Recherche de microblogs : quels critères pour raffiner les résultats des moteurs usuels de RI ?

Firas Damak — Karen Pinel-Sauvagnat — Guillaume Cabanac

Université de Toulouse – IRIT UMR 5505 CNRS
118 route de Narbonne, F-31062 Toulouse cedex 9
prenom.nom@irit.fr

RÉSUMÉ. *Depuis quelques années, les services de microblogs, comme Twitter, attirent l'attention des internautes. Cet attrait peut s'expliquer par la facilité et la rapidité avec laquelle les internautes peuvent partager des informations, le plus souvent en temps réel. Les microbloggeurs, en parallèle de leur publication de microblogs, cherchent également souvent à collecter des informations récentes sur leurs derniers sujets d'intérêt. Trouver les meilleurs résultats pour un sujet demeure dépendant des caractéristiques des microblogs (comme par exemple la longueur très faible des messages, la qualité du langage utilisé, la fiabilité du diffuseur…). Dans cet article, nous proposons d'évaluer l'impact de certains critères sur la qualité des résultats renvoyés sur un sujet donné. Nous considérons trois groupes de critères : ceux liés au contenu des microblogs, ceux liés à leur hypertextualité et enfin ceux liés aux auteurs des microblogs. Nos résultats expérimentaux utilisent la collection TREC Tweets 2011 et montrent l'intérêt de critères tels que l'hypertextualité.*

ABSTRACT. *The last few years saw the advent of microblogging platforms, such as Twitter. Such an appeal may be due to platforms' features allowing to easily and quickly share information in real-time. Microbloggers, while posting microblogs, search for fresh information related to their interests. Finding good results concerning the given subjects needs to consider the characteristics of microblogs, such as short length of messages, poor syntax and credibility of the sender. In this paper, we evaluate the impact of some microblog features on search result quality. We consider three groups of features: content-based, hypertext-based, and author-based. Our experiments on the TREC Tweets 2011 collection show the interest of hypertext-based features.*

MOTS-CLÉS : *Microblog, Twitter, moteur de recherche*

KEYWORDS: *Microblog, Twitter, search engine*

CORIA 2012, pp. 317–328, Bordeaux, 21-23 mars 2012

1. Introduction

Les plate-formes de *microblogging* sont des réseaux sociaux au travers desquels les utilisateurs (microbloggeurs) partagent des messages de faible longueur. Outre la publication de tweets, les utilisateurs cherchent également des informations récentes sur un sujet particulier. À ce niveau, deux questions peuvent être posées : les moteurs de recherche usuels sont-ils toujours utiles pour ce nouveau type de média ? Quelles doivent être les caractéristiques d'un moteur de recherche de *microblogs* ?

Les motivations des internautes pour chercher des *microblogs* sont diverses (Teevan *et al.*, 2011). Certaines sont similaires à la recherche sur le web (comme par exemple la recherche d'actualités), et d'autres sont spécifiques à la recherche de *microblogs* (comme par exemple la recherche temps réel ou d'informations sociales). Les *microblogs* ont certaines caractéristiques spécifiques. En considérant la plate-forme Twitter, par exemple, les messages sont limités à 140 caractères, utilisent une syntaxe spécifique (`RT` pour re-tweet, `#` pour hashtag, `@` pour mention). Les messages sont similaires aux SMS de par leur syntaxe appauvrie et la prépondérance de mots écrits phonétiquement. Enfin, les *microblogs* peuvent référencer des sources externes à l'aide d'hyperliens. Par conséquent, un moteur de recherche de *microblogs* doit prendre en compte leurs spécificités ainsi que de nouvelles exigences des utilisateurs en termes de fraîcheur, de nouveauté d'information et d'importance du réseau social, par exemple. Ces différentes caractéristiques sont concrètement prises en compte dans les moteurs de recherche actuels en combinant des critères pour mesurer la pertinence des *microblogs* vis-à-vis d'un besoin en information. Par exemple, en considérant le facteur importance de l'auteur, les critères associés pourraient être le nombre de tweets de l'auteur et le nombre de ses disciples/abonnés (Nagmoti *et al.*, 2010). Nous pouvons également considérer le nombre de fois qu'un utilisateur a été mentionné ou bien le score de l'auteur selon un algorithme semblable à PageRank basé sur des relations de rediffusion des messages (Ben Jabeur *et al.*, 2011).

Dans cet article, nous proposons d'évaluer l'impact de certains critères sur la qualité des résultats restitués pour un sujet donné. Nous considérons trois groupes de critères : ceux liés au contenu des *microblogs*, ceux liés à leur hypertextualité et enfin ceux liés aux auteurs des *microblogs*. Nous basons notre travail sur la collection d'évaluation proposée pour la tâche Microblog de TREC 2011. Cet article est organisé comme suit. La section 2 donne quelques clés sur le fonctionnement des plate-formes de *microblogging*. La section 3 dresse un état de l'art des approches actuelles pour la recherche d'information (RI) dans les *microblogs*. La section 4 détaille les critères que nous nous proposons d'évaluer. La section 5 présente la méthodologie expérimentale mise en œuvre. La section 6 commente les résultats obtenus. Enfin, nous concluons cet article et délimitons des perspectives à notre travail dans la section 7.

2. Qu'est-ce que le *microblogging* ?

Les plate-formes de *microblogging* sont les réseaux sociaux les plus récents du Web 2.0, et peuvent être considérées comme une nouvelle forme de blogs, où les informations diffusées sont plus courtes mais aussi publiées plus souvent et plus rapidement. Les utilisateurs publient des mises à jour ou des statuts dans leurs profils sous forme de

messages de faible longueur, souvent limités à 140 caractères, ce qui réduit le temps nécessaire pour rédiger le contenu. Parmi les plate-formes de *microblogging* exitantes, on peut citer Identi[1], Google+[2] ou encore Twitter[3]. Twitter, le service le plus populaire de *microblogs*, a vécu une croissance exponentielle ces dernières années. Nous allons nous appuyer sur cet exemple pour décrire le fonctionnement des plate-formes de *microblogging*.

De nombreux usages ont fait leur apparition avec l'écriture des tweets, qui ont peu à peu évolués pour devenir des « normes de balisage » :

– `@` suivi par un nom d'utilisateur permet d'indiquer qu'on mentionne ou s'adresse à une personne particulière,

– `RT` indique que le message est un retweet : le mécanisme de reweet permet aux utilisateurs de rediffuser des tweets qu'ils trouvent intéressants parmi les tweets publiés par leurs amis (par exemple, `RT @mashable Top 10 Twitter Trends This Week http://on.mash.to/eA2jY5`),

– `#` suivi par un mot est un hashtag. Un hashtag indique un mot important que le système peut utiliser pour permettre une recherche par navigation. Les hashtags permettent de catégoriser les tweets selon un contexte (événement, lieu, etc.). Par exemple : `...par des personnalités extérieures, c'est le #doctorat", défend Cédric Villani.`

– Les tweets peuvent également contenir des URL. Ces hyperliens sont souvent raccourcies avec des services tels que `bit.ly` et `tinyurl.com`, en raison du nombre limité de caractères pour chaque tweet.

Un utilisateur de Twitter A peut suivre le flux de tweets envoyés par un utilisateur B sans avoir besoin de sa permission. Les relations entre utilisateurs des réseaux sociaux sont appelées des abonnements. Si A est abonné à B, alors A est appelé disciple de B et reçoit automatiquement toutes les publications de B. Les relations d'abonnement peuvent être dans un seul sens, mais également dans les deux sens si B s'abonne à son tour à A. Un microbloggeur peut également envoyer un message direct et privé à l'un de ses amis.

Pour conclure, les plate-formes de *microblogging* (Twitter en particulier), représentent un nouveau type de média en pleine évolution grâce à un ensemble de caractéristiques spécifiques :

– de fonctionnalité, telles que le partage d'information temps réel, les abonnements sans restriction, etc. Ces nouvelles fonctionnalités ont généré de nouvelles habitudes comme le suivi des célébrités, la réalisation de campagnes électorales, l'analyse de l'humeur et des avis des gens temps réel, la participation à distance à des conférences.

– de forme, telles que la longueur faible des messages, l'utilisation du jargon du net, une syntaxe spécifique, etc. Ces critères sont aujourd'hui très étudiés par le monde de la

1. `http://www.identi.ca`
2. `http://www.plus.google.com`
3. `http://www.twitter.com`

recherche, afin de trouver des moyens pour les exploiter, et ainsi rendre les plateformes plus performantes en tant que moyen de partage et de source d'information.

3. État de l'art sur la recherche dans des microblogs

Les travaux de la littérature peuvent être regroupés en deux catégories. Certains travaux portent sur l'analyse statistique des caractéristiques des *microblogs*. Ainsi ont notamment été étudiées la réciprocité entre les microbloggeurs (Java *et al.*, 2007) et la structure topologique et géographique du réseau social induit (Huberman *et al.*, 2008). D'autres travaux portent sur les tâches de recherche d'information au sens large (accès à l'information) à partir des *microblogs*. Notre travail étant lié à ce second point, nous détaillons les différents types de recherche dans les sections suivantes.

3.1. *Recherche de microbloggeurs*

Les plate-formes de *microblogging* sous-tendent des réseaux sociaux. Aussi, plusieurs travaux se sont focalisés sur l'identification des utilisateurs les plus populaires. TwitterRank (Weng *et al.*, 2010) est une approche inspirée de l'algorithme PageRank (Brin *et al.*, 1998) qui mesure la popularité des utilisateurs de Twitter. Le score de chaque utilisateur est mesuré en fonction des scores de ses disciples. Cette approche prend en considération les similarités des sujets discutés entre les utilisateurs, ainsi que la structure des liens d'abonnements. Cependant, en analysant les habitudes de diffusions d'information dans Twitter, Lee *et al.* (2010) ont découvert que la diffusion d'information atteint son paroxysme durant la première période de son apparition. Par conséquent, ils ont proposé une approche considérant l'ordre temporel d'adoption de l'information pour détecter le meilleur diffuseur d'information.

3.2. *Classification thématique des microblogs*

Des travaux se sont intéressés à la classification thématique de tweets, et par extension, d'utilisateurs. Cela permet, notamment, de construire des filtres thématiques sur les flux d'information. Une première solution pour ce type de problème est de regrouper les *microblogs* en fonction des hashtags qu'ils contiennent (Efron, 2010). Par ailleurs, les hashtags peuvent être extraits des résultats de recherche pour étendre ensuite la requête initiale. Ramage *et al.* (2010) ont utilisé une implémentation étiquetée de LDA (Latent Dirichlet Allocation) afin d'extraire des tags et de les utiliser pour caractériser les utilisateurs et les *microblogs*. Par ailleurs, des informations spatio-temporelles des tendances ont été exploitées dans (Song *et al.*, 2010) afin d'identifier les tags co-occurents. Ces tags sont utilisés par la suite pour regrouper les tweets dans des classes. Cependant, cette approche génère une distribution de termes liés thématiquement plutôt que des sujet significatifs. Enfin, Bernstein *et al.* (2010) ont considéré que les sujets présentés sous forme de distributions de termes ne sont pas très utiles, et ont proposé un algorithme pour détecter précisément les sujets des *microblogs*. Ce dernier consiste à détecter les entités nommées dans un *microblog* et les soumettre à un moteur de recherche. Le sujet du *microblog* correspondra alors au terme le plus important dans les résultats, calculé à travers un algorithme de pondération (TF·IDF (Sparck Jones, 1988)).

3.3. *Recherche de microblogs*

Les travaux s'intéressant à la recherche de *microblogs* peuvent se diviser en deux classes principales :

La première classe est constituée des recherches qui réalisent uniquement l'ordonnancement des *microblogs* résultant d'un moteur de recherche . Ces travaux considèrent que le fait d'ordonner les résultats dans un ordre chronologiquement inverse ne garantit pas que les meilleurs résultats apparaissent en tête de liste. Dans (Nagmoti *et al.*, 2010), un ensemble de critères, tel que la présence d'une URL et la popularité d'un auteur en fonction du nombre de publications et du nombre de ses abonnés, ont été calculés. Les scores des critères sont ensuite simplement sommés afin de produire un score final.

La deuxième classe contient les travaux qui cherchent à extraire ainsi qu'à ordonner les *microblogs* dans un même temps. Duan *et al.* (2010) ont proposé une approche qui collecte et ordonne les *microblogs* par apprentissage. Ils ont utilisé le modèle BM25 (Robertson *et al.*, 1994) pour calculer la pertinence du contenu, ainsi que des critères spécifiques à Twitter, tels que la fréquence des hashtags et le nombre de fois un *microblog* a été rediffusé, et d'autres pour mesurer l'importance des auteurs tels que le nombre de leurs abonnés. Dans le même objectif, un modèle de langue a été proposé dans (Massoudi *et al.*, 2011). Les auteurs ont également défini une approche dynamique pour étendre les requêtes, qui suppose que les termes publiés au même moment que la requête sont *a priori* les plus pertinents.

Généralement, toutes les approches qui tentent d'extraire de l'information des *microblogs* utilisent des critères différents pour compenser l'impuissance des systèmes de recherche à gérer les nouvelles spécificités de ce média. La majorité de ces critères ont été traités bien avant l'apparition des *microblogs*. Parmi eux on peut citer la longueur des textes (Singhal *et al.*, 1996) ainsi que le réseau des liens hypertextes (Kleinberg, 1999). Cependant, très peu de travaux ont évalué la valeur ajoutée par ces critères dans les cas des *microblogs*. Aussi, dans cet article, nous proposons d'évaluer l'impact de certains critères sur la qualité des résultats restitués pour un sujet donné.

4. Quels critères pour raffiner les résultats de RI ?

Nous décrivons dans cette section les sept critères que nous considérons, divisés en trois groupes. Nous utiliserons les notations suivantes par la suite :

– q est la requête (composée de mots-clés et caractérisée par une date),

– C_q est le corpus des tweets publiés avant la date de la requête,

– T_q est l'ensemble des tweets restitués par un moteur de recherche donné calculant la pertinence par rapport à q uniquement sur le contenu des tweets ($T_q \subseteq C_q$),

– t est un tweet $\in T_q$ sur lequel on applique un critère.

4.1. *Critères basés sur le contenu des tweets*

Nous avons considéré trois critères issus de l'état de l'art et relatifs au contenu des tweets.

– Popularité du thème du tweet (Duan *et al.*, 2010) : ce critère estime la popularité d'un tweet dans T_q. On suppose qu'un tweet est populaire si plusieurs autres tweets ont un contenu similaire. La similarité $sim(\vec{t_i}, \vec{t_j})$ entre chaque paire de tweets est calculée avec un modele vectoriel qui considère également la fréquence des termes de la requête dans le tweet (Cohen *et al.*, 2007). $\vec{t_i}$ étant le vecteur contenant les termes du tweet t_i :

$$f_1(t_i, q) = \frac{\sum_{t_{j \neq i} \in T_q} sim(\vec{t_i}, \vec{t_j})}{|T_q| - 1} \quad [1]$$

– Longueur du tweet (Duan *et al.*, 2010) : intuitivement, plus un message est long, plus il contient de l'information. Nous avons calculé ce critère en comptant le nombre de termes dans un tweet. On note $l(t_i)$ le nombre de termes dans un tweet $t_i \in T_q$. Ce critère est calculé de la manière suivante :

$$f_2(t_i, q) = \frac{l(t_i)}{max_{t_j \in T_q} l(t_j)} \quad [2]$$

– Hashtags dans le tweet (Duan *et al.*, 2010) : Les microbloggeurs peuvent catégoriser ou suivre des sujets à l'aide des hashtags. On note la fréquence d'un hashtag dans le corpus C_q par $freq(h)$:

$$f_3(t_i) = \sum_{h \in t_i} freq(h) \quad [3]$$

4.2. *Critères basés sur l'hypertextualité*

Nous considérons deux critères additionnels qui peuvent indiquer la qualité de l'information publiée dans les tweets :

– Présence d'une URL dans le tweet (Nagmoti *et al.*, 2010; Zhao *et al.*, 2011) : partager des URL est une manière de confirmer l'information publiée dans un tweet. Ceci permet également d'attirer l'attention sur un contenu présent sur le web. Ainsi, on fait l'hypothèse que la présence d'une URL indique que le tweet a un caractère informatif renforcé. Ce critère est binaire :

$$f_4(t_i, q) = \begin{cases} 1 & \text{si } t_i \text{ contient une URL} \\ 0 & \text{sinon} \end{cases} \quad [4]$$

– Fréquence de l'URL dans le corpus : ce critère permet de calculer la popularité des URL publiées dans un tweet dans le corpus C_q. On note par $frq(url)$ le nombre de fois qu'une URL apparaît dans le corpus C_q :

$$f_5(t_i) = \sum_{url \in t_i} frq(url) \quad [5]$$

4.3. *Critères basés sur la popularité des auteurs*

Nous avons considéré deux critères spécifiques aux auteurs de *microblogs*.

– Nombre de tweets de l'auteur (Nagmoti *et al.*, 2010) : l'objectif de ce critère est de valoriser les tweets publiés par des auteurs actifs par rapport aux tweets publiés par des auteurs moins actifs. L'idée de ce critère est que les utilisateurs actifs ont plus de valeur en tant que sources d'information que des utilisateurs moins actifs. On note par $a(t_i)$ l'auteur du tweet t_i et $n(a(t_i))$ le nombre de tweets publiés par l'auteur du tweet t_i dans le corpus C_q.

$$f_6(t_i) = n(a(t_i)) \quad [6]$$

– Nombre de références de l'auteur (Zhao *et al.*, 2011) : plus un auteur est mentionné, plus il est populaire. $m(a(t_i))$ indique combien de fois un auteur du tweet t_i a été mentionné dans le corpus C_q :

$$f_7(t_i) = m(a(t_i)) \quad [7]$$

Certains critères présentés semblent contradictoires aux habitudes de la recherche d'information, tels que la longueur des messages qui est habituellement considérée comme facteur négatif, ou la fréquence des URLs et des hashtags qui expriment dans le cas des *microblogs* de l'importance et de la valeur ajoutée. Ceci s'explique par la faible longueur des *microblogs* (au plus 140 caractères dans Twitter) par rapport aux documents usuels, incitant les auteurs des microblogs à décrire exactement l'information souhaitée, sans avoir la possibilité d'entrer dans les détails.

5. Démarche d'évaluation

Cette section détaille la tâche microblog de TREC 2011, puis le système que nous avons mis en place, et enfin notre protocole d'évaluation.

5.1. *Tâche Microblog de TREC 2011*

Il s'agit, pour un moteur de recherche, de fournir les tweets dont le contenu satisfait un besoin (*topic*) sous forme de mots clés. La collection de test Tweets2011 comprend :

– 16 millions de tweets exprimés dans diverses langues et publiés sur Twitter entre le 23 janvier 2011 et le 8 février 2011. Chaque tweet est caractérisé par son identifiant (`ID`), son auteur et sa date de publication,

– 49 *topics* dont chacun est composé de plusieurs balises. La balise `title` décrit le besoin exprimé à un moment donné (`querytime`),

– les jugements de pertinence (*qrels*) associées aux 49 *topics*. La pertinence de chaque tweet est ternaire : non pertinent, moyennement pertinent et hautement pertinent.

De façon usuelle, les résultats d'un moteur de recherche sont évalués selon le score des documents. Ce n'est pas le cas dans la tâche Microblog, qui promeut la recherche temps réel (*real-time search*). Cela se traduit par une préférence pour les tweets les plus proches temporellement du *topic*. Au niveau de la procédure d'évaluation, cette contrainte est mise en œuvre en réordonnant les résultats (*runs*) d'un moteur de recherche en fonction de leur proximité temporelle à la requête (le champ `sim`, score de similarité du *run* est recalculé en fonction). Enfin, le *run* est évalué avec le programme trec_eval fourni par le NIST. Une seule mesure officielle a été considérée : la précision à 30 documents (P@30). Le classement des systèmes a été réalisé sur la P@30 moyennée, la MAP étant uniquement donnée à titre indicatif.

5.2. *Notre système*

La première étape de notre système consiste à indexer et extraire les top-N tweets pertinents pour chaque requête (figure 1). Nous nous sommes basés sur la version du moteur de recherche open-source Lucene[4] fournie par les organisateurs de TREC.

4. `http://lucene.apache.org`

Notre système se voulant « *real-time* », ne sont indexés pour chaque requête que les tweets ayant été publiés avant la date de la requête. Nous passons ensuite au traitement de ces résultats afin de produire des scores pour les critères. Le score final d'un tweet (équation 9 avec α = 0,5) est calculé en combinant le score de Lucene et les scores des critères (équation 8). Le score des critères est calculé par une simple combinaison linéaire $f()$ et sans pondération. On réalise différentes normalisations de sorte que $f_n(t_i, q) \in [0, 1]$ et $f_{sources}(t_i, q) \in [0, 1]$.

$$f_{criteres}(t_i, q) = f(f_1(t_i, q), f_2(t_i, q), f_3(t_i), f_4(t_i, q), f_5(t_i), f_6(t_i), f_7(t_i)) \quad [8]$$

$$score(t_i, q) = \alpha * lucene(t_i, q) + (1 - \alpha) * f_{criteres}(t_i, q) \quad [9]$$

Les tweets sont enfin filtrés pour ne garder que les tweets en anglais.

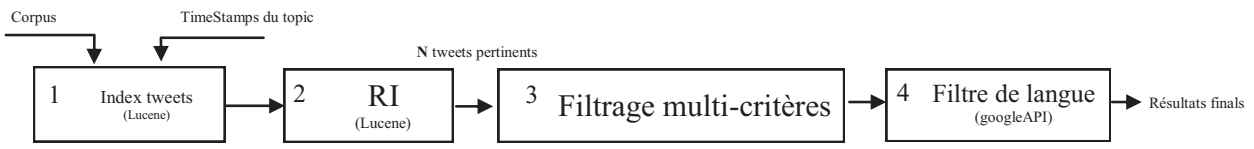

Figure 1. *Processus d'indexation et de recherche mis en œuvre dans notre système*

5.3. *Évaluation de notre système*

Les résultats présentés par la suite sont évalués en fonction d'un classement sur le seul score de pertinence, contrairement à la tâche Microblog qui évalue en réordonnant les résultats sur la date des tweets au préalable, ce qui ne rend pas compte de la qualité des critères. Nous utilisons les 5000 premiers résultats renvoyés par Lucene.

6. Résultats

Dans cette section, nous comparons l'apport des différents critères. Puis, nous positionnons nos résultats par rapport aux résultats officiels de la tâche Microblog à TREC 2011. Enfin, nous généralisons nos conclusions en faisant abstraction de Lucene.

6.1. *Comparaison des différents critères*

Le tableau 1 montre les résultats obtenus en considérant les critères décrits dans la section 4 un par un. Un astérisque indique que la différence est statistiquement significative selon le test t de Student (1908) pairé et bilatéral avec $p < 0{,}05$.

Comme nous pouvons le constater, et de façon surprenante, tous les critères, sauf un (f_4), conduisent à une dégradation des résultats. Concernant f_4 (présence d'une URL dans le tweet), on observe une hausse des résultats (+12,5 % sur la P@30).

Le tableau 1 ne nous permet cependant pas de voir les interactions entre les différents critères : peuvent-ils se compléter afin d'améliorer les résultats ? Afin d'étudier ce dernier point, il faudrait effectuer toutes les combinaisons possibles des critères entre eux (246 runs). Pour simplifier cette étude des interactions entre les critères, nous avons

Système	MAP	P@30
Lucene	0,3141	0,3544
Lucene+f_1	0,2280*	0,3027*
Lucene+f_2	0,2241*	0,2701*
Lucene+f_3	0,2078*	0,2823*
Lucene+f_4	**0,3348**	**0,3986***
Lucene+f_5	0,2570*	0,3238
Lucene+f_6	0,1724*	0,1850*
Lucene+f_7	0,2475*	0,3245*

Tableau 1. *Apport de chaque critère par rapport à Lucene*

Système	MAP	P@30
Lucene	0,3141	0,3544
Lucene+G1	0,2464*	0,2925*
Lucene+G2	**0,3168**	0,3687
Lucene+G3	0,1589*	0,1837*
Lucene+G1+G2	0,3144	**0,3810**
Lucene+G1+G3	0,2315*	0,2592*
Lucene+G2+G3	0,2718*	0,3109*
Lucene+G1+G2+G3	0,2990	0,3565

Tableau 2. *Apport de chaque groupe de critères et de leurs combinaisons par rapport à Lucene*

décidé d'observer le comportement des différents groupes qu'ils forment : Groupe G1 (f_1, f_2 et f_3), Groupe G2 (f_4 et f_5) et Groupe G3 (f_6 et f_7). Les résultats sont décrits dans le tableau 2. Les combinaisons qui améliorent la précision à 30 sont les groupes qui contiennent G2, contenant le critère f_4. Ceci tend à confirmer que le critère f_4 a un intérêt dans notre système. On remarque également que la combinaison de G1 et G2 donne de meilleurs résultats que G2 seul. Ceci montre que des critères de G1 peuvent améliorer les résultats que lorsqu'ils sont combinés avec d'autres.

6.2. *Comparaison avec les résultats officiels*

Nous avons comparé notre meilleur run (Lucene + f_4) avec les résultats officiels de la tâche Microblog de TREC 2011 (Ounis *et al.*, 2011). Les résultats figurent dans le tableau 3. Afin que la comparaison soit équitable, seuls sont présents dans le tableau les runs officiels automatiques n'utilisant pas de ressources externes et fonctionnant en temps réel, c'est-à-dire n'utilisant pas d'évidence future (données produites après la date de la requête). Nous rappelons que lors de l'évaluation officielle, les tweets doivent être ordonnés par ordre chronologique inverse. Notre run est coupé à 30 résultats afin d'éviter le biais introduit par le tri chronologique, assimilable à l'introduction d'un critère indépendant de la pertinence qui introduit un aléas non souhaitable. Ainsi, les résultats du tableau 3 diffèrent sur la MAP par rapport au tableau 1. Les résultats sans coupe de notre run sont également présentés dans le tableau. On note l'effet négatif sur les résultats du tri chronologique des tweets, et ce sur les deux mesures.

Ce run nous aurait permis de nous placer à la 5^e position des participants selon la P@30. Ces résultats améliorent notre participation officielle dont les détails sont donnés dans (Damak *et al.*, 2011).

6.3. *Généralisation des résultats*

Les résultats que nous avons obtenus et présentés dans les sections précédentes sont liés à la performance du moteur Lucene : ce sont sur les tweets renvoyés par Lucene que nous appliquons les critères. On pourrait donc penser que le score final d'un tweet dépend fortement du score de Lucene. C'est pourquoi nous avons cherché à

Groupe	Run	MAP	P@30
isi	isiFDL	0,1923	0,4551
FUB	DFReeKLIM30	0,2348	0,4401
CLARITY_DCU	clarity1	0,2139	0,4211
Purdue_IR	myrun2	0,2003	0,3993
IRIT	Lucene + f_4 coupé à 30	0,1843	0,3986
IRIT	Lucene + f_4 coupé à 1000	0,1549	0,1272
IRIT	Run officiel	0,1940	0,2565

Tableau 3. *Comparaison avec les résultats officiels de TREC 2011*

généraliser nos résultats précédents, en mettant en place une évaluation indépendante de Lucene. Pour ce faire, nous avons sélectionné 5000 tweets avec Lucene, desquels nous avons enlevé le score associé. Ensuite, nous avons ajouté à cet ensemble les tweets pertinents manquants obtenus à partir des jugements de pertinence officiels (*qrels*). Comme il semble obligatoire d'avoir au moins un critère basé sur le contenu de la requête, la contribution de Lucene a été remplacée par un score très simple : le pourcentage de termes de la requête présents dans le tweet (run étiqueté Base). Le score final de chaque tweet est ensuite calculé selon la formule 9 dans laquelle le score de Lucene est remplacé par Base. Les résultats généralisés sur l'apport des critères un par un sont présentés dans le tableau 4.

Système	MAP	P@30
Base	0,1785	0,2184
Base+f_1	0,1629*	0,2034
Base+f_2	0,1155*	0,1531*
Base+f_3	0,1214*	0,1755*
Base+f_4	**0,2019***	**0,2449***
Base+f_5	0,1610*	0,2095
Base+f_6	0,0980*	0,1190*
Base+f_7	0,1481*	0,2054

Tableau 4. *Apport des critères en faisant abstraction de Lucene*

Système	MAP	P@30
Base+G1	0,1268*	0,1687*
Base+G2	**0,1864***	**0,2286**
Base+G3	0,1013*	0,1218*
Base+G1+G2	0,1705*	0,2075
Base+G1+G3	0,1272*	0,1633*
Base+G2+G3	0,1552*	0,1850*
Base+G1+G2+G3	0,1617*	0,1952*

Tableau 5. *Apport des groupes de critères et de leurs combinaisons en faisant abstraction de Lucene*

Nous constatons une nouvelle fois que le critère f_4 semble avoir un intérêt, les autres dégradant les résultats. Si l'on considère maintenant les différents groupes de critères (tableau 5), le meilleur groupe est G2, contenant le critère f_4. Ces résultats concernant les groupes de critères sont en accord avec les résultats obtenus en section 6.1.

6.4. *Discussion*

La conclusion principale de ces expérimentations est que la présence de liens hypertextes dans les tweets semble être un indicateur de pertinence, en complément

bien sûr de la pertinence de leur seul contenu. L'ensemble de nos résultats peut être généralisé sur toutes les plateformes de microblogging vu qu'elles utilisent souvent la même syntaxe (hashtags, mention. . .) et ont les mêmes caractéristiques (faible longueur des messages, l'inclusion des URL. . .).

Concernant maintenant les mesures d'évaluation utilisées, nous avons constaté qu'il n'y a pas au moins 30 tweets pertinents par topic. Le système idéal atteindrait une P@30 de 0,7619. Par ailleurs, la P@30 étant une mesure ensembliste, elle ne tient pas compte du classement des résultats.
Pour ces deux raisons, la MAP, ou toute autre mesure mesure sensible au rang, nous semblerait plus appropriée afin de classer les participations officielles.

Enfin, la « fraîcheur » des tweets est un facteur important dans leur pertinence, dont nous nous sommes abstraits dans ces expérimentations. Nous envisageons bien entendu à terme d'intégrer ce facteur dans notre système. Cependant, nous pensons que l'évaluation telle que faite par TREC en forçant le tri chronologique des tweets est trop restrictive, la tâche de recherche se traduisant en fait ici par une tâche de filtrage.

7. Conclusion et perspectives

La recherche de *microblogs* a été considérée dans cet article, en s'appuyant sur la collection de test fournie par la tâche Microblog de TREC 2011. Nous avons introduit et évalué des critères permettant d'affiner les résultats d'un moteur de recherche usuel (Lucene). Les expérimentations réalisées soulignent l'importance de critères liés à l'hypertextualité des *microblogs* en complément de ceux liés à leur seul contenu. En termes de perspectives, nous envisageons dans un premier temps d'approfondir l'étude du rôle des URL dans les *microblogs* en passant à l'analyse du contenu des documents cibles. Nous allons également étudier l'impact d'autres facteurs, tels que la fraîcheur ou encore l'introduction de critères externes (sites de news, etc.).

8. Bibliographie

Ben Jabeur L., Tamine L., Boughanem M., « Un modèle de recherche d'information sociale dans les microblogs : cas de Twitter », *Conference sur les Modeles et l'Analyse des Reseaux : Approches Mathematiques et Informatique (MARAMI), Grenoble*, octobre, 2011.

Bernstein M., Suh B., Hong L., Chen J., Kairam S., Chi E., « Eddi : interactive topic-based browsing of social status streams », *ACM symposium on User interface software and technology*, ACM, ACM, New York, NY, p. 303-312, 10/2010, 2010.

Brin S., Page L., « The anatomy of a large-scale hypertextual Web search engine », *Comput. Netw. ISDN Syst.*, vol. 30, n° 1-7, p. 107–117, 1998.

Cohen D., Amitay E., Carmel D., « Lucene and Juru at TREC 2007 : 1-Million Queries Track. », *TREC'07*, 2007.

Damak F., Jabeur L. B., Cabanac G., Pinel-Sauvagnat K., Lechani L., Boughanem M., « IRIT at TREC Microblog 2011 », *TREC'11 : 20th Text Retrieval Conference*, NIST, November, 2011.

Duan Y., Jiang L., Qin T., Zhou M., Shum H.-Y., « An empirical study on learning to rank of tweets », *COLING '10 : Proceedings of the 23rd International Conference on Computational Linguistics*, p. 295–303, 2010.

Efron M., « Hashtag retrieval in a microblogging environment », *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, ACM, New York, NY, USA, p. 787–788, 2010.

Huberman B. A., Romero D. M., Wu F., « Social networks that matter : Twitter under the microscope », *First Monday*, vol. 14, n° 1, p. 1–9, 2008.

Java A., Song X., Finin T., Tseng B., « Why we twitter : understanding microblogging usage and communities », *WebKDD'07 : Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, p. 56–65, 2007.

Kleinberg J. M., « Authoritative Sources in a Hyperlinked Environment », *Journal of the ACM*, vol. 46, n° 5, p. 604–632, 1999.

Lee C., Kwak H., Park H., Moon S., « Finding influentials based on the temporal order of information adoption in twitter », *WWW'10 : Proceedings of the 19th international conference on World wide web*, ACM, New York, NY, USA, p. 1137–1138, 2010.

Massoudi K., Tsagkias E., de Rijke M., Weerkamp W., « Incorporating Query Expansion and Quality Indicators in Searching Microblog Posts », *ECIR 2011 : 33rd European Conference on Information Retrieval*, Springer, Dublin, p. 362–367, 2011.

Nagmoti R., Teredesai A., De Cock M., « Ranking Approaches for Microblog Search », *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, IEEE Computer Society, Washington, USA, p. 153–157, 2010.

Ounis I., Lin J., Soboroff I., « Overview of the TREC-2011 Microblog Track », 2011.

Ramage D., Dumais S. T., Liebling D. J., « Characterizing Microblogs with Topic Models. », *ICWSM'10*, 2010.

Robertson S. E., Walker S., « Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval », *ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, p. 232–241, 1994.

Singhal A., Buckley C., Mitra M., « Pivoted document length normalization », *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '96, ACM, New York, NY, USA, p. 21–29, 1996.

Song S., Li Q., Zheng N., « A spatio-temporal framework for related topic search in microblogging », *Proceedings of the 6th international conference on Active media technology*, AMT'10, Springer-Verlag, Berlin, Heidelberg, p. 63–73, 2010.

Sparck Jones K., *A statistical interpretation of term specificity and its application in retrieval*, Taylor Graham Publishing, London, UK, UK, p. 132–142, 1988.

Student, « The Probable Error of a Mean », *Biometrika*, vol. 6, n° 1, p. 1–25, 1908.

Teevan J., Ramage D., Morris M. R., « #TwitterSearch : a comparison of microblog search and web search », *Proceedings of the fourth ACM international conference on Web search and data mining*, WSDM '11, ACM, New York, NY, USA, p. 35–44, 2011.

Weng J., Lim E.-P., Jiang J., He Q., « TwitterRank : finding topic-sensitive influential twitterers », *WSDM'10 : Proceedings of the third ACM international conference on Web search and data mining*, ACM, New York, NY, USA, p. 261–270, 2010.

Zhao L., Zeng Y., Zhong N., « A weighted multi-factor algorithm for microblog search », *Proceedings of the 7th international conference on Active media technology*, AMT'11, Springer-Verlag, Berlin, Heidelberg, p. 153–161, 2011.