

# #AIDS Analyse Information Dangers Sexualité : caractériser les discours à propos du VIH dans les forums de santé

Yves Mercadier<sup>1</sup>, Jérôme Azé<sup>1</sup>, Sandra Bringay<sup>1,2</sup>, Viviane Clavier<sup>3</sup>,  
Erick Cuenca<sup>1</sup>, Céline Paganelli<sup>4</sup>, Pascal Poncelet<sup>1</sup>, Arnaud Sallaberry<sup>1,2</sup>

<sup>1</sup> LIRMM, Université de Montpellier, CNRS, Montpellier, France, [prenom.nom@lirmm.fr](mailto:prenom.nom@lirmm.fr)

<sup>2</sup> AMIS, Université Paul Valéry, Montpellier, France

<sup>3</sup> GRESEC EA 608 — Université Grenoble Alpes, France

<sup>4</sup> LERASS EA 827, Université Paul Valéry, F34000, Montpellier, France

**Résumé :** **Contexte :** Les forums de discussion consacrés au VIH remplissent trois fonctions. Tout d’abord, ils sont utilisés par les internautes comme sources d’information pour se renseigner sur les traitements, les risques d’infection liés au sida ou le « vivre avec » la maladie ou le virus. Ensuite, ils servent de média pour les institutions de santé ou les associations afin de communiquer des informations de prévention sur le VIH. Enfin, ils apparaissent comme sources de connaissances pour les professionnels de santé (médecins, soignants) pour comprendre les comportements liés au VIH, et pour les professionnels de la prévention, pour modérer les forums et adapter les campagnes de prévention aux différents publics.

**Objectif :** Notre objectif dans cet article est de proposer un processus d’ingénierie des connaissances complet, permettant : 1) de collecter des messages dans des médias sociaux ; 2) de classer ces messages de manière semi-automatique selon le genre, le niveau d’expertise, le niveau informationnel, le type de risque ainsi que les émotions exprimées ; 3) de visualiser ces nouvelles connaissances dans une représentation originale prenant en compte la temporalité et la hiérarchie sous-jacente à la classification. Cette visualisation pourrait permettre aux gestionnaires de sites de forums et aux professionnels de santé de naviguer dans le flot de messages pour suivre l’évolution de l’importance de ces thématiques.

**Méthodes :** Notre approche combine une démarche qualitative et quantitative. Nous avons travaillé sur plus de 226 252 messages issus du forum Sida Info Service. Deux chercheurs en sciences de l’information et de la communication ont élaboré une grille d’analyse, puis ont annoté 4 481 messages selon cette grille. Ces données ont été utilisées pour apprendre des classificateurs qui ont permis d’étiqueter l’ensemble des messages du site à notre disposition. Afin de définir les meilleurs classificateurs, nous avons comparé l’efficacité des méthodes de classification traditionnelles statistiques et plusieurs architectures d’apprentissage profond. Une fois les messages étiquetés, nous avons utilisé une visualisation de type *streamgraph*, combinée avec un outil de navigation hiérarchique, pour visualiser l’évolution de ces annotations dans le temps.

**Résultats :** Les résultats sont prometteurs et montrent l’efficacité des méthodes d’apprentissage profond pour caractériser les messages des forums de manière automatique. La méthode de visualisation mise en place permet d’explorer les résultats de ces méthodes et ainsi faciliter l’accès aux connaissances.

**Mots-clés :** Classification automatique de textes, Apprentissage profond, Visualisation, VIH

## 1 Introduction

À l’heure où 3,9 milliards d’individus dans le monde sont connectés à internet (soit 51% de la population mondiale), et où 2,91 milliards d’entre eux sont inscrits sur des réseaux sociaux (soit 29% de la population mondiale)<sup>1</sup>, les pratiques pour s’informer et communiquer ont considérablement évolué et sont marquées par les innovations technologiques récentes et la place occupée par le numérique. En 2017, 74% des français accèdent tous les jours à Internet (95% des 18-24 ans) et passent en moyenne 1 heure 15 par jour sur les réseaux sociaux.

Les travaux présentés dans cet article se déroulent dans le cadre du projet #AIDS (Analyse Information Dangers Sexualité) qui réunit une équipe de chercheurs pluridisciplinaire en sciences de l’information et de la communication et en informatique, ainsi que des acteurs de la santé et de la prévention. Ce projet vise à analyser semi-automatiquement les contenus

---

1. <https://www.blogdumoderateur.com/chiffres-reseaux-sociaux/>

échangés sur les médias sociaux (forums de discussion, réseaux sociaux, plateformes de microblogging) consacrés au VIH pour en extraire de la connaissance. Avec plus de 35 millions de morts à ce jour, le VIH continue de représenter un problème mondial majeur de santé publique. En 2016, 1 million de personnes sont décédées d'une ou des causes liées au VIH dans le monde<sup>2</sup>. Toujours en 2016, en France, 6 003 personnes ont découvert leur séropositivité.

Les médias sociaux consacrés aux VIH s'adressent principalement aux personnes infectées et aux malades dans une démarche de prévention ou d'accompagnement mais également aux internautes soucieux de connaître les risques de contamination. Les publications peuvent prendre plusieurs formes comme des dispositifs de prévention, d'accompagnement, d'amélioration de la santé mentale et physique, des espaces d'échanges où sont évoqués les traitements, le "vivre avec" la maladie ou le virus, et plus largement des questions de la vie de tous les jours ou encore donner lieu à des usages sexuels d'Internet. Un grand nombre de ressources en ligne sont disponibles, à toute heure. De nombreux blogs comme *Je VIH avec*<sup>3</sup> alimentés par des médecins, des porteurs du virus, des associations de malades sont présents sur le web. De nombreux groupes de discussion sont consacrés à ces thématiques sur des réseaux sociaux destinés aux patients comme Carenity<sup>4</sup> ou encore sur le site généraliste Doctissimo<sup>5</sup>. On dénombre, au 1er mars 2018, une centaine de groupes facebook français en lien avec le VIH. Plus de 500 comptes twitter mentionnent les hashtags #VIH ou #SIDA dans l'intitulé de leur compte, au 1er mars 2018. Ces comptes émanent d'associations, de particuliers, d'institutions ou de médecins. Ils portent pour la plupart sur les traitements ou sur la vie quotidienne avec le VIH. Dans ce travail, nous nous focaliserons sur les messages qui nous ont été fournis par les gestionnaires du forum associé au site Sida Info Service (SIS)<sup>6</sup>.

Si les médias sociaux ont largement été étudiés, peu de travaux ont jusqu'à présent été menés sur les médias sociaux consacrés au VIH et en particuliers les forums de santé. L'originalité de ce travail est de considérer, d'une part ces médias comme des dispositifs informationnels sur lesquels se développent de nouvelles formes d'échanges et qui véhiculent des informations n'ayant pas leur place dans d'autres espaces (cabinet médical, discussions familiales, médias traditionnels...). D'autre part, les discours produits sur ces médias reflètent des pratiques ordinaires relatives aux questions de sexualité, de consommations de produits illicites et d'addictions et permettent de détecter des pratiques émergentes, de nouvelles conduites à risque. **Notre objectif dans cet article est de proposer une chaîne de traitements incluant une phase de classification et un outil de visualisation pour mieux analyser et comprendre les contenus de ces discours. Ces informations sont essentielles pour la modération du site.**

Notre approche combine une démarche qualitative et quantitative. Nous avons travaillé sur plus de 226 252 messages issus du forum SIS. Deux professionnels de l'information et de la communication ont élaboré une grille d'annotation, puis ont annoté 4 481 messages selon cette grille. Ces données ont été utilisées pour apprendre des classifieurs supervisés qui ont permis d'étiqueter dans un deuxième temps, les 226 252 messages de manière automatique. Une fois les messages étiquetés, nous avons utilisé une approche *MultiStream* (Cuenca *et al.*, 2018) adaptée aux séries temporelles multiples. L'originalité est ici de pouvoir visualiser simultanément la hiérarchie associée aux annotations et leur évolution dans le temps.

Concernant la phase de classification automatique de textes il s'agit d'un sujet classique de Traitement Automatique de la Langue (TAL), qui consiste à assigner des catégories prédéfinies à des documents en texte libre. Les approches de classification classiques se concentrent sur le choix du meilleur classifieur (e.g. SVM ou régression logistique) et sur la définition des meilleures caractéristiques prises en entrée de ces classifieurs. La plupart des techniques sont

2. <http://www.who.int/mediacentre/factsheets/fs360/fr/>

3. <http://je-vih-avec.over-blog.com/>

4. <https://www.carenity.com/>

5. <http://www.doctissimo.fr/>

6. <https://www.sida-info-service.org/>

basées sur les mots, les lexiques et sont spécifiques à une tâche particulière. Ces modèles ont été appliqués avec succès sur de très hautes caractéristiques dimensionnelles parfois éparses. Dernièrement, pour de nombreuses tâches de classification de textes, les méthodes d'apprentissage profond se sont révélées efficaces et ont permis de faire des progrès importants dans des domaines tels que la reconnaissance de formes (pattern recognition) (Baccouche *et al.*, 2011) ou la bio-informatique (Min *et al.*, 2016). Cette tendance s'est confirmée avec le succès des word embeddings (Mikolov *et al.*, 2010, 2013) et des méthodes d'apprentissage profond (Socher *et al.*, 2013). L'apprentissage profond permet l'apprentissage automatique de représentations à plusieurs niveaux. (Collobert *et al.*, 2011) ont démontré qu'une architecture d'apprentissage profond, même simple, surpasse la plupart des approches classiques pour des tâches variées de TAL telles que la reconnaissance d'entités nommées (NER) (Joi, 2015), le Parsing (Vinyals *et al.*, 2015; Zhu *et al.*, 2013), l'étiquetage de rôles sémantiques (SRL) (Semantic Role Labeling) (He *et al.*, 2017; Srivastava *et al.*, 2015), le marquage POS (Andor *et al.*, 2016; Kumar *et al.*, 2016), la classification de sentiments (Rosenthal *et al.*, 2015; Nakov *et al.*, 2016; Kalchbrenner *et al.*, 2014; Kim, 2014), la traduction automatique (Sukhbaatar *et al.*, 2015). Depuis, de nombreux algorithmes complexes d'apprentissage profond ont été proposés pour résoudre ces tâches difficiles. **Dans cet article, nous allons entre autres comparer l'efficacité des méthodes de classification traditionnelles et plusieurs architectures d'apprentissage profond.**

De nombreux travaux en visualisation portent sur la représentation de données textuelles<sup>7</sup> (Kucher & Kerren, 2015) et/ou temporelles<sup>8</sup> (Aigner *et al.*, 2011). Une approche prometteuse, permettant de représenter des données temporelles issues de textes, a initialement été proposée par (Havre *et al.*, 2000) sous le nom de *ThemeRiver*. Généralisée sous le nom de *StreamGraphs*, cette approche consiste à superposer des flots représentant l'évolution d'une valeur numérique (le plus souvent issue de textes) dans le temps (Byron & Wattenberg, 2008). De nombreuses évolutions ont été proposées, e.g. (Cui *et al.*, 2011; Sun *et al.*, 2014; Wu *et al.*, 2014). La principale limite de ces approches est la difficulté de représenter un grand nombre de flots et un long intervalle de temps. La méthode utilisée dans cet article appelée *MultiStream* (Cuenca *et al.*, 2018), pallie ces problèmes en décrivant une approche de type focus+contexte combinant un ensemble de vues interactives. **Dans cet article, nous montrons comment cette visualisation aide le chercheur en science de l'information et de la communication à analyser l'ensemble des messages du site.**

L'article est organisé en quatre sections. Dans la section 2, nous présentons notre processus. Dans la section 3, nous présentons l'évaluation de la chaîne de traitements proposée. Dans la section 4, nous discutons ces résultats et ouvrons des perspectives.

## 2 Méthodes

La figure 1 illustre la méthode proposée, structurée en 4 étapes : (1) Collecte et nettoyage de données, (2) Annotation manuelle par deux experts, (3) Prédiction automatique de catégories, (4) Visualisation temporelle et hiérarchique de ces catégories.

### 2.1 Étape 1 : Collection et nettoyage des données

Nous avons effectué une analyse de contenu de 226 252 messages en français postés sur le forum SIS qui nous ont été fournis par les gestionnaires du forum<sup>9</sup>. Ces messages ont été anonymisés en supprimant les pseudonymes, noms, prénoms et localités. Puis les prétraitements suivants ont été appliqués : suppression des ponctuations et des caractères spéciaux,

7. <http://textvis.lnu.se/>

8. <https://vcg.informatik.uni-rostock.de/~ct/timeviz/timeviz.html>

9. Accord CNIL, Certificat d'enregistrement 2-17031

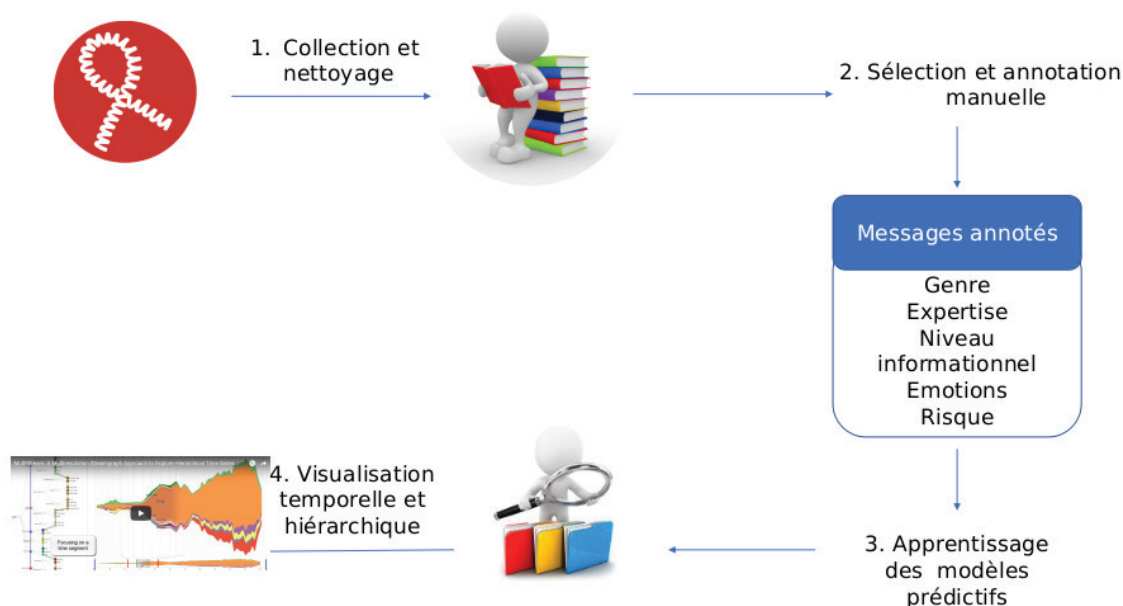


FIGURE 1 – Processus en 4 étapes

changement des majuscules en minuscules, suppression des mots trop fréquents pour apporter de l'information à partir d'une liste de stop words et enfin stématisation consistant à ne conserver que le radical des mots, pour regrouper sous le même radical tous les mots d'une famille.

## 2.2 Étape 2 : Annotation manuelle

Deux chercheurs en sciences de l'information et de la communication impliqués dans le projet, ont élaboré une grille d'annotation permettant de caractériser les types d'informations véhiculées par les messages postés sur les médias sociaux à partir d'une typologie testée précédemment sur des forums de discussion de santé et en tenant compte de la temporalité de l'écriture, des thématiques abordées, des micro-échanges (Paganelli & Clavier, 2014; Clavier & Paganelli, 2017). 4 481 messages ont été annotés manuellement par les deux experts<sup>10</sup> (sans croisement des annotations). Les chercheurs ont systématiquement annoté l'un des forums général de SIS (le plus volumineux), qui comporte plusieurs dizaines de fils de discussion, sans faire de sélection *a priori* des thématiques. La grille s'organise selon 5 dimensions :

1. le **genre** de l'auteur décliné en femme et homme ;
2. le **niveau d'expertise** de l'auteur. On distingue les messages des modérateurs, correspondant à des savoirs, de ceux des patients, qui parfois citent leur source ou rapportent des discours ;
3. le **niveau informationnel** du message. Les messages peuvent correspondre à des partages (d'expérience ou de connaissances), à des questions formulées explicitement ou non, à des témoignages faisant référence au vécu, à l'expérience, à des informations spécialisées (qui rapportent notamment des discours professionnels) ou scientifique (contenu supposé objectif dont la validité ne s'appuie pas sur le vécu), à des conseils (informations pratiques, procédures, protocoles) ;
4. les **émotions** exprimées par les participants en considérant un spectre émotionnel : la panique, l'angoisse, la peur, l'inquiétude, la réassurance (le fait de rassurer quelqu'un) ou le fait d'être rassuré.

10. Avec l'aide de Amal Jaouzi, stagiaire, doctorante en sciences de l'information et la communication.

5. le **risque**. Un message peut faire référence à un risque encouru par les internautes lors d'un événement passé, présent ou à venir, déclencheur d'une sensation de prise de risque. Les types de situation qui déclenchent une sensation de risque n'ont pas été annotés compte tenu du fait que ces situations sont bien connues des associations et figurent dans la rubrique « Questions fréquentes ». En revanche, nous avons laissé la possibilité à l'annotateur d'introduire un commentaire libre sur l'éventuelle prise d'alcool et de drogue durant l'événement déclencheur.

Un message peut donc être décrit selon ces 5 dimensions. Dans la Table 1, nous résumons la liste des classes définies après la phase d'annotation manuelle, ainsi que le nombre de messages annotés avec ces étiquettes. Chaque message peut être associé à zéro, une ou plusieurs classes. Ces classes seront utilisées comme sorties pour la prédiction de catégories.

TABLE 1 – Sortie de la prédiction

Index	Dimensions d'analyse	Classe	Effectif
C1	genre	Femme	36
C2	genre	Homme	67
C3	expertise	modérateur	46
C4	expertise	savoir patient	222
C5	expertise	sources citées	19
C6	expertise	discours rapporté	8
C7	information	demande	402
C8	information	informations scientifiques ou pratiques	661
C9	information	apport-réponse	874
C10	information	témoignage	284
C11	information	information vide	1144
C12	information	conseils	159
C13	émotion	peur	66
C14	émotion	inquiétude	109
C15	émotion	panique	34
C16	émotion	général	28
C17	émotion	angoisse	142
C18	émotion	réassurance	57
C19	émotion	rassuré	22
C20	risque	anticipé	8
C21	risque	passé	92
C22	risque	actuel	1

### 2.3 Étape 3 : Modèles pour la prédiction de catégories

Dans cette étude, nous explorons l'utilisation de méthodes d'apprentissage profond pour prédire les 22 catégories des messages collectés. Nous avons comparé les résultats des modèles d'apprentissage profond aux méthodes de classification traditionnelles. Dans cette section, nous détaillons comment chaque modèle est construit.

**Modèles.** Les modèles utilisés dans notre étude sont résumés dans le tableau 2. Les modèles d'apprentissages classiques comparés sont de type SVM (LSVC, SSVC), Naïve Bayes (MNB, BNB), Régression linéaire (SGDC), Arbre de décision (DT, AB, RF), K-plus-proches-voisins (KNC, PA). Dans ce qui suit, nous discutons plus précisément les modèles d'apprentissage profond. Tous nos modèles sont construits à partir de la même couche d'entrée et de la même couche de sortie. La couche d'entrée est une couche de type embedding permettant de caractériser chaque mot du document analysé par un vecteur de dimension 300. La couche de sortie contient elle, autant de neurones qu'il y a de classes dans notre ensemble de données. Ensuite, nous décrivons les couches intermédiaires de nos différents modèles de



réseaux neuronaux. Nous invitons le lecteur à consulter des revues plus détaillées telles que celle de (Schmidhuber, 2015).

TABLE 2 – *Modèles de Classification : classiques vs apprentissage profond*

Classiques	Apprentissage profond
LinearSVC (LSVC)	Basic neural network (BNN)
SigmoidSVC (SSVC)	Multi-layer perceptrons (MLP)
MultinomialNB (MNB)	Deep Multi-layer perceptrons (DMLP)
BernoulliNB (BNB)	Long short-term memory (LSTM) RNN
SGDClassifier (SGDC)	Bidirectional Long short-term memory (BLSTM) RNN
Decision Tree (DT)	Gated Recurrent Unit (GRU) RNN
AdaBoost (AB)	Long-term Recurrent Convolutional Networks (LRCN)
Random Forest (RF)	Convolutional neural network (CNN)
KNeighborsClassifier (KNC)	Serial Convolutional neural network (SCNN)
	Multi Channel Convolutional neural network (MCNN)

- *Basic neural network (BNN)* connecte directement la couche d'entrée avec la couche de sortie.
- *Multilayer Perceptrons (MLP)* est le modèle le plus simple d'apprentissage profond. Il est réalisé par une couche cachée de neurones entre la couche d'entrée et de sortie.
- *Deep Multilayer Perceptrons (DMLP)* est basé sur 20 couches cachées.
- *Long Short-Term Memory (LSTM)* est un sous-type de RNN. Les réseaux neuronaux récurrents permettent l'étude de séquences de données. Ces réseaux, composés de plusieurs couches, estiment leurs sorties en fonction des états de leurs couches précédentes en utilisant une mémoire intermédiaire. LSTM est basé sur des blocs de mémoire qui sont utilisés comme unités dans la couche récurrente pour capturer des dépendances à plus longue portée.
- *Bidirectional Long Short-Term Memory (BLSTM)* permet d'utiliser l'information après et avant les données étudiées par le réseau au temps  $t$ .
- *Gated Recurrent Unit (GRU)* est un réseau de neurones permettant une étude plus rapide que le LSTM tout en conservant les qualités de celui-ci.
- *Long-term Recurrent Convolutional Networks (LRCN)* associe en série un réseau convolutif et un réseau LSTM.
- *Convolutional Neural Network (CNN)* sont structurés par deux opérations : convolution puis max-pooling. La convolution est basée sur plusieurs filtres combinés pour extraire les nombreuses propriétés associées aux données. La seconde opération compresse les résultats de l'opération précédente pour extraire des informations denses.
- *Serial Convolutional neural network (SCNN)* place séquentiellement quatre réseaux convolutionnels les uns après les autres.
- *Multi Channel Convolutional neural network (MCNN)* est basé sur six réseaux de convolution branchés sur la couche d'entrée. Chaque réseau de convolution possède un noyau différent des autres. Ensuite, nous associons la sortie des six réseaux de convolution à une seule couche pour construire une sortie mono-couche entièrement connectée.

**Entrées des modèles :** Les entrées des modèles sont les messages. Nous utilisons deux pré-traitements distincts pour les deux types de méthodes de classification étudiées. Pour les méthodes de classification traditionnelles, nous appliquons une vectorisation basée sur la mesure Tf-Idf (Ramos, 1999). Pour les méthodes d'apprentissage profond, les données sont préparées différemment. Nous conservons les 5000 mots les plus fréquents (hors stop words) et nous représentons chaque document par une séquence de mots ce qui permet de conserver l'ordre des mots.

**Sorties des Modèles :** Les effectifs dans les classes sont très déséquilibrés. Nous avons donc mis en place deux stratégies que nous appelons *précise* et *imprécise*. Dans la straté-

gie *précise*, l'objectif est de décrire finement un message, par exemple selon la dimension émotion, en l'associant à l'étiquette *peur* ou *angoisse*. Dans la stratégie *imprécise*, le niveau de description attendu est plus général et le message pourrait être associé à une *émotion négative*. Nous avons donc regroupé manuellement, après discussion avec les experts, certaines classes. Par ailleurs, un grand nombre de messages ne sont pas étiquetés comme le montre le tableau 1. Pour le niveau d'expertise, les émotions et le risque, nous avons considéré une classe *Autres* correspondant à des messages non étiquetés. Nous effectuons sur ces échantillons un ré-échantillonnage de la taille de la classe majoritaire (non étiquetée). Une description des classes finales est donnée dans la table 3.

TABLE 3 – Classes utilisées dans les modèles selon les deux stratégies précise et imprécise

Stratégie	Précise (Nombre de posts)	Imprécise (Nombre de posts)
<b>Genre</b>	<i>genre_1</i> C1 (57) C2 (28)	
<b>Niveau d'expertise</b>	<i>expe_1</i> C3 (222) C4 (46) C5/C6 (74)	<i>expe_2</i> C3/C4/C5/C6 (342) Autres (342)
<b>Niveau informationnel</b>	<i>info_1</i> C7 (402) C8 (661) C9 (874) C10 (284) C11 (1144) C12 (159)	<i>info_2</i> C7 (402) C8/C9/C10/C12 (1978) C11 (1144)
<b>Émotion</b>	<i>emot_1</i> C13/C14 (65) C15/C17 (544) C18/C19 (30)	<i>emot_2</i> C13/C14/C15/C16 (609) C18/C19 (30) Autres (609)
<b>Risque</b>		<i>risq</i> C20/C21/C22 (101) Autres (101)

**Partitionnement des données et entraînement** Afin d'évaluer les différents modèles, nous appliquons une validation croisée en dix plis. L'ensemble de données est divisé en dix sous-ensembles. Nous utilisons neuf sous-ensembles pour la phase d'apprentissage et un sous-ensemble pour la phase de validation. Nous avons répété ce processus dix fois. Pour chaque pli, nous utilisons un sous-ensemble de différentes phases de validation et nous avons calculé une métrique pour évaluer la performance.

Pour les algorithmes de classification classiques, nous utilisons l'outil `sklearn`<sup>1</sup> avec les paramètres par défaut. Pour toutes les architectures d'apprentissage profond, nous avons utilisé les paramètres par défaut pour la taille du Mini-batch (c.-à-d. Le nombre d'instances d'entraînement à considérer en même temps), la dimension d'intégration (embedding dimension, chaque mot est décrit par un vecteur de dimension  $n$ ), le nombre d'époch (le nombre d'itérations sur l'ensemble d'entraînement), la fonction d'activation et le taux d'abandon (Dropout ratio, ratio d'unités cachées à désactiver dans chaque formation de Mini-batch). Ensuite, nous utilisons une configuration avec un Mini-batch de taille 32, une dimension d'intégration de 300, une couche cachée de taille 256, un nombre d'époch de 25, la fonction d'activation `selu` et un Dropout de 0,3.

1. <http://scikit-learn.org/stable/>

**Mesure de la performance** Pour tous les modèles, nous nous sommes basés sur une métrique couramment utilisée dans le domaine de la classification. Chaque sortie est un vecteur de dimension  $n$  correspondant aux  $n$  classes à prédire. Nous mesurons la qualité des classificateurs avec la métrique *exactitude* (accuracy). Celle-ci exprime le nombre de classes prédites justes en regard du nombre total de prédictions de classes réalisées. Nous avons choisi d'utiliser une métrique stricte en considérant une sortie comme juste uniquement si l'intégralité des classes prédites sont exactes.

### Exemple 1

Considérons 3 documents décrits par un vecteur d'annotations pour 4 classes. Chaque élément  $i$  de ce vecteur correspond à la présence (1) ou à l'absence (0) de la classe  $i$  pour le document.

*Document1*  $\rightarrow$  0010

*Document2*  $\rightarrow$  1000

*Document3*  $\rightarrow$  0100

Considérons maintenant un modèle retournant 3 vecteurs de prédictions de ces 4 classes. Chaque élément  $i$  de ce vecteur correspond à la prédiction (1) ou à l'absence de prédiction (0) de la classe  $i$  pour le document.

*prédiction1*  $\rightarrow$  0010

*prédiction2*  $\rightarrow$  0010

*prédiction3*  $\rightarrow$  1100

Seule la première prédiction sera considérée comme juste pour les trois prédictions émises. L'exactitude du modèle sera égale à :  $acc = \frac{1}{3} = 0.333$

## 2.4 Étape 4 : Visualisation temporelle et hiérarchique

Le flux de messages peut être représenté sous la forme d'une série temporelle multiple, c'est-à-dire un ensemble de variables quantitatives associées à un même intervalle temporel. Au cours des dernières années, les visualisations de type *Streamgraphs* (Byron & Wattenberg, 2008) ont été largement utilisées pour représenter l'évolution de plusieurs séries temporelles. Cependant, les *Streamgraphs* posent deux problèmes majeurs : (1) Comment gérer un long intervalle de temps ? (2) Comment gérer un grand nombre de séries temporelles ?

Pour résoudre le problème (1), des méthodes de type vue d'ensemble+détail ou encore focus+contexte peuvent être mises en place (voir (Munzner, 2014) pour une introduction). Les méthodes vue d'ensemble+détail consistent à combiner deux vues, une montrant l'ensemble des données de façon agrégée, l'autre montrant en détail un sous ensemble sélectionné. Les méthodes de type focus+contexte consistent à intégrer les parties agrégées (contexte) et les parties détaillées (focus) dans une seule et même vue. La plus connue de ces méthodes est sans doute le *fish-eye* qui consiste à appliquer une distortion sur la visualisation de façon à agrandir la zone du focus en réduisant la zone du contexte.

Pour résoudre le problème (2), les séries temporelles peuvent être agrégées selon leur proximité pour former une structure hiérarchique visualisable sous forme d'arbre. Cet arbre peut aider à sélectionner les niveaux de détail affichés.

La méthode que nous utilisons ici (Cuenca *et al.*, 2018) repose sur 3 vues permettant de traiter les problèmes susmentionnés. Une première vue (voir Fig. 2 cadre 1) permet de voir les séries agrégées sur l'ensemble de l'intervalle de temps, et de sélectionner un sous-intervalle (vue d'ensemble). Une seconde vue (voir Fig. 2 cadre 2) affiche les séries sur l'intervalle sélectionné (détail). Cette vue inclut un *fish-eye*, i.e. une zone dans laquelle les séries sont détaillées (focus), entourée par des zones dans lesquelles les séries sont agrégées (contexte). La troisième vue (voir Fig. 2 cadre 3) permet de visualiser l'arbre d'agrégation, dans notre cas la hiérarchie des classes présentées dans la table 1 et de sélectionner les niveaux devant être utilisés pour les zones agrégées et pour les zones détaillées.



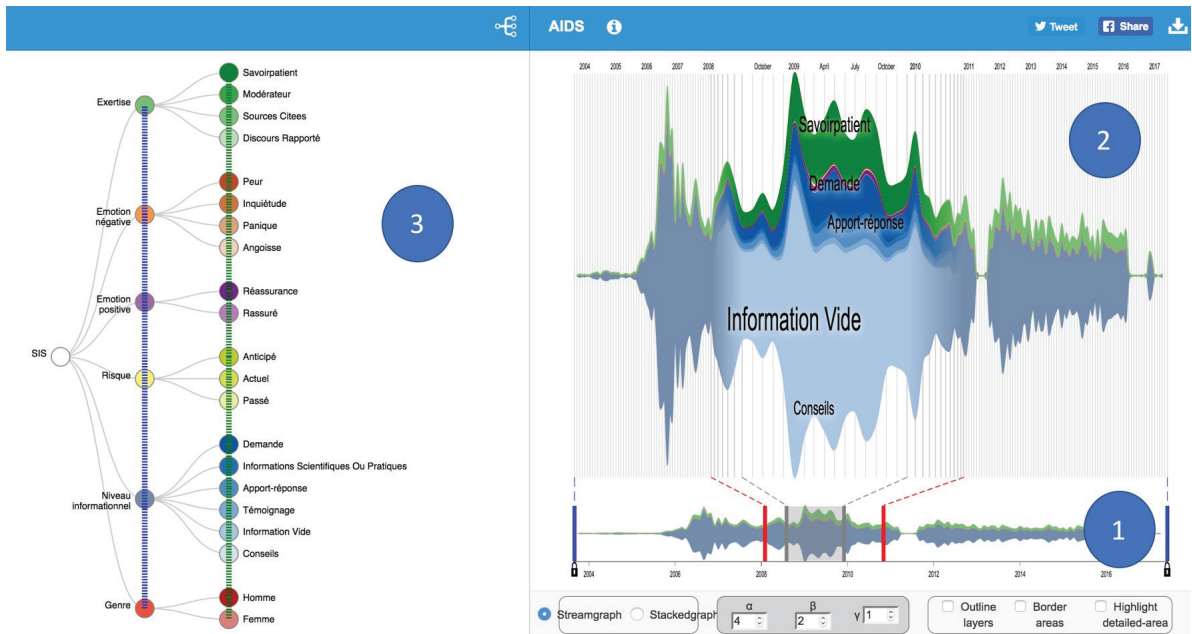


FIGURE 2 – Visualisation des séries temporelles avec un focus sur la période 2008-2011.



FIGURE 3 – Visualisation de la dimension "Niveau informationnel" dans laquelle les messages relatifs à la classe "Information vide" ont été filtrés grâce à l'arbre d'agrégation.

### 3 Résultats

#### 3.1 Comparaison des modèles de classification

TABLE 4 – Comparaison des performances des classifieurs.

Stratégie	Précise				Imprécise			
jeux	emot_1	genre_1	expe_1	info_1	emot_2	expe_2	info_2	risq
LSVC	0.318	0.621	0.735	0.255	0.533	0.659	0.488	<b>0.638</b>
SSVC	<b>0.325</b>	0.624	0.735	0.242	0.532	0.422	0.000	0.440
MNB	0.020	0.631	0.701	0.175	0.472	0.662	0.406	0.603
BNB	0.058	<b>0.660</b>	0.725	0.280	<b>0.563</b>	<b>0.672</b>	0.371	0.603
SGDC	0.309	0.640	0.711	0.238	0.518	0.630	<b>0.490</b>	0.608
DT	0.244	0.533	0.664	0.206	0.468	0.625	0.362	0.564
AB	0.300	0.563	0.718	0.248	0.493	0.600	0.450	0.633
RF	0.0	0.650	<b>0.752</b>	0.0	0.116	0.379	0.0	0.440
KNC	0.227	0.533	0.677	<b>0.323</b>	0.439	0.589	0.281	0.461
BNN	0.718	0.945	0.966	0.372	0.841	0.894	0.883	0.883
MLP	0.662	0.945	0.966	0.356	0.745	0.901	0.909	<b>0.921</b>
DMLP	0.627	0.658	0.788	0.308	0.801	0.342	0.293	0.376
LSTM	<b>0.827</b>	<b>0.990</b>	<b>0.969</b>	<b>0.564</b>	<b>0.882</b>	<b>0.922</b>	<b>0.940</b>	0.906
BLSTM	0.776	0.927	0.953	0.521	0.855	0.906	0.898	0.902
GRU	0.502	0.684	0.884	0.355	0.441	0.805	0.745	0.805
LRCN	0.795	0.954	0.926	0.516	0.832	0.881	0.893	0.907
CNN	0.713	0.954	0.963	0.408	0.841	0.900	0.883	0.897
SCNN	0.741	0.954	0.959	0.436	0.829	0.905	0.893	0.892
MCNN	0.734	0.963	0.956	0.382	0.847	0.905	0.881	0.892

Le tableau 4 synthétise les résultats des différents classifieurs sur les différentes dimensions et selon les 2 stratégies. Comme l'on pouvait s'y attendre, les résultats pour les stratégies de classification précises sont inférieurs aux résultats pour les stratégies de classification imprécises. Par exemple, pour la dimension relative au niveau informationnel, l'exactitude pour la classification info\_1 (stratégie précise) est au mieux de 0,564 alors que pour info\_2 (stratégie imprécise) est de 0,906. Pour chaque type de classification, les meilleurs classifieurs traditionnels ont obtenu des résultats inférieurs aux classifieurs en apprentissage profond. Par exemple, pour la classification du genre, le meilleur classificateur traditionnel obtient une exactitude de 0,660 alors que le meilleur classifieur en apprentissage profond obtient une exactitude de 0,990. Pour 7 types de classification sur 8, la meilleure architecture est LSTM. Ce résultat est cohérent avec la littérature. Par exemple, Johnson & Zhang (2016) obtiennent également des résultats performants sur des textes avec LSTM. La classification du genre, de l'expertise, du risque et du niveau d'expérience obtiennent en apprentissage profond des exactitudes supérieures à 0,9 dans le cas imprécis et même parfois précis. Les résultats sont plus mitigés pour les émotions et le niveau informationnel qui correspondent aux classes très déséquilibrées, en particulier avec les classifieurs classiques. Par exemple, le classifieur RF ne fonctionne pas sur les classes déséquilibrées telles que emot\_1 ou emot\_2. Il serait intéressant de réannoter un nombre suffisant d'exemples pour rééquilibrer ces classes.

#### 3.2 Visualisation

La visualisation obtenue à partir des 226 252 messages annotés automatiquement à partir du meilleur modèle (LSTM) est disponible sur le site du projet Multistream<sup>11</sup>. La figure 2 permet de repérer un pic d'activité du forum en 2009 et une majorité d'information de type

11. <http://advanse.lirmm.fr/multistream/> - jeu de données AIDS

"Information vide" révélant la part importante de messages de convivialité ou de soutien, les conseils étant également très recherchés dans des forums modérés par des professionnels de la prévention. La figure 3 est obtenue après sélection des classes à afficher dans l'arbre d'agrégation, en filtrant les messages de type "Information vide". On remarque alors que l'information de type "Demande" domine le pic de messages de 2009, ce qu'il n'était pas possible de visualiser avant le filtrage. Ces deux figures illustrent le processus de navigation dynamique dans le flux de messages, rendu possible par l'outil *MultiStream*.

Nous allons détailler dans la suite les 5 dimensions identifiées lors de la phase d'annotation manuelle (le genre, le niveau d'expertise, le niveau informationnel, le risque et les émotions) et illustrer avec des exemples ce qu'il est possible de repérer dans ces flots d'informations, et l'exploitation éventuelle qui peut en être faite.

Concernant la dimension *Genre*, on remarque qu'il y a plus de messages d'hommes que de femmes. Cette situation est spécifique au VIH, puisque les femmes sont souvent plus présentes sur les forums de santé (Paganelli & Clavier, 2011). Dans le cas du sida, les hommes ayant des relations sexuelles avec d'autres hommes (HARSAH), en termes de prévalence, sont les plus touchés par le VIH, ils sont aussi la cible des campagnes de prévention et sont bien informés des risques d'exposition au sida (De Oliveira, 2012). Ces résultats nous laissent penser que les forums constituent des médias alternatifs et complémentaires (Renahy & Chauvin, 2006) indispensables pour des hommes plutôt moins acculturés aux messages de prévention, qui se renseignent pour des prises de risques plus exceptionnelles (par exemple dans le cadre de relations extra-conjugales hétéro ou homosexuelles).

Concernant la dimension *Niveau d'expertise*, les messages annotés montrent que se développent sur les forums de nouvelles formes d'expertise, hors de la sphère médicale et relayées par le biais des participants ou des modérateurs. Ces informations sont de première importance pour les professionnels (de santé et de la prévention) qui peuvent ainsi appréhender les connaissances qui circulent sur le VIH. Ainsi, la majorité des messages correspondent à des "savoirs patients" et dans une moindre proportion à des messages de modérateurs ou à des messages contenant des sources citées et des discours rapportés. Dans le cas du VIH, les patients apparaissent comme experts de leur propre maladie, de leur cas singulier bien qu'ils ne soient pas spécialistes de la maladie en général ni de la médecine. Ces derniers ont été conduits à côtoyer des professionnels de santé dont ils rapportent les propos, ce qui légitime les conseils qu'ils donnent aux autres participants : *"Merci de m'avoir répondu, c'est vrai vous m'avez bien rassurée ms je ne peux m'empêcher de penser qu'il y a un risque, j'en ai parlé à 2 médecins qui travaillent avec moi, ils me disent que le risque est minime ms qu'il y en a un, ils ont l'habitude de soigner des séropositifs puisqu'ils accueillent des patients ayant eu un accident exposant au sang comme on dit chez nous"*. Les forums sont aussi le moyen de transmettre des sources spécialisées qui participent à la circulation des savoirs sur le VIH : *"En tout cas pour tout ce qui est des modes de transmission du vih tu trouveras tout ce qu'il te faut savoir ici : [http://www.sida-info service.org/forums ... hp4?t=3257\[ ... \]](http://www.sida-info service.org/forums...hp4?t=3257)."*

Concernant le *Niveau informationnel*, la majorité des messages sont vides, au sens où ils renvoient à des messages de convivialité (ex. *"Bonjour tous"*, *"@ bientôt"*, *"Merci++"*). Une fois ces messages filtrés, deux types d'informations apparaissent : des demandes ou des apports d'information associés à ces demandes. Ces messages concernent essentiellement des informations pratiques ou scientifiques : *"Bonjour, J'ai lu que les anticorps apparaissent généralement entre 20 et 45 jours après un risque. Ca veut donc dire que ||nom|| n'est très fiable qu'après 6 semaines ? Je sens que je vais avoir la réponse sybilline "1 mois=bonne indication, 3 mois= sur", mais j'aimerais savoir ce que ca veut dire bonne indication (je suis négatif a 4 semaines avec des ganglions et des courbatures partout... Et pouvezvous aussi me dire si le test ||nom|| detecte le VIH-1 et le VIH-2 ? Merci pour vos reponses!!"*, et dans une moindre mesure des témoignages relatant le vécu des porteurs du virus. Les participants au forum directement concernés par le VIH cherchent en premier lieu des informations relatives à cette maladie, aux symptômes, aux traitements. Ils cherchent également à mieux comprendre le discours des experts en amont d'un rendez-vous médical ou en aval pour trouver la définition

de termes spécialisés ou confirmer le discours des professionnels de santé. Pour les proches non directement concernés par le virus, la recherche d'information permet de découvrir des conseils pratiques pour accompagner une personne malade.

Concernant le *risque*, il faut distinguer une dimension objective du risque, incarnée dans une connaissance plus ou moins partagée des facteurs de prise de risque, qui se confronte à des perceptions variables du risque en lien notamment avec l'expérience personnelle de l'individu, son environnement social, affectif et culturel. La dimension subjective du risque motive l'essentiel des demandes : *"Bonjour, ma copine est séropositive depuis peu, mon 1er test est négatif. J'attends 3 mois pour confirmer le résultat. Comme j'entends tout et n'importe quoi à propos des fellations, je voulais savoir si je risquais un risque ou pas lors d'une fellation dans mon cas. Merci pour votre réponse. C'est nouveau pour moi et je gère assez mal la situation."* Ainsi, les trois quarts des messages portent sur un risque passé, peu de personnes posant des questions avant une situation présumée risquée. Dans les forums où priment l'émotion, c'est la panique, le stress qui dominent. Cette situation engendre de la part des répondants des messages de réassurance, la catégorie d'émotion qui est la plus représentée en termes de pourcentage d'apports d'information. En effet, de nombreux messages d'encouragements, par exemple lors de l'attente d'un dépistage viennent soutenir les membres ayant pris un risque. Peu de messages ont cependant été étiquetés selon le risque lié à l'alcool et à la prise de substance psychotrope. Ce constat est certainement lié au fait qu'un forum institutionnel de type SIS n'est pas le lieu pour évoquer les pratiques à risques autres que celles liées à la sexualité.

### 3.3 Limites de l'étude

Notre approche préliminaire présente de nombreuses limites que nous détaillons dans cette section.

Concernant la phase d'annotation, la grille utilisée pourrait être retravaillée. La dimension relative au *Niveau Informationnel* pourrait être scindée en deux pour distinguer les messages relevant de l'interaction (demande ou apport de réponse à une question) des messages uniquement destinés à l'apport d'information (Information Scientifique et pratique, témoignage et conseil) ou encore les messages vides au niveau informationnel. Cela est tout à fait possible car la visualisation utilisée permet d'afficher plusieurs niveaux de hiérarchie.

Concernant la phase de classification, la principale limite est que nous avons utilisé uniquement l'exactitude comme métrique pour évaluer la performance des classifieurs. Le rappel et la précision sont deux mesures importantes pour évaluer la qualité d'un classifieur et devrait être prises en considération. Une autre limitation porte sur l'interprétation de l'apprentissage profond. Ces modèles sont des "boîtes noires". Ils ne fournissent pas d'explication même si la prédiction est efficace (Shwartz-Ziv & Tishby, 2017). Cependant, malgré un nombre de messages manuellement annotés limité, cette étude a montré que la performance des modèles traditionnels était beaucoup plus faible que la performance de l'apprentissage profond. De nouvelles techniques d'apprentissage profond sont étudiées pour faciliter l'interprétation de tels modèles (Liu *et al.*, 2017; Lipton, 2016) et pourraient s'avérer judicieuses dans ce contexte.

Nous avons également identifié des limites liées à la temporalité utilisée pour la visualisation de l'ensemble des données du site SIS. Actuellement, la temporalité est le mois. Or, la période temporelle d'interaction est généralement le jour ou la semaine dans le contexte des échanges sur les forums. Une possibilité de focus sur une période plus courte permettrait de repérer des épisodes liés à des questions sur une prise de risque et le délai pour se faire dépister. En effet, 48h après une prise de risque, il existe un traitement préventif qui peut être suggéré par les internautes. Les résultats du premier test permettant de savoir s'il y a eu contamination sont obtenus entre 6 semaines et 90 jours. Cette période s'accompagne de messages contenant des émotions négatives, de la réassurance et dans la majorité des cas se finit par des messages de la personne rassurée comme nous avons pu l'observer dans de nombreux fils de discussion. De même, la visualisation pourrait être améliorée par l'affichage d'informations supplémentaires comme l'heure des messages. Cette information est essentielle notamment



lorsque l'on s'intéresse aux émotions exprimées dans les messages. Par exemple, les angoisses s'expriment généralement la nuit. La quantification de ces phénomènes est importante pour le site SIS en terme de planification de leurs équipes de modérateurs.

Pour finir, une dernière limite de notre étude est liée à la généralisation de nos résultats. En effet, la tâche est très spécifique au sujet d'étude, le VIH, au type de textes que sont les messages des forums, ce qui rend difficile la généralisation de notre approche. Toutefois, notre méthodologie et les résultats peuvent être utilisés comme référence pour d'autres études sur l'identification automatique de catégories à partir de données sociales.

#### 4 Conclusion et perspectives

Notre étude a souligné l'efficacité des architectures d'apprentissage profond pour prédire des thèmes sur des données de forum. La visualisation originale sous forme de StreamGraph permet d'explorer de manière efficace tout en quantifiant les différentes catégories de messages. La faisabilité de notre approche peut conduire à de nouvelles applications en santé basées sur les médias sociaux destinés aux patients et aux professionnels de santé.

Nos résultats montrent clairement le besoin d'études capables d'analyser automatiquement des forums et d'en extraire des informations utiles. Nous proposons l'utilisation de l'apprentissage automatique et de la visualisation interactive pour relever ces défis. Cette étude reste préliminaire. Une étude plus approfondie sur les différents types de préparation des données, paramètres des algorithmes, modèles d'apprentissage permettrait d'affiner l'interprétation des résultats de la phase 3. En particulier, différents regroupement de classes et l'interprétation des liens entre classes serait pertinent. Quels sont les sentiments associés à une prise de risque ? Est-ce que les réponses apportées dans les fils sont rassurantes ou satisfaisantes ? Comment s'informent les participants dans les forums ? Pour finir, une étude poussée sur l'utilité et l'utilisabilité de la visualisation présentée en phase 4 est également nécessaire. Nous suggérons 4 perspectives.

Premièrement, nous prévoyons d'entreprendre une analyse à grande échelle en utilisant une collection de médias sociaux plus large (Autres forums, Facebook, Twitter, ...). Cette analyse inclura l'application de méthode d'apprentissage non supervisé de type Latent Dirichlet Allocation (LDA) (Pennacchiotti & Gurumurthy, 2011; Wang *et al.*, 2012) pour extraire les thèmes émergents des discussions et l'exploration du style linguistique des différents utilisateurs (Zhan *et al.*, 2017; Zeng & Tse, 2006; Wang *et al.*, 2014). Une attention particulière sera portée sur l'identification d'une typologie des risques encourus en lien avec le VIH (e.g. consommation de drogue, comportements sexuels atypiques, etc).

Deuxièmement, nous pensons que lorsque les ensembles de données sont petits, l'apprentissage est difficile. Une amélioration significative serait la mise en œuvre de techniques d'apprentissage actif (Olsson, 2009). En effet, dans ce type de tâche, il est important d'optimiser les informations disponibles afin que les systèmes de classification puissent les utiliser le plus efficacement possible pendant la phase d'apprentissage tout en préservant l'acquisition de nouveaux échantillons étiquetés (Ducoffe & Precioso, 2015). L'utilisation de la visualisation pour guider les annotateurs vers des messages à annoter pourrait s'avérer également intéressante.

Troisièmement, au sein d'un ensemble de données suffisamment important, nous pouvons tirer parti des modèles d'apprentissage automatique pour utiliser des fonctionnalités plus complexes pour caractériser les utilisateurs qui postent ces messages. Nous suggérons de mettre l'accent sur les groupes d'utilisateurs, y compris les professionnels de la santé, les célébrités, le grand public et les associations. Cela nous amènera à comprendre quel groupe d'utilisateurs est important, peut jouer le rôle d'influenceur, les incitant à partager leurs messages, à les aimer et à leur répondre.

Pour finir, nous prévoyons d'étudier la distribution temporelle des messages pour nous concentrer sur la dynamique des thématiques au fil du temps. Nous pouvons étudier les corrélations temporelles entre les réactions des internautes et les événements du monde réel comme les soirées de type Sidaction. Cette analyse exploratoire pourrait aider à identifier les



facteurs contribuant à la sensibilisation. Au-delà, nous pouvons également analyser la répartition géographique des messages.

Ce type d'étude est importante pour convaincre les parties prenantes, les professionnels de la santé et le grand public de s'impliquer et d'utiliser le Web 3.0 comme intelligence collective pour repousser les maladies telles que le VIH.

## 5 Remerciements

Ce travail s'intègre dans le projet #AIDS et a été soutenu par une subvention ANRS<sup>12</sup> en 2016. Les auteurs souhaitent remercier les gestionnaires du site Sida-Info-Service pour le partage des données et ses participants pour leur engagement à combattre le VIH.

## Références

- (2015). *Joint Named Entity Recognition and Disambiguation*.
- AIGNER W., MIKSCH S., SCHUMANN H. & TOMINSKI C. (2011). *Visualization of Time-Oriented Data*. Springer.
- ANDOR D., ALBERTI C., WEISS D., SEVERYN A., PRESTA A., GANCHEV K., PETROV S. & COLLINS M. (2016). Globally normalized transition-based neural networks. cite arxiv :1603.06042.
- BACCOUCHE M., MAMALET F., WOLF C., GARCIA C. & BASKURT A. (2011). Sequential deep learning for human action recognition. In *Proceedings of the Second International Conference on Human Behavior Understanding*, HBU'11, p. 29–39, Berlin, Heidelberg : Springer-Verlag.
- BYRON L. & WATTENBERG M. (2008). Stacked Graphs - Geometry & Aesthetics. *IEEE Transactions on Visualization and Computer Graphics*, **14**(6), 1245–1252.
- CLAVIER V. & PAGANELLI C. (2017). Une approche méthodologique croisée du traitement des données de la recherche : le cas d'un corpus d'échanges issus de médias sociaux dans le domaine de la santé. In *Colloque COSSI 2017, Méthodes et stratégies de gestion de l'information par les organisations : des "Big Data" aux "Thick Data"*, 85ème congrès de l'ACFAS, Université McGill.
- COLLOBERT R., WESTON J., BOTTOU L., KARLEN M., KAVUKCUOGLU K. & KUKSA P. (2011). Natural language processing (almost) from scratch. *J. Mach. Learn. Res.*, **12**, 2493–2537.
- CUENCA E., SALLABERRY A., WANG F. Y. & PONCELET P. (2018). MultiStream : A Multiresolution Streamgraph Approach to Explore Hierarchical Time Series. *IEEE Transactions on Visualization and Computer Graphics*, (to appear).
- CUI W., LIU S., TAN L., SHI C., SONG Y., GAO Z., QU H. & TONG X. (2011). TextFlow : Towards Better Understanding of Evolving Topics in Text. *IEEE Transactions on Visualization and Computer Graphics*, **17**(12), 2412–2421.
- DE OLIVEIRA J.-P. (2012). *Communication publique et formes de gouvernabilité contemporaines de l'Etat : le cas de l'homosexualité dans les campagnes de prévention du sida en France (1987-2007)*. PhD thesis, Université Stendhal, Grenoble.
- DUCOFFE M. & PRECIOUSO F. (2015). QBDC : query by dropout committee for training deep supervised architecture. *CoRR*, **abs/1511.06412**.
- HAVRE S., HETZLER E. & NOWELL L. (2000). ThemeRiver : Visualizing Theme Changes over Time. In *Proceedings of the IEEE Symposium on Information Visualization*, p. 115–123 : IEEE.
- HE L., LEE K., LEWIS M. & ZETTEMAYER L. (2017). Deep semantic role labeling : What works and what's next. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*.
- JOHNSON R. & ZHANG T. (2016). Supervised and semi-supervised text categorization using lstm for region embeddings. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48, ICML'16*, p. 526–534 : JMLR.org.
- KALCHBRENNER N., GREFFENSTETTE E. & BLUNSON P. (2014). A convolutional neural network for modelling sentences. In *ACL (1)*, p. 655–665 : The Association for Computer Linguistics.
- KIM Y. (2014). Convolutional neural networks for sentence classification. In A. MOSCHITTI, B. PANG & W. DAELEMANS, Eds., *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, p. 1746–1751 : ACL.

12. <http://www.anrs.fr/fr>

- KUCHER K. & KERREN A. (2015). Text visualization techniques : Taxonomy, visual survey, and community insights. In *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis)*, p. 117–121.
- KUMAR A., IRSOY O., ONDRUSKA P., IYYER M., BRADBURY J., GULRAJANI I., ZHONG V., PAULUS R. & SOCHER R. (2016). Ask me anything : Dynamic memory networks for natural language processing. In M. F. BALCAN & K. Q. WEINBERGER, Eds., *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, p. 1378–1387, New York, New York, USA : PMLR.
- LIPTON Z. C. (2016). The mythos of model interpretability. *CoRR*, **abs/1606.03490**.
- LIU W., WANG Z., LIU X., ZENG N., LIU Y. & ALSAADI F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, **234**, 11 – 26.
- MIKOLOV T., KARAFIÁT M., BURGET L., CERNOCKÝ J. & KHUDANPUR S. (2010). Recurrent neural network based language model. In T. KOBAYASHI, K. HIROSE & S. NAKAMURA, Eds., *INTERSPEECH*, p. 1045–1048 : ISCA.
- MIKOLOV T., SUTSKEVER I., CHEN K., CORRADO G. S. & DEAN J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. BURGESS, L. BOTTOU, M. WELLING, Z. GHAHRAMANI & K. Q. WEINBERGER, Eds., *Advances in Neural Information Processing Systems 26*, p. 3111–3119. Curran Associates, Inc.
- MIN S., LEE B. & YOON S. (2016). Deep learning in bioinformatics. *CoRR*, **abs/1603.06430**.
- MUNZNER T. (2014). *Visualization Analysis and Design*. A.K. Peters visualization series. A K Peters.
- NAKOV P., RITTER A., ROSENTHAL S., SEBASTIANI F. & STOYANOV V. (2016). Semeval-2016 task 4 : Sentiment analysis in twitter. In *Proceedings of the 10th International Workshop on Semantic Evaluation*, p. 1–18.
- OLSSON F. (2009). *A literature survey of active machine learning in the context of natural language processing*. Rapport interne T2009 :06.
- PAGANELLI C. & CLAVIER V. (2011). Le forum de discussion : une ressource informationnelle hybride entre information grand public et information spécialisée. *Yasri-Labrique Eleonore. Les forums de discussion : agoras du XXIe siècle ? Théories, enjeux et pratiques discursives, L'harmattan (collection Langue et Parole)*, p. 39–55.
- PAGANELLI C. & CLAVIER V. (2014). S'informer via des médias sociaux de santé : quelle place pour les experts ? **23**, 141–143.
- PENNACCHIOTTI M. & GURUMURTHY S. (2011). Investigating topic models for social media user recommendation. In *Proceedings of the 20th International Conference Companion on World Wide Web*, WWW '11, p. 101–102, New York, NY, USA : ACM.
- RAMOS J. (1999). Using tf-idf to determine word relevance in document queries.
- RENAHY E. & CHAUVIN P. (2006). Internet uses for health information seeking : A literature review. *Revue Epidémiologique de Santé Publique*, **54**(3), 263–275.
- ROSENTHAL S., NAKOV P., KIRITCHENKO S., MOHAMMAD S., RITTER A. & STOYANOV V. (2015). Semeval-2015 task 10 : Sentiment analysis in twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, p. 451–463.
- SCHMIDHUBER J. (2015). Deep learning in neural networks : An overview. *Neural Networks*, **61**, 85–117. Published online 2014; based on TR arXiv :1404.7828 [cs.NE].
- SHWARTZ-ZIV R. & TISHBY N. (2017). Opening the black box of deep neural networks via information. *CoRR*, **abs/1703.00810**.
- SOCHER R., PERELYGIN A., WU J. Y., CHUANG J., MANNING C. D., NG A. Y. & POTTS C. (2013). Recursively deep models for semantic compositionality over a sentiment treebank. In *EMNLP*, volume 1631, p. 1642.
- SRIVASTAVA R. K., GREFF K. & SCHMIDHUBER J. (2015). Training very deep networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'15, p. 2377–2385, Cambridge, MA, USA : MIT Press.
- SUKHBAATAR S., SZLAM A., WESTON J. & FERGUS R. (2015). End-to-end memory networks. In C. CORTES, N. D. LAWRENCE, D. D. LEE, M. SUGIYAMA & R. GARNETT, Eds., *Advances in Neural Information Processing Systems 28*, p. 2440–2448. Curran Associates, Inc.
- SUN G., WU Y., LIU S., PENG T.-Q., ZHU J. J. H. & LIANG R. (2014). EvoRiver : Visual Analysis of Topic Cooperation on Social Media. *IEEE Transactions on Visualization and Computer Graphics*, **20**(12), 1753–1762.
- VINYALS O., KAISER L. U., KOO T., PETROV S., SUTSKEVER I. & HINTON G. (2015). Grammar as a foreign language. In C. CORTES, N. D. LAWRENCE, D. D. LEE, M. SUGIYAMA & R. GARNETT, Eds., *Advances in Neural Information Processing Systems 28*, p. 2773–2781. Curran Associates, Inc.

- WANG S., J.PAUL M. & DREDZE M. (2014). Exploring health topics in chinese social media : an analysis of sina weibo. *AAAI Workshop on the World Wide Web and Public Health Intelligence*, **23**, 20–23.
- WANG Y., AGICHTEIN E. & BENZI M. (2012). Tm-lda : Efficient online modeling of latent topic transitions in social media. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, p. 123–131, New York, NY, USA : ACM.
- WU Y., LIU S., YAN K., LIU M. & WU F. (2014). OpinionFlow : Visual Analysis of Opinion Diffusion on Social Media. *IEEE Transactions on Visualization and Computer Graphics*, **20**(12), 1763–1772.
- ZENG Q. T. & TSE T. (2006). Viewpoint paper : Exploring and developing consumer health vocabularies. *JAMIA*, **13**(1), 24–29.
- ZHAN Y., LIU R., LI Q., LEISCHOW S. & ZENG D. (2017). Identifying topics for e-cigarette user-generated contents : a case study from multiple social media platforms. *J Med Internet Res*, **19**(1), e24.
- ZHU M., ZHANG Y., CHEN W., ZHANG M. & ZHU J. (2013). Fast and accurate shift-reduce constituent parsing.