
Un modèle de langue pour l'estimation de la fraîcheur et la pertinence des documents Web

Mariam Bambia ^{*,**} — Mohand Boughanem ^{**} — Rim Faiz ^{*}

^{*} LARODEC, Université de Tunis, Le Bardo, Tunisie

mariembambia@yahoo.fr

Rim.Faiz@ihec.rnu.tn

^{**} IRIT, 118 Route de Narbonne, Toulouse, France

bougha@irit.fr

RÉSUMÉ. Un système de recherche d'information performant doit satisfaire les différents types de besoins des utilisateurs visant une variété de catégories de requêtes. Ces catégories comprennent les requêtes sensibles au temps où le contenu récent est l'exigence principale de l'utilisateur. Cependant, l'utilisation des caractéristiques temporelles des documents pour mesurer leur fraîcheur reste une tâche difficile étant donné que ces caractéristiques ne sont pas représentées avec précision dans les documents récents. Dans cet article, nous proposons un modèle de langue qui estime la pertinence et la fraîcheur des documents vis-à-vis des requêtes sensibles au temps réel. Notre approche modélise la fraîcheur en exploitant des sources d'information fraîches, plus précisément des termes extraits à partir des tweets récents et thématiquement pertinents par rapport à la requête de l'utilisateur. Nos expérimentations montrent que l'extraction des termes frais à partir de Twitter améliore les résultats de recherche requis par l'utilisateur.

ABSTRACT. An effective information retrieval system must satisfy different users search intentions expecting a variety of query categories, comprising recency sensitive queries where fresh content is the major user's requirement. However, using temporal features of documents to measure their freshness remains a hard task since these features may not be accurately represented in recent documents. In this paper, we propose a language model which estimates the topical relevance and freshness of documents with respect to real-time sensitive queries. In order to improve freshness ranking, our approach models freshness by exploiting terms extracted from recently posted tweets topically relevant to each real-time sensitive queries. Our experiments show that there is a clear advantage of using twitter to extract fresh keywords.

MOTS-CLÉS : Requêtes sensibles au temps, Modèles de langue, Mots-clés frais.

KEYWORDS: Real-time sensitive queries, Language models, Fresh keywords.

1. Introduction

Un système de recherche d'information performant doit être en mesure de fournir des résultats pertinents pour différentes catégories de requêtes (par exemple, les requêtes non-sensibles au temps, les requêtes sensibles à la localisation, les requêtes sensibles au temps réel) lancées par différents utilisateurs ayant des attentes différentes. Cependant, il existe des catégories de requêtes dont le contenu et la pertinence des documents peuvent varier au fil du temps. Cela est particulièrement important pour les requêtes sensibles au temps, où les résultats requis ne doivent pas être uniquement pertinents au sujet de la requête, mais doivent contenir également des informations fraîches répondant aux exigences de l'utilisateur [(Dai *et al.*, 2011), (Moon *et al.*, 2010)]. En effet, la fraîcheur est classée en quatrième position parmi les neuf raisons principales motivant les internautes à préférer un document Web par rapport aux autres documents répondant à leur requête¹. Par exemple, un utilisateur qui a soumis une requête sur une émission sportive cherche généralement le contenu télévisuel le plus frais, comme les matchs les plus récents et les dernières actualités du sport.

Toutefois, un événement qui a été diffusé la semaine dernière et qui a été classé comme étant pertinent pour une requête donnée, peut sans doute devenir moins important par rapport aux nouveaux événements frais et relatifs à cette thématique. Pour cette famille de requêtes, on peut proposer aux utilisateurs des options avancées pour filtrer explicitement les résultats de recherche à une période spécifique. Cette alternative présente deux limites principales : 1) plusieurs informations sont nécessairement spécifiées par l'utilisateur. Celui-ci est amené à investir l'effort au moyen de reformulation, bien que son besoin en contenu frais est immédiat, 2) seule une partie limitée du Web comme les flash infos peut répondre à des requêtes sensibles au temps réel, même si l'utilisateur spécifie explicitement la dimension temporelle, 3) extraire les caractéristiques temporelles des pages Web reste une tâche difficile, car elles peuvent ne pas être représentées correctement dans les documents frais. Par exemple, la requête "Débats présidentiels" où les documents pertinents sont sensibles aux jours, voire même aux heures, requiert la mesure de l'âge véritable des documents pertinents pour l'estimation de leur fraîcheur. L'exécution de ce type de requêtes complique la procédure de récupération du contenu le plus frais et l'estimation de cette notion de la fraîcheur des informations.

Le besoin en contenu frais est exprimé de plus en plus par les les internautes interagissant à travers les microblogs, les actualités et les réseaux sociaux. Par conséquent, les médias sociaux semblent être des sources d'informations intéressantes pour estimer la fraîcheur des documents Web (Damak *et al.*, 2012), (Ben Jabeur *et al.*, 2012), (Massoudi *et al.*, 2011), (Huo et J. Tsotras, 2010) and (Pasca, 2008). Dans cet article, nous estimons que la fraîcheur d'un document se réfère à un critère

1. <http://consumersunion.org/wp-content/uploads/2013/05/a-matter-of-trust.pdf>

qui dépend des occurrences des mots-clés frais dans ce document. Les mots-clés frais sont des termes qui font le Buzz et sont communément partagés dans le Web social et pertinents par rapport à la thématique spécifiée dans la requête. Dans notre cas, nous supposons que les mots-clés frais sont les termes qui apparaissent dans des ressources publiées récemment. Nous considérons que les microblogs, les dernières nouvelles et les tendances d'actualités sont les meilleures sources qui peuvent fournir ces mots-clés.

Les questions de recherche principales abordées dans cet article sont :

- Comment exploiter l'information sociale afin d'estimer la fraîcheur des documents afin de répondre aux requêtes sensibles au temps réel ?
- Comment agréger les deux critères de fraîcheur et de pertinence ?

Nous proposons un modèle de langue unifié pour l'estimation de la fraîcheur et la pertinence des documents Web. Cet article est structuré comme suit. La section 2 dresse un état de l'art sur les modèles proposés en Recherche d'Information (RI) pour l'estimation de la fraîcheur. La section 3 détaille le modèle de langue proposé. Section 4 présente la méthodologie expérimentale mise en œuvre et analyse les résultats obtenus. Finalement, nous concluons cet article et délimitons les perspectives.

2. Etat de l'art

Il existe plusieurs définitions de la fraîcheur des documents Web dans la littérature. Ces définitions diffèrent selon les mesures et les moyens utilisés pour caractériser les documents Web. La plupart des travaux antérieurs de classement de fraîcheur exploitent une variété de caractéristiques de la fraîcheur (par exemple le TimeStamp) pour le classement des documents. Ces caractéristiques se réfèrent à des signaux temporels et à des propriétés relatives aux requêtes et aux documents.

Selon (Dong *et al.*, 2010a), la notion de la fraîcheur introduit l'idée de l'extraction de l'âge d'un document. Ils ont proposé des caractéristiques de fraîcheur pour représenter la fraîcheur (par exemple Timestamp, Linktime, Webbuzz). Cependant, ils ont intégré les URLs fraîches extraites à partir de Twitter dans un système Web général. Ils ont utilisé ainsi un algorithme d'apprentissage automatique afin de prédire le classement approprié des résultats de recherche pour les requêtes, en se basant sur l'étiquette <requête, url>. Ils ont rajouté ensuite des caractéristiques extraites de Twitter, telles que les caractéristiques textuelles représentant l'ensemble des tweets contenant une URL donnée, afin d'identifier les URL fraîche (Dong *et al.*, 2010b).

(Karkali *et al.*, 2012) déclare un document comme étant frais s'il n'a pas été découvert auparavant par l'utilisateur et s'il est assez frais à l'égard de ses attentes et son historique de navigation. Ils ont proposé une variante de BM25 afin d'extraire

les mots-clés pertinents pour les intérêts des utilisateurs en tenant compte de son historique de navigation. Ils ont recueilli un ensemble de mots-clés des pages Web récemment parcourues par l'utilisateur, introduit la fréquence de chaque terme sur une période donnée et attribué un poids à chaque mot-clé afin d'en extraire les plus frais.

(Wang *et al.*, 2012) ont défini un ensemble de caractéristiques temporelles (par exemple le TimeStamp et le StoryAge de l'URL) permettant d'extraire toutes les dates mentionnées dans l'URL afin d'estimer sa fraîcheur et sa pertinence. (Inagaki *et al.*, 2010) ont déterminé un ensemble de caractéristiques temporelles, telles que « Webbuzz » qui représente le taux de mise à jour de l'URL et qui reflète la popularité des pages Web. Ils ont eu recours au taux de clics pondéré en fonction du temps afin de prélever la pertinence de l'URL pour une requête en calculant le rapport entre le nombre de sessions et celui des clics.

Dans le but d'estimer la variation temporelle de la pertinence des documents pour les requêtes sensibles au temps, (Moon *et al.*, 2012) ont analysé le retour des clics qui dépend du classement des résultats observés par l'utilisateur. (Dai *et al.*, 2011) ont introduit un framework supervisé qui exploite le profil temporel des requêtes en intégrant des caractéristiques de classement pour améliorer la fraîcheur des résultats de recherche. (Lee et Kim, 2013) ont conçu un modèle de clics pour détecter un point de temps, à partir duquel l'intention de recherche pour une requête donnée change.

En outre, plusieurs approches ont été proposées pour identifier et détecter les requêtes sensibles au temps [(Dakka *et al.*, 2012), (Dong *et al.*, 2010a) and (Diaz, 2009)]. Cependant, extraire les caractéristiques temporelles des pages Web constitue une tâche difficile vu qu'elles ne peuvent pas être représentées avec précision dans les documents récents.

Dans notre approche, nous supposons que la fraîcheur, en particulier les thématiques "fraîches", peut être fournie par des sources d'informations en temps réel telles que les actualités, les microblogs et les réseaux sociaux. Ces sources sont susceptibles d'englober des contenus potentiellement utiles à exploiter pour mesurer la fraîcheur des documents.

3. Un modèle de langue estimateur de fraîcheur et de pertinence

La pertinence est une mesure abstraite de la façon dont un document répond aux besoins de l'utilisateur représentés par une requête. Plusieurs travaux de recherche ont exploité les modèles de langue pour modéliser cette pertinence. Ces modèles se basent principalement sur le contenu des documents et exploitent la distribution des

mots dans les documents.

Sous sa version de base, un modèle de langue traite une requête comme étant un échantillon textuel et il estime la probabilité que la requête soit générée par le modèle du document. Plus précisément, la pertinence d'un document d vis-à-vis d'une requête Q est donnée par :

$$P(Q, d) = \prod_{t \in Q} P(t|d) \quad [1]$$

Afin d'éviter le problème de fréquences nulles qui se produit lorsqu'un document ne contient pas un terme de la requête, nous avons utilisé dans l'Équation 1 l'estimateur de lissage Dirichlet (Zhai et Lafferty, 2004), μ est la taille de l'échantillon sur $[0, 1[$:

$$P(t|d) = \frac{tf(t,d) + \mu P(t|c)}{|d| + \mu} \quad [2]$$

Les deux probabilités $P(t|d)$ et $P(t|c)$ sont estimées en utilisant l'estimateur du maximum de vraisemblance. Équation 2 permet d'estimer seulement la pertinence du document d à la requête Q .

Afin de favoriser les documents frais, nous supposons que la fraîcheur est décrite par un ensemble de termes frais connus a priori $F = \{f_1, f_2, \dots, f_n\}$. On tente ensuite d'évaluer la proximité des termes des documents de ceux qui sont frais.

Sur le plan théorique, la probabilité d'un terme peut être réécrite comme suit :

$$P(Q|d) = \prod_{t \in Q} \sum_{f \in F} P(t|f) * P(f|d) \quad [3]$$

Nous optons à mesurer $\hat{P}(t|f)$, qui permet d'estimer le degré d'indépendance entre le terme t et le terme frais f et de définir comment un terme t est proche de f (Equation 4). Cette probabilité peut être estimée à l'aide de la co-occurrence ou la probabilité conditionnelle classique (Bai *et al.*, 2006). Dans notre cas, nous l'estimons tout en prenant en considération le nombre de documents dans les quels t et f apparaissent simultanément. La formulation proposée n'est pas une probabilité pure. Nous avons utilisé le coefficient de Dice, mais on suppose que ce facteur permet d'approcher cette probabilité.

$$\hat{P}(t|f) = \frac{NbDoc(t,f)}{NbDoc(t)+NbDoc(f)} \quad [4]$$

Avec :

$Nbdoc(t)$: Nombre de documents contenant le terme t ,

$NbDoc(f)$: Nombre de documents contenant le terme frais f ,

$NbDoc(t, f)$: Nombre de documents contenant t et f .

Concernant $P(f|d)$, il est estimé en utilisant l'équation 5.

Nous obtenons :

$$P(f|d) = \frac{freq(f,d)}{|d|} \quad [5]$$

Avec :

$freq(f, d)$: La fréquence du terme frais f dans le document d ,

$|d|$: La somme des fréquences de tous les termes dans le document.

$$P(t|C) = \frac{tf(t)}{|C|} \quad [6]$$

Avec :

$tf(t)$: la fréquence du terme t dans tous les documents de la collection,

$|C|$: la somme des fréquences de tous les termes dans la collection,

$P(t|C)$ est la probabilité estimée d'observer le terme t dans la collection. Il constitue le rapport de la fréquence du terme t au le nombre des termes dans toute la collection des documents (Equation6).

4. Expérimentations

Nous menons une série d'expérimentations afin de comparer le modèle de langue proposé avec deux méthodes de référence : le modèle HBM25 proposé par (Karkali *et al.*, 2012) et le modèle surpondéré introduit par (Dong *et al.*, 2010b). Notre objectif consiste à réordonner une liste de documents tout en tenant compte des deux critères de classement : la fraîcheur et de la pertinence.

4.1. Données

Pour construire notre ensemble de données, nous avons collecté 1000 requêtes à partir des grands titres des actualités à partir de l'API «BBC»². En d'autres termes, chaque requête représente un grand titre extrait à partir de « BBC News ». Ces requêtes sont recueillies pendant quatre jours à partir de 08/03/2014 au 11/03/2014, 250 requêtes par jour. Ces requêtes sont susceptibles d'être sensibles aux jours, voire même aux heures. Chaque requête est soumise au moteur de recherche Google. Nous avons sélectionné ensuite les 40 premiers documents (pages Web). L'ensemble des pages Web sont recueillies pour chaque requête afin de former notre collection de documents (notée C). Nous visons à réordonner ces documents en fonction de leur fraîcheur et de leur pertinence. Afin de former notre collection de termes frais, nous avons utilisé l'API « Twitter 4j »³ pour extraire les tweets les plus récents en soumettant des hashtags représentant chaque requête. Nous avons sélectionné les tweets publiés une heure avant l'événement (l'actualité) jusqu'à 12 heures après l'événement. Ensuite, nous avons extrait les X premiers termes les plus fréquents à partir de ces tweets. Nous avons considéré différentes valeurs de $X=\{10, 20, 40, 100\}$ en fonction de leurs fréquences dans les tweets.

Dans la pratique, nous avons employé le moteur de recherche Lucene⁴ intégrant le lemmatiseur Porter et éliminant une liste de mots vides. Afin d'évaluer l'efficacité des résultats, nous avons proposé à 50 personnes de notre université une série de requêtes à évaluer. Chaque évaluateur a jugé 20 requêtes. Ils ont été invités à évaluer les documents en fonction de leur fraîcheur et leur pertinence. Chaque paire requête-document a été évalué le même jour que l'échantillonnage de la requête. Nous définissons deux qualités de pertinence : pertinent et non pertinent. Ensuite, nous leur demandons de juger la fraîcheur de chaque paire requête-document : frais ou non-frais.

Le tableau 1 montre que nous avons exprimé les annotations de pertinence et de fraîcheur séparées dans une seule note représentée par une valeur définie. Nous agrégeons les qualités de pertinence et de fraîcheur, de telle sorte qu'un document soit étiqueté pertinent s'il est frais par rapport à la thématique abordée.

Degré de Pertinence/ Degré de Fraîcheur	Frais	Non Frais
Pertinent	2	1
Non Pertinent	0	0

Tableau 1. *Étiquettes agrégées selon le degré de pertinence et de fraîcheur*

2. <https://developer.bbc.co.uk/>

3. <http://www.twitter4j.org/>

4. <http://www.lucene.apache.org/>

Nous avons utilisé la mesure d'évaluation Pertinence et Fraîcheur Cumulées et Actualisées (DCRF) qui permet de pénaliser les erreurs de classement tout en s'appuyant sur les jugements de pertinence et de fraîcheur effectués par les évaluateurs :

$$DCRF_n = \sum_{i=1}^n \frac{RF_i}{\log_2(i+1)} \quad [7]$$

Où i est la position dans la liste des documents réordonnés, n est le nombre de documents et RF_i est la note de la fraîcheur et de la pertinence selon les étiquettes agrégées et éditées par les évaluateurs (Tableau 1). Nous utilisons la MAP calculée en fonction des valeurs RF_i .

Vu que la précision dépend de l'accord entre les évaluateurs, nous avons eu recours au coefficient « Fleiss Kappa ». Cette mesure est envisagée afin de fournir une mesure quantitative de l'ampleur de l'accord entre les évaluateurs (J. Viera et M. Garrett, 2005).

4.2. Modèles de référence

Nous comparons notre modèle avec deux modèles de référence à savoir le modèle HBM25 et le modèle pondéré. Les trois approches tentent de réordonner les résultats de recherche initiale (les premiers documents retournés par Google).

Le modèle HBM25 est une variante de la fonction de pondération des termes BM25 présentée par (Karkali *et al.*, 2012). Il affecte à chaque terme d'un document un poids en se basant sur sa fréquence dans l'historique des pages parcourues par l'utilisateur sur une période donnée. Afin de mettre en œuvre cette méthode, nous avons défini un corpus de tweets envoyés par chaque utilisateur sur une période de temps, puis, nous avons introduit le modèle temporel de chaque terme figurant dans la collection des documents. Ensuite, nous avons employé la fonction de pondération des termes afin d'en extraire les mots-clés les plus frais.

Le second modèle de référence (TWM) est le modèle surpondérée proposé par (Dong *et al.*, 2010a). Ils ont utilisé des caractéristiques de fraîcheur représentées dans les documents récents, tels que les dates, pour fournir leurs évidences temporelles. Pour chaque requête, nous recueillons les paires requête sensible au temps-documents et les paires requête non sensible au temps-document. Ensuite, nous avons eu recours à la fonction de perte utilisée par (Dong *et al.*, 2010a) afin de générer des poids pour chaque paire requête-document en se basant sur le temps de soumission des requêtes et sur les dates figurant dans la collection des documents.

5. Résultats

Le tableau 2 énumère les *MeanDCRF* (MDCRF) et les *MAP* des trois approches pour les quatre jours d'évaluation. On remarque que plus l'événement est ancien, plus la *MAP* et la *MDCRF* des modèles de référence diminuent, tandis que le modèle *FREL* est toujours plus puissant en terme de *MAP* et *DCRF*. Cependant, un document peut devenir éventé quelques heures après, ce qui affecte sa pertinence par rapport à une requête sensible au temps. La *MAP* obtenue pour notre modèle *FRel* surpasse celles calculées pour les modèles HBM25 et TWM de plus de 20% chaque jour. Cela montre que l'extraction de mots-clés frais à partir des tweets récents est potentiellement indispensable pour améliorer le classement de fraîcheur pour les requêtes sensibles au temps réel.

Afin d'évaluer l'accord entre les différents évaluateurs, nous avons utilisé le coefficient Fleiss Kappa. Nous obtenons un Fleiss Kappa de 73,60% sur la fraîcheur des documents et de 78% sur la pertinence des documents, ce qui indique l'existence d'un accord substantiel entre les évaluateurs sur les deux critères.

Le tableau 2 montre que notre méthode obtient les *MDCRF* les plus élevées, soit 60% au cours des 4 jours, alors que les MDCRF des autres méthodes sont à moins de 50%.

Notre méthode surpasse le modèle surpondéré (Dong *et al.*, 2010a). Ceci peut être expliqué par l'hypothèse définie par ce modèle : un document est étiqueté « très frais » que si son contenu est créé le même jour de la requête (événement). Toutefois, les caractéristiques de fraîcheur dont les valeurs ne sont pas enregistrées en temps opportun, peuvent être inexactes et peuvent dégrader la pertinence de certains documents récents.

En ce qui concerne la méthode HBM25, certains termes sont pénalisés parce qu'ils apparaissent fréquemment dans la plupart des pages vues par l'utilisateur au cours d'une période donnée, alors qu'ils peuvent être frais quelques jours après.

Le Tableau 3 illustre un exemple des cinq documents récupérés à partir de la liste de recherche retournée par Google et réordonnés par notre modèle pour la requête 'Malaysian MH370 FLIGHT'. Selon les jugements de fraîcheur, nous soulignons que notre modèle favorise le contenu le plus pertinent et le plus frais en haut de la liste des documents réordonnés.

		HBM25	TWM	FRel
Jour 1	MAP	41.2	50.21	78
	MDCRF	50.01	39.30	64.20
Jour 2	MAP	41	50.10	69.22
	MDCRF	42.12	43.79	87.31
Jour 3	MAP	40.15	46.50	68.74
	MDCRF	36.30	42	68.92
Jour 4	MAP	41.22	35	75.60
	MDCRF	38.44	34.27	75.30

Tableau 2. *Comparaison de FRel avec les modèles de référence selon la MAP et la DCRF*

	rang	Document	Pertinence	Fraîcheur
(a) Résultats de recherche initiale retournés par le moteur de recherche Google				
rang dans (b)	1(4)	http://en.wikipedia.org/wiki/Malaysia_Airlines_Flight_370	pertinent	non-frais
	2(5)	http://www.malaysiaairlines.com/mh370	non pertinent	frais
	3(12)	http://www.independent.co.uk/news/world/australasia/\mh370-airline-boss-claims-missing-flight-did-not-crash-into-indian-ocean-\9790455.html http://www.redbullstratos.com/about-pilot-biography	non pertinent	non-frais
	4(2)	http://www.bbc.com/news/world-asia-26503141	pertinent	frais
	5(8)	http://www.atlantico.fr/dossier/disparition-\vol-mh370-malaysia-airlines-1015555.html	non pertinent	non-frais
(b) Reclassement des résultats par FRel				
rang dans (a)	1(9)	http://timesofindia.indiatimes.com/the-mystery-of-Malaysia-Airlines-flight-MH370/specialcoverage/32010069.cms	pertinent	frais
	2(4)	http://www.bbc.com/news/world-asia-26503141	pertinent	frais
	3(6)	http://www.redbullstratos.com/\the-team/felix-baumgartner/	pertinent	frais
	4(1)	http://en.wikipedia.org/wiki/Malaysia_Airlines_Flight_370	pertinent	non-frais
	5(2)	http://www.malaysiaairlines.com/mh370	non pertinent	frais

Tableau 3. Un exemple d'amélioration de classement selon la fraîcheur en se comparant aux résultats de recherche retournés par Google. La requête est "Malysian Flight", la durée de lancement de la requête est de 16 :31 à 20 :59

Nous avons également évalué l'impact du nombre de mots-clés frais. La Figure 1 montre que plus le nombre des mots-clés frais existants dans un document est élevé et varie au fil du temps, plus le *DCRF* et la *MAP* augmentent.

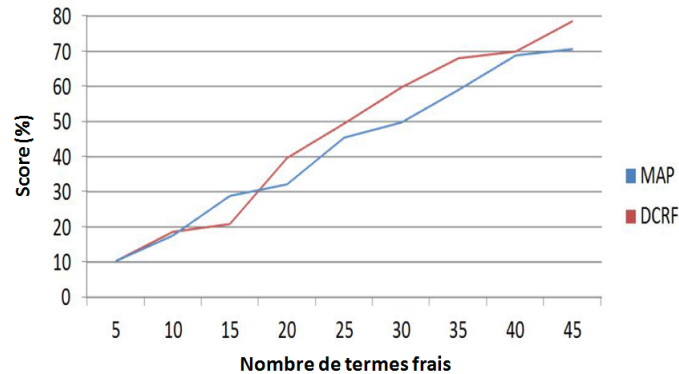


Figure 1. L'effet du nombre de mots-clés frais sur la performance des scores

6. Conclusion

Nous proposons un modèle de langue qui permet d'estimer la pertinence et la fraîcheur des documents Web à l'égard des requêtes sensibles au temps en se basant sur les mots-clés frais et pertinents aux thématiques extraits à partir des tweets les plus récents. Les résultats obtenus soulignent que notre approche surpasse les modèles de référence. Les expérimentations réalisées montrent qu'il est indispensable d'utiliser l'information sociale pour mesurer la fraîcheur des documents afin de répondre aux requêtes sensibles au temps. Nous avons également montré l'importance d'utiliser Twitter comme étant une source pour l'extraction des termes frais récurrents dans les microblogs et décrivant les événements les plus récents. Nos résultats indiquent ainsi l'importance des modèles de langue pour améliorer le classement de fraîcheur des documents Web. Les résultats encourageants de notre modèle ouvrent plusieurs orientations futures : Nous envisageons dans un premier temps d'intégrer notre modèle dans les systèmes de recommandation afin de proposer des programmes de télévision les plus pertinents et les plus frais répondant aux différentes attentes des utilisateurs. Un autre axe de recherche sera d'étendre la personnalisation par d'autres caractéristiques de classement, tel que le comportement social de l'utilisateur, et d'exploiter les annotations pour traiter la fraîcheur des émissions TV.

Bibliographie

Bai J., yun Nie J., Cao G., « Context-dependent term relations for information retrieval », *Proc. Empirical Methods in Natural Language Processing (EMNLP '06)*, p. 551-559, 2006.

- Ben Jabeur L., Tamine L., Boughanem M., « Featured tweet search : Modeling time and social influence for microblog retrieval », *Proceedings of International Conference on Web Intelligence*, China, p. 166-173, 2012.
- Dai N., Shokouhi M., Davison B., « Learning to Rank for Freshness and Relevance », *Proceeding of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, p. 95-104, 2011.
- Dakka W., Gravano L., G. Ipeirotis P., « Answering General Time-Sensitive Queries », *Proceedings of the IEEE Transactions on Knowledge and Data Engineering*, vol. 24, p. 220-235, 2012.
- Damak F., Pinel-Sauvagnat K., Guillaume C. G., « Recherche de microblogs : quels critères pour raffiner les résultats des moteurs usuels de RI ? », *Conférence francophone en Recherche d'Information et Applications (CORIA)*, Bordeaux, France, LABRI, p. 317-328, mars, 2012.
- Diaz F., « Integration of news content into web results », *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, ACM, New York, NY, USA, p. 182-191, 2009.
- Dong A., Chang Y., Zheng Z., Mishne G., Bai J., Zhang R., Buchner K., Liao C., Diaz F., « Towards recency ranking in web search », *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, ACM, New York, NY, USA, p. 11-20, 2010a.
- Dong A., Zhang R., Kolari P., Bai J., Diaz F., Chang Y., Zheng Z., Zha H., « Time is of the essence : improving recency ranking using Twitter data », *Proceedings of the 19th international conference on World Wide Web*, WWW '10, ACM, New York, NY, USA, p. 331-340, 2010b.
- Huo W., J. Tsotras V., « Temporal top-k search in social tagging sites using multiple social networks », *Proceedings of the 15th International Conference on Database Systems for Advanced Applications*, p. 498-504, 2010.
- Inagaki Y., Sadagopan N., Dupret G., Dong A., Liao C., Chang Y., , Zheng Z., « Session Based Click Features for Recency Ranking », *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, 2010.
- J. Viera A., M. Garrett J., « Understanding Interobserver Agreement : The Kappa Statistic », *Family Medecine Research Series*, vol. 37, n° 5, p. 360-363, 2005.
- Karkali M., Plachouras V., Vazirgiannis M., Stefanatos C., « Keeping Keywords Fresh : A BM25 Variation for Personalized Keyword Extraction », *Proceedings of the 2nd Temporal Web Analytics Workshop*, p. 17-24, 2012.
- Lee S., Kim D., « A click model for time-sensitive queries », *Proceedings of the 22nd International World Wide Web Conference*, WWW '13, p. 147-148, 2013.
- Massoudi K., Tsagkias M., Rijke M., Weerkamp M., « Incorporating Query Expansion and Quality Indicators in Searching Microblog Posts », *Proceedings of the 33rd European Conference on IR Research*, Dublin, Ireland, p. 362-367, 2011.
- Moon T., Chu W., Lihong L., Zheng Z., Chang Y., « Online learning for recency search ranking using real-time user feedback », *ACM Transactions on Information Systems*, vol. 30, n° 4, p. 20, 2010.
- Moon T., Li L., Liao C., Zheng Z., Chang Y., « An online learning framework for refining recency search results with click feedback », *ACM Transactions on Information Systems*, vol. 30, n° 4, p. 20, 2012.

- Pasca M., « Towards temporal web search », *Proceedings of the 2008 ACM symposium on Applied computing*, ACM, New York, NY, USA, p. 1117-1121, 2008.
- Wang H., Dong A., Li L., Chang Y., « Joint Relevance and Freshness Learning From Click-throughs for News Search », *Proceedings of the 21st International World Wide Web Conference Committee*, p. 579-588, 2012.
- Zhai C., Lafferty J., « A Study of Smoothing Methods for Language Models applied to Information Retrieval », *ACM Transactions on Information Systems*, vol. 2, n^o 2, p. 179-214, 2004.