

Génération automatique de HashTags

Guillaume Tisserant
LIRMM & Awadac
tisserant@lirmm.fr

Mathieu Roche
TETIS & LIRMM
mroche@cirad.fr

Violaine Prince
LIRMM
prince@lirmm.fr

Résumé

Les hashtags sont des mots-clés que les utilisateurs de réseaux sociaux choisissent de mettre en avant dans leurs messages. Ils ont été popularisés sur le réseau social Twitter, qui a permis à ses utilisateurs de sélectionner des HashTags à suivre et d'afficher l'ensemble des messages contenant un HashTag suivi. Ils sont aujourd'hui utilisés sur les principaux réseaux sociaux, tels que Facebook, Google+, Diaspora*, et sont un facteur important de la diffusion de l'information sur Internet. Dans cet article, nous proposons une méthode fondée sur des informations statistiques, syntaxiques et sémantiques pour générer des HashTags.

1 Introduction

Avec l'usage grandissant d'Internet, une quantité d'information de plus en plus importante se trouve à notre disposition. La difficulté n'est donc plus seulement de multiplier les ressources auxquelles nous pouvons accéder mais de trier les informations auxquelles nous accédons. Pour permettre à ses utilisateurs de sélectionner le contenu qui l'intéresse, certains

réseaux sociaux, comme *Twitter*, proposent aux utilisateurs de sélectionner des HashTags, qui correspondent à leurs centres d'intérêts et affichent dans leur flux les messages contenant des HashTags suivis. Les utilisateurs voient les HashTags contenus dans les messages de leur flux, et peuvent les ajouter à leur liste de HashTags suivis.

Toutefois, le choix de HashTags à suivre ou à utiliser pour indexer un de ses tweets peut être difficile à réaliser : un HashTag trop générique va indexer le message dans un flux très important de données, il ne sera donc pas mis en valeur alors qu'un HashTag trop spécifique risque de ne pas être suivi. L'exploitation des HashTags demande donc un effort plus important à l'utilisateur. La création d'un système automatique de recommandation de HashTags est donc une solution intéressante pour faciliter l'accès aux ressources par les utilisateurs.

2 Problématique

2.1 Qu'est ce qu'un HashTag?

Les HashTags sont des termes que les utilisateurs des réseaux sociaux, en particulier Twitter, choisissent de mettre en avant dans leurs messages en les faisant précéder du symbole #.

Un HashTag peut avoir plusieurs significations. Il peut servir à référencer un tweet par rapport à un sujet ou à participer à une discussion en cours sur le sujet associé au HashTag (Huang et al., 2010). Par exemple, les HashTags #2012, #Elysée, et #Elysée2012 ont servi à indiquer qu'un tweet évoquait la campagne présidentielle de 2012. Mais les HashTags servent aussi à marquer son appartenance à une communauté ou une idéologie (Conover et al., 2011). Twitter, pendant les dernières élections présidentielles en France, a vu se multiplier des noms de candidats (#Eva pour Eva Joly, #NS pour Nicolas Sarkozy), des partis (#UMP, #PS) ou des slogans (#Placeaurope) utilisés comme HashTags.

Récemment, de nombreux travaux ont montré l'intérêt d'exploiter les HashTags dans le cadre de tâches de fouille de texte. (Conover et al., 2011), par exemple, utilisent les HashTags comme descripteurs pour la classification et mettent en avant le fait qu'ils sont plus pertinents que les autres termes. (Ozdikis et al., 2012) utilisent les HashTags pour faire du clustering. Ils montrent que les HashTags permettent un meilleur partitionnement des tweets. Ils montrent aussi que l'enrichissement sémantique pour des tâches de partitionnement est de meilleure qualité en se fondant sur les HashTags plutôt que sur les termes. Cela montre que les HashTags sont des données à la fois exploitables par des êtres humains et par des programmes automatiques.

2.2 Comment générer des HashTags?

La recommandation de HashTags est un domaine encore peu exploré (Kywe et al., 2012). La majorité des approches s'appuient sur des méthodes statistiques, comme (Zangerle et al., 2011) qui utilisent la pondération *TF-IDF* ou (Godin et al., 2013) qui exploitent le modèle

Latent Dirichlet Allocation. Ces approches se fondent sur l'idée qu'un HashTag contient une information qui a pour but d'indexer un tweet par rapport à un sujet. Mais un HashTag possède aussi une sémantique importante évoquée dans le tweet. Générer des HashTags depuis les tweets demande donc de détecter des termes qui soient à la fois sémantiquement intéressants et marqueurs d'une opinion ou d'une appartenance à un groupe. Nous allons, dans cet article, proposer une méthode à la fois statistique et sémantique, permettant de détecter les termes les plus discriminants pour l'indexation et les plus intéressants sémantiquement.

3 Analyse

Pour développer notre méthode de génération de HashTags, nous avons commencé par analyser les termes utilisés en tant que HashTags dans un corpus de tweets, et nous les avons comparé à des termes issus d'une analyse statistique d'un corpus de tweets, et des termes issus d'une ressource sémantique en rapport avec notre corpus. Dans cette section, nous présenterons ces différentes ressources, ainsi que notre corpus.

3.1 Les ressources utilisées

Pour comprendre quels termes pouvaient être des HashTags potentiellement intéressants, nous sommes partis d'un corpus de tweets politiques, et nous avons comparé les termes utilisés en tant que HashTags avec des termes statistiquement discriminants, et avec une liste de termes politiquement clivants, mettant en avant des opinions politiques. La méthode statistique pour favoriser les termes discriminants provient d'une méthode appelée *GenDesc* (Tisserant et al., 2014) et la ressource sémantique

utilisée pour sélectionner les termes politiquement clivants provient du GWAP (Game With A Purpose) *PolitIt* (Tisserant, 2015).

3.1.1 Le corpus

Pour tester nos méthodes et faire nos mesures statistiques, nous avons utilisé un sous-ensemble du corpus POLOP (Political Opinion Mining) (Bouillot et al., 2012). POLOP rassemble des tweets écrits en français par des élus de différents partis politiques pendant la campagne présidentielle de 2012. Nous avons travaillé sur un sous-ensemble du corpus de 2500 tweets équitablement répartis entre 5 partis politiques. Ces mouvements sont l'*UMP*, le *MoDem*, le *PS*, le *Front de Gauche*, et *EELV*. Les tweets sélectionnés font en moyenne 81 caractères. Le plus petit tweet fait 31 caractères, le plus long en fait 140.

3.1.2 GenDesc

GenDesc (Tisserant et al., 2014) est une méthode que nous avons développée pour répondre à des problématiques de classification de textes. L'objectif de notre méthode est de détecter les termes les moins discriminants et de les remplacer par des descripteurs plus génériques de façon à obtenir une meilleure représentation en vue de son utilisation par un algorithme de classification automatique. Nous utilisons une formule statistique que nous appelons D pour privilégier les termes les plus discriminants :

$$D(x) = \frac{\text{occClasse}(x)}{\text{occCorpus}(x)}$$

où $\text{occClasse}(x)$ est le nombre d'occurrences de x dans la classe qui le contient le plus et $\text{occCorpus}(x)$ représente le nombre d'occurrences de x dans l'intégralité du corpus.

La mesure D s'est révélée pertinente pour des tâches de classification (Tisserant et al.,

2014). Nous allons montrer comment cette mesure, en détectant les termes les plus discriminants des tweets, peut être utilisée pour une tâche de génération de HashTags. Les *classes* considérées seront les différents mouvements politiques.

3.1.3 PolitIt

PolitIt est un GWAP qui traite de la "polarité politique" des termes. Dans ce jeu, une interface propose des termes aux joueurs qu'ils doivent associer à un des six courants politiques proposés : *Extrême-gauche*, *Ecologie*, *Gauche modérée*, *Droite modérée*, *Droite*, *Extrême droite*. Lorsqu'ils considèrent qu'aucun courant ne correspond mieux que les autres, ils peuvent passer à un autre terme sans donner de réponse. Les données obtenues grâce aux parties jouées permettent de rattacher des termes aux centres d'intérêts des courants politiques. Par exemple, l'entité nommée *Adam Smith* est associée par les joueurs à la *droite*. Cela ne veut pas forcément dire que le descripteur représente un concept défendu par le courant politique rattaché. Par exemple, le terme *nucléaire* est rattaché au courant politique *Ecologie*. Cela s'explique par le fait que, bien qu'ils s'y opposent, le nucléaire est un sujet de préoccupation important pour les mouvements écologistes. Ces données peuvent donc être adaptées pour détecter les termes clivants dans notre corpus de tweets politiques.

3.2 Les termes sélectionnés par chaque mesure

La Table 1 montre un échantillon des termes appartenant à une des catégories. Les termes de *PolitIt* présentés sont ceux ayant le plus grand nombre de réponses attachant le terme au même courant politique. Les HashTags mis en avant sont les plus utilisés dans le corpus. Les ter-

mes en provenance de *GenDesc* sont ceux ayant la valeur de la mesure D la plus élevée. Une version plus complète de ce tableau est donnée dans (Tisserant, 2015).

<i>PolitIt</i>	HashTag	<i>GenDesc</i>
centriste	air	puteaux
dieu	éducation	metz
assurance	marseille	mélenchon
elf	jdd	besançon
bourse	karachi	front
rpr	interview	laurent
rtt	optimisme	edf
ss	hollande	nazaire
terre	crise	nucléaires
altermondialiste	sénat	démanteler

Table 1: Mots n'apparaissant que dans un seul des trois ensembles

Nous voyons qu'une partie des termes retournés par *GenDesc* semblent les plus difficiles à rattacher à un mouvement politique. Le terme *démanteler* ou le prénom *Laurent*, par exemple, sont impossibles à analyser sans leur contexte. Toutefois, en prenant en compte le contexte, certains de ces termes deviennent politiquement orientés. *Metz*, par exemple, fait référence à un meeting qui a eu lieu pendant la campagne législative de 2012, au moment où le corpus de tweets a été constitué.

Une partie des HashTags nécessite aussi un contexte pour les rattacher à un mouvement politique. Toutefois, même privés de leur contexte, nous pouvons considérer qu'ils représentent des concepts politiques. Par exemple, si nous prenons les termes *éducation* ou *sénat*, nous pouvons considérer qu'ils appartiennent au champ lexical de la politique, même s'ils ne peuvent être rattachés à un mouvement politique spécifique.

Les termes de *PolitIt* peuvent en grande partie être rattachés à un courant politique, même hors de tout contexte. Certains termes comme *RPR* ou *altermondialiste* font même directement référence à des courants politiques. Toutefois, il ne faut pas forcément en déduire que ces termes vont être utilisés par les courants auxquels ils sont reliés. Par exemple, le terme *SS* est rattaché à l'*extrême droite* dans *PolitIt*, mais il est principalement utilisé sur Twitter par des sympathisants d'*extrême gauche* pour parler du *Front National* de manière péjorative.

4 Contribution

Nous allons dans cette Section présenter deux méthodes de génération de HashTags. Chaque méthode sera évaluée quantitativement et qualitativement sur un corpus de tweets politiques.

Nous proposons dans la Section 4.1 une méthode pour sélectionner, à partir de tweets, des termes candidats pour être des HashTags. Puis, dans la Section 4.2, nous proposerons une méthode permettant de générer des HashTags composés de plusieurs mots.

4.1 Génération de HashTags simples

Nous avons vu précédemment que les termes provenant à la fois de *GenDesc* et *PolitIt* sont des HashTags potentiels intéressants. Nous allons nous appuyer sur ce constat pour proposer une première méthode de génération de HashTags.

4.1.1 Méthodologie

L'idée est d'exploiter les termes détectés par *GenDesc* et ceux provenant de *PolitIt* pour générer des HashTags. Les termes provenant de *GenDesc* sont discriminants pour les tweets. Cela indique qu'ils peuvent donner des HashTags intéressants pour marquer politiquement un tweet. Les termes provenant de *PolitIt*

sont des termes ayant une sémantique politique forte. À ce titre, ils ont de fortes chances d'être un marqueur d'attachement à un mouvement politique.

Pour vérifier notre hypothèse, nous avons mené l'expérimentation suivante : Nous avons sélectionné 25 termes pour chaque catégorie (*GenDesc*, *PolitIt* et $GenDesc \cap PolitIt$). Les termes de *GenDesc* étant ceux avec la mesure *D* la plus importante, et les termes de *PolitIt* sont ceux ayant été rattachés le plus grand nombre de fois au même courant politique. Pour l'intersection de *GenDesc* et *PolitIt*, nous avons pris les termes ayant la mesure *D* la plus importante qui sont rattachés à plus de 50 % à un même courant politique dans *PolitIt*.

4.1.2 Résultats

Pour vérifier si les HashTags générés pouvaient s'avérer pertinents, nous avons observé, grâce au site hashtags.org¹, s'ils sont aujourd'hui utilisés comme HashTag. Les mesures ont été effectuées en Juillet 2014, soit plus de deux ans après la construction du corpus. Les HashTags sont considérés comme utilisés régulièrement s'ils ont été utilisés plus de cent fois en moyenne par jour. Les résultats de l'expérimentation sont donnés en Table 2.

<i>GenDesc</i>	<i>PolitIt</i>	$GenDesc \cap PolitIt$
52 %	76 %	92 %

Table 2: Tag utilisés fréquemment

4.1.3 Analyse

Nous remarquons qu'à peine plus de 50 % des termes avec une valeur de la mesure *D* élevée sont utilisés comme des HashTags, alors que ceux provenant de *PolitIt* sont à 76 % utilisés comme HashTags. Cela permet de

¹www.hashtags.org

montrer que l'information sémantique issue de GWAP est plus pertinente que l'information statistique pour la génération de HashTags. Toutefois, nous voyons que l'information statistique peut être pertinente pour la génération de HashTags. En effet, en prenant l'intersection de *GenDesc* et *PolitIt*, nous obtenons un meilleur résultat (92 %) qu'en utilisant les termes de *PolitIt* (76 %).

Le fait que la combinaison des deux méthodes soit plus efficace que l'utilisation des méthodes isolées vient du fait que *GenDesc* et *PolitIt* apportent des informations différentes et complémentaires :

- *GenDesc* nous permet de savoir qu'un terme est marqueur d'une classe politique,
- son apparition dans *PolitIt* montre qu'il appartient au champ lexical de la politique.

4.2 Génération de HashTags composés

Nous avons vu dans la section précédente que nous pouvions générer des HashTags pertinents composés d'un unique terme en combinant des informations statistiques et sémantiques. Mais une partie importante des HashTags sont en réalité composés de plusieurs mots.

4.2.1 Méthodologie

La problématique de génération de HashTags composés de plusieurs mots est plus complexe que celle de génération de HashTags simples. En effet, il faut pouvoir proposer des combinaisons de termes représentant des hashtags potentiels, et ensuite sélectionner ceux dont la combinaison offre une sémantique intéressante permettant d'identifier rapidement la thématique du tweet.

Nous avons décidé de nous appuyer sur l'utilisation de patrons syntaxiques pour effectuer une sélection de termes candidats. Puis,

pour sélectionner les termes les plus pertinents pour être utilisés en tant que HashTags, nous avons utilisé les informations provenant de *PoliltIt* et *GenDesc*.

Étape 1 : Patrons syntaxiques

La première étape de la génération consiste à choisir un ensemble de syntagmes candidats. Nous avons choisi d'utiliser des patrons syntaxiques pour sélectionner des syntagmes candidats. Cette approche est proche de celle adoptée par certaines méthodes d'extraction de terminologie (Aussenac-Gilles et al., 2000). Nous avons recouru à trois patrons syntaxiques classiques (Daille, 1994), présentés dans la Table 3.

Patron syntaxique	Exemple
NOM - ADJECTIF	listes électorales transition énergétique
ADJECTIF - NOM	haute surveillance affreux dictateur
NOM - PREPOSITION -NOM	syndicalisme de lutte gaz de schiste

Table 3: Patrons syntaxiques et exemples de termes associés présents dans le corpus.

Pour détecter les patrons syntaxiques dans les tweets, nous avons choisi d'utiliser l'étiqueteur grammatical SYGFRAN (Chauché, 1984).

Étape 2 : Filtre statistique endogène

Nous avons ensuite appliqué un filtre statistique endogène sur nos candidats, pour ne conserver que les syntagmes dont au moins un était considéré comme pertinent par *GenDesc*. Ce filtre permet de supprimer les HashTags n'étant pas considérés comme discriminants par notre mesure statistique. Ainsi, des syntagmes comme "*journal de campagne*" ou "*texte à trous*" vont être supprimés de la liste des HashTags candidats.

Étape 3 : Filtre sémantique

Nous avons ensuite appliqué un filtre sémantique, pour ne garder que les couples contenant des termes appartenant à *PoliltIt*. Ce filtre nous permet de ne conserver que des syntagmes représentant des concepts politiques. Ainsi, certains syntagmes candidats comme "*fdg créé*" ou "*projet irresponsable*" vont être écartés de la liste des HashTags candidats.

Étape 4 : Filtre statistique exogène

Un certain nombre de HashTags générés à partir des patrons syntaxiques ne représentent pas de concepts. Le nombre important de ce type de HashTags s'explique, en partie, par la mauvaise construction grammaticale des tweets, qui a tendance à induire en erreur l'analyseur syntaxique. Nous avons donc utilisé un filtre statistique exogène pour détecter la pertinence de l'association de termes. L'idée est de nous appuyer sur un corpus différent qui est à la fois indépendant et de taille supérieure pour y mesurer la fréquence d'apparition des syntagmes sélectionnés. Ce filtre a pour but de supprimer des syntagmes ne représentant pas forcément un concept, comme "*consommation collaborative*" ou "*petitjournal politesse*". Nous avons choisi d'utiliser internet comme corpus pour ce filtre. Nous avons mesuré la fréquence d'apparition des syntagmes grâce au moteur de recherche Bing, en considérant le nombre de résultats retournés (Turney, 2001). Ce filtre nous a permis de supprimer des groupes de termes apparaissant rarement ensemble sur Internet, et qui ne représentent pas forcément une sémantique intéressante.

4.2.2 Résultats

Après application du processus, nous avons relevé qu'un grand nombre de HashTags générés étaient en rapport direct avec des évènements survenus à la période où le corpus

a été constitué. Nous avons choisi d'évaluer chaque HashTag en prenant en compte deux types d'informations sémantiques :

- la représentation d'un **concept** du champ lexical politique.
- le marquage d'une **orientation** politique.

Des exemples de HashTags illustrant ces notions sont donnés dans la Table 4.

Concept politique	<i>#voteutile</i> <i>#pouvoirdachat</i>
Orientation politique	<i>#buffetsurcanalplus</i> <i>#gaucheàbastia</i>
Concept politique et Orientation politique	<i>#drapeaurouge</i> <i>#alliancecentriste</i> <i>#agriculturepaysanne</i>
HashTag apolitiques	<i>#nouvellechanson</i> <i>#texteàtrous</i>

Table 4: Exemple de HashTags générés et classés en fonction de leur sémantique politique

Nous avons évalué notre méthode de génération de HashTags composés. Pour cela, nous avons annoté manuellement 40 HashTags pour chaque type de filtre utilisé (*GenDesc*, *PolitIt*, *GenDesc* \cap *PolitIt*, *GenDesc* \cap *PolitIt* \cap Web). Les résultats sont donnés dans la Table 5.

4.2.3 Analyse

La Table 5 nous montre que le filtre fondé sur *GenDesc* est efficace pour écarter les HashTags générés n'ayant pas d'*orientation politique*. Au contraire, le filtre fondé sur *PolitIt*, se montre plus efficace pour supprimer les HashTags ne représentant pas un concept politique. L'utilisation des deux filtres combinés permet d'obtenir un pourcentage de HashTags représentant un concept politique supérieur à

	Sémantique Politique	Orientation Politique	Sémantique Politique \cap Orientation Politique
Aucun	27.5 %	5 %	5 %
<i>GenDesc</i>	42.5 %	50 %	27.5 %
<i>PolitIt</i>	55 %	22.5 %	17.5 %
<i>GenDesc</i> \cap <i>PolitIt</i>	62.5 %	50 %	32.5 %
<i>GenDesc</i> \cap <i>PolitIt</i> \cap Web	80 %	52.5 %	47.5 %

Table 5: Pourcentage de HashTags générés représentant un concept politique ou une orientation politique en fonction des filtres utilisés. La ligne *Aucun* correspond à l'ensemble des HashTags sélectionnés grâce aux *patrons syntaxiques*.

n'importe lequel des deux filtres utilisés seul. Le recours au filtre utilisant *Bing* combiné aux deux autres filtres permet d'améliorer encore la qualité des HashTags générés. **La combinaison des trois filtres correspondant à l'application du processus dans sa globalité permet de générer des HashTags dont 80 % sont porteurs d'une sémantique politique et 47.5 % sont à la fois porteurs d'une sémantique politique et marqueur d'une orientation politique.**

Le fait que 80 % des HashTags générés avec l'utilisation des trois filtres représentent un concept politique indique que l'algorithme ne propose que 20 % de HashTags réellement non pertinents. Par ailleurs, plus de la moitié des HashTags générés sont porteurs d'une *orientation politique*. La combinaison de filtres que nous proposons nous permet donc de générer des HashTags pertinents, à la fois porteurs de sens et d'une orientation politique.

5 Conclusion

Nous avons décrit dans cet article deux méthodes de génération de HashTags. Nous avons vu que les données statistiques comme les données sémantiques permettaient de développer des méthodes de génération de HashTags. Nous avons développé une méthode mêlant informations sémantiques, informations syntaxiques, et approches statistiques s'appuyant sur des données endogènes et exogènes. Nous avons montré que la combinaison de ces méthodes permet d'obtenir de meilleurs résultats que chacune des méthodes utilisée séparément.

Toutefois, nos travaux sur la génération de HashTags contiennent plusieurs limites. Nous nous sommes placés dans un contexte où nous avons connaissance des thématiques des tweets. Nous pensons étendre notre méthode à un cadre non supervisé, sans connaissances *a priori* des thématiques abordées dans le corpus ni des différentes opinions exprimées.

Une autre limite importante de nos travaux vient du faible nombre de patrons syntaxiques utilisés pour la sélection de HashTags candidats. Or de nombreux HashTags ont des structures complexes, non représentées par ces patrons. Nos futurs travaux s'appuieront sur les syntagmes verbaux et l'association de termes ou de n-grammes de caractères extraits à partir des tweets.

References

- Aussenac-Gilles, N., Biébow, B., Szulman, S., et al. (2000). Modélisation du domaine par une méthode fondée sur l'analyse de corpus. In *Actes Ingénierie des Connaissances (IC)*, pages 93–104.
- Bouillot, F., Poncelet, P., Roche, M., Ienco, D., Bigdeli, E., and Matwin, S. (2012). French presidential elections: what are the most efficient measures for tweets? In *Proc. of the Workshop on Politics, elections and data*, pages 23–30. ACM.
- Chauché, J. (1984). Un outil multidimensionnel de l'analyse du discours. In *Proc. of Int. Conf. on Computational Linguistics*, pages 11–15.
- Conover, M., Gonçalves, B., Ratkiewicz, J., Flammini, A., and Menczer, F. (2011). Predicting the political alignment of twitter users. In *Proc. of Conference on Social Computing (SocialCom)*.
- Daille, B. (1994). *Approche mixte pour l'extraction de terminologie: statistique lexicale et filtres linguistiques*. PhD thesis.
- Godin, F., Slavkovikj, V., De Neve, W., Schrauwen, B., and Van de Walle, R. (2013). Using topic models for twitter hashtag recommendation. In *Proc. of Int. conference on World Wide Web companion*, pages 593–596.
- Huang, J., Thornton, K. M., and Efthimiadis, E. N. (2010). Conversational tagging in twitter. In *Proc. of Conference on Hypertext and hypermedia*, pages 173–178. ACM.
- Kywe, S. M., Hoang, T.-A., Lim, E.-P., and Zhu, F. (2012). On recommending hashtags in twitter networks. In *Proc. of Int. Conference on Social Informatics, SocInfo*, pages 337–350.
- Ozdikis, O., Senkul, P., and Oguztuzun, H. (2012). Semantic expansion of hashtags for enhanced event detection in twitter. In *Proc. of International Workshop on Online Social Systems*.
- Tisserant, G. (2015). *Généralisation de données textuelles adaptée à la classification automatique*. PhD thesis, Univ. Montpellier.
- Tisserant, G., Prince, V., and Roche, M. (2014). Gendesc: Vers une nouvelle représentation des données textuelles. *Num. sp. "Fouille de Données Complexes"*, *RNTI*, pages 127–146.
- Turney, P. (2001). Mining the web for synonyms: Pmi-ir versus LSA on TOEFL. In *Proc. of European Conference on Machine Learning*, pages 491–502.
- Zangerle, E., Gassler, W., and Specht, G. (2011). Recommending#-tags in twitter. In *Proc. of Workshop on Semantic Adaptive Social Web (SASWeb 2011)*. *CEUR*, volume 730, pages 67–78.