
Recherche d'Information Sociale en Langue Arabe : Cas de Facebook

Ismail Badache¹ — Aya Abu-Thaher² — Mariam Hamdan² — Lara Abu-Jaish²

¹Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France

²Faculty of Engineering and Technology, Birzeit University, Palestine

RÉSUMÉ. Cet article propose une approche de recherche d'information (RI) en langue arabe sur Facebook, qui exploite toutes les traces des utilisateurs (ex. polarité, partage, j'aime, haha) laissées sur des publications Facebook pour estimer leur importance sociale. Notre objectif est de montrer comment ces signaux peuvent jouer un rôle vital dans l'amélioration de la recherche en langue arabe sur Facebook. Premièrement, des polarités (positive ou négative) portée par les signaux textuels (ex. commentaires) et non textuels (ex. les réactions j'adore et triste) ont été identifiées pour chaque publication Facebook. Par conséquent, la polarité de chaque commentaire exprimé sur une publication donnée, est estimée sur la base d'un modèle neuronal de sentiment en langue arabe. Deuxièmement, des signaux en fonction de leur complémentarité ont été regroupés en utilisant des algorithmes de sélection. Troisièmement, des algorithmes de learning to rank ont été appliqués pour re-ordonner les résultats de recherche de Facebook en fonction des groupes de signaux sélectionnés. Enfin, des expérimentations sont réalisées sur 13500 publications Facebook, collectées à partir de 45 requêtes en langue arabe. Les expérimentations révèlent des résultats prometteurs pour la RI en langue arabe sur Facebook.

ABSTRACT. This paper proposes an approach on Facebook Search in Arabic language, which exploits all traces of users (e.g. polarity, share, like, haha) left on Facebook publications to estimate their social importance. Our goal is to show how these signals can play a vital role in improving Arabic Facebook search. First, the identification of the polarities (positive or negative) carried by the textual signals (e.g. comments) and non-textual ones (e.g. the reactions love and sad) for a given Facebook publication. Therefore, the polarity of each comment expressed on a given Facebook post, is estimated on the basis of a neural model of sentiment in Arabic language. Secondly, grouping signals according to their complementarity using selection algorithms. Third, the application of learning to rank algorithms to re-rank Facebook search results based on the selected groups of signals. Finally, experiments are carried out on 13500 Facebook posts, collected from 45 topics in Arabic language. Experiments reveal promising results.

MOTS-CLÉS : Recherche sur Facebook, Analyse de sentiments, Contenus générés par l'utilisateur

KEYWORDS: Facebook Search, Sentiment Analysis, User Generated Content

1. Introduction

Les médias sociaux ont largement contribué au lancement du "printemps arabe". Depuis lors, la pénétration des médias sociaux n'a cessé de croître. Le nombre d'utilisateurs de Facebook dans le monde arabe est estimé à 83.7 millions¹. Ce mouvement reflète la démocratisation des moyens de production et d'interaction dans le Web (contenu généré par les utilisateurs) grâce aux nouvelles technologies. Parmi ces moyens de plus en plus accessibles à un large public, on peut citer les réseaux sociaux, les blogs, les microblogs, etc. Le contenu généré par les utilisateurs (en anglais User-Generated Content : UGC) se réfère à un ensemble de données (ex. commentaires, publications, réactions et interactions via des signaux sociaux, etc) dont le contenu est principalement, soit produit, soit directement influencé par les utilisateurs finaux.

Une des tâches populaires en RI consiste à retrouver un ensemble de documents pertinents vis-à-vis d'un besoin en information exprimé dans une requête. Pour cela, des approches efficaces existent depuis de nombreuses années qui exploitent deux classes de critères pour trier les documents répondant à une requête donnée. La première classe, la plus exploitée, dépend de la requête, elle concerne tous les critères liés aux statistiques des termes telles que leur fréquence dans le document (*tf*), et dans la collection de documents (*idf*). La deuxième classe concerne les critères indépendants de la requête, qui mesurent l'importance a priori du document. On y trouve par exemple, le nombre de liens entrants vers un document (Kraaij *et al.*, 2002), Page-Rank (Brin et Page, 1998), la localité thématique du document (Davison, 2000), la présence d'URL dans le document (Westerveld *et al.*, 2001), ses auteurs (Macdonald et Ounis, 2006) et les signaux sociaux (Badache et Boughanem, 2017b).

Cet article se situe dans le contexte de la RI sociale (RIS) en langue arabe et s'intéresse plus particulièrement à l'exploitation des contenus générés par les utilisateurs sur Facebook (J'aime, Partage, Commentaire Positif, Commentaire Négatif, J'adore, Haha, Grrr, Wouah et Triste) dans le processus de RI en langue arabe sur Facebook. Afin de concevoir notre approche, des tâches fondamentales sont effectuées. D'abord, l'identification des polarités (positive ou négative) pour tous les commentaires laissés sur une publication (*Post*) donnée en utilisant un modèle d'analyse de sentiments neuronal en langue arabe. Ensuite, nous utilisons des algorithmes de sélection d'attributs pour identifier les meilleurs signaux (critères) et groupes de signaux. Enfin, nous évaluons l'impact de ces critères sur l'estimation de la pertinence des résultats retournés par le moteur de recherche de Facebook. Plus précisément, nous tentons de sélectionner les critères les plus efficaces et de les combiner avec des approches d'apprentissage d'ordonnement (*Learning to Rank*) pour améliorer la RI sur Facebook. Les principales contributions abordées dans cet article sont doubles :

(C1). Évaluer l'impact des signaux sociaux ainsi que l'analyse du sentiment des commentaires sur la recherche en langue arabe sur Facebook. Par conséquent, nous essayons de répondre aux questions de recherche suivantes : a) Quels sont les meilleurs signaux et groupes de signaux adaptés à cette tâche ? ; b) Quel est l'impact de ces critères sociaux sur les performances du moteur de recherche de Facebook ?

1. <https://www.internetworldstats.com>

(C2). La construction d'une collection de test (documents, requêtes, qrels) issue de Facebook. Cette collection est utile pour l'évaluation des systèmes de RI sociale en langue arabe. Des études expérimentales orientées utilisateurs - *user studies* - ont été menées pour collecter les jugements de pertinence.

L'article est organisé comme suit. La section 2 présente certains travaux connexes. La section 3 détaille notre approche. La section 4 présente notre collection de test que nous même avons créé. L'évaluation expérimentale est présentée dans la section 5. Enfin, la section 6 conclut l'article en annonçant des perspectives.

2. Contexte et travaux connexes

Dans cette section, nous introduisons tout d'abord quelques notions générales sur les UGCs (particulièrement les signaux sociaux), ensuite nous présentons une vue d'ensemble sur la RI sociale, ainsi nous articulons le lien de notre travail avec les travaux de l'état de l'art qui exploitent les UGCs comme source d'évidence additionnelle pour mesurer la pertinence d'un document donné. Enfin, comme le processus de notre approche de RIS nécessite l'utilisation d'une méthode d'analyse du sentiment en langue arabe, certains travaux dans ce domaine seront présentés.

2.1. Contenus générés par l'utilisateur

En général, chaque réseau social a ses propres UGCs (voir le tableau 1) avec des règles de fonctionnement différentes. La popularité des UGCs, en particulier dans le contexte des médias sociaux, a donné naissance à de nombreux nouveaux problèmes en recherche d'information (Bouadjenek *et al.*, 2016). Plus précisément, comment exploiter ces contenus sociaux en faveur de la RI est une question ouverte. Dans la suite, nous présentons une vue d'ensemble sur la recherche d'information sociale.

Type	Exemple	Réseaux sociaux
<i>Vote</i>	J'aime, +1	Facebook, LinkedIn, Google+, StumbleUpon
<i>Message</i>	Tweet, Publication (Post)	Facebook, Google+, LinkedIn, Twitter
<i>Partage</i>	Partage, Re-tweet	Google+, Twitter, Buffer, Facebook, LinkedIn
<i>Tag</i>	Bookmark, Pin	Delicious, Pinterest, Diigo, Digg
<i>Commentaire</i>	Commentaire, Répondre	Facebook, Google+, LinkedIn, Twitter
<i>Émotion</i>	J'adore, Haha, Wouah Triste, Grrr	Facebook
<i>Réaction Évènement</i>	Thankful (Fête des mères)	Facebook
<i>Relation</i>	Abonnés, Amis	Facebook, Twitter

Tableau 1 : Liste des différents types d'UGCs (signaux sociaux) actuels

2.2. RI sociale : Vue d'ensemble

En 2012, Jaime Teevan², chercheur à Microsoft, donne une définition décrivant la recherche d'information sociale comme suit : "*Social search is an emerging re-*

2. <https://www.microsoft.com/en-us/research/video/social-search-panel/>

search area that explores how social interactions and social data can enhance existing information-seeking experiences, as well as enable new information retrieval scenarios. This session will showcase different models of social search, including 1) the use of social data to augment search, 2) social data as new information to be searched, and 3) social interaction and collaboration as part of the search process.”

Les approches de RI sociale ont étendu les modèles traditionnels avec différentes caractéristiques sociales afin de satisfaire des motivations sociales derrière les besoins d’information de l’utilisateur. En outre, de nouvelles approches sociales entrent en vue afin de répondre aux nouveaux besoins en information et en RI initiées par les pratiques sociales sur le Web. Nous considérons en effet la RI sociale selon 3 axes : 1) le premier axe concerne la recherche d’information de nature sociale. Il s’agit de trouver des informations sociales qui répondent à l’utilisateur. On distingue par exemple la recherche d’information dans les blogs, microblogs (Damak, 2014 ; Luo *et al.*, 2012) et la recherche de conversations (Magnani *et al.*, 2012) ; 2) le deuxième porte sur l’exploitation des contenus sociaux pour améliorer la RI, dans laquelle l’information sociale est utilisée afin d’améliorer le processus de recherche d’information, par exemple, les tags dans les folksonomies ont été trouvés utiles pour améliorer la recherche Web et la recherche personnalisée (Bullock *et al.*, 2018), le reclassement (re-ranking) des résultats de recherche (Badache et Boughanem, 2017b ; Badache et Boughanem, 2017a ; Badache et Boughanem, 2015 ; Buijs et Spruit, 2014) et 3) le troisième paradigme concerne la recherche d’information effectuée par plusieurs personnes, recherche collaborative (Yue et He, 2018).

À notre connaissance, aucun travail n’a été réalisé sur l’exploitation des signaux sociaux pour la recherche en arabe sur Facebook. Les travaux les plus liés à notre travail incluent (Khodaei et Shahabi, 2012 ; Buijs et Spruit, 2014 ; Chelaru *et al.*, 2014 ; Chan *et al.*, 2016 ; Orso *et al.*, 2017 ; Badache et Boughanem, 2017b). Ces travaux s’intéressent à l’exploitation des caractéristiques sociales pour améliorer la RI en anglais sur le Web et sur les réseaux sociaux. L’approche que nous proposons dans cet article se situe à la fois dans le contexte des deux premiers points précédents. Elle est dans la même lignée que ces travaux, c’est-à-dire exploiter les caractéristiques sociales autour des publications (documents) Facebook pour améliorer le tri de pertinence des résultats de recherche. Cependant, notre travail se différencie de l’état de l’art sur les points suivants. D’abord, notre approche concerne la RIS en langue arabe sur Facebook. Nous supposons que les communautés arabophone sont influencées par leur environnement culturel, et donc réagissent différemment sur les publications Facebook par rapport à d’autres communautés. Autrement, les traces sociales laissées par des utilisateurs arabes influencent la pertinence pour un besoin d’information d’un utilisateur de culture arabe. Une analyse de sentiment des commentaires laissés par les utilisateurs sur une publication donnée est nécessaire. Ensuite, nous utilisons des algorithmes d’apprentissage d’ordonnancement (*Learning To Rank*) combinés avec des techniques de sélection d’attributs. Plus précisément, nous estimons l’importance sociale d’une publication Facebook en exploitant ces traces sociales (j’aime, partage, polarité du commentaire : positive ou négative, j’adore, haha, grrr, wouah et triste) pour améliorer le moteur de recherche en langue arabe de Facebook.

2.3. Approches d'analyse de sentiments

L'analyse du sentiment a fait l'objet de très nombreuses recherches antérieures, en se basant sur des approches supervisées et non supervisées. Dans les approches non supervisées, nous pouvons citer les approches basées sur les lexiques telles que l'approche développée par (Turney, 2002) ou bien des méthodes basées sur des corpus comme les travaux de (Mohammad *et al.*, 2013). Au rang des approches supervisées, nous pouvons citer (Pang *et al.*, 2002) qui comme nombre de travaux perçoivent la tâche d'analyse de sentiments comme une tâche de classification et utilisent donc des méthodes comme les SVM (*Support Vector Machines*) ou les réseaux bayésiens. D'autres travaux récents sont basés sur les RNN (*Recursive Neural Network*) tels que les travaux de (Socher *et al.*, 2013).

À notre connaissance, peu de travaux ont été réalisés pour l'analyse du sentiment en langue arabe. Ceci peut être expliqué par le manque de collections standards (*datasets*). (Almas et Ahmad, 2007 ; Farra *et al.*, 2010) proposent une méthode linguistique s'appuyant sur un ensemble de patrons permettant d'extraire les polarités d'un document financier. (Abdulla *et al.*, 2014a) construisent manuellement un lexique contenant 4815 mots. Leur système calcule le nombre de mots positifs et négatifs dans un texte afin de générer sa polarité globale. (Al-Kabi *et al.*, 2014) ont mis en place un outil qui détermine la subjectivité, la polarité d'une opinion et son intensité. Ils utilisent deux lexiques généraux et seize lexiques spécifiques. (Abdulla *et al.*, 2014b) proposent une approche statistique de détection de subjectivité et de polarité dans les réseaux sociaux en utilisant des attributs morphologiques. (Bayoudhi *et al.*, 2015) comparent trois classifieurs : SVM, Naïve Bayes et un réseau de neurones simple. (El-Halees *et al.*, 2011) est le premier à avoir proposé un système hybride pour l'analyse d'opinions pour l'arabe. Il propose une hiérarchie séquentielle de classifications combinées. (Ibrahim *et al.*, 2015) utilise un lexique de 5244 adjectifs, un lexique de 3296 idiomes pour améliorer la classification de phrases avec un SVM. (Refae et Rieser, 2016) appliquent une approche hybride pour la prédiction de l'intensité de la polarité dans les tweets. Ils ont utilisé particulièrement la régression logistique pour prédire les scores initiaux qui sont ajustés en appliquant des règles extraites à partir d'un lexique de polarité.

Plusieurs travaux récents appliquent des techniques d'apprentissage profond pour l'analyse d'opinion (Barhoumi *et al.*, 2017 ; Dahou *et al.*, 2016). (Barhoumi *et al.*, 2017) utilisent les représentations continues de documents combinées avec un Perceptron MultiCouche (PMC, en anglais (*MultiLayer Perceptron MLP*)) tandis que (Dahou *et al.*, 2016) utilisent un CNN (*Convolutional Neural Network*). (Barhoumi *et al.*, 2018) illustrent une comparaison pertinente entre plusieurs systèmes de détection de sentiments en langue arabe, expérimentés sur le corpus LABR (Large-scale Arabic Book Review dataset)³, ⁴. Ils montrent que les meilleurs résultats ont été obtenus par (Dahou *et al.*, 2016) avec l'utilisation d'un CNN (77.39% d'exactitude). Le

3. <http://www.mohamedaly.info/datasets/labr>

4. <https://github.com/mohamedadaly/LABR>

deuxième meilleur système est celui de (ElSahar et El-Beltagy, 2015), ils ont construit un grand lexique arabe multi-domaines pour l'analyse de sentiments. Les données ont été recueillies à partir de divers commentaires (*reviews*) de sites Web (hôtels⁵, films⁶, produits⁷ et restaurants⁸).

Comme le propos de cet article est d'exploiter des critères sociaux pour améliorer la recherche en langue arabe sur Facebook, nous avons besoin de mesurer la polarité des commentaires et la considérer comme un facteur additionnel dans l'estimation de pertinence vis-à-vis d'une requête. Nous avons donc utilisé l'approche proposée par (Dahou *et al.*, 2016) que nous décrivons dans la section 3.1

3. RI en langue arabe sur Facebook

Notre approche est basée à la fois sur les traces classiques (ex. les fréquences des signaux j'aime et partage, etc) et sur les traces émotionnelles (ex. les fréquences des réactions j'adore et triste, etc) ainsi que sur l'analyse de sentiments dans les commentaires exprimés par les utilisateurs sur chaque publication (document) Facebook. Nous notons que les réactions de Facebook permettent aux utilisateurs d'exprimer des émotions plus nuancés par rapport aux signaux classiques. L'objectif de notre approche consiste à améliorer la pertinence des résultats retournés par le moteur de recherche en langue arabe de Facebook en exploitant l'ensemble des traces sociales issues de ce dernier. Ces traces sociales (ou signaux) sont considérées comme des connaissances a priori à prendre en compte dans le processus de recherche en langue arabe sur Facebook. La figure 1 illustre le processus de *learning to rank* de notre approche.

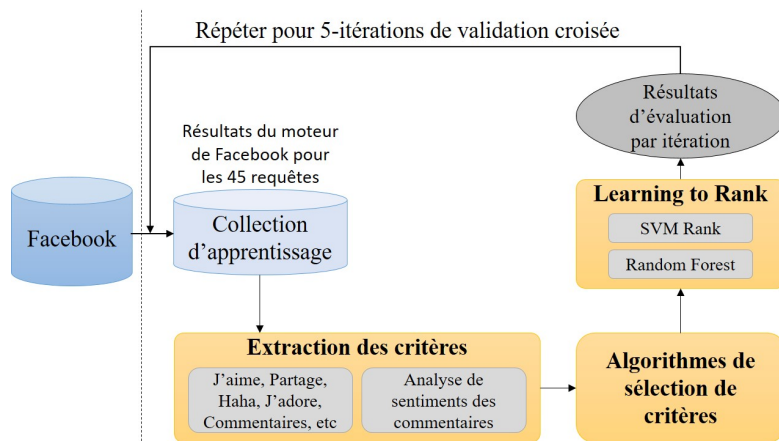


Figure 1 : Processus de *learning to rank* en utilisant les techniques de sélection

5. <https://www.tripadvisor.com/>

6. <https://www.elcinema.com/>

7. <https://uae.souq.com/ae-ar/>

8. <http://www.qaym.com/>

3.1. Analyseur de sentiments portés par les commentaires en langue arabe

Dans ce travail, le sentiment porté par un commentaire sur une publication (document) Facebook est estimé en utilisant le modèle proposé par (Dahou *et al.*, 2016) dont son implémentation est publiquement disponible⁹. (Dahou *et al.*, 2016) ont proposé une approche basée sur le CNN (*Convolutional Neural Network*) pour identifier la polarité des opinions des commentaires (*reviews*) en langue arabe. Lorsque l'on considère la sémantique des mots, il a été montré que les plongements de mots neuronaux (*word embedding*) capturent des similarités sémantiques entre les mots (Mikolov *et al.*, 2013). De telles représentations distribuées qui plongent les mots dans un espace vectoriel dense sont apprises de façon efficace sur de grandes collections. (Dahou *et al.*, 2016) ont donc utilisé la méthode du *word embedding* ("plongement de mots" ou "plongement lexical" en français)¹⁰ à l'aide d'un corpus de 3.4 milliards de mots choisis parmi un corpus de 10 milliards de mots analysés sur le Web. Ensuite, une architecture CNN similaire à (Kim, 2014) a été entraînée en utilisant les plongements de mots neuronaux pour classifier la polarité de sentiments au niveau d'une phrase sans aucune considération au niveau d'aspect (sujet sur lequel porte le sentiment). Ils ont entraîné le modèle *word2vec* (Mikolov *et al.*, 2013) de type *Skip-gram* (SKIP-G) et *Continuous Bag Of Words* (CBOW) sur des pages Web. Leurs expériences ont montré que CBOW est plus performant. Les résultats de l'expérience montrent que leur schéma surpasse les méthodes existantes sur plusieurs jeux de données disponibles au public (Aly et Atiya, 2013 ; Nabil *et al.*, 2015 ; Refaee et Rieser, 2014 ; Abdulla *et al.*, 2013 ; ElSahar et El-Beltagy, 2015).

3.2. Sélection des meilleurs critères de pertinence

La difficulté dans notre approche réside dans la sélection des groupes de critères de pertinence. Tant que nous n'avons pas essayé toutes les combinaisons des critères de pertinence possibles, il est impossible de juger convenablement et de façon rigoureuse la complémentarité des critères de pertinence, et d'identifier les meilleures combinaisons. Dans cette étape de pré-traitement, nous nous sommes ainsi appuyés sur des techniques de sélection d'attributs pour déterminer les meilleurs critères à prendre en compte dans une tâche de RI sur Facebook.

Les techniques de sélection d'attributs visent à identifier et enlever le maximum d'information redondante et non pertinente en amont d'un processus à base d'apprentissage (Hall et Holmes, 2003). Elles permettent également de sélectionner de manière automatique les sous-ensembles de critères de pertinence permettant d'avoir les meilleurs résultats. Cette phase a fait ressortir plusieurs ensembles de facteurs. Ensuite, nous avons évalué l'efficacité de ces ensembles en les appliquant sur des techniques d'apprentissage d'ordonnement (*learning to rank*) dans un contexte de RI.

9. <https://pan.baidu.com/s/1eS2mxCe#list/path=%2F>

10. https://fr.wikipedia.org/wiki/Word_embedding

3.3. Combinaison des algorithmes d'apprentissage avec les techniques de sélection

Dans cette étape, nous évaluons certains algorithmes d'apprentissage (*learning to rank*) en utilisant les critères identifiés par les techniques de sélection. L'objectif est double : d'une part, nous cherchons à valider si la sélection des critères améliore effectivement les résultats d'une tâche de RI en langue arabe sur Facebook. D'autre part, nous cherchons à mesurer la performance de certains algorithmes d'apprentissage d'ordonnement combinés avec les techniques de sélection dans ce type de tâches.

Nous nous sommes inspirés des travaux de (Hall et Holmes, 2003). Ils ont étudié l'efficacité de certaines techniques de sélection d'attributs en les confrontant avec des algorithmes d'apprentissage. Étant donné que la performance des critères diffère d'une technique d'apprentissage à une autre, nous avons identifié les meilleures techniques de sélection d'attributs permettant de retrouver les critères les plus performants en fonction des algorithmes d'apprentissage d'ordonnement à utiliser.

4. Collection de test : Facebook

4.1. Données collectées

A notre connaissance, il n'existe pas à ce jour de collection de test standard en langue arabe, contenant des documents (*Posts*) issus de Facebook, des requêtes et des signaux sociaux, pour évaluer l'efficacité des systèmes de RI en langue arabe sur Facebook. De ce fait, dans le but d'expérimenter l'efficacité de notre approche, nous avons collecté 13500 publications (*Posts*) en arabe extraites du réseau social Facebook via son API et aussi en utilisant du *parsing*, entre le 16 et le 28 Janvier 2018. Ces publications ont été collectées à partir d'un profil nouvellement créé via le moteur de recherche de Facebook pour 45 requêtes que nous même avons défini. Pour chaque publication, nous avons collecté également ses signaux classiques (*j'aime*, *partage*, *commentaire*) et ses signaux émotionnels (*j'adore*, *triste*, *grrr*, *haha*, *wouah*). Le tableau 2 donne quelques statistiques sur notre collection.

Champ	Somme	Min	Max	Moyenne
Publications	13500	-	-	-
J'aime	2031958	0	32025	151
Partage	2329934	0	16781	173
Commentaire	2717589	0	24306	201
Triste	63970	0	80	5
Grrr	95752	0	119	7
J'adore	397679	0	496	29
Haha	246715	0	308	18
Wouah	171234	0	213	13
Commentaire Positif	1527546	0	13750	113
Commentaire Négatif	1134831	0	10063	84

Tableau 2 : Les chiffres des données issues de Facebook (pour 45 requêtes en arabe)

4.2. Jugements de pertinence par les utilisateurs

Nous avons défini 45 requêtes en langue arabe (le tableau 2 montre un exemple de ces requêtes). La figure 3 présente des statistiques sur la nature des requêtes.

Requête	Traduction du besoin d'information
الطفل السوري عمران	Ce qui est arrivé à l'enfant syrien <i>Omran</i> , qui a été bombardé avec son frère à son domicile à Alep, et la réaction mondiale à cette tragédie.
اضراب جامعة بيرزيت	Les réactions sur la grève à l'université de Birzeit (Palestine) et son impact sur le processus éducatif.
قتل السفير الروسي	Pourquoi l'ambassadeur de Russie a-t-il été tué en Turquie? Et quelles sont les opinions publiques arabes et autres concernant sa mort?
رقصة الزومبا	Quels sont les avantages de la Zumba et a-t-elle un rapport avec la perte de poids?

Figure 2 : Exemples de requêtes d'évaluation en langue arabe

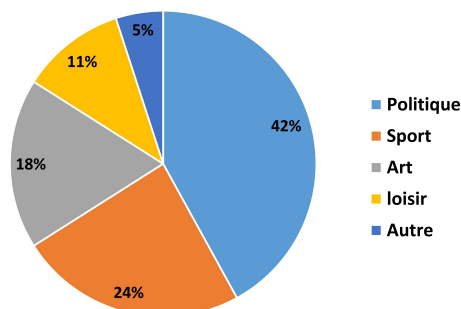


Figure 3 : Statistiques sur la nature des 45 requêtes

Pour obtenir les jugements de pertinence des documents, nous avons impliqué 6 participants. Il y avait 3 femmes et 3 hommes, la tranche d'âge était entre 22 et 33 ans. Tous les participants étaient de la discipline informatique dont 3 sont des étudiants, 1 doctorant, 2 enseignants. Nous avons demandé à chaque participant de choisir au moins 7 requêtes de notre ensemble de requêtes, et leur avons demandé d'évaluer les 300 premiers documents affichés pour une requête donnée en utilisant une échelle d'évaluation à 3 points (non-pertinent, peu pertinent et pertinent). Nous notons que chaque requête est jugée par 3 utilisateurs. Pour éviter tout biais, aucun des critères sociaux n'a été affiché avec les documents, mais tous les contenus textuels, images ou vidéo (selon la publication Facebook) sont affichés pour faciliter la tâche de jugement.

Nous avons analysé le degré d'accord entre les jugements de pertinence pour chaque requête avec la mesure Kappa Cohen k . Cet indicateur prend en compte la proportion d'accord entre les évaluateurs et la proportion de l'accord attendu entre les évaluateurs par hasard. La mesure de Kappa est égale à 1 si les évaluateurs sont complètement d'accord, 0 s'ils ne sont d'accord que par hasard. k est négatif si l'accord entre évaluateurs est pire que l'aléatoire. Comme nous avons trois évaluateurs par requête, la valeur Kappa a été calculée pour chaque paire d'évaluateurs, puis leur moyenne a été calculée. Nous avons constaté que la mesure de l'accord varie de 0.45 à 0.90. La mesure moyenne d'accord entre les évaluateurs est de 75%, ce qui correspond à un accord fort.

5. Impact des critères sociaux sur la RI en langue arabe dans Facebook

Dans cette étude, nous nous sommes appuyés sur des techniques de sélection d'attributs pour déterminer les critères sociaux (parmi ceux cités dans le tableau 3) les plus importants pour la tâche d'apprentissage d'ordonnancement de pertinence.

c_i	Critère	Description
c_1	J'aime	Nombre de <i>J'aime</i> sur le document
c_2	Partage	Nombre de <i>Partage</i> sur le document
c_3	Commentaire	Nombre total des <i>Commentaire</i> sur le document
c_4	Triste	Nombre de <i>Triste</i> sur le document
c_5	Grrr	Nombre de <i>Grrr</i> sur le document
c_6	J'adore	Nombre de <i>J'adore</i> avec notation
c_7	Haha	Nombre de <i>Haha</i> sur le document
c_8	Wouah	Nombre de <i>Wouah</i> sur le document
c_9	Commentaire Positif	Nombre de <i>Commentaire Positif</i> sur le document
c_{10}	Commentaire Négatif	Nombre de <i>Commentaire Négatif</i> sur le document

Tableau 3 : Liste des signaux Facebook exploités

Le tableau 3 présente les 10 critères que nous avons considérés pour estimer la pertinence des publications sur Facebook par rapport à une requête exprimée en langue arabe. La nature des critères c_1 jusqu'au critère c_{10} est un simple comptage, par exemple les critères c_1 , c_2 et c_3 représentent le nombre de J'aime, de Partage et de Commentaire générés sur le document, respectivement. Les critères c_4 , c_5 , c_6 , c_7 et c_8 représentent les réactions émotionnelles de Facebook. Concernant les deux derniers critères c_9 et c_{10} , ils représentent le nombre d'opinions exprimées sur le document selon leur polarité (positive ou négative), respectivement. Ces deux critères sont calculés en se basant sur un modèle d'analyse de sentiments en langue arabe, présenté dans la section précédente 3.1.

5.1. Identification des critères les plus efficaces

Afin de comprendre l'impact réel des différents signaux sociaux, nous avons évalué l'impact de chacun d'eux en utilisant des algorithmes de sélection d'attributs. Le but est de déterminer les meilleurs signaux à exploiter dans le modèle de recherche d'information. Les algorithmes de sélection d'attributs visent à identifier et supprimer le maximum d'information inutile, redondante et non pertinente en amont d'un processus à base d'apprentissage (Hall et Holmes, 2003). Ils permettent également de sélectionner de manière automatique les sous ensembles de critères de pertinence permettant d'avoir les meilleurs résultats. Nous avons utilisé Weka¹¹ (dernière version stable 2018 : 3.8.3), un outil open-source écrit entièrement en Java et qui rassemble un bon ensemble de techniques d'apprentissage et des techniques de sélection d'attributs.

Nous avons procédé ainsi : les premiers 300 documents (publications) pour chaque requête (45 requêtes au total) ont été restitués avec le moteur de recherche de Facebook. Les scores de tous les critères (signaux sociaux) de pertinence ont été associés

11. <http://www.cs.waikato.ac.nz/ml>

à chaque document. Nous avons identifié les documents pertinents et les documents non pertinents selon les Qrels. L'ensemble obtenu contient 13500 documents dont :

- 2971 documents (publications Facebook) pertinents.
- 10529 documents (publications Facebook) non pertinents.

Les classes de ces ensembles sont déséquilibrées, or lorsque le nombre d'éléments d'une classe dans une collection d'apprentissage dépasse considérablement les autres échantillons des autres classes, un classifieur tend à prédire les échantillons de la classe majoritaire et peut ignorer complètement les classes minoritaires (Yen et Lee, 2006). Pour cette raison, nous avons appliqué une approche de sous-échantillonnage de manière aléatoire (en réduisant le nombre d'échantillons qui ont la classe majoritaire) pour générer des collections équilibrées composées de :

- 2971 documents (publications Facebook) pertinents.
- 2971 documents (publications Facebook) non pertinents.

Dans notre cas, les algorithmes de sélection d'attributs consistent à attribuer un score à chaque critère en fonction de sa signification vis-à-vis la classe de pertinence (pertinent et non pertinent). Ces algorithmes retournent un classement d'importance des critères en fonction du nombre de fois qu'un critère donné a été sélectionné par l'algorithme dans la validation croisée. Nous notons que nous avons utilisé pour chaque algorithme le paramétrage par défaut fourni par Weka.

Nous avons appliqué une validation croisée à 5 itérations pour 10 critères, c'est-à-dire $n = 10$. Le tableau 4 présente les critères sélectionnés par les algorithmes de sélection d'attributs. Un critère sélectionné (choisi) par l'algorithme de sélection est un critère désigné par un "+" et un critère non sélectionné est désigné par un "-".

Algorithme	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
CfsSubsetEval	+	+	-	-	-	+	-	+	+	+
WrapperSubsetEval	+	+	-	-	-	+	-	-	+	+
ConsistencySubsetEval	+	+	+	+	+	+	+	+	+	+
FilteredSubsetEval	+	+	-	+	+	+	-	-	+	+
ChiSquaredAttributeEval	+	+	+	+	+	+	+	+	+	+
FilteredAttributeEval	+	+	+	+	+	+	+	-	+	+
GainRatioAttributeEval	+	+	-	-	-	+	-	+	+	+
InfoGainAttributeEval	+	+	+	+	+	+	+	+	+	+
OneRAttributeEval	+	+	+	+	+	+	+	-	+	+
ReliefAttributeEval	+	+	-	+	+	+	+	+	+	+
SVMAttributeEval	+	+	-	+	+	+	+	-	+	+
SymmetricalUncertEval	+	+	-	+	+	+	-	-	+	+
Total	12	12	5	9	9	12	7	6	12	12

Tableau 4 : Les critères sélectionnés par les algorithmes de sélection d'attributs

Le tableau 4 montre que les critères c_{10} : *Commentaire Négatif*, c_9 : *Commentaire Positif*, c_1 : *J'aime*, c_2 : *Partage* et c_6 : *J'adore* sont les plus sélectionnés et les mieux classés par rapport aux autres critères. Les critères c_4 : *Triste*, c_5 : *Grrr* sont modérément favorisés par les algorithmes de sélection d'attributs, à l'exception des algorithmes *CfsSubsetEval*, *WrapperSubsetEval* et *GainRatioAttributeEval* qui ne l'ont

pas sélectionné. Les critères c_7 : *Haha*, c_8 : *Wouah* ne sont sélectionnés que par 7 et 6 algorithmes, respectivement. Enfin, le critère le plus faible et le plus désavantagé est le c_3 : *Commentaire*, il est seulement sélectionné par 5 sur 12 algorithmes.

5.2. Apprentissage d'ordonnement de pertinence sensible au critères sociaux

D'autres expérimentations ont été menées en exploitant ces critères dans des approches supervisées basées sur des modèles d'apprentissage d'ordonnement. Nous avons utilisé les instances (les publications Facebook) des 45 requêtes comme ensembles d'apprentissage. Nous avons ensuite utilisé deux algorithmes d'apprentissage d'ordonnement. Ce choix s'explique par le fait qu'ils ont souvent montré leur efficacité en recherche d'information : RankSVM (Joachims, 2006) et Random Forests (Mohan *et al.*, 2011). Concernant RankSVM, nous utilisons l'implémentation¹² avec ses paramètres par défaut proposée par *Joachims* (Joachims, 2006). Alors que pour Random Forests, nous avons utilisé l'implémentation¹³ proposée par Weka. Nous avons mis l'option "*max depth*" à 0 (illimité) et le nombre d'arbres à 100. L'entrée de chaque algorithme est un vecteur de critères (voir tableau 3), soit tous les critères ou seulement les critères sélectionnés par un algorithme de sélection précis. Les algorithmes d'apprentissage d'ordonnement prédisent le rang de pertinence des résultats de recherche. Enfin, nous avons appliqué une validation croisée pour 5 itérations (*5-folds cross-validation*).

La question à ce stade est liée à la spécification du vecteur de critères d'entrée pour les algorithmes d'apprentissage, soit on prend tous les critères, soit on garde uniquement ceux sélectionnés par les techniques de sélection d'attributs. Dans ce cas, avec quels algorithmes d'apprentissage ces derniers seront combinés.

Afin de prendre en compte les critères choisis par les algorithmes de sélection dans des modèles d'apprentissage, nous avons réalisé plusieurs expérimentations pour identifier les meilleures techniques de sélection d'attributs permettant de retrouver les critères les plus performants en fonction des techniques d'apprentissage d'ordonnement. Nous avons trouvé les meilleurs couples des algorithmes d'apprentissage d'ordonnement et des techniques de sélection d'attributs suivants :

- L'ensemble des critères sélectionnés par *CfsSubsetEval* (CFS) et *WrapperSubsetEval* (WRP) sont appris par le modèle RankSVM et Random Forests.
- L'ensemble des critères sélectionnés par *ReliefFAttributeEval* (RLF) sont appris par le modèle Random Forests.
- L'ensemble des critères sélectionnés par *SVMAttributeEval* (SVM) sont appris par le modèle RankSVM.

Nous rappelons que la phase des algorithmes de sélection d'attributs (*CfsSubsetEval*, *WrapperSubsetEval*, *SVMAttributeEval*, *ReliefFAttributeEval*) a fait ressortir les ensembles de critères suivants (voir le tableau 5).

12. http://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html

13. <http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/RandomForest.html>

Algorithmes de sélection	Critères sélectionnés
CfsSubsetEval (CFS)	$c_1, c_2, c_6, c_8, c_9, c_{10}$
WrapperSubsetEval (WRP)	$c_1, c_2, c_6, c_9, c_{10}$
SVMAttributeEval (SVM)	$c_1, c_2, c_4, c_5, c_6, c_7, c_9, c_{10}$
ReliefAttributeEval (RLF)	$c_1, c_2, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}$

Tableau 5 : Ensembles des critères sélectionnés par CFS, WRP, SVM et RLF

Afin de vérifier la significativité des résultats par rapport aux résultats du moteur de recherche de Facebook (modèle de base), nous avons effectué le t-test de Student. Nous avons attaché * (forte signification) et ** (très forte signification) aux résultats de la table 6 quand $p\text{-value} < 0.05$ et $p\text{-value} < 0.01$, respectivement.

Modèle de RI		P@5	P@10	nDCG	MAP
Facebook (modèle de base)		0.1911	0.1721	0.2513	0.1002
Algo. d'apprentissage	Algo. de sélection	P@5	P@10	nDCG	MAP
RankSVM	CfsSubsetEval (CFS)	0.2133*	0.1944*	0.2955*	0.1204*
	WrapperSubsetEval (WRP)	0.1992	0.1802	0.2674	0.1076
	SVMAttributeEval (SVM)	0.2627**	0.2441**	0.3939**	0.1654**
	Tous les critères	0.2254*	0.2066*	0.3196*	0.1314*
Random Forests	CfsSubsetEval (CFS)	0.2395*	0.2046*	0.2955*	0.1149*
	WrapperSubsetEval (WRP)	0.2072	0.1883	0.2834	0.1149
	ReliefAttributeEval (RLF)	0.2920**	0.2735**	0.4522**	0.1921**
	Tous les critères	0.2526**	0.2340**	0.3738**	0.1563**

Tableau 6 : Les résultats de $P@\{5, 10\}$, nDCG et MAP du *Learning to Rank*

Le tableau 6 compare les différentes configurations de notre approche en termes de précision@ k ($k \in \{5, 10\}$), nDCG et MAP, obtenus par RankSVM et Random Forests exploitant les critères ressortis de l'étude utilisant les techniques de sélection d'attributs : CFS, WRP, SVM et RLF. Nous remarquons globalement, avec la prise en compte des critères sociaux, les résultats obtenus sont significativement meilleurs que ceux obtenus par le modèle de base de Facebook. Les résultats sont discutés ci-dessous pour chaque algorithme.

5.2.1. Résultats obtenus par RankSVM

Les résultats obtenus par RankSVM en utilisant l'algorithme de sélection *SVMAttributeEval* (SVM), où seulement les deux critères c_3 et c_8 n'ont pas été sélectionnés, sont meilleurs par rapport à ceux obtenus en utilisant (CFS, WRP ou tous les critères). Nous avons enregistré des taux d'amélioration de 57% et 65% en termes de nDCG et MAP, respectivement, par rapport au modèle de base. En utilisant CFS qui sélectionne seulement 6 critères $c_1, c_2, c_6, c_8, c_9, c_{10}$, et WRP qui sélectionne encore moins de critères $c_1, c_2, c_6, c_9, c_{10}$, les résultats baissent avec des taux de -25% et -32% en termes de nDCG, respectivement. Par conséquent, les critères non sélectionnés c_4, c_5, c_7 et c_8 sont donc fructueux pour RankSVM. En outre, avec la sélection de tous les critères, RankSVM obtient de meilleurs résultats par rapport à ceux obtenus avec CFS et WRP lorsque certains critères sont ignorés. En effet, certaines requêtes telles que (traduction : *l'enfant syrien Omran*) et (traduction : *blocus de Gaza*) ont enregistré la plus haute précision lorsque les critères c_4 : *Triste* et c_5 : *Grrr* sont pris en compte (avec 0.8957 et 0.9324 en termes de $P@10$, respectivement). Les critères c_8 : *Wouah* et c_7 : *Haha* sont plus efficaces avec des requêtes qui représentent des informations

bizarres, passionnantes ou marrante. Enfin, même si l’algorithme RankSVM est un peu coûteux en termes de temps d’exécution, il reste favorisé pour obtenir des résultats significatifs. Nous avons remarqué que RankSVM combiné avec l’algorithme de sélection (SVM) a obtenu le deuxième meilleur résultat après les résultats obtenus par Random Forests combiné avec l’algorithme de sélection (RLF).

5.2.2. Résultats obtenus par Random Forests

Selon le tableau 6, les résultats confirment que l’arbre de décision Random Forests est le modèle le plus approprié quand il est combiné avec l’algorithme de sélection *ReliefAttributeEval* (RLF), il prend en compte tous les critères, à l’exception du critère c_3 : *Commentaire*, de manière plus efficace que les autres configurations (des taux d’amélioration de 80% et 92% en termes de nDCG et MAP par rapport au modèle de base, respectivement). Les taux d’amélioration par rapport au modèle de base en utilisant CFS et WRP sont relativement faibles (18% et 13% en termes de nDCG, respectivement). Nous remarquons également que Random Forests (combiné avec l’algorithme de sélection RLF) dépasse la meilleure configuration de RankSVM (combiné avec l’algorithme de sélection SVM) avec un taux de 15% et 16% en termes de nDCG et MAP, respectivement. En outre, les améliorations sont également fortement significatives pour la configuration prenant tous les critères avec Random Forests (classée 3^{ème} après Random Forest avec RLF et RankSVM avec SVM).

6. Conclusion

Cet article propose une approche supervisée de RI en langue arabe dans Facebook basée sur des critères sociaux propres à Facebook. Certains critères sont un simple comptage (j’aime, triste, haha, etc), tandis que d’autres représentent une polarité des commentaires (positive ou négative). Nous avons utilisé des techniques de sélection d’attributs (critères) combinées avec des algorithmes de *learning to rank*. L’évaluation menée sur la collection issue Facebook montre que Random Forests prenant en entrée les critères sélectionnés par RLF est la configuration la plus fructueuse pour estimer l’ordre de pertinence des résultats. De plus, les algorithmes d’apprentissage d’ordonnement basés sur les critères les plus pertinents selon les algorithmes de sélection sont généralement meilleurs comparés à ceux obtenus lorsque les algorithmes de sélection sont ignorés. Enfin, nous notons que nous sommes conscients que l’évaluation de notre approche est encore limitée. La principale faiblesse de notre approche est sa dépendance à la qualité du modèle d’analyse de sentiments. Une étape de traitement essentielle pour une RI sociale en langue arabe efficace consiste à utiliser un *stemmer* pour l’arabe dialectal. D’autres expérimentations à plus grande échelle sur d’autres collections sont également nécessaires. Ceci étant même avec ces éléments simples, les premiers résultats obtenus nous encouragent à investir davantage cette piste.

7. Bibliographie

Abdulla N. A., Ahmed N. A., Shehab M. A., Al-Ayyoub M., « Arabic sentiment analysis : Lexicon-based and corpus-based », *AECT*, IEEE, p. 1-6, 2013.

- Abdulla N. A., Ahmed N. A., Shehab M. A., Al-Ayyoub M., Al-Kabi M. N., Al-Rifai S. Y., « Towards Improving the Lexicon-Based Approach for Arabic Sentiment Analysis », *IJITWE*, vol. 9, n° 3, p. 55-71, 2014a.
- Abdulla N. A., Al-Ayyoub M., Al-Kabi M. N., « An extended analytical study of Arabic sentiments », *IJBDI*, vol. 1, n° 1/2, p. 103-113, 2014b.
- Al-Kabi M. N., Gigieh A. H., Alsmadi I. M., Wahsheh H. A., Haidar M. M., « Opinion mining and analysis for Arabic language », *IJACSA*, vol. 5, n° 5, p. 181-195, 2014.
- Almas Y., Ahmad K., « A note on extracting sentiments in financial news in English, Arabic & Urdu », *2nd WS on Computational Approaches to Arabic Script-based Languages*, 2007.
- Aly M. A., Atiya A. F., « LABR : A Large Scale Arabic Book Reviews Dataset », *ACL*, 2013.
- Badache I., Boughanem M., « A priori relevance based on quality and diversity of social signals », *SIGIR*, ACM, p. 731-734, 2015.
- Badache I., Boughanem M., « Emotional social signals for search ranking », *SIGIR*, ACM, p. 1053-1056, 2017a.
- Badache I., Boughanem M., « Fresh and Diverse Social Signals : Any Impacts on Search ? », *CHIIR, Oslo, Norway, March 7-11, 2017*, p. 155-164, 2017b.
- Barhoumi A., Camelin N., Estève Y., « Des représentations continues de mots pour l'analyse d'opinions en arabe : une étude qualitative », *TALN*, p. 215, 2018.
- Barhoumi A., Estève Y., Aloulou C., Belguith L. H., « Document Embeddings for Arabic Sentiment Analysis », *LPKM, Kerkennah (Sfax), Tunisia*, 2017.
- Bayoudhi A., Ghorbel H., Belguith L. H., « Sentiment Classification of Arabic Documents : Experiments with multi-type features and ensemble algorithms », *PACLIC*, 2015.
- Bouadjenek M. R., Hacid H., Bouzeghoub M., « Social networks and information retrieval, how are they converging ? A survey, a taxonomy and an analysis of social information retrieval approaches and platforms », *Inf. Syst.*, vol. 56, p. 1-18, 2016.
- Brin S., Page L., « The Anatomy of a Large-Scale Hypertextual Web Search Engine », *Computer Networks*, vol. 30, n° 1-7, p. 107-117, 1998.
- Buijs M., Spruit M. R., « The Social Score - Determining the Relative Importance of Webpages Based on Online Social Signals », *KDIR, Rome, Italy*, p. 71-77, 2014.
- Bullock B. N., Hotho A., Stumme G., « Accessing Information with Tags : Search and Ranking », *Social Information Access - Systems and Technologies*, p. 310-343, 2018.
- Chan L., Jhaver S., Soni S., « PostScholar : Surfacing Social Signals in Google Scholar Search », *CSCW '16 Companion*, ACM, New York, NY, USA, p. 17-20, 2016.
- Chelaru S., Orellana-Rodriguez C., Altingövde I. S., « How useful is social feedback for learning to rank YouTube videos ? », *World Wide Web*, vol. 17, n° 5, p. 997-1025, 2014.
- Dahou A., Xiong S., Zhou J., Haddoud M. H., Duan P., « Word Embeddings and Convolutional Neural Network for Arabic Sentiment Classification », *COLING*, p. 2418-2427, 2016.
- Damak F., Étude des facteurs de pertinence dans la recherche de microblogs. (Study of salient factors for microblog search), PhD thesis, Paul Sabatier University, Toulouse, France, 2014.
- Davison B. D., « Topical locality in the Web », *SIGIR*, p. 272-279, 2000.
- El-Halees A. *et al.*, « Arabic opinion mining using combined classification approach », 2011.
- ElSahar H., El-Beltagy S. R., « Building Large Arabic Multi-domain Resources for Sentiment Analysis », *CICLing*, p. 23-34, 2015.

- Farra N., Challita E., Assi R. A., Hajj H. M., « Sentence-Level and Document-Level Sentiment Mining for Arabic Texts », *ICDMW*, p. 1114-1119, 2010.
- Hall M. A., Holmes G., « Benchmarking Attribute Selection Techniques for Discrete Class Data Mining », *IEEE Trans. on Knowl. and Data Eng.*, vol. 15, n° 6, p. 1437-1447, 2003.
- Ibrahim H. S., Abdou S. M., Gheith M., « Sentiment Analysis For Modern Standard Arabic And Colloquial », *CoRR*, 2015.
- Joachims T., « Training linear SVMs in linear time », *ACM SIGKDD*, p. 217-226, 2006.
- Khodaei A., Shahabi C., « Social-Textual Search and Ranking », *Proceedings of the First International Workshop on Crowdsourcing Web Search, Lyon, April 17, 2012*, p. 3-8, 2012.
- Kim Y., « Convolutional Neural Networks for Sentence Classification », *EMNLP*, 2014.
- Kraaij W., Westerveld T., Hiemstra D., « The Importance of Prior Probabilities for Entry Page Search », *SIGIR, Tampere, Finland*, p. 27-34, 2002.
- Luo Z., Osborne M., Petrovic S., Wang T., « Improving Twitter Retrieval by Exploiting Structural Information », *AAAI, Toronto, Ontario, Canada*, 2012.
- Macdonald C., Ounis I., « Voting for candidates : adapting data fusion techniques for an expert search task », *ACM CIKM, Arlington, Virginia, USA*, p. 387-396, 2006.
- Magnani M., Montesi D., Rossi L., « Conversation retrieval for microblogging sites », *Inf. Retr.*, vol. 15, n° 3-4, p. 354-372, 2012.
- Mikolov T., Chen K., Corrado G., Dean J., « Efficient Estimation of Word Representations in Vector Space », *CoRR*, 2013.
- Mohammad S., Kiritchenko S., Zhu X., « NRC-Canada : Building the State-of-the-Art in Sentiment Analysis of Tweets », *SemEval@NAACL-HLT*, p. 321-327, 2013.
- Mohan A., Chen Z., Weinberger K. Q., « Web-Search Ranking with Initialized Gradient Boosted Regression Trees », *Yahoo! Learning to Rank Challenge, held at ICML*, p. 77-89, 2011.
- Nabil M., Aly M. A., Atiya A. F., « ASTD : Arabic Sentiment Tweets Dataset », *EMNLP*, 2015.
- Orso V., Ruotsalo T., Leino J., Gamberini L., Jacucci G., « Overlaying social information : The effects on users' search and information-selection behavior », *Inf. Process. Manage.*, 2017.
- Pang B., Lee L., Vaithyanathan S., « Thumbs up ? Sentiment Classification using Machine Learning Techniques », *EMNLP, Philadelphia, PA, USA*, 2002.
- Refaee E., Rieser V., « An Arabic Twitter Corpus for Subjectivity and Sentiment Analysis », *LREC, Reykjavik, Iceland*, p. 2268-2273, 2014.
- Refaee E., Rieser V., « iLab-Edinburgh at SemEval-2016 Task 7 : A Hybrid Approach for Determining Sentiment Intensity of Arabic Twitter Phrases », *SemEval*, p. 474-480, 2016.
- Socher R., Perelygin A., Wu J., Chuang J., Manning C. D., Ng A. Y., Potts C., « Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank », *EMNLP*, 2013.
- Turney P. D., « Thumbs Up or Thumbs Down ? Semantic Orientation Applied to Unsupervised Classification of Reviews », *ACL, Philadelphia, PA, USA.*, p. 417-424, 2002.
- Westerveld T., Kraaij W., Hiemstra D., « Retrieving Web Pages Using Content, Links, URLs and Anchors », *TREC, Gaithersburg, Maryland, USA*, 2001.
- Yen S.-J., Lee Y.-S., *Under-Sampling Approaches for Improving Prediction of the Minority Class in an Imbalanced Dataset*, Springer Berlin Heidelberg, p. 731-740, 2006.
- Yue Z., He D., « Collaborative Information Search », *Social Information Access - Systems and Technologies*, p. 108-141, 2018.