Characterising concepts of interest leveraging Linked Data and the Social Web

Fabrizio Orlandi
DERI, NUI Galway
Galway, Ireland
Email: fabrizio.orlandi@deri.org

Pavan Kapanipathi, Amit Sheth
Knoesis Center, Wright State University
Dayton, OH, USA
Email: {pavan, amit}@knoesis.org

Alexandre Passant
seevl.net, MDG Web ltd
Dublin, Ireland
Email: alex@seevl.net

***Abstract*—Extracting and representing user interests on the Social Web is becoming an essential part of the Web for personalisation and recommendations. Such personalisation is required in order to provide an adaptive Web to users, where content fits their preferences, background and current interests, making the Web more social and relevant. Current techniques analyse user activities on social media systems and collect structured or unstructured sets of entities representing users' interests. These sets of entities, or user profiles of interest, are often missing the semantics of the entities in terms of: (i) popularity and temporal dynamics of the interests on the Social Web and (ii) abstractness of the entities in the real world. State of the art techniques to compute these values are using specific knowledge bases or taxonomies and need to analyse the dynamics of the entities over a period of time. Hence, we propose a real-time, computationally inexpensive, domain independent model for concepts of interest composed of: popularity, temporal dynamics and specificity. We describe and evaluate a novel algorithm for computing specificity leveraging the semantics of Linked Data and evaluate the impact of our model on user profiles of interests.**

I. Introduction

The Social Web has become essential part of the everyday user interaction with the Web. Users on the Social Web interact with each other, create/share content and express their interests on different social websites with many user accounts and different purposes. On each of these systems, personal information is usually recorded in order to provide to the user recommendations and personalisation. User profiles are then built on these websites representing the interests of the users. These interests are normally sets of *entities* that uniquely identify concepts or instances of the real world and for which the user has demonstrated some sort of interaction. In order to provide personalisation it is necessary for a system to compute the relevance of each entity for the user. Extensive research has been conducted in this regard and in this work we present a novel characteristic of entities of interest that can be applied to many different scenarios.

One of the main use cases for our work is about personalisation of a social web stream, as described in Section III. For instance, on Twitter, information overload[1] has led to a series of research problems in terms of personalisation [1][22][10], recommendation, filtering [2][18], etc. Each of these problems are being solved by various innovative approaches leveraging the named entities present in the content. For example, [1][22] have extracted entities from user generated content to build user profiles that reflect the interest of users on the Social Web. Personalisation and/or recommendation approaches are recently being further enhanced by the introduction of semantic web technologies and specifically by the use of Linked Open Data (LOD) [6] as a knowledge base. Although entities from LOD are leveraged to make the Social Web more meaningful by adding semantics, the pragmatics of it have not been completely addressed. In other words, the way entities are being used and spoken about by users on the Social Web has still not been completely explored. In order to personalise a user's social web stream according to entities of interest, it is necessary to determine not only their relevance for a specific user or their similarity with other entities, but also their semantics and pragmatics. Hence, in this paper we identify three important dimensions that can be used to characterise entities according to their meaning and usage given by social web users. These dimensions are (i) *popularity*, (ii) *temporal dynamics* and (iii) *specificity*. Trends and popularity of entities are well studied by researchers whereas *specificity* is a new attribute that we introduce in this work.

DEFINITION: specificity (or abstractness, as opposite), we define it as the level of abstraction that an entity has in a common conceptual schema shared by humans. Human knowledge can be organised in taxonomies where concepts and instances (in general, entities) are categorised and related to each other with broader/narrower relations. These relations for instance reflect and determine the specificity of the entities in a hierarchical classification system. Entities positioned at high positions in a taxonomy are considered less specific (or broader/more generic) than entities positioned in lower positions of the taxonomy (hence closer to the leaves of the hierarchy). According to this definition, the entity *Alternative Rock Music* is more specific (or has a higher degree of specificity) than the entity *Music*.

In this work we present a novel approach to automatically determine the specificity of entities and hence to improve personalisation on the Social Web. In particular, in order to address the challenge given by our particular use case we have a few requirements to consider: (i) the computation of the specificity measure has to be done in **real-time** in order to cope with the intensive volume of social web data to analyse, (ii) it has to be constantly **up to date** with the events of the real world, and (iii) it has to be **domain** and **knowledge base independent**. For this reason we propose our new measure

The work presented in this paper is funded by Science Foundation Ireland under grant number SFI/08/CE/I1380 (Líon 2), and by an IRCSET scholarship.

[1] http://blog.twitter.com/2011/03/numbers.html

for characterising the specificity of entities, called *"DRR" (Distinct Relations Ratio)*, which is based on Linked Data.

To summarise, the contributions of this work are as follows: (i) we introduce a new attribute "specificity" for entities of interest that enhances the dynamic semantics of entities; (ii) we employ a Linked Data based approach to measure the specificity of entities which is evaluated by comparing our method against a gold standard (created with user ratings) and against an alternative state of the art approach based on the DMOZ taxonomy; (iii) we propose other relevant measures (popularity and trend) as complementary measures; (iv) we evaluate the impact of our measures on user profiles of interests through a user study. In this paper, following an introduction of the related work in Section II, we present our motivation/scenario in Section III. Then in Section IV we define the specificity attribute and explain our approach for measuring it. The experiments conducted for evaluating our measure are presented in Section VI, while the evaluation of the impact of our model on user profiles is in Section VII. Finally we conclude with detailing our future work in Section VIII.

II. Related Work

Entities from Linked Open Data have been consistently leveraged as background knowledge in various domains. Particularly in the context of social networks/data analysis, LOD has been leveraged to improve user profiling [22], to provide context for recommendations [23][21], to enhance filtering social media streams [2][16], etc. Analysis of these entities/resources have recently started gaining momentum. Searching semantic web data is one of the prominent tasks that requires analysing entities/resources on LOD [8][11][20]. Swoogle [8] employs PageRank to determine the relative importance between resources in a dataset. On the other hand [20] utilizes Wikipedia as a graph structure for the same purpose. Another attribute that is determined by analysing entities is their relatedness to each other [12]. Here the authors leverage Wikipedia with spreading activation algorithm to determine the mutual importance between entities. Thalhammer et al. [28] exploit the pragmatic features of entities and the graph structure of Linked Open Data to analyse and provide summaries of entities that can be leveraged by semantic web search engines. Although there has been ample work in determining and analysing different features of entities, our work stands apart by introducing a new attribute *"specificity"* and a novel approach for measuring it, together with an evaluation of their impact on user profiles.

Research on developing innovative trend detection techniques to be applied on various online components has been performed for a long time. Online components include web blogs [13][3], news media [25], social networks [17] and Wikipedia [7]. However, we focus on presenting the state of the art on trend detection in social networks. Online applications such as Trendsmap[2], Whatsthetrend[3], and TwitterMonitor [17], detect trending keywords/hashtags on Twitter with spatio-temporal-thematic facets. Academic research on detecting trends has come a long way [4][14]. In order to detect trending topics/keywords, Benhardus et al. [4] use simplistic measures such as frequency and *tf-idf*, and Irani et al. [14] detect trending terms on Twitter using statistical techniques and classify tweets based on the corresponding trends. Similar research has been performed by analysing pragmatic features such as user page views and edits on pages on Wikipedia [7][26].

III. Motivation and Scenario

Consumption of social data is an ongoing problem due to information overload. Considering for example social media streams of updates such as Twitter[4], one of the most popular social websites of the moment, they first introduced keyword search and later the official support for hashtags (a technique driven by Twitter users) as the mechanism to reduce information overload. Extensive effort is being made both by industry and academia to innovate and develop techniques to reduce information overload on Twitter [2][18]. Filtering personalised Twitter streams based on user's interests is one of them. In our previous work [16], in order to deliver a personalised Twitter stream, we (1) developed an automatic approach to generate user profiles based on the entities mentioned in the content that has been generated by a user on multiple social networks; (2) extracted metadata (including DBpedia resources[5]) for every incoming tweet from the Twitter stream in real-time; (3) delivered tweets that mention entities semantically similar to those in a user's profile to that user. Matching narrow interests such as "Semantic Web" to deliver tweets worked well because of the usage (pragmatics) of the entity "Semantic Web" on the Social Web. However, entities such as "Music"/"India"/"Barack Obama" generated more tweets, in turn causing information overload that could not be handled by the user. This problem introduced the need of determining the pragmatics (*"specificity"* measure) of an entity on the Social Web and to categorise entities/interests based on the number of tweets generated on the Social Web. Furthermore, we hypothesize that complementing this measure with the informativeness of a tweet will reduce information overload for generic entities. The higher the *"specificity level"* of an entity (e.g. "Semantic Web"), the higher the *"informativeness score"* of the tweets delivered and vice-versa.

These hypothesis are not only valid for personalisation of Twitter data but also for any other personalisation system for the Social Web where the quality of the recommendations is based on the accuracy of the entities selected to represent user preferences. In this section we present a possible real use case for our work representing a typical scenario of a semantic (entity-based) recommender system on the Social Web. We developed our solutions having this scenario in mind, where the real-time nature of a system and the need for a large and continuously updated background knowledge are key. However, as described in Section II, the applications for the measures proposed in this work are numerous, from information filtering and personalisation, to analysis of social tagging behaviour [5], to automatic ontology/taxonomy development and concept recommendation to name a few.

IV. A Real-time Domain-independent Measure of Concept Abstractness

Our main motivation for this work is to investigate a methodology for efficiently computing the level of specificity (or abstractness) of a particular real-world entity or concept

[2]http://trendsmap.com/
[3]http://whatthetrend.com/
[4]http://twitter.com
[5]http://dbpedia.org

that is uniquely representable on the Web. State of the art approaches [5], using a particular set of background knowledge, lack of many relevant features that are useful for the Social Web use-case example, as described previously.We introduce a novel real-time measure for identifying the specificity of entities in the real world just using the Linked Data links structure. For *"specificity"* we refer to the definition we provide in Section I. This measure expresses how abstract an entity is in a common conceptual schema and does not refer to the popularity of the term. A real-world entity can at the same time be very generic but not very popular in Social Media systems (e.g. "Classical Music") or can also be both very specific and very popular (e.g. a Pop/Rock song of the moment). This is why for characterising and ranking the relevance of the entities of interest we need to combine this dimension with popularity features (see Section V). To note also that according to [5] popularity and abstractness have a certain degree of correlation (i.e. popular tags are on average more abstract).

It is also necessary to keep in mind that specificity is a measure that can be personal and contextual. According to psychological studies [27] it is shown that categorisation in conceptual hierarchies by humans can be influenced by experience. Even specificity is then subject to this kind of bias, as users with high experience in one domain might perceive the entities related to that domain as less specific than users without experience. However, this might suggest that this measurement is an estimation and not an exact absolute value, and automatic methods to compute it would always have a certain percentage of error.

Several state of the art approaches, in order to compute the specificity, utilize a taxonomy of concepts which are categorized and organised in a hierarchical structure. The more the entities are categorised in a position of the hierarchy close to the top (or the root) the more they are considered generic. Hence the specificity of the entities increases when going from the root to the leaves of the categorisation tree. This approach works well in many situations and is clearly justified by the fact that a taxonomy is by definition organized by supertype-subtype relationships, also called *generalization-specialization* relationships. The problem comes when the knowledge base that has to be used for a particular use-case needs to be (i) continuously updated with the evolution of the entities and the events in the real-world, (ii) organised in a tree structure, (iii) universal (not restricted to a particular domain) and (iv) suitable for real-time computation. Many large and available knowledge bases have been used in research for this purpose such as Wikipedia/DBpedia, DMOZ, WordNet, OpenCyc[6], etc. but they do not satisfy all the aforementioned requirements. *Wikipedia/DBpedia* for example is continuously updated and very large but its category structure is not a hierarchy, it is a graph [24]. *DMOZ, Wordnet* and *OpenCyc* on the other hand present a hierarchical structure but are not continuously updated by a large collaborative mass of users who keep the knowledge base up-to-date.

Following these requirements we decided to directly use the potentialities offered by the Web of Data as background knowledge. Thus, instead of using measures for hierarchical structures we use graph measures on the Web of Data graph leveraging the Linked Data principles. Because the entities and concepts are represented on the Linked Data cloud as nodes of a network, common network properties can be measured. In particular we can consider the Linked Data network as a directed labelled graph.

A. Basic Notations

A directed graph, or *digraph*, D, consists of a set of *vertices* $V(D)$, a set of *edges* $E(D)$, and a function which assigns each edge e an ordered pair of vertices (u, v). Normally, u is called the *tail* of e, v the *head* of e, and u, v the *ends* of e. If there is an edge with tail u and head v, then we let (u, v) denote such an edge, and we say that this edge is directed *from u to v*. For completeness, to use the Semantic Web terminology, and citing the RDF/XML Syntax Specification[7], a "RDF graph has nodes and labeled directed arcs that link pairs of nodes and this is represented as a set of RDF triples where each triple contains a *subject* node, *predicate* and *object* node. Nodes are RDF URI references, RDF literals or are blank nodes. [...] Predicates are RDF URI references and can be interpreted as either a relationship between the two nodes or as defining an attribute value (object node) for some subject node". In this paper we use both the terminologies for describing our measures applied to the Linked Open Data cloud. In particular, considering our previously defined graph D and its sets $V(D)$ and $E(D)$, we can say that a tail u is also called subject, a head v is a object, and an edge e is a predicate or a property. In a labelled directed graph we can define a set of *triples* $T(D)$ which consists of three elements: two vertices connected by a edge, $((u, v) \in V(D), e \in E(D))$. If there is an edge e with tail u and head v, then we let $(u, e, v) \in T(D)$ denote such a *triple*. As typical properties on a graph we can define: the *outdegree* of a vertex v, denoted $deg^+(v)$, as the number of triples with tail v, and the *indegree* of v, denoted $deg^-(v)$ as the number of triples with head v. In this paper, we substitute *indegree* with the acronym "**IP**" **(Incoming Predicates)** and *outdegree* with "**OP**" **(Outgoing Predicates)**. Additionally, we can also define other two measures that we call *"distinct outdegree"* and *"distinct indegree"* of a vertex v, respectively as the number of distinct edges with tail v and the number of distinct edges with head v. In this paper we call these two measures "**ODP**" **(Outgoing Distinct Predicates)** and "**IDP**" **(Incoming Distinct Predicates)**. They basically represent the number of ***distinct*** predicates that are connected to a subject or an object.

B. A Measure for Specificity

Having defined some basic properties of direct labelled graphs such as the Linked Data graph, we now describe the main measure proposed with this work and the other measures used in the evaluation as a comparison. Some of the common measures for importance or influence of nodes in a graph use the in/outdegree properties. For instance, graph analysis measures such as some centrality measures (degree centrality, eigenvector centrality, the PageRank, etc.) are frequently used to determine the importance of a vertex within a graph and they use the degrees of the nodes in their formulas. However importance, or influence, of a vertex is a different characteristic

[6]http://www.cyc.com/platform/opencyc

[7]http://www.w3.org/TR/REC-rdf-syntax/

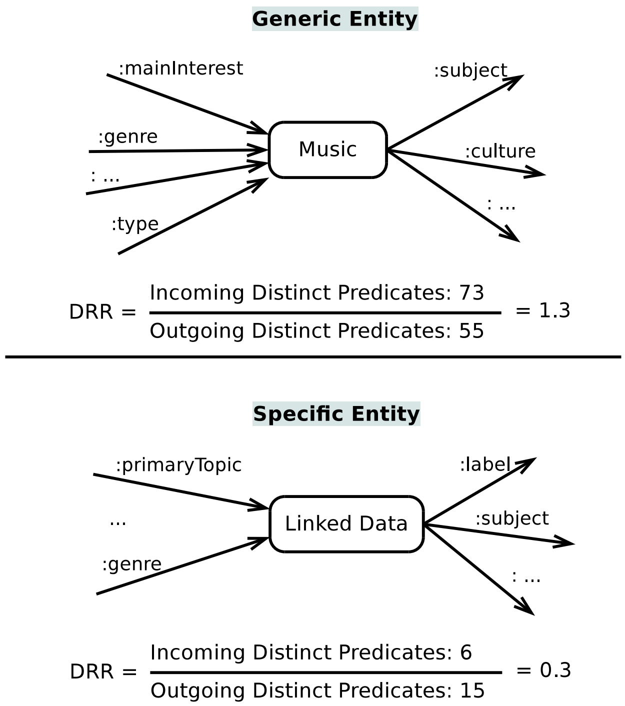

Fig. 1. Example of the DRR measure with two different entities.

compared to specificity. Moreover, in a directional labelled graph the labels of the edges have a meaning or a value assigned and hence this property should be taken into consideration. For example, a node that has many outgoing links but almost all of the same type is different from a node that has the same number of links but of many various types.

Analysing the predicates connecting entities on the LOD cloud (and in particular on DBpedia) we noticed that very specific entities have many different types of outgoing predicates compared to the incoming ones. Opposite behaviour for abstract concepts. From this observation then the hypothesis is that the ratio between Incoming Distinct Predicates (IDP) and Outgoing Distinct Predicates (ODP) characterises the specificity of entities. Thus, we formulate *our measure for specificity*:

$$\textbf{DRR}(DistinctRelationsRatio) = \frac{IDP}{ODP} \tag{1}$$

Where high DRR values mean abstract entities and low DRR values for specific terms. We compare our DRR measure with other similar measures such as: IP/OP, IP+OP, IP and different state of the art approaches (as described in Section VI). These alternative measures that we evaluate represent standard graph measures such as degree centrality in an undirected graph and in-degree and ratio between in- and out- degree values. The DRR measures shows the potential of a directed labelled graph such as the LOD graph.
The other very important advantage of the IDP/ODP measure is that it is simple and can be computed quickly. In fact, the datasets of entities on the LOD cloud are almost all of them indexed and can be queried on the Sindice project SPARQL endpoint[8]. Using a simple SPARQL query it is possible to interrogate the entities represented in RDF on the Web of Data and for instance get their in/out-degrees instantly.

[8]http://sindice.com/

V. Complementary Measures for Popularity and Trend on the Social Web

A. Popularity

Popularity expresses how much the entity is well known, shared or interesting to the majority of the people on the Web. This can be easily measured by looking at the frequency the entity is being mentioned on Social Media Systems and for this particular case we can use Twitter as the system where we evaluate our methodology. Naturally, just using Twitter as a source for this measure can provide biased results, however the large number of users on this microblogging platform and the extensive studies conducted on this source of information [19] demonstrate that the popularity of the entities being spoken about in the real-world is very close to their popularity on Twitter. What we suggest is a straightforward approach that utilizes the Twitter Search API to monitor the frequency an entity has been mentioned by users in a recent time frame. This is done with specific tools for named entity recognition and disambiguation on the resulting tweets from an initial search query in order to filter out ambiguous results. The result of this method is in our case an application that, given a DBpedia resource in input, returns a number representing the tweets per second being generated on Twitter about that entity (high numbers of tweets per second mean high popularity and vice versa). Another advantage of using this measurement on Twitter is that it allows a fast and real-time computation of the popularity, which is very important in our specific scenario. While this measure provides an instant picture of an entity at that specific point of time (a snapshot), it does not consider the temporal evolution of the popularity over time. Indeed, an entity or concept can be very popular at one specific point of time but not popular considering a longer period of time or vice versa.

B. Temporal Dynamics

The dynamics of the frequency of mention of entities on the Social Web over time provide accurate understanding on the popularity of terms, their evolution and their trend. Understanding these dynamics can be very useful for instance in recommendation systems. It would be possible to discard entities of interest from recommendations if they were popular only at a very specific point of time in the past (as they might be related to a specific past event). This measurement is particularly useful to characterise trendy entities which are very time-dependant and entities showing stable dynamics of popularity over time. For this measure we propose the usage of the Wikipedia page views. The number of Wikipedia page views for every day every year are publicly available and accessible through the Wikipedia MediaWiki API. This source of information provides an affective way for detecting the interests of the users on entities over time. In our experiments we use mean and standard deviation of the number of views for the past 30 days to distinguish concepts that are steadily popular over time or present relevant fluctuations in their dynamics. Following empirical experimentation, we define as *"stable"* a resource that has low standard deviation value for the page views of the last recent days (and "unstable" otherwise). Additionally, we define as *"trendy"* an entity that shows a clear increase of the number of page views only in the last recent days. In this case we use a simple method based on linear regression of the page views. In both cases a threshold has to be defined through empirical studies in order to perform the categorisation. Similarly to the other measures, this methodology is not computationally intensive, it is simple, and it is based on a large and continuously updated knowledge base.

VI. Evaluation of the Specificity Measure

To evaluate our approach for identifying in real-time the specificity of entities on the Web of Data we tested the measures described in Section IV-B on a set of 160 entities. The entities for our experiment were randomly selected from a large dataset of user profiles of interests generated for more than 50 different users. The profiles were automatically generated from the analysis of Facebook and Twitter user accounts as described in our previous work [22]. For each entity we computed and recorded the value of different measures (our DRR, and the non-distinct similar ones: IP/OP, IP+OP, IP) computed querying the Sindice SPARQL endpoint. We then compared those values with a gold standard generated by users classifying/rating the specificity levels of our test set of concepts. Additionally, we also reproduced a state of the art approach for measuring specificity based on the DMOZ hierarchical classification of the entities. We evaluate this other method against the gold standard and we then compare the accuracy of this method with our measure. The generation of the gold standard is described in the following section. The implementation of the DMOZ based method is detailed in Section VI-B and later in Section VI-C the evaluation and the results are examined.

A. Generation of the Gold Standard

Our gold standard has been generated through user manual annotation. The user evaluation set-up is composed of two interviews conducted at two different stages. **First**, we asked 5 evaluators (2 females and 3 males, different age groups and expertise) to **classify** each of the 160 entities in two categories: *Specific* or *Generic* entities. As suggested to the evaluators: *"the classification should indicate whether the entity is an abstract concept in the real world and can be further refined and specified into many other levels of detail (hence Generic) or if it corresponds to a well defined and narrow instance (Specific)"*.

At a **second** stage (2 weeks later) we asked the same users to give **a score** to the same entities according to their perceived level of specificity. Instead of a binary classification then we looked for a more fine grained value. The scale used for the scores goes **from 1 to 10** (only integer numbers) where 1 identifies very generic entities (or with a very low level of specificity) and 10 was given to very specific entities.

The **first round of evaluation** has been completed by the evaluators on average in 20 minutes, while the second type of evaluation took more time: around 30 minutes. The second stage of the user evaluation has been conducted after the feedback received from the evaluators at the first stage and after a preliminary analysis of the results. Briefly, according to the users, in several cases it was difficult to choose between only two levels of abstraction, as entities have different degrees of specificity. The results of the 5 evaluators for the first evaluation have been aggregated and the inter-rater agreement has been computed. We computed the Fleiss' generalised Kappa coefficient for 160 subjects, 5 raters and 2 categories and we obtained $K = 0.61$. This value, according for example to the scale for Kappa's significance by Rietveld and van Hout (1993) is considered as indicator of substantial agreement [9]. The 5 raters agreement for this classification process could then be used as a gold standard for the first evaluation.

At this stage the five evaluators classified 38% *Generic* concepts and 62% *Specific*. As the entities collected for the evaluation are randomly extracted concepts from user profiles of interests, it is reasonable to have such percentages. Ideally, if we think about taxonomies of concepts the number of those which are generic, and hence on top of a hierarchical classification system, are less than the specific ones which are closer to the leaves of the hierarchical tree.

For the **second evaluation**, as previously introduced in this section, the same five users were asked to rate the specificity of the same 160 entities on a 1 to 10 scale of integers. For this type of evaluation it was not appropriate to compute the Kappa coefficient for the inter-raters agreement, as the number of categories in this case was high (i.e. 10 categories). Hence, mean values and average standard deviation for the different ratings provided by the users were computed to estimate the agreement of the evaluators. In Table I we provide details about this part of the evaluation. An analysis of the results will be provided in Section VI-C.

Average Rate	7.03
Average Std. Dev.	1.45
Average Top30 High Std. Dev.	5.66
Average Top30 Low Std. Dev.	7.51

TABLE I. Second evaluation, rating: Details about the scores given by the five users.

Interesting to note that he average standard deviation of the ratings is 1.45 on a scale of 10 values, which is an acceptable value. Moreover, the average score given by the raters is 7.03, which confirms again the tendency highlighted by the first evaluation of having a higher percentage of specific concepts. Interestingly, the average score for the top 30 entities with highest, or lowest, standard deviation is respectively 5.66 and 7.51. This clearly means that entities with the highest disagreement among the evaluators have lower scores and hence are more generic. A behaviour observed also in the first evaluation.

The purpose of this second different experiment is first of all to analyse the raters agreement in two different tasks, as suggested also by the results of the first experiment. Moreover, with the fine-grained ratings provided by the users we could rank the specificity of the entities, use this ranking as our gold standard and compare it to the other different ranking strategies given by our Linked Data measures and the DMOZ categorisation. As previously described, especially in a use-case scenario where user profiles of interests need to be ranked and filtered for selecting the top most relevant and specific interests, it is beneficial to have fine-grained values allowing for specificity ranking methods. More details about the results are described in Section VI-C.

To evaluate the accuracy of the different ranking strategies we use the following prominent Information Retrieval ranking evaluation metric: the *Normalized Discounted Cumulative Gain (NDCG)* [15]. This evaluation metric supports graded judgments and penalizes error near the beginning of most relevant tags determined by our approach. NDCG is the normalized value of Discounted Cumulative Gain (DCG). The DCG accumulated at a particular rank position n is defined as:

$$DCG_n = rating_1 + \sum_{i=2}^{n} \frac{rating_i}{\log_2(i)} \quad (2)$$

where i is the rank of the result, $rating_i$ is the graded relevance of the result at position i. The Normalized DCG is then:

$$NDCG_n = \frac{DCG_n}{DCG_{ideal_n}} \quad (3)$$

where DCG_{ideal_n} is the DCG value computed with the benchmark ranking at position n. We use the ranking provided by the evaluators as our benchmark ranking or gold standard.

B. DMOZ Classification Method

We use a popular taxonomy such as DMOZ as a source for applying a state of the art method that can be evaluated against the gold standard and compared to our DRR measure. Here we explain how we used the DMOZ taxonomy to infer specificity levels of entities.

The Open Directory Project[9], also called DMOZ, combines the collaborative efforts of more than 96,877 volunteers helping to categorize the Web. ODP is one of the largest and most comprehensive human-edited Web page taxonomies. It is organized as a tree-structured taxonomy with over 1,014,849 categories and more than 5.1 million sites categorized[10]. The taxonomy powers core directory services for some of the most popular portals and search engines on the Web, including AOL Search, Google, etc. and it is also used in many research projects as a large-scale and structured background knowledge. Here we use the DMOZ taxonomy to manually assess the specificity of entities by looking at their position in the DMOZ hierarchical structure. We started with the assumption that entities classified in a hierarchy in a position close to the root are less specific (broader) than entities classified in positions close to the leaves. The ODP hierarchical tree is built with one root (*Top*) connected to 16 *Top Categories* (e.g. Arts, Science, Sports, etc.) expanding then into more than 1 million categories at different depth levels. A standard state of the art approach for identifying the level of specificity of entities is to match the entities to the corresponding category in the tree structure, and then count the number of levels separating the root node and the category node (Fig. 2). For our experiment we tried to match the 160 random DBpedia entities of our test dataset to the DMOZ taxonomy. As there is a difference between the two knowledge bases, we manually chose the closest match on DMOZ by identifying the category containing either the website clearly representing the entity, or the category with name almost identical to the entity name. Unfortunately we were able to match only 62 entities out of 160. The remaining entities had to be discarded as there was no equivalent on DMOZ or there were multiple possible categories to choose. This methodology, however, is comparable to the state of the art approaches that can automatically compute the specificity of terms using a structured and hierarchical background knowledge.

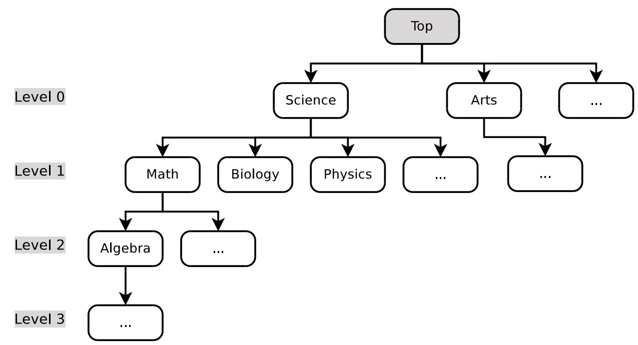

Fig. 2. Fraction of the ODP (DMOZ) taxonomy with our level numbers

To the 62 entities mapped on DMOZ we gave a score starting from the level of the 16 *Top Categories*. To this was given the level number 0, all the immediate subcategories were assigned the level 1 and so on, continuing increasing 1 level for each sub-category level (as depicted in Figure 2). Overall, for the 62 DMOZ entities, the average value is 4.1 with maximum value 9 and minimum 0. The entities that were categorised by the five evaluators as *Specific* in our first evaluation (see previous Section) on the DMOZ hierarchy got an average value of 5.2 with standard deviation 1.5, while the *Generic* ones got an average value of 2.7 with standard deviation 1.2. According to these average values we selected our threshold for classifying the concepts as either *Generic* or *Specific*. The threshold selected is the level number 4: entities categorised with a level lower than 4 were classified as *Generic*, and with a value greater or equal to 4 were classified as *Specific*. This classification provided us 42 specific entities and 20 generic out of the total 62. This classification has been compared with the user based classification of the first evaluation and the DMOZ scores have been compared (as a specificity ranking strategy) with the user-based benchmark ranking and our Linked Data automatic rankings. More details in the following Section VI-C.

C. Analysis of the Results

In this section we analyse the results of the two experiments conducted in order to evaluate the performance of our DRR measure compared to the gold standard. We additionally evaluate the performance of the other link-based measures (IP/OP, IP+OP, IP) and a state of the art approach using DMOZ as a background knowledge.

1) First Evaluation: Classification.: On our evaluation dataset of 160 random DBpedia entities of interest we performed the classifications with the different methodologies as explained previously in this section. Afterwards, the precision of the DMOZ classification and the Linked Data measures have

[9]ttp://www.dmoz.org

[10]From the ODP website, accessed in January 2013

been computed against the manual classification performed by the 5 human evaluators. With *agreement* here we intend the number of entities classified in the same way by the two methods over the total number of entities of the dataset. In Table II we show the results of this stage of evaluation.

DMOZ	DRR	IP/OP	IP+OP	IP
0.839	**0.841**	0.700	0.700	0.725

TABLE II. FIRST EVALUATION: AGREEMENT OF THE DIFFERENT METHODS COMPARED TO THE MANUAL USER CLASSIFICATION.

As we can see from the results the DRR measure and the DMOZ classification have similar performance compared to the manual classification. For around 84% of the entities the two strategies classified the entities in the same way as the human evaluators. All the other LOD-based measures perform clearly worse in this classification task (around 10% worse) as they correctly match only 70/72% of the manually classified concepts. To note again that for the DMOZ method less entities are evaluated (only 62) because of the mismatch between DMOZ and DBpedia. According to these initial results our automatic measure has comparable performance with state of the art approaches such as those using a taxonomy like DMOZ as a background knowledge. The clear advantages in using Linked Data is that the background knowledge is extended on a Web scale, it is always updated with the quick evolution of the Social Web, it does not need to be pre-processed or stored and simple measures like the DRR can be computed in real-time.

2) Second Evaluation: Ranking.: Despite the positive results obtained by the first evaluation, the experiment continued with a different scope: the capability of a method to rank the specificity of a set of entities. This revealed to be necessary after the feedback received by the users on the complexity of the first classification task and their need to express a more fine-grained score for the specificity. This evaluation tests the performance of our methods in ranking specificity of concepts compared to state of the art approaches and user-based rating.

For all the aforementioned methods we use the NDCG metric described in Section VI-A and we perform the ranking experiment on a subset of 50 randomly chosen entities from our complete test dataset of 160 entities. This is because of the intrinsic reduction in reliability of the NDCG measure when computed on a high rank position (effect of logarithmic reduction factor of DCG). All the NDCG measures have been computed using the human rating as gold standard (ideal ranking). In Table III we summarise the NDCG values obtained for the different methods at some rank positions p.

It is clear that our DRR method that uses distinct properties is performing better than the other methods. In particular for the first 20 rank positions, where the ranking is on average almost 5% better. To note that the NDCG for the IP measure has been computed but not shown in the table as it is very close to the IP+OP one. The random method shown in the table is just a random ranking function that we evaluated as a comparison. As for the DMOZ method we had to compute the NDCG values with two different strategies. Since the DMOZ method does not provide a fine grained score to the entities but only maximum 10 possible values (unlike the other methods that are then more suitable for rankings), multiple equal scores were given to groups of entities. We then had to rank the entities first following the DMOZ method and then, for the entities sharing the same score, rank them again according to two possible rankings given by the gold standard. Therefore, following the human ranking we were able to provide the *worst* possible DMOZ ranking (*DMOZ-*) and the *best* possible one (*DMOZ+*) for the same entities. Even in this second evaluation our proposed method for characterising the specificity of entities using the DRR measure is performing better than all the other evaluated methods.

NDCG	DMOZ-	DMOZ+	DRR	IP/OP	IP+OP	random
p=10	0.902	0.923	**0.968**	0.911	0.897	0.725
p=20	0.924	0.933	**0.966**	0.921	0.928	0.774
p=50	0.965	0.975	**0.986**	0.972	0.965	0.898

TABLE III. SECOND EVALUATION: NDCG AT DIFFERENT RANK POSITIONS p FOR ALL METHODS USING MANUAL HUMAN RANKING AS GOLD STANDARD.

VII. EVALUATION ON USER PROFILES OF INTERESTS

So far we have proposed a model and a set of measures for characterising entities of interest and we have evaluated the accuracy of our novel measure for concept abstractness (specificity). In this section we present an evaluation of the impact of this model directly on user profiles of entities of interest. We generated user profiles, as described in our previous work [22], for 27 users (volunteers for our user study). To each user we asked to rate the relevance of 30 entities of interest according to their personal preferences. The entities of interest were generated and ranked for their user profile according to their activities on Facebook and Twitter and their number of mention in their social data (occurrence based weighting strategy). For more details on the profile generation we refer to [22]. In total we collected 794 user ratings (not 810 because some users evaluated less than 30 entities) on a scale from 1 (low relevance) to 5 (high relevance), on a total of 529 distinct DBpedia resources as interests. For every entity we computed our measures as described in Sections V and IV, and we analysed the average user score grouping by each different feature.

Type of entity	Tot. Entities	AVG Score	Std.dev.
All	794	3.34	1.47
Non-Specific	297	3.66	1.39
Non-Popular	410	3.40	1.46
Stable	663	3.37	1.47
Non-Trendy	778	3.35	1.47
Stable & Non-Trendy	659	3.38	1.47
Non-Popular & Non-Specific	134	**3.84**	1.39

TABLE IV. EVALUATION OF THE AVERAGE USER SCORES (ON A 1 TO 5 SCALE) GROUPED BY TYPE OF ENTITY OF INTEREST.

As we can see from Table IV the entities of interest categorised as "Non-Specific" (which have high values for our Specificity measure) provide an improvement on the user score of almost 8% on the average score for all the interests. This means that users perceive concepts of interest which are more abstract as more relevant for their user profiles. This improvement has been confirmed by the tests for statistical significance performed. The other measures (Popularity, Temporal Stability and Trend) do not show significant improvement on the average score. We also show the effect of aggregation of two types of interests: "Stable plus non-Trendy ones" and "non-Popular plus Non-Specific" concepts. The latter give best results with more than 12% improvement over the average user score and demonstrates the validity and complementarity of these two measures. To note that the thresholds chosen for the binary classification of every measure are the average of all the values for the measure.

With this user study we provided insight on the effect of these dimensions on user profiles, however in our future work we plan to measure the impact of these features on recommendations. We believe there is a difference between the relevance a user perceives on an entity of interest and the one on an object recommended according to an interest in a user profile. For instance, our hypothesis is that users prefer, on the one hand, to be categorised with abstract interests and, on the other hand, to receive recommendations related to specific interests. We plan to verify this hypothesis and also demonstrate that the specificity measure can be used for refining any recommender system, by assigning different weights to specific or generic entities of interest.

VIII. Conclusions

In this paper we proposed and evaluated a novel approach for characterising important dimensions of entities of interest: specificity, popularity and temporal dynamics. We showed how this features are relevant in general for social web users and in particular for our specific use case which is about personalisation of a microblog stream in real-time. Specificity in particular is extremely relevant for recommendations and personalisation on the Social Web as we showed that user interests can be ranked also according to their conceptual level of abstraction. Trend and popularity of concepts on the Social Web can be considered complementary to the specificity and enrich the semantics and the pragmatics of entities. A new lightweight and domain-independent approach for automatically identifying specificity of entities in real-time has been detailed (the DRR measure) and it is based on the Linking Open Data cloud. An evaluation of the validity of the DRR has been performed by comparing the results of our measure against a gold standard (generated with user ratings) and a traditional state of the art approach based on the DMOZ taxonomy. The results of the evaluation clearly show that our method outperforms the other methods and its performance in ranking the specificity of entities is close to the the gold standard. Finally, we also evaluated the validity and the impact of our methodology on user profiles of interests through a user study with 27 users and almost 800 ratings. The study mainly suggests that abstract entities provide better scores when user profiles are evaluated by the users themselves and that specificity and popularity positively complement each other.

As future work we plan to apply and evaluate these dimensions of entities to a system for filtering the public Twitter stream and to measure the impact of the presented features on social media recommender systems. Moreover, we will improve the DRR measure to increase its accuracy and to expand the semantic features that can be considered on the Web of Data.

References

[1] F. Abel, Q. Gao, G. Houben, and K. Tao, "Analyzing user modeling on twitter for personalized news recommendations," *UMAP*, 2011.

[2] F. Abel, C. Hauff, G. Houben, R. Stronkman, and K. Tao, "Semantics + filtering + search = twitcident. exploring information in social web streams," in *Hypertext*, 2012.

[3] E. Amitay, D. Carmel, M. Herscovici, R. Lempel, and A. Soffer, "Trend detection through temporal link analysis," *J. Am. Soc. Inf. Sci. Technol.*, vol. 55, no. 14, Dec. 2004.

[4] J. Benhardus, "Streaming trend detection in twitter," *Natural Language Processing and Information Retrieval*, 2010.

[5] D. Benz, C. Körner, A. Hotho, G. Stumme, and M. Strohmaier, "One tag to bind them all: Measuring term abstractness in social metadata," in *ESWC 2011*, 2011.

[6] C. Bizer and et al., "Linked data – the story so far," 2009.

[7] M. Ciglan and K. Nørvåg, "Wikipop: personalized event detection system based on wikipedia page view statistics," ser. CIKM, 2010.

[8] L. Ding, T. Finin, A. Joshi, R. Pan, R. S. Cost, Y. Peng, P. Reddivari, V. Doshi, and J. Sachs, "Swoogle: a search and metadata engine for the semantic web," ser. CIKM, 2004.

[9] B. D. Eugenio, "On the usage of Kappa to evaluate agreement on coding tasks," *Proceedings of LREC*, 2000.

[10] J. Fiaidhi, S. Mohammed, and A. Islam, "Towards identifying personalized twitter trending topics using the twitter client rss feeds," *Journal of Emerging Technologies in Web Intelligence*, vol. 4, no. 3, 2012.

[11] T. Franz, A. Schultz, S. Sizov, and S. Staab, "Triplerank: Ranking semantic web data by tensor decomposition," ser. ISWC, 2009.

[12] A. Freitas, J. G. Oliveira, S. O'Riain, E. Curry, and J. C. P. Da Silva, "Querying linked data using semantic relatedness: a vocabulary independent approach," ser. NLDB, 2011.

[13] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blogspace," ser. WWW, 2004.

[14] D. Irani, S. Webb, C. Pu, and K. Li, "Study of trend-stuffing on twitter through text classification," in *Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS)*, 2010.

[15] K. Järvelin and J. Kekäläinen, "Ir evaluation methods for retrieving highly relevant documents," in *ACM SIGIR*, 2000.

[16] P. Kapanipathi, F. Orlandi, A. Sheth, and A. Passant, "Personalized filtering of the twitter stream," in *SPIM Workshop at ISWC*, 2011.

[17] M. Mathioudakis and N. Koudas, "Twittermonitor: trend detection over the twitter stream," ser. SIGMOD, 2010.

[18] P. Mendes, A. Passant, and P. Kapanipathi, "Twarql: tapping into the wisdom of the crowd," in *I-SEMANTICS 2010*, 2010.

[19] G. Miller, "Social scientists wade into the tweet stream," *Science*, vol. 333, no. 6051, 2011.

[20] R. Mirizzi, A. Ragone, T. Di Noia, and E. Di Sciascio, "Ranking the linked data: the case of dbpedia," ser. ICWE, 2010.

[21] S. Noor and K. Martinez, "Using social data as context for making recommendations: an ontology based approach," in *1st Workshop on Context, Information and Ontologies*, ser. CIAO '09, 2009.

[22] F. Orlandi, J. Breslin, and A. Passant, "Aggregated, interoperable and multi-domain user profiles for the social web," in *I-SEMANTICS*, 2012.

[23] A. Passant, "dbrec: music recommendations using dbpedia," in *9th international semantic web conference*, ser. ISWC'10, 2010.

[24] S. P. Ponzetto and M. Strube, "Knowledge derived from Wikipedia for computing semantic relatedness," *Journal of Artificial Intelligence Research*, 2007.

[25] K. Rajaraman and A. Tan, "Topic detection, tracking, and trend analysis using self-organizing neural networks," in *5th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, ser. PAKDD '01, 2001.

[26] J. Ratkiewicz, S. Fortunato, A. Flammini, F. Menczer, and A. Vespignani, "Characterizing and modeling the dynamics of online popularity," *Physical review letters*, vol. 105, no. 15, 2010.

[27] J. W. Tanaka and M. Taylor, "Object categories and expertise: Is the basic level in the eye of the beholder?" *Cognitive Psychology*, 1991.

[28] A. Thalhammer, I. Toma, J. A. Roa-Valverde, and D. Fensel, "Leveraging usage data for linked data movie entity summarization," *CoRR*, 2012.