Be In The Know: Connecting News Articles to Relevant Twitter Conversations

Bichen Shi, Georgiana Ifrim, and Neil Hurley

Insight Centre for Data Analytics
University College Dublin
Dublin, Ireland
{bichen.shi,georgiana.ifrim,neil.hurley}@insight-centre.org

Abstract. In this paper we propose a framework for tracking and automatically connecting news articles to Twitter conversations as captured by Twitter hashtags. For example, such a system could alert journalists about news that get a lot of Twitter reactions, so they can investigate those conversations for new developments in the story, promote their article to a set of interested consumers, or discover general sentiment towards the story. Mapping articles to hashtags is nevertheless challenging, due to different language styles of articles versus tweets, the streaming aspect, and user behavior when marking tweet-terms as hashtags. We track the IrishTimes RSS-feed and a focused Twitter stream over a two months period, and present a system that assigns hashtags to each article, based on its Twitter echo. We propose a machine learning approach for classifying article-hashtag pairs. Our empirical study shows that our system delivers high precision and recall for this task.

Keywords: news tracking, social media, Twitter, hashtag recommendation

1 Introduction

Since its start in 2006, Twitter has established itself as an alternative media source. Increasingly, Twitter conversations and calls to action that mobilize masses have dedicated hashtags, as showcased by recent world events, e.g., #ArabSpring, #Syria, #freethe7. Twitter hashtags thus lead to the formation of ad hoc publics around specific themes and topics without the need for the users to be otherwise explicitly connected [2].

Hashtags can convey information about the community that uses them or the sentiment of the messages they group. For an outsider, or even for an insider that doesn't continuously track the massive Twitter activity, it is close to impossible to stay in the know when it comes to the right hashtags or users to follow, for current and developing news stories. Nevertheless for journalists in particular, it is vital to get to the right hashtags quickly, in order to be able to follow new developments on topics of interest. Data analytics techniques can provide tools that link news stories to the relevant Twitter conversations.

Automatically mapping news articles to appropriate hashtags (where a hashtag is seen as a group of tweets forming a conversation around it) can be very challenging. This is due to different language styles used in the two types of data (e.g., clean, long

Table 1. A news article and initially retrieved hashtags (before learning algorithm is applied).

News Article	Retrieved Hashtags	Hashtag Category
Headline: *FG fears day of reckoning over Enda Kennys Seanad gamble*	#seanad, #enda, #irelandsaysno	Relevant (Specific)
Sub-headline: *There is deep concern within the Fine Gael ranks that its populist referendum campaign misfired so badly*	#ireland, #rtept, #news	Relevant (General)
	#caughtrotten, #whip	Relevant (Abusive)
	#mentalhealth	Irrelevant

articles versus messy, short tweets), the fast paced streaming aspect of both news and tweets (matching two streams moving at different speeds), as well as user behaviour when coining certain tweet-terms as hashtags. To showcase the third issue, in Table 1 we present an example news article and the categories we identified for the hashtags retrieved for it, in an initial pre-processing stage. The article is about Irish politics: the 2013 referendum to adopt a unicameral parliamentary system by abolishing one of the current two houses of parliament, the Seanad. The hashtags retrieved for this article in an initial pre-processing step, range from highly specific and relevant, to general but still relevant, to abusive but potentially relevant, to irrelevant. We can see from this example that an approach that can accurately filter irrelevant hashtags and rank relevant hashtags can deliver value by connecting to the right Twitter conversations.

In this paper we propose a large-scale, real-time framework for connecting news articles from mainstream media to their echo on the Twitter stream. We discuss the data collection process for continuously gathering, processing and connecting a stream of news articles and a focused Twitter stream relevant to the tracked news stories. We analyze relevant features and propose a machine learning approach for ranking hashtags for a given news story. Our experiments show that our system can achieve high precision and recall on this task. The rest of the paper is organized as follows. Section 2 discusses related work and our contributions. In Section 3 we explain the data collection process, while in Section 4 we describe the process of modeling hashtag ranking as a learning problem. In Section 5 we discuss our results and Section 6 concludes with directions for future work.

2 Related Work

Recent years have seen an explosion of research work analyzing social media (e.g., most prominently the micro-blog Twitter) and the connection between traditional media and this new form of reporting. Among the diverse investigations of Twitter data, two categories are most relevant to this paper.

Hashtag Recommendation. Tag recommendation for tagging systems such as Last.fm and Delicious has been studied in a number of works such as [7] that applies topic modelling using Latent Dirichlet Allocation (LDA) to the problem. Focusing in particular on hashtag retrieval over a Twitter corpus, in [5], language modelling is used to find hashtags given a keyword query. A model of each hashtag is learned from the set of tweets that contain the tag as a multinomial distribution over terms. Hashtags are ranked according to the Kullback Leibler divergence of their corresponding model to the query model. In [4] the issue of recommending hashtags to untagged tweets is addressed. An

LDA topic model is used to categorise tweets into topics and a translation probability maps topics to hashtags. The method is modified in [3] by replacing standard LDA with the topic model of [14].

News and Tweets. Work that investigates the connection between news and Twitter includes [11]. Given a set of tweets that specifically mention the URL of a given article, this work focuses on a method to filter this set into a subset of most interesting tweets. The authors use four indicators of interestingness, namely informativeness, opinionatedness, popularity and authority to filter the initial set. TweetMogaz [8], a system for microblog search and filtering, aims to find tweets relevant to regional news. It relies on a curated list of *key players* from which to collect an initial set of relevant tweets. The initial set is augmented, by firstly extracting a set of keywords from news sites and searching for tweets containing these keywords. The *keyword* tweets are filtered by training a classifier using the *key player* tweets as positive examples and a set of random tweets as negative examples. Other works investigated automatic news detection from tweets [12], recommending news articles using tweets [9], forecasting the popularity of news using Twitter [1], or enhancing news articles with information extracted from Twitter, such as *comment tweets* [6].

Our work differs from the above research in a number of ways. In particular, we address hashtag recommendation in a streaming context, with a requirement that the model be updated on a daily basis. Rather than applying topic modeling on a large, static Twitter corpus, containing potentially many diverse topics, we attempt to filter irrelevant tweets directly by using the news articles to be hashtagged in order to focus the data collection from the Twitter stream. Nevertheless, unlike other work on connecting articles and microblogs, we avoid seeding our data collection with a curated user group or with tweets that specifically mention the articles in question (via the URL). As discussed later, our automatic-keyword Twitter stream allows for a wide set of tweets to be gathered, while ensuring that the collection contains relevant tweets with high probability. Our search strategy provides sufficient breadth to allow high recall in gathering relevant hashtags, while avoiding being drowned in a vast sea of Twitter noise. We alternate this high recall with a high precision oriented step, by using a learning approach to rank the retrieved hashtags for each article.

Our contributions are as follows: (1) we propose a focussed Twitter data collection strategy based on automatic keyword extraction from news articles; (2) we formulate a large-scale, real-time learning process for assigning hashtags to news articles; (3) we deliver a system for matching a daily news stream and a relevant Twitter conversation stream [10].

3 Data Collection

We collect data from two sources between October 7, 2013 and November 30, 2013, RSS Feeds of news articles and a focused Twitter stream.

News Articles from RSS Feeds. We gathered the news articles streamed on the Irish Times RSS feeds, by polling the RSS feeds every 5 minutes, yielding a total of 4,862 unique articles, around 170 articles per day. The Irish Times is an Irish mainstream media outlet, that covers Irish news and high impact world news. Each article has a

Table 2. Processed keyword set by permutation.

Original keywords/phrases	Final keywords/phrases
enda kenny	enda kenny
fine gael	fine gael
fg	fg fears
fears	fg seanad
seanad	fears seanad

headline, a one paragraph description that summarises the article (sub-headline), and the article body.

We extract representative keywords for each downloaded article, by parsing the headline and sub-headline, part-of-speech tagging this text, and extracting nouns and named entities using shallow parsing techniques and heuristics (e.g., we extract Aer Lingus, Enda Kenny, etc.). We do not use the article-body for keyword extraction, since it poses risks of topic drift and noise. For example, for the news article in Table 1 we extract the keywords *enda kenny, fine gael, fg, fears, seanad.*

Focused Twitter Stream. Since we are interested in continuously streaming news and corresponding tweets, we use the Twitter Streaming API[1], which can be employed with either keywords (words or phrases), geographical boundary boxes or user ID. We gather Twitter streaming data by using a dynamic[2] set of keywords extracted from the stream of RSS news articles every 30 mins each day. Each article's keywords are streamed for 24 hours.

Additionally, we noticed that in order to get relevant tweets, it helps if we constrain each tweet returned by the Twitter API to contain at least two article keywords. We achieve this by splitting our original keyword set, into individual keywords, and creating all possible permutation pairs as our final keyword set, with the constraint that we freeze named entities. For example, for the article in Table 1, we process the keyword set *enda kenny, fine gael, fg, fears, seanad* by keeping the named entities and permuting the single keywords to form pairs, as shown in Table 2. We apply this process every 30 minutes to *all* the RSS articles downloaded up to that point in time, pool the keywords together, and re-connect with the Streaming API using the updated keyword list.

Through this process we aim to retrieve a large set of relevant tweets while not being restricted to a set of manually curated user lists, locations or keywords. The problem of retrieving relevant tweets to a set of news has been pointed out in recent research [6] with ad-hoc retrieval techniques achieving low Recall (0.5). Prior work relies mostly on tweets where the url of the article is explicitly provided, therefore obtaining a clean but potentially small set of tweets. Our initial tweet-retrieval process gathers a large set of potentially relevant tweets (23.3 million unique tweets), which we carefully filter in the following machine learning process.

[1] https://dev.twitter.com/docs/streaming-apis

[2] Dynamic refers to the list of extracted keywords being updated every 30 mins with the new keywords of incoming articles

4 Learning Algorithm for Scoring Hashtags

In this section we discuss the process of modelling hashtag selection as a learning problem. We parse the stream of news articles and the Twitter stream in a 24 hours timewindow, in order to extract and rank hashtags for each news article. For tweet-processing, we remove stop words, punctuation, URLs and user names, and apply stemming to the remaining terms. For each day, and each news article, we separate the tweets of the corresponding Twitter stream *per article*, based on a shallow matching of tweet keywords and article-keywords (as extracted for the Streaming API and showcased in Table 2). This results in a local tweet-bag per article, that can be analyzed for extracting hashtags and hashtag information, e.g., frequency, tf.idf profile describing the hashtag as reflected in its tweet-bag. Next, we form article-hashtag pairs, and compute features of each pair useful for discriminating whether a hashtag is relevant to a given article.

Features. We extract four features for each article-hashtag pair: two features that characterize the local hashtag profile (based on tweets in the article-tweet bag), while the other two characterize the global hashtag profile (based on tweets in the entire Twitter stream that are retrieved till that time point), useful for describing specific versus general hashtags.

One of the first features we select is the *local cosine similarity* between the tf.idf profile of the article and the local tf.idf profile of hashtag (as extracted from the tweets mentioning that hashtag in the article tweet-bag, by considering tweets as an *articles* and calculate tf.idf weight of them). To avoid noise in the article tf.idf profile, rather than selecting terms from the full article-body, we only select them from the headline and sub-headline, but compute their tf.idf weight using the entire article (stop words removed, stemming). Additionally, we extract the *local frequency* of the hashtag (i.e., the number of tweets in the article tweet-bag, mentioning that hashtag). The hashtag frequency captures whether a hashtag is actively used.

We compute the *global cosine similairy* between the local and the global hashtag tf.idf profile (as extracted from tweets in the entire historical Twitter stream[3], mentioning that hashtag), to asses how much does the global hashtag profile diverge from the local profile. Note that globally, the same hashtag may refer to different events, or a hashtag may be preferred over a time window to refer to a certain event, and then slowly discarded or outweighed by other hashtags. Therefore, using local and global features for each hashtag, addresses the issue of time-of-use and scope of a hashtag. we also extract the *global frequency* of the hashtag (i.e., the number of tweets in the entire Twitter stream, mentioning that hashtag).

For each article-hashtag pair, we now have four features describing how relevant a hashtag may be for a given article. We normalize all four features to the $[0, 1]$ interval. Next, we discuss how to use these features and a set of manually labeled article-hashtag pairs for learning to identify relevant hashtags.

Labeled Data. In order to build a classification algorithm for recognizing relevant hashtags, we need labeled article-hashtag pairs. We selected two days at random from the two month dataset, October 23, and November 23, 2013, extracted all the article hashtag pairs and their features as described above, and asked two annotators to manually

[3] Historical refers to the time the article was published

label each pair. We asked the annotators to decide which of the three scenarios applies to each pair: (1) a hashtag is *specific and relevant* to the topic of the news article, (2) *general and relevant*, or (3) *irrelevant*. For abusive hashtags, the annotators were advised to decide which of the three scenarios the hashtag belongs to depending on the local context. For the purpose of our experiments, we merged the first two classes into simply relevant (a positive example in binary classification) or irrelevant (negative example). The inter-annotator agreement was 80%. We used the subset of examples where both annotators agreed for training/testing a classification algorithm.

Besides of manually labeled data, we also gathered tweets containing both the Irish Times article's URL and hashtags. These tweets naturally form Article-User Hashtags pairs and can be used as a form of ground truth, by assuming all user assigned hashtags are relevant to the article.

Classification Algorithm We train and test our approach by employing a series of Weka [13] classification algorithms using default settings, including Logistic, kNN, Decision Tree, Multilayer Perceptron etc. The algorithm only sees the examples as described by the four features, and can learn thresholding strategies on the provided features. For example, to classify a hashtag as relevant for a given article, a classification algorithm may learn (from the training set) that the cosine feature should be higher than 0.5 and the hashtag frequency should be close to 1. Additionally, most classifiers provide a score describing the likelihood that a hashtag is relevant to the article. We use this classification score to rank hashtags for each article.

5 Evaluation

In order to evaluate our overall strategy for retrieving, learning, and ranking hashtags, we evaluate classifiers using three different settings: **Small**, **Medium** and **Large**. The evaluation settings and results are shown in Table 3.

5.1 Results: Small Experiment

For the small experimental setting, we use the manual labelling of article-hashtag pairs for two random days: October 23, 2013 (874 examples) for training, and November 23, 2013 (1,122 examples), for testing.

Baselines. We first evaluate two simple baseline techniques. On the test set (November 23, 2013), we select the top-3 hashtags per article (257 pairs out of 1,122), using the *highest local hashtag frequency* and the *highest local cosine similarity*. We evaluate the precision of these two baseline. Since we only take the top-3, recall is not applicable in this case.

Learning Approach. We now evaluate the classifier's ability to retrieve all the hashtags deemed relevant by our annotators as well as its ability to rank them before the irrelevant ones. We experimented with a series of Weka classifiers, with default parameter settings. MultilayerPerceptron, Logistic (regularised logistic regression) and Kstar (K-nearest neighbours with entropy-based-distance) delivered the best results, as shown in Table 3. We note that all three classifiers have high precision (0.85), recall (0.80) and AUC (0.92), showing that the classifier ranks relevant hashtags before irrelevant ones.

Table 3. The evaluation results of **Small**, **Medium** and **Large** experiment settings.

Settings	Training Set		Testing Set		Results					
						Baseline		Learning Approach		
	Dataset	Size	Dataset	Size		Most Frequent Top3	Highest Cosine Top3	Multilayer Perceptron	Logistic	Kstar
Small	Oct 23	847	Nov23	1122	Accuracy	-	-	**84.6%**	84.4%	83.9%
					Precision	0.548	0.634	0.850	**0.876**	0.861
					Recall	-	-	**0.807**	0.770	0.774
					F1	-	-	**0.846**	0.844	0.839
					AUC	-	-	0.921	**0.924**	0.911
Medium	Oct 23	847	Article-User Hashtags	1147	Recall	0.503	0.644	**0.781**	0.756	0.704
	Nov 23	1122						0.792	**0.808**	0.787
	Oct 23 & Nov 23	1996						**0.845**	0.797	0.776
Large	Oct 23 & Nov23 & Article-User Hashtags	3143	Randomly Selected Article-Hashtag Pairs	1029	Precision	-	-	-	0.869	-
					Precision@1	-	-	-	0.900 ([0.871, 0.929], $p < 2.2e-16$)	-
					NDCG@3	-	-	-	0.877 ([0.850, 0.904], $p < 2.2e-16$)	-

The AUC is particularly important, since ultimately it is useful to rank the hashtags of each document, from most relevant to least relevant. Additionally, the Logistic classifier is a linear model that can be easily interpreted, scales to large data and its classification scores are true probabilities. The Logistic model deemed all four features as important (non-zero weights), with the local cosine feature getting the highest weight, followed by the frequency based features, and ending with the global cosine.

5.2 Results: Medium Experiment

In this setting, we use Article-User Hashtags data, which is gathered from tweets containing both the Irish Times articles' URL and hashtags and can be used as a form of ground truth, by assuming all user assigned hashtags are relevant to the article. The articles with user assigned hashtags are a subset of the total streaming articles set. Since we assume all the test examples are relevant (belongs to the positive class), in this setting the Accuracy is the same as Recall, and Precision is 1 by default.

Baselines. We still employ the two baseline techniques: On the test set (Article-User-Hashtags), we select the top-3 hashtags per article (257 pairs out of 1,122), using the *highest local hashtag frequency* and the *highest local cosine similarity*. We evaluate the recall of these two baseline.

Learning Approach. As training data we analyze three settings, October 23, 2013, November 23, or both days together as training, and Article-User-Hashtag pairs as test (1,147 test examples). Note that we assume that all the user-assigned hashtags are relevant, which may not necessarily be the case, since sometimes users also assign spurious hashtags, e.g., #annoying #omg. Our algorithm may consider such hashtags as irrelevant.

Table 3 shows Recall over all article-user-hashtags retrieved by our learning algorithm as being relevant (classification score above 0.5). We note that when we increase the amount of training data, by combining the October and November examples, the accuracy of MultilayerPerceptron stands at 84.5%, a similar value to that of the small setting experiment.

5.3 Results: Large Experiment

In this setting, we use the labeled examples of October 23, November 23, and the Article-User Hashtags data as training data, and around 1000 randomly selected article-hashtag pairs extracted from the RSS and Twitter streams (no labels) as testing data.
Baselines. No baselines are employed in this setting due to the expensive cost of human evaluation.
Learning Approach. Due to cost of manual evaluation, we only use the Logistic classifier in this setting, for the reason that it performs well in the previous two settings, and it runs in linear time, so the scalability would not be an issue.

We apply the trained Logistic classifier to all the article-hashtag pairs. We randomly select 422 articles that get at least one relevant hashtag (based on classification score above 0.5), and manually asses the relevant article-hashtag pairs (1,029 pairs), using 0, 1 and 2 relevancy scores. As shown in Table 3, we evaluate both the filtering quality, i.e., the classification across all article-hashtag pairs (to asses the Precision over the pairs classified as relevant), as well as the hashtag *ranking quality per article*, using information retrieval metrics. For the article oriented metrics, we use Precision@1 and NDCG@3 and average them across all articles.

We note that the precision for the filtering step (binary classification into relevant/irrelevant) is fairly high (Precision 0.86), and similar to what we have seen in the previous experiments. When we evaluate the quality of ranking of hashtags for each article, we see a similar result: the Precision@1 is 0.9, while the NDCG@3 which penalizes relevant hashtags ranked at low ranks, is 0.87.

5.4 Discussion

In order to make the whole methodology more explicit, in Table 4 we show some example articles from our annotated sample of the **Large** setting, their extracted keywords, their (up to top-3) ranked hashtags, together with the features extracted for the corresponding pair, the classifier score and the annotator relevance score. We observe that for local as well as international news (first 3 articles), the hashtags assigned and ranked by classifier score are relevant and quite specific (e.g., #ecb, #walshwhiskeydistillery).

We found three main reasons why an article does not get any (relevant) hashtag: the article-keyword extraction process is faulty (due to part-of-speech tagging errors or due to the fact that the extracted keywords are too generic); there is no discussion on Twitter about that particular news story; the tweets relevant to an article do not contain any hashtags. The aspect of assigning noisy or irrelevant hashtags can be mitigated to some extent by tuning the classifier threshold (here we used the default classification score of 0.5). Additionally, the four features describing each article-hashtag pair could

Table 4. Example results from our two-month annotated sample.

Article Headline	Article Keywords	Hashtag	LFr	LCo	GFr	GCo	ClassifScor	RelScor
Tech titans in town for Dublin Web Summit	dublin, dubstarts, summit, tech, web	#websummit	1.00	0.35	0.58	0.82	0.92	2
		#tech	0.23	0.45	0.70	0.37	0.89	1
		#web	0.42	0.41	0.40	0.52	0.82	2
Whiskey distillery to create 55 jobs for Co Carlow	carlow, co, walsh, distillery, whiskey	#whiskey	1.00	0.73	0.16	0.56	0.99	2
		#carlow	0.90	0.64	0.16	0.53	0.99	2
		#walshwhiskey-distillery	0.66	0.61	0	1	0.97	2
ECB's Draghi moves to ease fears on interest rates	banks, draghi, ecb	#ecb	1.00	0.50	0.39	0.42	0.97	2
		#draghi	0.54	0.58	0.19	0.69	0.96	2
		#news	0.00	0.46	0.89	0.29	0.90	1
Climate change watchdog must be robust and independent, says report	advisory, climate, expert, fiscal	#delleir	1.00	0.27	0.00	1.00	0.60	0
		#job	0.00	0.30	0.80	0.35	0.53	0
		#delcfe	0.89	0.26	0.00	1.00	0.51	0
Europe bank payouts capped as capital bar keeps rising	capital, europe	#europe	0.79	0.35	0.55	0.02	0.84	1
		#travel	0.66	0.27	0.68	0.38	0.72	0

be enhanced, e.g., using user authority to re-weight tweets, filtering spammy hashtags (e.g., #ff, #followback). Regarding the lack of hashtags in the tweet-bag of an article, in such cases we could employ recent techniques for extracting informative tweets [11], or adapt our approach for the problem of assigning Twitter users (rather than hashtags) relevant to a given news article. The manual annotation for the learning approach is also potentially noisy, since at times it is quite difficult to decide whether a hashtag is relevant or not, without considerable background knowledge. In this respect we plan to employ crowd sourcing platforms such as Crowdflower, in order to obtain larger and possibly cleaner labeled datasets, but even then, the annotators require considerable background knowledge for labelling (e.g, political climate in Ireland).

6 Conclusion

In this work we present a framework for connecting news articles to their relevant Twitter conversations, as semantically grouped by Twitter hashtags. We discuss the aspect of continuously tracking a stream of news and tweets, and present an approach for obtaining a large focused Twitter stream, automatically seeded by a dynamic keyword set extracted from the articles. Furthermore, we model the problem of hashtag assignment as a classification problem, and analyze a framework for hashtag retrieval and appropriate features and data for building a hashtag classifier. We evaluate our methods and show that our approach achieves high precision for this task.

Future Work. We plan to improve our approach by avoiding using manually labelled data (which is expensive) and instead using Article-User Hashtag data (which is free to obtain) as training data. This means changing the two class classification problem into one class (without negative examples), or two class with noisy/random negative data. The Article-User Hashtag data collected through the real-time Twitter stream could help re-training the classifier and keeping it up-to-date, to avoid potential concept drift.

We plan to extend our study to track several RSS news feeds and Twitter conversations, and test a prototype with journalists. We also intend to investigate applications of our methods to clustering of articles in hashtag space, story tracking and event detection.

An early prototype Insight4News system[4] integrates the techniques described here, and demonstrates the real-time, large-scale nature of our proposed hashtag recommendation process. Unlike the 24 hours time window used in this paper, the Insight4News system updates hashtag recommendation for articles every 15min, and processes around 300 articles pre day.

Acknowledgements This work was supported by Science Foundation Ireland under grant 07/CE/I1147.

References

1. Roja Bandari, Sitaram Asur, and Bernardo A Huberman. The pulse of news in social media: Forecasting popularity. In *ICWSM*, 2012.
2. Axel Bruns and Jean E Burgess. The use of twitter hashtags in the formation of ad hoc publics. 2011.
3. Zhuoye Ding, Xipeng Qiu, Qi Zhang, and Xuanjing Huang. Learning topical translation model for microblog hashtag suggestion. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 2078–2084. AAAI Press, 2013.
4. Zhuoye Ding, Qi Zhang, and Xuanjing Huang. Automatic hashtag recommendation for microblogs using topic-specific translation model. In Martin Kay and Christian Boitet, editors, *COLING (Posters)*, pages 265–274. Indian Institute of Technology Bombay, 2012.
5. Miles Efron. Hashtag retrieval in a microblogging environment. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 787–788. ACM, 2010.
6. Alok Kothari, Walid Magdy, Ahmed Mourad Kareem Darwish, and Ahmed Taei. Detecting comments on news articles in microblogs. *ICWSM 2013*, 2013.
7. Ralf Krestel, Peter Fankhauser, and Wolfgang Nejdl. Latent dirichlet allocation for tag recommendation. In *Proceedings of the third ACM conference on Recommender systems*, pages 61–68. ACM, 2009.
8. Walid Magdy. Tweetmogaz: A news portal of tweets. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '13, pages 1095–1096, New York, NY, USA, 2013. ACM.
9. Owen Phelan, Kevin McCarthy, Mike Bennett, and Barry Smyth. Terms of a feather: Content-based news recommendation and discovery using twitter. In *Advances in Information Retrieval*, pages 448–459. Springer, 2011.
10. Bichen Shi, Georgiana Ifrim, and Neil Hurley. Insight4news: Connecting news to relevant social conversations. In *ECML/PKDD*, 2014.
11. Tadej Štajner, Bart Thomee, Ana-Maria Popescu, Marco Pennacchiotti, and Alejandro Jaimes. Automatic selection of social media responses to news. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 50–58. ACM, 2013.
12. Ilija Subašić and Bettina Berendt. Peddling or creating? investigating the role of twitter in news reporting. In *Advances in Information Retrieval*, pages 207–213. Springer, 2011.
13. Ian H Witten, Eibe Frank, and Mark A Hall. *Data Mining: Practical Machine Learning Tools and Techniques: Practical Machine Learning Tools and Techniques*. Elsevier, 2011.
14. Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. Comparing twitter and traditional media using topic models. In *Advances in Information Retrieval*, pages 338–349. Springer, 2011.

[4] `http://insight4news.ucd.ie/insight4news/`