Modelling Techniques for Twitter Contents: A step beyond classification based approaches.

Angel Castellanos, Juan Cigarrán and Ana García-Serrano

Natural Language Processing and Information Retrieval Group, UNED
Juan del Rosal, 16 (Ciudad Universitaria), 28040 Madrid, Spain.
{acastellanos,juanci,agarcia}@lsi.uned.es

Abstract. In this paper we present our first participation at RepLab Campaign. Our work is focused in two contributions. The first one is the use of an IR method to address Polarity and Filtering tasks. These two tasks can be seen as the same problem: to find the most relevant class to annotate a given tweet. For that, we applied a classical IR approach, using the tweet content as query against an index with the models of the classes used to annotate tweets. To model these classes we propose the use of the Kullback Leibler Divergence (KLD), in order to extract their most representative terminology. Different data and ways to model these data (through KLD) are also proposed. The second contribution is related to the Topic Detection task. Instead a clustering based technique; we propose the application of Formal Concept Analysis (FCA) to represent the contents in a lattice structure. To extract topics from the lattice, we applied a FCA concept: stability. According to the results, our IR based approach has been proven as very satisfactory for the Polarity task, while for the Filtering task, it seems to be less suitable. On the other hand FCA modelling has been demonstrated as a promising methodology for Topic Detection, achieving high successful results.

Keywords: Formal Concept Analysis, Stability, Kullback Leibler Divergence, Content Modelling, ORM, POS Tagging.

1 Introduction

In this paper we summarize our participation in the 2013 edition of the RepLab Campaign [1]. RepLab Campaign is focused on the Online Reputation Management (ORM) task; that is, the reputation monitoring of entities and persons on the Web, and more concretely in Twitter. Our participation focuses on three of the RepLab Tasks: in the filtering and polarity tasks and, by other hand, in the topic detection task.

The first two tasks (filtering and polarity) are usually addressed through classification approaches: a data set is used to train classification systems and learns a set of classes (related/unrelated, positive/neutral/negative) allowing the classification of new contents. More sophisticated approaches, based on probabilistic techniques, have been also recently proposed to address filtering and polarity tasks; one of them, maybe the most widely used, is Topic Modelling.

Both tasks (filtering and polarity) can be seen technically as the same task: given a tweet to annotate, find the most similar class. For that, instead of the common state of the art approaches, we propose the application of an IR based annotation that, given a tweet to annotate, uses its content as query against an index containing the content models of the classes to annotate the tweet. To generate these content models, we apply a divergence based technique (Kullback Leibler Divergence) to find the most representative terminology of each class. We have previously applied this technique for content modelling, outperforming other content modelling techniques [3], and also for content modelling for polarity and sentiment detection [4].

The other task in which we have participated this year is the topic detection task. As in the previous tasks, classification approaches has been commonly used to address the detection of topics. However, this approach poses a problem: often new topics unseen in the training data appear along the time, making useless to detect them the learnt classes. To solve that, unsupervised techniques based on clustering have been proposed. But, even these techniques have many problems with the issue of topics diversity.

Given this problems, we propose a Formal Concept Analysis (FCA) based approach. FCA allows the modelling of the contents (tweets) according to their attributes (terminology) in a lattice structure. FCA also allows the adaptation for detecting new topics while take advantage of knowledge provided by the training data. Once the content was modelled through FCA, clusters/topics should be selected; however, the number of concepts (possible topics) generated by FCA is potentially quite high. In order to select proper cluster/topics, we applied the concept of stability, coming from FCA field.

The rest of the paper is organized as follows: In section 2 we present the IR based approach applied for the Polarity and Filtering tasks, in section 3 we expose the novel application of FCA for the Topic Detection task, in section 4 we present the results of each task and, finally, in section 5 we present our conclusions and the feasible future work.

2 IR-based approach for Polarity and Filtering Tasks

As we said before, filtering and polarity tasks can be as a classification problem. In filtering task, tweets have to be classified in RELATED and UNRELATED, while in polarity task they have to be classified in POSITIVE, NEGATIVE and NEUTRAL. So, we proposed the same approach for both tasks. Instead of common approaches, based on classification, we propose an IR-based approach. If we considered the contents of the tweets to be classified and the contents of the classes (gathered from the tweets in the training set annotated with them), these tasks can be seen as an IR task, using the tweet content as query against an index containing the class contents. Then, the classification will be dependent on the results of these queries. The work done for both tasks includes:

- **Annotation.** A well-known problem in the use of tweets is the scarcity of information. To limit the impact of this problem, tweet contents have been processed in

order to identify some features (hashtags, named entities, adjectives), which have been added to the information used to model the class.

- **Modelling.** To represent each of the classes with their representative terminology, the contents of the tweets annotated with them in the training set have been modelled. The modelling technique is based on the comparison of class contents with the content of the rest of the class/es, by applying the Kullback-Leibler Divergence as weighting function [8]. The application of a divergence-based technique intends to identify the terminology that better differentiate one class from the rest. The different models generated for each class (see sections below) have been indexed taking into account the relevance of each term in the model, according KLD formulation.
- **IR-based Classification.** To classify tweets, their contents have been used as query against the indexed models. Each tweet will be classified into the class with the highest relevance according the results returned by the query.

Specific details of these steps for each task, and the executed runs for each of them are presented in the sections below.

2.1 Filtering

This task is focused on classify a set of tweets as related/unrelated to an entity. Our approach is based on modelling the related and unrelated content to identify the more representative terminology for every class in the collection. For that, we have experimented with different data sources: a) Wikipedia entity pages; b) Content of the set of tweets related to the entity and c) Content of the external webs which appear in the related tweets. Furthermore, we annotated Twitter data with some features, helpful to represent the entities: a) Named Entities (this annotation has been carried out through Stilus Core[1] tool) and b) Hashtags, given that they are usually used to identify a specific topic in Twitter.

Our modelling is based on the comparison between terms of a specific content with terms present in the rest of the contents of the collection. In this context, it results in comparing the related (unrelated) terms of an entity with the related (unrelated) terms of the rest of the entities. Modelling the entities in this way, we will be able to say that a tweet is related to an entity if, using the tweet content as query, the IR system returns the entity model. It could be also interesting to identify the more representative terms of the related contents of an entity according to their unrelated contents. So, for each entity we have also modelled their related and unrelated content following this approach, denoted from here as Related vs. Unrelated (RvsU) modelling.

As our modelling is based on compare entities, since there are 4 domains in the collection (university, automotive, music and banking) some domain-specific words could be identified as entity-specific words (e.g. car and wheel can be set as representative of the automotive entities). To cope with this, we proposed a domain-

[1] http://api.daedalus.es/stiluscore-info

specific modelling (in contrast with the "generic modelling"): each entity is modelled by comparing it only with the entities of its domain.

Besides of modelling experimentation, we have experimented with different IR-based methods. Firstly, given some tweet contents, we query against an index containing the related models of each entity: is the tweet related to a given entity? Nevertheless, looking at detail the collection there are much more tweets related than unrelated (about the 75% of the tweets). Taking that into account we propose an inverse approach by querying against the unrelated models: is not this tweet related with the entity? The idea is to consider all the tweets as related, except those for which we have solid evidences to the contrary.

Even so, some tweets are undoubtedly related with the entity, thus there is no need to check if the tweet is not related. This situation is addressed by querying against the Wikipedia models: since Wikipedia contents are very accurate, if a tweet is related to these contents, it will be related to the entity with a high probability.

Taking into account all of these considerations we have conducted the following experiments, summarized in Table 1:

Table 1. Filtering Runs

Run	Content used to Model	Modelling
filtering_1	Wikipedia Entity Pages	Specific
filtering_2	Wikipedia Entity Pages	Generic
filtering_3	Content of the Related Tweets	Specific
filtering_4	Content of the Related Tweets	Generic
filtering_5	External Webs Content	Generic
filtering_6	Hashtags in Content of the Related Tweets	Generic
filtering_7	NER in Content of the Related Tweets	Generic
filtering_8	Content of the Unrelated Tweets	Generic
filtering_9	Content of the Unrelated Tweets	Generic RvsU
filtering_10	Content of the Unrelated Tweets	Generic RvsU + Wiki Filter

2.2 Polarity

This task is focused on identify the polarity of a tweet for the reputation of an entity. We have followed the same approach as in the filtering task, modelling polarity values (POSITIVE, NEGATIVE and NEUTRAL) of each entity with its contents related, in the training set. In the same way that in the filtering task, we have experimented with different types of information, modelling techniques and classification approaches. To model each polarity we have used one single source: the content of the related tweets. From this source we have gathered: a) Tweet Contents and b) Adjectives identified in these tweet contents, using Stilus Core[2]. We have also applied generic and specific modelling, as in the filtering task, and what we have called most similar modelling. With this technique, given a tweet to be annotated, it searches for the most similar tweet in the training set and it uses its polarity as annotation. If there is not

[2] http://api.daedalus.es/stiluscore-info

similar tweet, the first approach is applied. The intuition is that if two tweets are similar, their polarity has to be the same. With these considerations, we have developed the following runs:

Table 2. Polarity Runs

Run	Content used to Model	Modelling
polarity_1	Tweet Content	Specific
polarity_2	Tweet Content	Generic
polarity_3	Tweet Content	Generic
polarity_4	Tweet Content	Specific
polarity_5	Tweet Adjectives	Most Similar Specific
polarity_6	Tweet Adjectives	Most Similar Generic

3 Detecting Topics through a FCA-based Approach

Topic Detection task is focused on, given a stream of tweets related to an entity; identify topics in this stream. Usually this kind of task is addressed with a classification-based approach, but this approach is not valid for this task, because topics in the new tweets may be not related to the training topics. All we know is: in the past these topics appeared in the tweets; now there is a set of new tweets, try to take advantage of the prior knowledge to detect topics in the new tweets.

The best way to address this task is a clustering-based approach. However, a clustering approach also has some drawback: How many clusters are? How can the systems take into account the prior knowledge? Does the running of the systems has to be fixed by the data in the training set or they have to show a certain degree of adaptability? These entire considerations make the Monitoring task an specially challenging task. To solve the clustering drawbacks we proposed a novel approach, especially suitable for the context of this task, based on Formal Concept Analysis (FCA). FCA can be seen as a powerful tool to automatically structure and classify all the resources retrieved and enriched from the Internet. This theory fits on a lattice-based clustering approach improving information access and exploratory tasks on pure Information Retrieval (IR) scenarios [5-7].

3.1 FCA Highlights

FCA is a mathematical theory [12] of concept formation derived from lattice and ordered set theories that provide a theoretical model to organize *formal contexts*: collections of *objects* related with sets of *attributes*. The main construct of the theory is the *formal concept*. A formal concept is a pair (*O, A*) with *O* is a set of objects (the *extend* of the formal concept), and *A* a set of attributes (the *intend* of the formal concept). In addition, *O* and *A* are connected as follows:

- If an object *o* in *O* is tagged with an attribute *a*, then *a* must is included in *A* (i.e, the intend of the formal concept includes all the attributes shared by the objects in the extend).

- Conversely, if an object *o* is tagged with all the attributes in *A*, then *o* must be included in *A* (i.e., the extend of the formal concept includes all those objects filtered out by the intend).

Formal concepts can be ordered by their extends. More formally, $(O,A) \subseteq (O',A') \Leftrightarrow O \subseteq O'$; in this case (O',A') is called a *super-concept* of (O,A) and, conversely, (O,A) a *sub-concept* of (O',A'). The order that results can be proved to be a *lattice*, which is called the *concept lattice* associated to the formal context.

In a concept lattice, two interesting kinds of formal concepts are *object concepts* and *attribute concepts*. Indeed:

- The *object concept* associated with an object *o* is the most specific concept including *o* in its extend. In order to construct it, it is possible to include in its intend all the attributes of *o*, and to include in its extend, in addition to *o*, all those objects tagged exactly with the same attributes than *o*.
- Conversely, the *attribute concept* associated with the attribute *a* is the most generic concept including *a* in its intend. It can be constructed in a dual way to an object concept: (i) add all the objects tagged by *a* to the extend, and (ii) in addition to *a*, add all the attributes shared by those objects to the intend.

3.2 Modelling

As FCA allows modelling a set of objects according to their attributes, in order to adapt this context to the Monitoring task, the tweets are identified as the objects, terminology of the tweets is identified as the attributes and, consequently, formal concepts can be identified as topics, containing a set of tweets according to a set of common terminology.

Since the modelling performance is highly dependent on the terminology, before applying this modelling we have pre-processed the contents in the next way: we have removed generic and domain stop-words; we have stemmed the terms; we have disambiguated the named entities, unifying them in common labels (e.g. *bmw_m3* and *m3* will be considered the same entity); and, finally, we have expand the terminology with the identified hashtags (i.e. if there is a hashtag #m3, all the tweets with the term m3 will be expand with the hashtag #m3). All of this pre-processing pursues expand the content of the tweets in order to facilitate the finding of relationships between contents.

Although in the theoretical model all the tweet terms can be considered as attributes, in the real scenario this would generate an unmanageable lattice with a huge number of concepts. For that we have applied an algorithm for filtering attributes according to their representativeness [7]

3.3 Topic Annotation

In spite of the attribute reduction, the number of generated formal concepts is very high to consider every concept as a cluster. So, it remains the decision of what concepts are suitable to represent a topic. In this sense, a desirable characteristic of the

concepts is the cohesion between their objects. Otherwise it would indicate that this concept it isn't really a cluster but an aggrupation of different topics/clusters.

To reflect how much each concept in the lattice fits with this requirement (object cohesion), we propose the use of the stability concept. Stability was first introduced in [9] in relation to hypotheses generated from positive and negative examples, and it was extended to formal concepts in [10]. In [11] they present an algorithm to calculate it based on an original concept lattice. Briefly explained, the stability of a concept (i.e. also known intentional stability) indicates how much the concept intent depends on particular objects of the extent. In other words, the stability of a concept is the probability of preserving its intent after leaving out an arbitrary number of objects. Thus, a high stability value indicates that the concept represents a cohesive set of tweets or, what is the same, it can represent a proper cluster.

We have experimented with different stability values as threshold to select the clusters (all the concepts with a stability value higher than the threshold will be taken as cluster), from 40% to 90%. We have also experimented with two values for attribute selection in the reduction algorithm presented in the previous section. More concretely, the experiments developed are:

Table 3. Topic Detection Runs

Run	Attribute Reduction Threshold	Stability Threshold
topic_detection_1	1%	90%
topic_detection_2	1%	80%
topic_detection_3	1%	70%
topic_detection_4	1%	60%
topic_detection_5	1%	40%
topic_detection_6	5%	90%
topic_detection_7	5%	80%
topic_detection_8	5%	70%
topic_detection_9	5%	60%
topic_detection_10	5%	40%

4 Results

4.1 Filtering

In the Table 4 it is shown the results achieved by our experiments in this task. Results are expressed in terms of Reliability Sensitivity and F measure [2]. For this task we proposed the experimentation with 3 different data sources: Wikipedia (filtering_1 and filtering_2), contents of the external webs in the tweets (filtering_5), and Twitter; in this sense, besides of the tweets contents (filtering_3 and filtering_4), we also used named entities (filtering_7) and Hashtags (filtering_6) in the tweets.

As general comments we can point out the low performance of the proposed approaches, if we compare them with the best approach. Looking in detail the results; regarding to the data type, the best results is obtained with the external webs contents,

followed by Wikipedia and finally Twitter contents; within Twitter contents, the named entities achieves the highest performance, followed by twitter raw contents and finally Hashtags. One aspect to remark here is that the performance obtained with the external webs content is driven by the improvement in the sensitivity value; that is, the use of these contents increases the coverage of the annotation process.

Table 4. Filtering Results

Run	Description	Reliability	Sensitivity	F(R,S)
BEST_APPROACH	**X**	**0,7288**	**0,4507**	**0,4885**
filtering_1	WP-Specific-Modelling	0,1443	0,2482	0,1341
filtering_2	WP-Generic-Modelling	0,1483	0,2606	0,1406
filtering_3	Twitter-Content-Specific-Modelling	0,1467	0,2155	0,1151
filtering_4	Twitter-Content-Generic-Modelling	0,1511	0,2190	0,1206
filtering_5	ExternalLinks-Content	0,1440	**0,2927**	0,1468
filtering_6	Twitter-Hashtags	0,1527	0,2245	0,1084
filtering_7	Twitter-NER	0,1631	0,2169	0,1287
filtering_8	Unrelated-Modelling	0,3095	0,1878	0,1598
filtering_9	Unrelated-RvsU-Modelling	0,2899	0,2184	**0,1738**
filtering_10	Unrelated-RvsU-Modelling-WikiFilter	**0,3521**	0,1198	0,1085

On the other hand, the use of specific modelling doesn't improve the performance of the generic modelling, either using Wikipedia data (filtering_1 and filtering_3), or Twitter data (filtering_2 and filtering_4). This confirms the results that we obtained when we experimented with these approaches in the training step.

Taking into consideration the use of unrelated models, this modelling obtains a significant improvement of the results offered by the run used as baseline of these approaches (filtering_4). This improvement is mainly due to the improvement of the Reliability value, that is, they are more precise. Among these results, Related vs. Unrelated based modelling (filtering_9) is the best performing method. Only the approach using the Wikipedia Filter doesn't get the baseline result despite of the improvement of the Reliability value, due to the low Sensitivity value.

4.2 Polarity

In the Table 5 it is shown the results of the Polarity runs, expressed in terms of Reliability, Sensitivity and F-Measure [2]. In this task we experiment with 2 different ideas. The first one was based on the data used to model: tweet contents (polarity_3 y polarity_4) and adjectives in the tweets (polarity_5 and polarity_6). In this sense it is remarkable the very low performance obtained by the adjective based approach; looking in more detail these results, the low performance can be explained with the extremely low sensitivity value. That is, the adjective based approaches only annotate a small number of tweets; however almost without error, as it can be seen in the Reliability value.

Table 5. Polarity Results

Run	Description	Reliability	Sensitivity	F(R,S)
BEST_APPROACH	**X**	**0,7288**	**0,4507**	**0,4885**
polarity_1	Most-Similar-Approach-Specific-Modelling	0,3337	**0,3093**	**0,3169**
polarity_2	Most-Similar-Approach-Generic-Modelling	0,3328	0,3044	0,3115
polarity_3	Twitter-Content-Generic-Modelling	0,3162	0,2967	0,2956
polarity_4	Twitter-Content-Specific-Modelling	0,3210	0,2943	0,2958
polarity_5	Twitter-Adjectives-Specific-Modelling	**0,9369**	0,0054	0,0099
polarity_6	Twitter-Adjectives-Generic-Modelling	0,9337	0,0062	0,0111

The other approach that was proposed in this task is the use of the most similar modelling. The results of this approach outperform the baseline modelling results according to all of the measures. Finally, like in the filtering task, the application of specific modelling doesn't offer any improvement. In fact, the results for generic and specific approaches can be considered as equal.

4.3 Topic Detection

Table 6 shows the results obtained for the experiments sent to the topic detection task. Results are expressed in terms of Reliability Sensitivity and F measure [2]. Some interesting general considerations are that in general their reliability values are very high, some of them in the Pareto Frontier of the Reliability-Sensitivity Curve (topic_detection_6 and topic_detection_10).

Going into detail, these results can be divided according to the threshold applied to the attribute reduction algorithm (1 and 5 %). The best performance according to F measure is obtained by the first set of runs which use a threshold equal to 1% (topic_detection_1 – topic_detection_5). If we focused on the stability value applied (from 90% to 40%), as the stability value decreases also F-Measure decreases, but the Reliability value increases. This latter behaviour relies in the fact that, the lower is the stability value, the lower the number of generated clusters is; so, it makes sense this precision improvement. If we look at the performance of runs which use a threshold equal to 5%, the results barely differs between them. Given that the threshold for the attribute selection algorithm is higher, less attributes for the application of FCA algorithm are taken into account; leading on a general reduction in the stability value of the generated concepts.

In spite of good Reliability values, the general performance of our application is quite low (according F-Measure). But here there is something affecting the performance of our proposal. Previously to the application of FCA, we filtered out the unrelated tweets. However we didn't the filtering goldstandard at the time of sending the runs, so we had to apply one of our filtering approaches; which as it can be seen before, it doesn't have very accurate results.

Table 6. Topic Detection Results

Run	Description	Reliability	Sensitivity	F(R,S)
BEST_APPROACH	**X**	**0,4624**	**0,3246**	**0,3252**
topic_detection_1	Attribute-Treshold-1-Stability-90	0,6735	**0,1092**	**0,1711**
topic_detection_2	Attribute-Treshold-1-Stability-80	0,6806	0,1061	0,1669
topic_detection_3	Attribute-Treshold-1-Stability-70	0,6930	0,1026	0,1624
topic_detection_4	Attribute-Treshold-1-Stability-60	0,6958	0,1018	0,1615
topic_detection_5	Attribute-Treshold-1-Stability-40	0,7470	0,0969	0,1560
topic_detection_6	Attribute-Treshold-5-Stability-90	0,8331	0,1076	0,1548
topic_detection_7	Attribute-Treshold-5-Stability-80	0,8331	0,1076	0,1548
topic_detection_8	Attribute-Treshold-5-Stability-70	0,8333	0,1076	0,1547
topic_detection_9	Attribute-Treshold-5-Stability-60	0,8333	0,1076	0,1547
topic_detection_10	Attribute-Treshold-5-Stability-40	**0,8338**	0,1075	0,1546

In order to address this problem, we used the filtering goldstandard as a filter, once it was released by the organizers, and we obtain the results shown in the Table 7. The table shows only the experiments using a stability value of 90%, the value which a higher performance achieves. Both runs outperform the sent runs, so the low performance of our approach can be attributed to the low performance of the filtering step, previous to the FCA modelling. However the improvement is much clearer for the enhanced_run_1; in fact this result would be placed in the first third of the overall RepLab results. This seems to indicate that a threshold equal to 5% is too restrictive and it leaves out an important part of the knowledge contained in the tweet terminology.

Table 7. Topic Detection Enhanced Runs Results

Run	Description	Reliability	Sensitivity	F(R,S)
BEST_APPROACH	**X**	**0,4624**	**0,3246**	**0,3252**
enhanced_run_1	Attribute-Treshold-1-Stability-90	**0,6615**	0,1940	**0,2336**
enhanced_run_2	Attribute-Treshold-5-Stability-90	0,6184	**0,2469**	0,1730

As a final comment, we want to remark one strange result obtained during the evaluation step. We can see that if we considered only two clusters (the root of the lattice as one cluster and the rest of the lattice as another cluster) the results are surprisingly good (see Table 8); in fact they improve the performance of the best approach.

Table 8. 2 Cluster Approach Results

Run	Description	Reliability	Sensitivity	F(R,S)
BEST_APPROACH	**X**	**0,4624**	**0,3246**	**0,3252**
2_cluster_run_1	Attribute-Treshold-1-Stability-90	0,5510	**0,3537**	**0,3477**
2_cluster_run_2	Attribute-Treshold-1-Stability-40	**0,5556**	0,3483	0,3459

5 Conclusions and Future Work

The work done in the RepLab campaign was divided in two sides. First, we participated in the Filtering and Polarity tasks by proposing the application of an IR approach to annotate tweets, instead of common classification approaches. For the Topic Detection task our work was focused on the application of Formal Concept Analysis in order to model tweet contents and to detect a set of topics in these contents. The results obtained by our IR approach are opposite. While for the Filtering task we don't achieve satisfactory results, for the Polarity task our results are quite satisfactory, comparing them with the rest of presented approaches.

Analysing in more detail results, all of our ideas to enhance filtering models was confirmed, outperforming baseline results. In general, the use of specific information (named entities, content of external webs, Wikipedia) seems to be better than only the use of tweet contents for this task. Also remark that the use of unrelated contents was the best approach; as we supposed, on a collection where the contents was mostly related to the entities, looking only for the unrelated contents is more precise.

Focusing on Polarity task, our approach works well for the proposed task. The use of models generated through the most similar approach has proven to be more representative than baseline models, even though these models are more dependent on the coverage of training set and that the test set are greater than the training set(1500 vs 750 tweets per entity). Results obtained by the adjective based approaches are interesting; they achieved an almost perfect precision value; however the low coverage made that these approaches achieved an extremely low F measure results.

In relation with topic detection, results of the sent runs got a very satisfactory value in terms of precision; however their general performance was not so good. But, as we cited before, we had some problems with the pre-filtering step with our sent runs. Once this problem was solved, by using filtering goldstandard, our FCA-based approach results achieved a significant improvement, positioning them between the best performing approaches according F-measure and in the first place according Reliability. We want to specially remark our 2-cluster approach results, the best among all proposals, according F-measure. At this point, a proper analysis has to be done here to understand the reason for that and how to apply to our proposal.

As future work, it would be interesting the application of the lessons learnt in the filtering task for modelling enhancement (regarding to the data to use and the ways to model) to other modelling proposals. Focusing on the work done for the polarity task, results point out our approach as a promising way to address this task. Especially interesting would be the use of adjective based modelling as a previous step of other annotation approach, given their high performance in terms of precision.

Finally, FCA has been proven as a promising technique for the topic detection task. The results obtained by our enhanced runs demonstrate the validity of our approach for addressing the topic diversity. Also a further analysis has to be done in order to explain the results of 2 cluster approach and if it is reproducible in other contexts.

Acknowledgments. This work has been partially supported by the Regional Government of Madrid under Research Network MA2VIRMR (S2009/TIC-1542), and HOLOPEDIA (TIN 2010-21128-C02). Special thanks to Daedalus for licencing the utilization of Stilus Core.

References

1. Amigó, E., Carrillo de Albornoz, J., Chugur, I., Corujo, A., Gonzalo, J., Martín, T., Meij, E., de Rijke, M., Spina, D.: Overview of RepLab 2013: Evaluating Online Reputation Monitoring Systems. In: Proceedings of the Fourth International Conference of the CLEF initiative, CLEF 2013, Valencia, Spain. (2013)
2. Amigó, E., Gonzalo, J., Artiles, J., Verdejo, F.: Combining evaluation metrics via the unanimous improvement ratio and its application to clustering tasks. Journal of Artifical Intelligence Research, 42 (1): 689-718 (2011)
3. Castellanos, A.: Recomendación de contenidos digitales basada en modelos del lenguaje: Diseño, experimentación y evaluación. Master Thesis. UNED (2013)
4. Castellanos, A., Cigarrán, J., García-Serrano, A.: Using IR Techniques for topic-based sentiment analysis through divergence models. In: Workshop on Sentiment Analysis at SEPLN (2012)
5. Cigarrán, J., Gonzalo, J., Peñas, A., Verdejo, F.: Browsing search results via formal concept analysis: Automatic selection of attributes. In: Eklund, P. (eds.) Concept Lattices. LNAI, vol. 2961, pp. 74-87. Sydney, Australia. Second International Conference on Formal Concept Analysis (2004)
6. Cigarrán, J., Gonzalo, J., Peñas, A., Verdejo, F: Automatic selection of noun phrases as document descriptors in an fca-based information retrieval system. In: Ganter, B., & Godin, R. (eds.) Formal Concept Analysis, LNAI, vol. 3403, pp. 49-63. Lens, France. Third International Conference (2005)
7. Cigarrán, J.: Agrupación de Resultados de Búsqueda Mediante Análisis Formal de Conceptos. PhD. Thesis. UNED (2008)
8. Kullback, S., Leibler. R.A.: On information and sufficiency. Annals of Mathematical Statistics, 22(1): 79-86 (1951)
9. Kuznetsov, S.O.: Stability as an estimate of the degree of substantiation of hypotheses derived on the basis of operational similarity. Journal of Automatic Documentation and Mathematical Linguistics (1990)
10. Kuznetsov, S.O: On stability of a formal concept. Annals of Mathematics and Artificial Intelligence, 49: 101-115 (2007)
11. Roth, C., Obiedkov, S., Kourie, D.: Towards concise representation for taxonomies of epistemic communities. Concept Lattices and Their Applications, pp. 240-255 (2008)
12. Wille, R.: Ordered Sets, chap. Restructuring lattice theory: An approach based on hierarchies of concepts, pp. 445-470. Prentice Hall (1982)