

User Profile Modelling in Online Communities

Miriam Fernandez¹, Arno Scharl², Kalina Bontcheva³, Harith Alani¹

¹Knowledge Media Institute, The Open University, UK
{m.fernandez, h.alani}@open.ac.uk

²Department of New Media Technology, MODUL University Vienna, Austria
scharl@modul.ac.at

³Department of Computer Science, University of Sheffield, UK
k.bontcheva@dcs.shef.ac.uk

Abstract. With the rise of social networking sites user information is becoming increasingly complex and sophisticated. The needs, behaviours and preferences of users are dynamically changing, depending on their background knowledge, their current task, and many other parameters. Existing ontology models capture demographic information as well as the users' activities and interactions in online communities. These vocabularies represent the raw data, but actionable knowledge comes from filtering these data, selecting useful features, and mining the resulting information to uncover the most salient preferences, behaviours and needs of the users. In this paper we propose reusing and re-engineering ontological resources to provide a broader representation of users and the dynamics that emerge from the virtual social environments in which they participate.

Keywords: Semantic Web, Social Web, User Profile

1 Introduction

It is crucial for service providers to adequately understand the *needs*, *preferences* and *behaviours* of their users to ensure that their services are delivered to the right people at the right time. However, achieving such understanding of the user, based on a wide range of inter-dependent attributes and implicit information, is a complex research task. The user's current situation, past history and social environment need to be combined and integrated. Data about the time and activity of users should be linked with the users' past information to understand their current situation; previous activities and interactions should be taken into account to interpret and fully understand this situation; relations with other people and other user behaviour in similar contexts should be also considered and captured.

With the emergence of the Social Web and social networking sites such as Facebook, Twitter, Google+ and YouTube, a vast amount of personal information is created on a daily basis. The scale of this personal and social context data has a huge potential to improve the coverage of user modelling approaches and enhance the effectiveness of adaptive systems.

Multiple efforts have emerged from the Semantic Web (SW) community to target this problem. Vocabularies in standard representation formats, such as RDF and OWL, have been developed, to model users and their social context. Examples of these vocabularies include FOAF – Friend of a Friend [6] and extensions like the Relationship Vocabulary [17], SIOC [2;9], OPO – Online Presence Ontology [15], or MOAT – Meaning of a Tag [7]. While these ontologies do indeed capture user interactions within online communities, they do not model more dynamic user aspects such as behavioural evolution within the community. The aforementioned vocabularies represent the raw data, but actionable knowledge comes from filtering the vocabularies, selecting useful features, and mining the profile data to uncover the most salient preferences, behaviours and needs of the users.

In this paper we present a user profile model that goes beyond capturing raw data from user activities and interactions to capture the interpretation of these data within particular contexts. To generate this user profile model, we reuse existing ontological resources for modelling users and their social context, and extend this knowledge with well-known features extracted from current social media analysis methods [19, 32, 33, 34, 35]. By modelling and storing these features we enable inferences to be made over a richer layer of data, allowing the dynamic learning of user preferences, needs and behaviours.

To generate user profiles, we have followed the NeOn methodology [18], and its guidelines for reusing and re-engineering ontological resources. According to this methodology, three main steps should be followed: (1) select the most suitable ontological resources to be reused; (2) carry out the ontological resource re-engineering process to modify the selected ontological resources, and (3) assess if the modified/new ontology fulfils the ontology requirement specification. The ontology requirement specification states why the ontology is being built, what its intended uses are, and which requirements the ontology should fulfil.

The main use case for the presented user profile model is a collective awareness platform currently being built as part of *DecarboNet* (www.decarbonet.eu), a research project that aims to increase environmental awareness, trigger behavioural change and track the resulting information diffusion patterns across various social networks. The collective awareness platform of DecarboNet will consist of a knowledge co-creation environment embedded into an existing media analytics platform available at www.ecoresearch.net/climate [41], an upcoming social media application in the tradition of games with a purpose, and a portfolio of analytic services to identify patterns in both the structure of social networks as well as the content communicated between the nodes of these networks. The dynamic user and context models of DecarboNet will help to integrate the observable data flows across these components. They will enable analysts to capture the role of users in social innovation processes, assess their environmental knowledge and information seeking behaviour, and measure their engagement level within the DecarboNet community.

The remainder of this paper is structured as follows: Section 2 presents the ontology requirements specification. Section 3 outlines the available ontological resources. Section 4 describes the re-engineering process and the proposed extensions and modifications to the selected ontological resources. Section 5 discusses the results and concludes the paper.

2 Requirements Specification

The goal of systems that use personal data is to gain the capability to adapt aspects of their functionality or appearance to the preferences and needs of their users. To do so, the system must have an internal representation (i.e. a model/profile) of the user. Approaches that generate these user profiles generally distinguish among: (i) *modelling* – which information defines the user? (ii) *representation* – which formats and structures are used to represent the user profile? (iii) *acquisition and update* – how the previous identified information is acquired and evolves over time?

In this paper we focus on the problem of *user modelling*. Standard semantic formats such as RDF and OWL have been selected to represent our proposed user profile. Regarding the problem of user profile acquisition and update, we provide a brief overview of how the proposed user profile is currently being acquired and updated. Table 1 presents the Ontology Requirements Specification for our proposed user model following the NeOn methodology [18]. This methodology proposes the development of a filling card, and more particularly, a set of competency questions to assess whether the ontology fulfils the requirements. The resulting filling card for our user profile ontology is displayed below:

Table 1: Ontology Requirements Specification

Purpose: The purpose of building the user profile ontology is to provide a reference model for capturing the dynamics of user profiles in online communities
Scope: The scope of this ontology is the user in the context of online communities
Implementation Language: The ontology is implemented in OWL
Intended Users: The intended users of this ontology are adaptive systems or social media analysis modules. No human users are intended for this ontology
Intended Uses: <ul style="list-style-type: none">- To dynamically infer, for a user, her exhibited behaviour within a particular online community and moment in time- To dynamically infer, for a user, her needs within a particular online community and moment in time- To dynamically infer, for a user, her preferences within a particular online community and moment in time- To infer, for a user, her personality from her previous actions across online communities
Ontology Requirements: <ul style="list-style-type: none">(a) Non-Functional requirements: none(b) Functional requirements: defined by four main competency questions:<ul style="list-style-type: none">- CQ1: What is the behaviour that user u adopts in the online community oc_x during the time period $t1-t2$?- CQ2: What are the needs of user u in the online community oc_x during the time period $t1-t2$?- CQ3: What are the preferences of user u in the online community oc_x during the time period $t1-t2$?- CQ4: What is the personality of user u?

As we can see in this filling card, the intended use of the ontology is to be able to infer, the *behaviour*, *needs*, *personality* and *preferences* of the user within a particular online community and moment in time. These concepts are discussed in detail in Section 4. User profiling based on the presented ontology and concepts aims to fulfil the Decarbonet requirements by enabling a structured analysis of behaviour patterns. Detecting user types (e.g. those likely to change their behaviours as a result of a specific intervention strategy) and updating dynamic user models on-the-fly will be used by the DecarboNet platform to guide data acquisition and filtering processes, form ad-hoc communities based on shared interests, devise effective engagement strategies, and provide tailored information services for citizens.

3 Ontology Selection

In this section we explore the ontologies developed so far to semantically model users particularly, in the context of online communities:

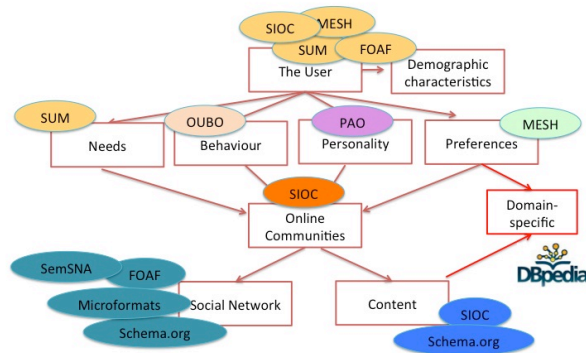


Fig. 1: Existing ontologies capturing the different user profile aspects in the context of online communities.

As we can see in **Fig. 1** our proposed user profile model aims to capture multiple aspects of the *user* and the online communities in which she participates. Among the aspects that the model aims to capture for the user we can highlight static elements, such as her demographic characteristics; but also more complex and dynamic elements, such as her needs, her behaviour, her personality and her topic (domain-specific) preferences. These aspects are inferred from the actions and interactions of the user within an online community. The online community provides information not only about the *social network* of the user (the people she interacts with) but also about the *content* she produces. Among the ontologies that aim to capture user information within the context of online communities we can highlight:

FOAF, the Friend Of A Friend vocabulary [6] describes people, their properties such as name, homepage, etc., and the social connections of different users by means of the foaf:knows relationship. This property allows people to be linked to one another across social web platforms.

The *Schema.org* vocabularies, [10] agreed among the major search engine providers (Google, Bing, Yahoo! and Yandex), are able to capture the knowledge about people and their social networks. They provide a collection of tags to define item types (Person, Place, Organisation, Review, Event, etc.) and social relations (knows, colleague, children, parent, sibling, relatedTo, etc.).

Microformats [11] provide vocabularies to describe people as well as their social connections. The hCard micro format [12] represents people and their attributes such as given name, family name, URL, email, etc. The XFN microformat [13] captures users' social networks by representing the relations between people via the 'rel' attribute (e.g.,).

Semantic Social Networks Analysis [14] (SemSNA) provides an understanding of the structure of networks, including richer representations of social links: cyclic path, directed path, betweenness, centrality, etc.

The *Online Presence Ontology* (OPO)[15] models the online presence of users. It proposes classes to describe the findability, noticeability or online status of users as well as their actions (working on a project, reading, listening, etc.)

The *SIOC* (Semantically Interlinked Online Communities) ontology [9], originally designed to capture the knowledge of discussion boards, models not only users and social interactions, as in previously mentioned works, but also content, and the reply-chain in which this content has originated. This ontology is based on, and reuses classes and relations from, several well-known ontologies such as the Friend Of A Friend (FOAF) vocabulary [6] and Dublin Core Metadata Terms (dcterms). [16]

Other works have also attempted to reuse some of these vocabularies to provide community-focused descriptions. One of these examples is the Facebook Open Graph Protocol, which can model and interlink users and objects within the Facebook social network. As opposed to this model, SIOC is not tailored to any particular social networking platform.

To the best of our knowledge, SIOC is the most complete and generic ontology developed to date to capture the knowledge of online communities. It does not only capture knowledge about users and their social interactions, as Microformats or FOAF do, but it also captures knowledge about the content and the content generation process. Additionally, as opposed to the Facebook Open Graph Protocol, its purpose is more generic and has not been designed with a particular online community in mind, building a crucial base for data integration and unification across different online communities. Given its popularity and adoption, SIOC is selected as the base of our proposed user model.

In addition to the previously presented ontologies, which capture demographic information about the users, as well as their actions and interactions within online communities; we have surveyed ontologies aiming to capture more complex user aspects. Capturing users' behaviour, personality, needs, or preferences can enable systems to provide better adaptations of their functionality or appearance.

While multiple ontologies can be found in the literature that aim to capture the domain or topic preferences of users for personalisation and recommendation [22, 36, 37], very few ontologies have been proposed to capture the behaviour of users, their needs or their personality.

Regarding behaviour modelling we can highlight the works of Ankolear et al. [38] and Rowe et al. [19]. Ankolear et al. [38] describe user roles in problem-solving

communities: bug fixer, bug reporter, contributor, developer, etc. While this work is focused on a specific type of online communities, Rowe et al. [19] propose a more generic model, the Open University Behaviour Ontology (OUBO); able to capture different user roles for online communities with different focuses.

Regarding user needs interpretation in the context of online communities, current research has focused on capturing user needs by: (i) applying well-known social theories such as Maslow's pyramid of needs or the self-determination theory to the world of online communities [27, 28, 29] or, (ii) explicitly asking users about their needs via questionnaires [30, 31]. The Semantic Web User Model (SWUM) [39] captures some of these user needs, as well as elements of the behaviour and personality of users.

Regarding the interpretation and understanding of users' personality in the context of online communities, current research has also focused on applying well-known social theories, such as the big-five personality traits [32, 33, 34]. The Personality Assessment Ontology (PAO) [40] captures this personality theory. Other ontologies like SWUM capture personality in the form of user characteristics such as "kind", "warm", "calm". While ontological models like SWUM capture needs and preferences of the users, they do not consider the dynamics of these user aspects. E.g., a user may exhibit different needs in different online communities or at different points in time. User's behaviours, needs and preferences are dynamic aspects and should be captured in context.

4 Dynamic User and Context Modelling

As we have seen in our previous section, existing vocabularies, either (i) capture raw data about the user and her social environment, but do not model more complex and dynamic aspects of the user (needs, personality, preferences, etc.) or (ii) they model more complex aspects of the user but they just capture a snapshot, from which the evolution over time, or in different communities, cannot be inferred. Our proposed user model aims to address these issues by reusing and extending some of the previously presented vocabularies.

4.1 Modelling User Actions and Interactions in Online Communities

To capture data about the user and her actions within online communities we have chosen SIOC as the base of our user profile model. SIOC makes use of the class *sioc:UserAccount*. This class reuses properties from other vocabularies, such as: *sioc:name*, which captures the name of the user, *dc:created*, which captures the time and date the user account was created, *sioc:creator_of*, which links the user to the content she generates or *foaf:knows*, which links the user with her social network.

To model the content creation process and the interaction of the user with other community members SIOC makes use of classes such as *sioc:Container*, *sioc:Thread* and *sioc:Post*. The class *sioc:Post* has the property *sioc:has_creator* that links the post with a particular user account. This class also has the property *sioc:hasParent*,

that links the Post with a particular Thread. The properties *sioc:reply_of* and *sioc:has_reply*, link the post to other posts, and the properties *sioc:content* and *sioc:created* capture the text of the content and the date/time it was posted. The class *sioc:Thread* is also linked to a particular *sioc:Container* (forum, blog, etc.) by the property *sioc:has_parent*. By reusing the SIOC classes and properties our model captures demographic information about the user as well as her actions within online communities (when the user posts a message, when she replies, etc.)

4.2 Modelling User Context

There are two types of context we wish to define to capture user dynamics and evolution: *location* and *time*. For the former we can use SIOC classes such as *sioc:Forum*, *sioc:Community*, etc., to represent the social virtual environment where the user, defined as an instance of *sioc:UserAccount*, is participating. To model time we reuse the class *oubo:TimeFrame* from the OUBO ontology [19]. The class *oubo:TimeFrame* defines a given time period in which users' features (see Section 4.3) are computed. We combine the temporal and location context aspects into a single context instance using the class *social-reality:C*. The class *social-reality:C* is reused from Hoekstra's work [20] and is used to represent a higher-level notion of context that can be used to include additional contextual information, apart from location and time.

4.3 Modelling User Behaviour

To capture and infer user behaviour, we propose an extension of the OUBO ontology [19]. This ontology uses SPIN rules to infer the role (*oubo:Role*) that a user (*sioc:UserAccount*) has in a given context (*social-reality:C*). To infer the role that a user assumes (Leader, Follower, Broadcaster, etc. [19, 35]) we need to capture fine-grained information about the user (user features). We propose to extend the OUBO ontology and to model user features under six different behavioural dimensions:

- *Popularity*: the popularity of a user measures whether the user is being liked, admired, or supported by many people.
- *Engagement*: the engagement of a user measures up to which level the user is committed to the community.
- *Initiation*: the initiation of a user measures how much the user instigates discussions and asks questions.
- *Contribution*: the contribution of a user measures the extent to which the user contributes or replies to threads initiated by other users.
- *Content Quality*: The content quality of a user measures her level of expertise and how useful her posted content is for the topic under discussion.
- *Focus Dispersion*: the focus dispersion of a user measures whether the user disperses his/her activity across many forums/sub communities/sub topics or concentrates his/her activity in a few forums/sub communities/sub topics.

User behavioural features can be computed using a variety of metrics. Table 2 presents some of the most common metrics used in the literature.

Table 2: Example of metrics to measure user behavioural dimensions.

User Dimension	Metrics
Popularity	<p><i>In-degree</i>: number of users that have replied to the user. A larger value indicates that the user is popular within the platform</p> <p><i>Post-replied ratio</i>: proportion of posts by the user u_i that yield a reply. This feature is used to gauge the popularity of users' content based on replies</p>
Engagement	<p><i>Out-degree</i>: proportion of users that the user has replied to. A larger number indicates that the user has contacted many different community members</p> <p><i>Bi-lateral neighbours ratio</i>: the proportion of neighbours where a reciprocal interaction has taken place - e.g. u_i replied to u_j and u_j replied to u_i. This measure allows the reciprocal engagement of the user to be captured where higher values demonstrate a tendency to interact</p> <p><i>Message-count</i>: total number of messages written by a user within a community</p>
Initiation	<p><i>Thread-initiation ratio</i>: measures the proportion of threads started by the user. Let P_s be set of thread starters authored by all users and $P_{s,i}$ be the set of thread starters authored by u_i. Thread-initiation ratio u_i is defined as $P_{s,i} / P_s$</p>
Contribution	<p><i>Thread-contribution ratio</i>: measures the proportion of thread replies that are created by the user. Let P_r be the total set of replies authored by all users and $P_{r,i}$ be the set of replies authored by u_i, thread-contribution ratio is defined as the contribution of u_i as $P_{r,i} / P_r$</p>
Content quality	<p><i>Average points</i>: measures the average points per post awarded to the user. This feature provides a measure of expertise of the user but is only valid for those communities that maintain explicit content quality ratings</p>
Focus Dispersion	<p><i>Forum entropy</i>: Let F_{u_i} be all the forums where the user u_i has posted and $p(f u_i)$ be the conditional probability of u_i posting in forum f. This can be derived using the post distribution of the user. Therefore the Forum Entropy of a given user is defined as (H_f)</p> $H_F(u_i) = -\sum_{j=1}^{ F_{u_i} } p(f_j u_i) \log p(f_j u_i)$

It is important to note that user features should always be considered in context, i.e., the value associated to these dimensions might vary along time, and within different online communities. The time-period and frequency used to compute these features strongly depends on the online community's activity level. A community with frequent and high activity levels will require computing user behavioural features over shorter time periods to appropriately capture behavioural evolution.

To model user behavioural dimensions and its associated set of features we extend the class *oubo:UserImpact* (see Fig. 2). This class aims to model the impact of the user in a certain time period, and in a certain online community. Our User Profile Ontology (UPO) extends the OUBO ontology with the class *upo:UserDimension*. This class captures the different user dimensions under which impact can be measured. Several subclasses have been defined so far to capture these dimensions, such as *upo:Popularity*, *upo:Engagement*, etc. Note that more subclasses can be added to increase the number of user dimensions. As previously explained, each of these dimensions can be measured using different metrics. To capture this notion we have created the class *upo:Metric* and its different subclasses, which represent the set of metrics described in Table 2: *upo:inDegree*, *upo:OutDegree*, etc. Each

upo:UserDimension is associated with one or more *upo:Metric* by the relation *upo:hasMeasureFunction*.

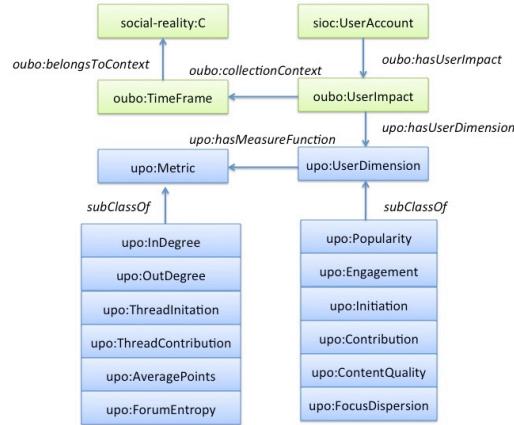


Fig. 2: Extensions proposed to capture the different behavioural dimensions

To infer the different roles that a user adopts over time we apply semantic rules encoded using SPIN (e.g., if popularity=high and contribution=high then role=leader). For more details of the role extraction process the reader is referred to [19]. Note that using the notion of context, features, and SPIN rules the proposed ontology fulfils **CQ1**, i.e., it can infer the behaviour (role) that user u adopts in an online community oc_x during a particular time period.

4.4 Modelling User Preferences

To capture and model user preferences semantically we build on our previous work and reuse parts of the MESH ontology [21]. This ontology has been used to model user preferences and has proven its effectiveness for personalisation and recommendation tasks [22]. Ontology concept-based preferences are more precise, and reduce the effect of the ambiguity caused by the use of keyword terms. For example, a preference stated as "*ProgrammingLanguage:java*" (this reads as the instance Java for the Programming Language class) lets the system understand unambiguously the preference of the user does not refer to the pacific island. Additionally, the multiple relations modelled in ontologies and their inference capabilities allow the inference of underlying user interest. For instance if a user is interested in *skiing*, *snowboarding* and *ice hockey* it can be inferred, with a certain degree of confidence, that the user is globally interested in *winter sports*.

To model user preferences we extend the class *sioc:UserAccount* with the properties *mesh:semanticInterest* and *mesh:itemRatings*. The property *mesh:semanticInterest* links the user with the class *mesh:SemanticInterest*. This class is modelled as a vector of *mesh:WeightedConcept* that represent the preferences of the user in terms of semantic concepts. A *mesh:WeightedConcept* class is represented by three main

properties *mesh:concept*, that captures the conceptual preference of the user, *mesh:weight*, that represents the preference score for that particular concept and *mesh:timestamp*, that represents the moment in time in which the user expressed interest for that particular concept.

To populate the preferences of our user profile model we make use of existing semantic annotators that are able to extract the subset of concepts expressed by the users in their posts. At the moment we make use of TextRazor to extract these concepts [23] but other systems, such as Alchemi API [24] or DBpediaSpotlight [25] could also be used. Note that TextRazor extracts concepts from DBpedia and FreeBase, to our knowledge, two of the most complete knowledge bases up to date. Note that concepts with a confidence score lower than 3, in a scale from 0.5-10, are discarded. The preference level of the user for the concept is based on a sentiment analysis of the content. The SentiCircles sentiment analysis approach is used to compute the sentiment of the extracted concepts [26].

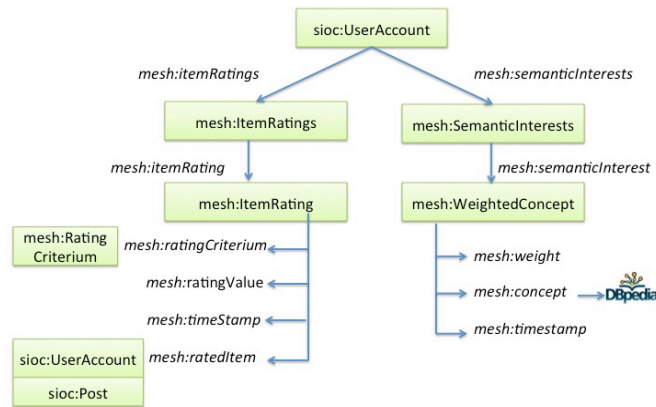


Fig. 3: Reused classes and properties of the MESH ontology to model preferences

In addition to the modelling of concept-based user preferences we also capture preferences in terms of ratings. In certain online communities users can provide ratings to express their preference for other users or for certain content. Preferences in terms of ratings are modelled with the class *mesh:ItemRatings*. This class is linked to a *sioc:UserAccount* via the property *mesh:itemRatings*. The class *mesh:ItemRatings* is a vector of *mesh:ItemRating*. This class, which represents a rating score is modelled using four main properties: *mesh:ratingCriterium*, which represents the criterion/method used to rate the items (score, stars, etc.); *mesh:ratingValue*, which represents the value assigned by the user, *mesh:ratedItem*, which represents the item for which a preference has been established, and *mesh:timestamp*, which represents the moment in time in which the user rated that particular item.

As in the case of behaviour modelling, preferences are also dynamic, i.e., only certain user preferences should be considered in each particular context or *sioc-reality:C*. To dynamically select user preferences we build on our previous approach [22]. More specifically, the selection of applicable preferences in a particular context *sioc-reality:C* is based on two main principles:

- If a concept keeps occurring along time, this concept is selected within the current context as a long-term preference of the user.
- If a concept occurrence is very high on the recent short period, this concept can be selected in the current context as a short-term preference of the user.

In our extension (see Fig. 4) we define the classes *upo:LongTermPreferences* and *upo:ShortTermPreferences* to capture long and short term preferences in a particular context, *socialReality:C*; and the classes *upo:ConsistentFrequencySelection* and *upo:HighFrequencySelection* to model the methods used to capture long and short term preferences respectively. Note that by modelling and applying these methods the ontology fulfils **CQ2**, i.e., it is able to infer the needs of user *u* in the online community oc_x during a particular time period.

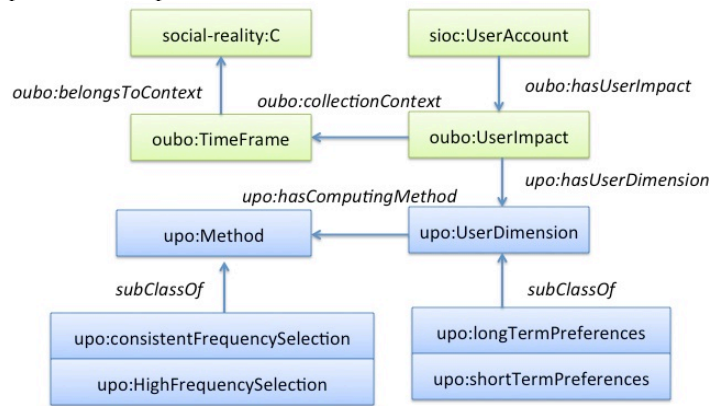


Fig. 4 Extension of the MESH ontology to capture dynamic preferences

4.5 Modelling User Needs

Our approach towards modelling and inferring user needs is based on the principle that needs are mirrored to certain online actions. For example, the action of commenting or replying to a post reflects the user's intention to help other users (*altruistic need*) as well as an aim to interact with other members of the community (*socialisation need*). In our model we aim to capture four main user needs that have been recurrently found in the literature [27, 28, 29, 30, 31]:

- *Information Need*: when a user initiates a discussion he or she is reflecting an information need. Users under this need are focused on solving their problems with the help of the community.
- *Socialisation Need*: when a user intentionally interacts with other users he reflects his need for socialisation.
- *Reputation need*: the reputation can be reflected on the number of points (ratings/likes/favourites) or replies the user receives from the community.
- *Altruistic Need*: the altruistic need, or need to help others, is reflected on the number of replies that a user provides to other people's questions. Users with a

high altruistic need share their knowledge with the community and spend their time and expertise to benefit others.

As in the case of user behaviour, several metrics can be used to represent needs (see Table 2). For example, the *Information Need* is linked with metrics such as: *Thread-initiation ratio* or *Self-reply ratio*. This last metric measures the number of replies given by user u_i in reply to his/her own threads. It is an indication of how strongly a user pursues obtaining an answer from the community. The *Socialisation Need* is reflected in metrics such as *out-degree*, which measures the proportion of users that the user has contacted/reply to. The *Reputation Need* is reflected in metrics such as the *average points/likes/favourites*, received by other users. The *Altruistic Need* is reflected in metrics such as *replies-count*, which measures the total number of replies written by a user within the community. Note that our proposed model can be extended to capture different user needs and different associated metrics. As shown in Fig. 5 the class *upo:UserNeed* and its corresponding subclasses are used to capture the defined user needs. Associated metrics are modelled under the class *upo:Metric*, such as *upo:SelfReplyRatio*. Note that we have decided not to reuse classes of the SWUM ontology, since this ontology captures needs in terms of features not measurable in the context of online communities, e.g., “sexual intimacy”.

To infer whether a user (*sio:UserAccount*) presents one particular need in a given context (*social-reality:C*), we apply semantic rules encoded in SPIN (e.g., if *upo:Thread-initiation=high* and *upo:SelfReplyRatio=high* then *upo:InformationNeed=high*). By following the same approach as for computing user behaviour [19] the ontology fulfils **CQ3**, i.e., it can infer the needs that user u adopts in an online community oc_x during a particular time period.

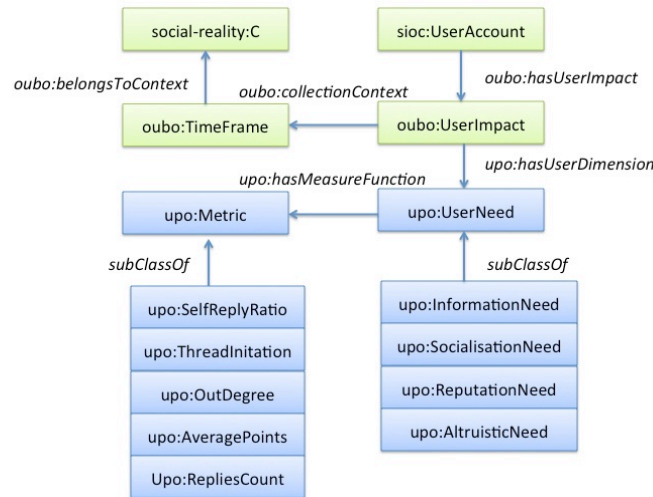


Fig. 5: Extensions to capture UserNeeds.

4.5 Modelling Personality

There is a body of research in online communities that has attempted to model and predict personality. These predictions are mainly based on the Big Five personality model [32, 33, 34], which defines personality in terms of five dimensions:

- *Openness to experience*: openness indicates the degree of intellectual curiosity.
- *Conscientiousness*: indicates a tendency to be organized and dependable.
- *Extroversion*: indicates sociability and the tendency to seek stimulation in the company of others.
- *Agreeableness*: indicates a tendency to be compassionate and cooperative.
- *Neuroticism*: indicates a tendency to experience unpleasant emotions easily.

To capture these personality dimensions, we reuse classes of the PAO ontology such as: *pao:Personality*, *pao:PersonalityDimension*, *pao:Agreeableness*, *pao:Conscientiousness*, *pao:Extroversion* and *pao:Neuroticism*. Recent studies have shown that the previous personality dimensions are reflected, and can be predicted, with certain degree of accuracy, from the online actions of users within online communities [32, 33, 34]. Quercia et al. [32], for example, predict users' personality in Twitter by using features such as "following", "followers" and "listed counts". These metrics are modelled in our profile as *upo:OutDegree*, *upo:InDegree*, *upo:FocusDispersion*. Note that research has consistently shown that people's personality scores are stable over time [15]. Therefore, personality in our model is not considered in context. To infer the levels of personality dimensions for each user u we define SPIN rules that capture the prediction model defined by Quercia et al. [32], e.g., (if *upo:OutDegree*=high and *upo:InDegree*=high then *pao:Extroversion*=high). By defining these rules the proposed model fulfils **CQ4**, i.e., it can infer the personality of a particular user u .

5 Discussion and Conclusions

This paper presents a semantic approach to user profile modelling that goes beyond collecting raw data from user activities in online communities. This approach captures the interpretation of these data within particular contexts, enabling the inference of user needs, behaviour and preferences - over time and for different online communities. The generated ontology has been made available online [43].

To generate the proposed user profile model we have reused and extended existing ontological resources. Following the NeOn methodology [18], we have assessed the generated user profile model by using four competency questions (see Section 2). These questions ensure that, by using the information captured within the proposed user profile model, we can infer the needs, preferences and behaviours of users within particular online communities and time frames. Personality, on the other hand, it is the only aspect of the user that is not considered in the context model, since research has repeatedly shown that personality scores are stable over time [15].

The problem of user modelling and representation has been tackled by different research areas apart from the Semantic Web, including Information Retrieval [5], Recommender Systems [1], Adaptive Hypermedia [4] and Pervasive Computing [3].

Researchers working in these areas has captured demographic features such as gender, age, nationality, etc.), and user context (e.g. social interactions, tasks, platforms, etc.). The representation of these data has evolved from traditional keyword-based representations (i.e. weighted feature vectors and weighted n-grams) to semantically enriched representations such as folksonomies, taxonomies and ontologies. Explicit (e.g. manual editing of user profiles or requesting documents that exemplify the user interests) and implicit (e.g. click-through data, opened documents, and browsing history) learning techniques have been used to capture this information.

Using an ontology and semantics to tackle the problem of user modelling offers a number of advantages: (i) the ontology provides a generic, reusable and machine understandable model for representing the concepts and properties required for describing user activities and measuring their evolution; (ii) due to the reuse of well-known vocabularies, our proposed user profile facilitates the integration of data from multiple social networking platforms; (iii) most importantly, the use of an ontology supports inferring mechanisms that can be used to calculate or derive user behaviour, needs, and preferences.

Future work within the DecarboNet project will advance existing methods to digest and distil information about a user's personal characteristics, opinions, and behaviour, encoded in user-generated content available from dynamic and heterogeneous evidence sources. Users will be able to inspect the user model and gain interactive means explore contextualised information spaces through tailored content services. This integrated and dynamic approach based on data across multiple systems and communities will help to better understand the emergence of collective awareness.

Acknowledgement. The research presented in this paper has been conducted as part of the DecarboNet project (www.decarbonet.eu), Grant Agreement No. 610829.

References

1. Adomavicius, G. and Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems. *IEEE Transactions on Knowledge and Data Engineering*, 17: 734-749
2. Breslin, J.G. and Decker, S. (2007). The Future of Social Networks on the Internet: The Need for Semantics. *IEEE Internet Computing*, 11(6): 86-90
3. Baldauf, M., Dustdar, S. and Rosemberg, F. (2007). A Survey on Context-Aware Systems. *International Journal of Ad Hoc and Ubiquitous Computing*, 2(4): 263-277.
4. Brusilovsky, P. and Millan, E. (2007). User Models for Adaptive Hypermedia and Adaptive Educational Systems. *The Adaptive Web - Methods and Strategies of Web Personalization*. Eds. P. Brusilovsky et al. Berlin: Springer. 3-53.
5. Tamine-Lechani, L., Boughanem, M. and Daoud, M. (2010). Evaluation of Contextual Information Retrieval Effectiveness: Overview of Issues and Research. *Knowledge and Information Systems*, 24(1): 1-34.
6. FOAF | www.foaf-project.org
7. Meaning of a Tag (MOAT) | moat-project.org
8. Relationship Vocabulary | www.vocab.org/relationship
9. SIOC Project | sioc-project.org
10. Schema.org | www.schema.org
11. Microformats | www.microformats.org

12. hcard Microformat | www.microformats.org/wiki/hcard
13. XFN microformat | www.gmpg.org/xfn
14. Erétéo, G., et al. Semantic social network analysis: A concrete case. Handbook of Research on Methods and Techniques for Studying Virtual Communities: Paradigms and Phenomena, pages 122–156, 2011.
15. Online Presence Ontology (OPO) | www.online-presence.net/ontology.php
16. DCTerms | www.dublincore.org/documents/dcmi-terms
17. Relationship Vocabulary | www.vocab.org/relationship
18. Suárez-Figueroa, Mari Carmen, et al. Ontology engineering in a networked world. Springer, 2012.
19. Rowe, M. et al. Community analysis through semantic rules and role composition derivation. Journal of Web Semantics: Science, Services and Agents on the World Wide Web 18.1 (2013): 31-47.
20. Hoekstra, Representing social reality in OWL 2, in: 7th International Workshop on OWL: Experiences and Directions (OWLED 2010), 2010
21. MESH user ontology | www.mesh-ip.eu/upload/MESH_user_ontology.zip
22. Cantador, Iván, et al. A multi-purpose ontology-based approach for personalised content filtering and retrieval. Advances in Semantic Media Adaptation and Personalization. Springer Berlin Heidelberg, 2008. 25-51.
23. TextRazor | www.textrazor.com
24. AchemyAPI | www.alchemyapi.com
25. DBpediaSpotlight | spotlight.dbpedia.org
26. Saif, H., Fernandez, M., He, Y. and Alani, H. (2014) SentiCircles for Contextual and Conceptual Semantic Sentiment Analysis of Twitter, 11th ESWC, Crete, Greece
27. Arena, 2012 | www.arena-media.co.uk/blog/2012/05/maslows-hierarchy-of-needs
28. Scholz, T. 2007 | es.slideshare.net/trebor/motivating-people-to-participate, slide 24
29. Antonios, 2010 | www.johnantonios.com/2010/02/06/the-social-media-hierarchy-of-needs
30. Schaefer, C. Motivations and usage patterns on social network sites. (2008).
31. Smock, A.D., et al. Facebook as a toolkit: A uses and gratification approach to unbundling feature use. Computers in Human Behaviour 27, 6 (2011), 2322–2329
32. Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. Our Twitter profiles, our selves: Predicting personality with Twitter. (Socialcom2011).
33. Golbeck, J., et al. Predicting personality from twitter. (SocialCom 2011).
34. Sumner, C. et al. Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In Machine Learning and Applications (ICMLA2012).
35. Chan, J., et al. Decomposing discussion forums and boards using user roles. In ICWSM 2010.
36. Abel, F., et al. Analyzing user modeling on twitter for personalized news recommendations. In User Modeling, Adaption and Personalization (pp. 1-12). Springer Berlin Heidelberg.
37. Szomszor, et al. Semantic modelling of user interests based on cross-folksonomy analysis (pp. 632-648). Springer Berlin Heidelberg.
38. Ankolekar, A. et al. Supporting online problem-solving communities with the semantic web. Proceedings of the 15th international WWW conference. New York, USA.
39. SWUM ontology | www.swum-ontology.org
40. Donohue, B., & Otte, J. N. Fall 2013 Ontological Engineering Prof. Barry Smith The Five-Factor Model and Personality Assessment Ontology
41. Scharl, A., et al. From Web Intelligence to Knowledge Co-Creation-A Platform to Analyze and Support Stakeholder Communication. IEEE Internet Computing 17(5): 21-29
42. Nyea, C.D. et al. Testing the measurement equivalence of personality adjective items across cultures. Journal of Research in Personality, July 2008
43. Upo ontology: <http://people.kmi.open.ac.uk/miriam/ontology/upo.zip>