you must authenticate yourself by using OAuth protocol. Fortunately the API client of Tim takes care most of the tasks and enables a fast and easy integration. Still you are required to create a new Twitter application before using the library.

So go to Twitter Applications Console, login by using your credentials, click on “Create new Application” button and fill in the form to register a new app. When you create it select the application and go to the “Details” tab (the first tab) and on the bottom of the page click the “Create my access token” button. Once you do this, go to the “OAuth tool” tab and note down the values: Consumer Key, Consumer secret, Access token and Access token secret.

Get your Datumbox API key

To access the Datumbox API sign up for a free account and visit your API Credentials panel to get your API Key.

Developing the Twitter Sentiment Analysis class

All we need to do in order to develop the tool is write a TwitterSentimentAnalysis class which uses the Twitter and Datumbox API Clients to fetch the gazouillis and evaluate their polarity.

Below you can see the code along with the necessary comments.

```
<?php  
class TwitterSentimentAnalysis {
```

```

protected $datumbox_api_key; //Your Datumbox API Key. Get it from http://www.datumbox.com/apikeys/view/

protected $consumer_key; //Your Twitter Consumer Key. Get it from https://dev.twitter.com/apps
protected $consumer_secret; //Your Twitter Consumer Secret. Get it from https://dev.twitter.com/apps
protected $access_key; //Your Twitter Access Key. Get it from https://dev.twitter.com/apps
protected $access_secret; //Your Twitter Access Secret. Get it from https://dev.twitter.com/apps

/**
 * The constructor of the class
 *
 * @param string $datumbox_api_key Your Datumbox API Key
 * @param string $consumer_key Your Twitter Consumer Key
 * @param string $consumer_secret Your Twitter Consumer Secret
 * @param string $access_key Your Twitter Access Key
 * @param string $access_secret Your Twitter Access Secret
 *
 * @return TwitterSentimentAnalysis
 */
public function __construct($datumbox_api_key, $consumer_key, $consumer_secret, $access_key, $access_secret){
    $this->datumbox_api_key=$datumbox_api_key;

    $this->consumer_key=$consumer_key;
    $this->consumer_secret=$consumer_secret;
    $this->access_key=$access_key;
    $this->access_secret=$access_secret;
}

/**
 * This function fetches the twitter list and evaluates their sentiment
 *
 * @param array $twitterSearchParams The Twitter Search Parameters that are passed to Twitter API. Read more
here https://dev.twitter.com/docs/api/1.1/get/search/gazouillis
 *
 * @return array
 */
public function sentimentAnalysis($twitterSearchParams) {
    $gazouillis=$this->getGazouillis($twitterSearchParams);

    return $this->findSentiment($gazouillis);
}

/**
 * Calls the Search/gazouillis method of the Twitter API for particular Twitter Search Parameters and returns the list of
gazouillis that match the search criteria.
 *
 * @param mixed $twitterSearchParams The Twitter Search Parameters that are passed to Twitter API. Read more
here https://dev.twitter.com/docs/api/1.1/get/search/gazouillis
 *
 * @return array $gazouillis
 */
protected function getGazouillis($twitterSearchParams) {
    $Client = new TwitterApiClient(); //Use the TwitterAPIClient

```

```

$Client->set_oauth ($this->consumer_key, $this->consumer_secret, $this->access_key, $this->access_secret);

$gazouillis = $Client->call('search/gazouillis', $twitterSearchParams, 'GET' ); //call the service and get the list of
gazouillis

unset($Client);

return $gazouillis;
}

protected function findSentiment($gazouillis) {
    $DatumboxAPI = new DatumboxAPI($this->datumbox_api_key); //initialize the DatumboxAPI client

    $results=array();
    foreach($gazouillis['statuses'] as $gazouillis) { //foreach of the gazouillis that we received
        if(isset($gazouillis['metadata']['iso_language_code']) && $gazouillis['metadata']['iso_language_code']=='en')
        { //perform sentiment analysis only for the English Gazouillis
            $sentiment=$DatumboxAPI->TwitterSentimentAnalysis($gazouillis['text']); //call Datumbox service to get the
            sentiment

            if($sentiment!=false) { //if the sentiment is not false, the API call was successful.
                $results[]=array( //add the gazouillis message in the results
                    'id'=>$gazouillis['id_str'],
                    'user'=>$gazouillis['user']['name'],
                    'text'=>$gazouillis['text'],
                    'url'=>'https://twitter.com/'.$gazouillis['user']['name'].'/'status/'.$gazouillis['id_str'],

                    'sentiment'=>$sentiment,
                );
            }
        }
    }

    unset($gazouillis);
    unset($DatumboxAPI);

    return $results;
}
?>

```

What we do is pass to the constructor the necessary keys for all the services. Then on the public sentimentAnalysis function we first call Twitter service in order to get the list of gazouillis which much our search parameters and then we call for each gazouillis the Datumbox service to get is polarity.

This is it! You ready to use this class to perform Sentiment Analysis on gazouillis and build your own Social Media Monitoring tool. You can find the complete PHP code of the Twitter Sentiment Analysis tool on Github.

Extra: Detailed Information about the Twitter Sentiment Analysis Classifier

This part is optional for those of you who are interested in learning how Datumbox's Twitter Sentiment Analysis works.

In order to detect the Sentiment of the gazouillis we used our Machine Learning framework to build a classifier capable of detecting Positive, Negative and Neutral gazouillis. Our training set consisted of 1.2 million gazouillis evenly distributed across the 3 categories. We tokenized the gazouillis by extracting their bigrams and by taking into account the URLs, the hash tags, the usernames and the emoticons.

In order to select the best features we used several different algorithms and at the end we chose the Mutual Information. Finally after performing several tests with various models and configurations we selected the Binarized Naïve Bayes as the best performing classifier for the particular problem (strangely enough Naïve Bayes beat SVM, Max Entropy and other classifiers which are known to perform usually better than NB). To evaluate the results we used the 10-fold cross-validation method and our best performing classifier achieves an accuracy of 83.26%.

Annexe III

Code source de l'application ASTAS et données collectées avec Excel

Le code source en *PHP*, *Javascript*, *HTML* et *SQL* formant le contenu de l'application ASTAS ainsi que l'ensemble de données collectées (disponibles sous format *Excel*), sont sauvegardées sur la plateforme Web *Git*. L'adresse est la suivante :

<https://github.com/mbasserge/astas>

Annexe IV

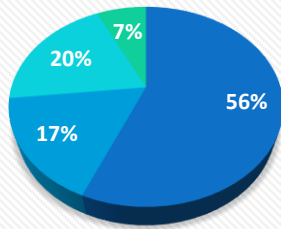
Résultats de l'analyse des sentiments

Résultats avec l'approche basée sur une machine d'apprentissage :

	Nb positifs	Nb négatifs	Nb neutres	Nb aucun	Sentiment
CHUM	40	12	14	5	59,15%
CHU Ste-Justine	32	9	29	10	65%
Hôpital Montréal pour enfant	186	71	107	19	57,96%

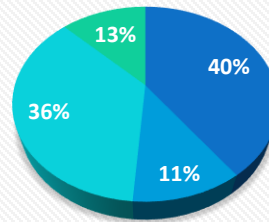
Tableau 15.Sentiments approche basée sur une machine d'apprentissage

Sentiments basés sur une machine d'apprentissage, pour le CHUM



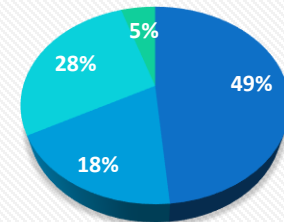
■ positives ■ negatives ■ neutres ■ nones

Sentiments basés sur une machine d'apprentissage, pour le CHU Ste-Justine



■ positives ■ negatives ■ neutres ■ nones

Sentiments basés sur une machine d'apprentissage, pour Hopital Children



■ positives ■ negatives ■ neutres ■ nones

Figure 21. Sentiments avec l'approche basée sur une machine d'apprentissage

Résultats avec l'approche basée sur un lexique :

	Nb positifs	Nb négatifs	Nb neutres	Nb None	Sentiment
CHUM	17	11	28	15	47,88%
CHU Ste-Justine	15	10	35	21	50%
Hôpital Montréal pour enfant	99	16	148	120	60,31%

Tableau 16.Sentiments avec approche basée sur un lexique

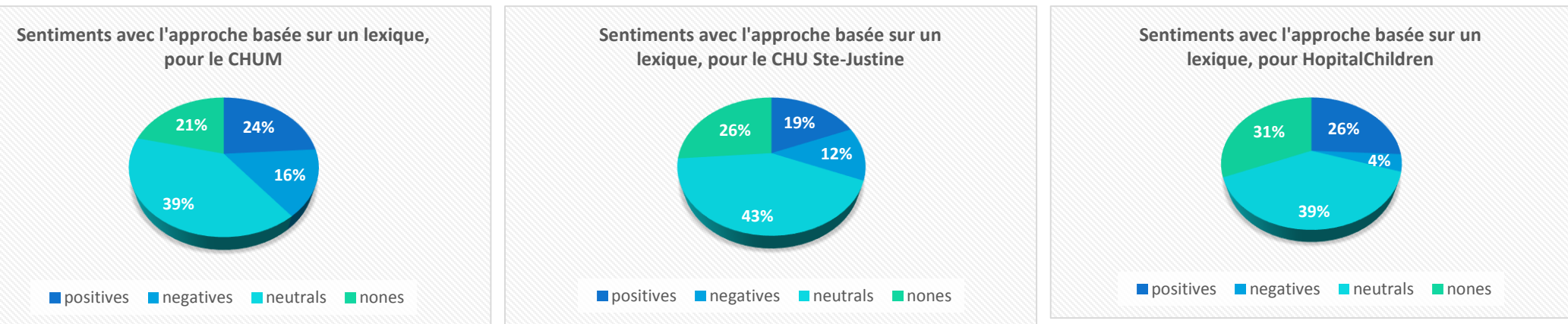


Figure 22.Sentiments avec l'approche basée sur un lexique

