Web Site Recommendation Modelling Assisted by Ontologies Networks

Edelweis Rohrer [1]
Regina Motz [1]
Alicia Díaz [2]

Abstract: Web site recommendation systems help to get high quality information. The modeling of recommendation system involves the combination of many features: metrics of quality, quality criteria, recommendation criteria, user profile, specific domain, among others. At the moment of the specification of a recommendation system it must be guaranteed a right interrelation of all of this features. In this paper we propose a ontology network based process for web site recommendation modeling. This ontology network conceptualizes the different domains (web site domain, quality assurance domain, user context domain, recommendation criteria domain, specific domain) in a set of interrelated ontologies. Basically, this work introduces the semantic relationships that were used to construct this ontology network. Moreover, it shows the usefulness of this ontology network for the detection of possible inconsistencies when specifying recommendation criteria. Particularly, this approach is illustrated for the health domain.

1 Introduction

Web site recommendation systems help to get high quality information. The modeling of recommendation system involves the combination of many features: metrics of quality, quality criteria, recommendation criteria, user profile, specific domain, among others. At the moment of the specification of a recommendation system it must be guaranteed a right

[1]Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Uruguay
{rmotz, erohrer}@fing.edu.uy
[2]LIFIA, Facultad de Informática, Universidad Nacional de La Plata, Argentina
alicia.diaz@lifia.info.unlp.edu.ar

interrelation of all of this features. In this paper we propose a ontology network based process for web site recommendation modeling.

The use of the web by common people as a repository where to find information, especially in the health area, increases drastically day by day. This is a very worrying reality because many of health websites do not contain data of good quality: precise, believable, relevant to users profile. There are several characteristics of website which make attention to quality issues necessity. Particularly, the lack of quality controls (editorial boards) at the stage of production, the dubious and alternative medicine products are now primarily offered on the Internet; and a " context deficit", leading to the situation that information does not necessarily have to be false to harm [1]. Furthermore, the fact that the web is a very dynamic medium, once a person has obtained misinformation, then, there is little change to be reversed by health professionals. In this sense, decentralized, intelligent recommender systems can automatically give an evaluation about the quality of the sources according to the consumers needs. Then, Quality-based recommendation systems are a help to get high quality health-related web sites for user needs.

Quality in websites is determined by several diverse factors, some of which are general, and therefore, considered for any type of sites and for any domain. Such features include, for example, navigation, user interface aspects, legibility (size of letter, colors, images), performance aspects (time it takes to access to the site content), the correct functioning of the site, its conformity with standards of language use or of accessibility like those described in normative such as the *Web Content Accessibility Guidelines* of the W3C [3]. There are quality models that take these features into consideration, some of them are for example WebQual [2] and WebQEM [3]. Particularly, in this work, we focus on the quality that arises of the information value that the site provides and its adequacy for the consumers context.

In this paper we specify a process driven by an ontology network for web site recommendation modeling, that leads to give a recommendation of suitability of web contents to a particular user who makes a specific query. This process has several steps, from the definition of user profiles, query contexts, quality and recommendation criteria to the quality assessment of the web contents and finally the recommendation itself. This process is supported by an ontology network, which during the execution of this process, the networked ontologies will help to discovering knowledge domain units in the web pages (i.e. based on health ontology and the specific health ontology), while in other cases, it helps to supporting quality or recommendation assessments. In the last cases, the ontology network can be used to both: assist in the modeling and specification and start-up of a recommendation system and check the correctness of the resulting system specification.

Regarding this last issue, the process proposed must assure that each step has all the required data, structured consistently. In this way, each step can be correctly executed so

[3]http://www.w3.org/WAI/GL/

the results obtained are as much accurate as possible. The ontology network that underlies the process helps to reach these goals, since it allows the ontologist to define a model with restrictions and also add rules to classify invalid quality and recommendation specifications and assessments. In addition to showing each step of the recommendation process, this paper explains how the ontology network helps in this regard.

Particularly, in this paper we will particularized to the field of health, however it can be extended to any other domains. We will define the `Salus` ontology network and the `Salus` recommendation process. `Salus` is the research project where this research fits.

This paper is organized as follow: Section 2 describes issues about web site quality assurance, model checking and related works. Then, in section 3, the `Salus` ontology network is introduced. After that, we discuss the process for websites recommendation based on the `Salus` ontology network, paying special attention to the configuration step. Following, in section 5 we give some examples to clarify the utility of the ontology network in validating the configuration of the process. Finally, we discuss conclusions and future perspectives.

2 Quality Assurance and Modeling Checking Background

There are, basically, two ways of defining data quality. The first one uses a scientific approach and defines data quality dimensions rigorously, classifying them as dimensions that are or are not intrinsic to an information system [4]. The second one is a pragmatic approach aimed at defining data quality in an operational fashion [5]. Wang et al. [4] identified four data quality dimensions: (1) intrinsic data quality; (2) contextual data quality, which defines the quality of the information within the context of the task; (3) data quality for data representation, which determines if the system presents the information in a concise, consistent, understandable way; (4) data quality regarding data access, which defines quality in terms of the role of the information system in the provision of the data. Within each dimension it is possible to identify several factors, including: for intrinsic data quality dimension: believability, accuracy, objectivity; for context dimension: value-added, relevancy, timeliness, completeness; for the representational dimension: interpretability, easy of understanding, concise representational; for the accessibility dimension: access security among others. To determine which of these dimension are relevant for a specific domain is a task that corresponds to be defined by the domain expert.

Regarding *Believability*, in [6] are introduced two definitions: *Believability which is the extent to which data is regarded as true and credible* and *Reputation, which is the extent to which data is highly regarded in terms of its source or content.* The former is a general definition that expresses the meaning of data believability, while the latter talks about data properties (source, content) to be considered to evaluate whether a document is believable. About this factor in health domain, it is important to take into account the existence of sites with certified quality labels, such as HON [4], WIS [5] and WMA [6], which means that documents linked by these sites will be evaluated with a higher level of quality than those that have no certification.

For the *Timeliness* dimension in [6], the following definition can be found: *Timeliness is the extent to which data is sufficiently up-to-date for the task at hand.* Regarding this factor, what perhaps may matters is measuring the freshness of published data, rather than the publication date, but this considerations depend on the context of the use of the information.

For the *Readability* dimension, in [7] were introduced different readability metrics created for different domains and user profiles. It sets the following definition: *Readability is what makes some texts easier to read than others*. The same work mentions the definition of G. Harry McLaughlin [8], creator of the SMOG readability formula: *The degree to which a given class of people find certain reading matter compelling and comprehensible.* There are a lot of readability formulas created for different authors, like FOG and SMOG grade levels, that reached good results when they were tested [7]. Here also, the decision on which formula to use must be taken for a domain expert.

Therefore, in order to develop a model for a web site recommendation system, the first step is to specify a formal model that represents the dimensions involved in the acquisition of the quality of web data as well as the different metrics that can be applied. Our approach to do this challenge is the design of an ontological model inspired in our previous work [9] on web data warehouse quality. Our proposal is to model a generic ontology for quality dimensions independent from the specific domain and from the different types of web data sources. Our generic ontology in spite of its high abstraction level is easily tailored to different user domains and different types of web data through its connection in the proposed ontology network.

In this paper, we mainly focus on the checking of the correctness of the resulting system specification. The quality assessment of web contents and the recommendation of them to users in a specific context, is strongly based on the correct specification of the quality dimensions to be assessed and the way the recommendation is carried out. Then, on one hand, it is very important how the administrator user determine the quality dimensions that are going to be considered to evaluate web contents, and what metrics are used to assess each dimension. On the other hand, the combination of user attributes, context features and quality levels associated to web contents is a key aspect to reach a recommendation that suggests the most suitable contents to the user. Semantic web technologies, such as ontologies, allows the ontologist to assist the administrator user, constraining the model and validating the data the user introduces, preventing possible mistakes he can make. OWL language and SWRL rules

[4] http://www.hon.ch/
[5] http://www.portalesmedicos.com/web_interes_sanitario/index.htm
[6] http://wma.comb.es/

are fundamental tools to reach that goal.

Regards distributed recommender systems, [10] treats the problem of social filtering techniques and collaborative filtering. To address this aspect, it proposes trust networks, along with trust propagation mechanisms, and taxonomy-driven profile generation and filtering.

With respect to the issue of correctness checking of recommender systems specification, a work about the consistency checking of systems with different paradigms, like classical relational databases and semantic web logic databases is addressed in [11]. It presents a list of possible sources of inconsistencies and another of solutions to them. It takes into account the use of ontologies and gives particular importance to the definition of the context, so a coherent subset of the ontologies can be considered. Also measures to estimate inconsistency are mentioned.

A more recent work is [12], that addresses the problem of consistency checking of models in large applications developed by a big number of developers. Given the large number of checks to be performed whenever a new modification to the model is introduced, this work presents an incremental inconsistency checker that aims to avoid performing a complete check of the model each time it evolves. This work results particularly interesting because, although it does not consider ontologies to face the consistency checking, it is based on a mechanism to validate models, applying the same philosophy that is addressed in our paper.

Our approach can be considered as a mix between the two previous approaches, but applied to the domain of website recommendation systems. In the next two sections we present the ontological model and how it is expressed in terms of an ontology network in order to specify a recommending process. Then, in section 5 we will show how the consistency of the resulting model can be checked.

3 `Salus` Ontology Network

The `Salus` ontology helps to obtain a reading recommendation of health-related web contents for a particular user. Specifically, it conceptualizes the different knowledge domains that are involved in a recommendation system in a shape of a ontology network.

An ontology networks is a collection of ontologies related together through a variety of different relationships such as mapping, modularization, version, among others [13]. Accordingly, a *networked ontology* is an ontology included in such a network, sharing relationships with other ontologies. Intuitively, this implies to define the ontologies' content, but also to define metadata information about the networked ontologies. Ontology metadata refers to the information which is attached to the ontology itself, not to its content and are critical in ontology networks. This ontology metadata would covers ontologies provenance, purpose and the relations with other ontologies and semantic resources.

There are some models that cover both the syntactic aspects of dealing with ontology relationships in networked ontologies and the semantic aspects of interpreting ontology networks and the relations between networked ontologies. For instance, the Ontology Metadata Vocabulary (OMV)[14] defines classes and relations to talk about authoring aspects, ontology type, purpose, etc; on the other hand the Collaborative Ontology Design Ontology (C-ODO) [15] is an ontology network that enables designers to talk about design entities (ontologies, modules, ontology elements, requirements, activities, tools, reusable knowledge, teams, people, etc.).

`Salus` ontology network conceptualizes the different domain related to a recommendation system. These domain are: specific health domain, the web site domain, the quality assurance domain, the user context domain and the recommendation domain. Each `Salus` networked ontology conceptualizes each of this domains:

Health ontology conceptualize a health domain. This ontology may be an already existing ontology like UMLS[7] which models for example the impact, treatment, risk factors, diagnostic, effects, phases of a disease. This ontology can be refined in terms of a specific disease i.e Alzheimer, and thus can exist the concept ”Alzheimer treatment”.

Web Site ontology conceptualizes the domain of web pages and, particularly, describes the web resources that will be consider to participate in a quality assessment. The main concepts of this ontology is the `web resource` and `web resource property`. A `web resource` is any resource which is identified by a URL; for instance, it can be instantiated as a *web page* which has attached a web content. `Web resource properties` models the properties that can be attached to a web resource. For instance, possible properties of a web content could be the author, the amount of words, etc. Among these properties there is a particular one, the `hasTopic` property that relates concepts (web resources) from the `Web Site` ontology with concept in the `Specific Health` ontology. The `hasTopic` property describes what a web resources talking about. This property always is presented for a *web page* resource. All of these properties should be retrieved through a specific information retrieval processes. Therefore, a recommendation system should have attached a repertoire of information retrieval processes.

Quality Assurance ontology conceptualize metrics, quality assurance specifications and quality assessments. *Metrics* are calculus defined base on web resource properties. A *quality assurance specification* describes the different quality dimensions; for instance readability, precision, believability, completeness, timeliness, etc. The quality assurance specification associate to each quality dimension the suitable metric calculus. A

[7] http://www.nlm.nih.gov/research/umls/

quality assessment models the assessment of a particular web resource (i.e. a web document) for a particular quality dimension through a specific metric. It also models the obtained quality level.

Context ontology describes the user profile and the query resource. The user profile describes *user properties* which could user age range, role, academic level, health domain expertise among others. The *query resource* represents the context of the query. The main concept of the query resource is the *query goal*.

Recommendation ontology describes the different criteria of recommendation for a particular context (user and query situation)and quality dimension and the obtained recommendation level.

Particularly, `Salus` is specific to the health field, but it could be adapted to another domains. It is achieve just by changing the health ontology by another specific domain ontology.

`Salus` networked ontologies are interrelated (see in the upper of figure 1) by three different relationships: `uses`, `extend` and `describes` relationships. The semantic of this relationship are the following:

- The `uses` relationship relates two ontologies by the import primitive. For example, this relationship occurs between the `Web Site` ontology and the `specific domain` ontology because of a web content topic can be any concept at the specific domain ontology. In The `Salus` ontology network, the specific domain ontology is the `Health` ontology and web content topics could be *treatment*, *diagnostic*, etc. In `Salus` , *Alzheimer Treatment* can be a topic of *Alzheimer Webpage*.
- The `extends` relationship describes a more specific ontology which is the specialization of a more general one. The more clear example is the Alzheimer ontology is an specialization of the `Health` ontology. For example at the `Health` ontology can be defined the concept: diagnostic, treatment, risk factors, etc, then these concepts can be specialized at the Alzheimer domain in the Alzheimer ontology.
- The `describes` relationship defines the relations between a model and its metamodel. For instance, the `Web Site` is an instantiation of the `Web Site Specification` ontology. The later is a meta-ontology for the former. `Webpages` are typical concepts at the `Web Site` ontology and model the class webpage concept. This class is an instance of `Web Resources` which are defined at the `Web Site Specification` ontology. Another example is the property `hasAuthor` that is defined at the `Web Site` as an instance of the `Web Resource Property` concept that was defined at the `Web Site Specification` ontology.

On the bottom part of the figure 1 is shown an example of the resulting knowledge base when a document *Alzheimer webpage* was assessed to be recommended to the user *Paul*. The content associated to the *Alzheimer webpage* has ”Alzheimer Treatment” and ”Alzheimer Diagnostic” as topics. In this example the recommendation assessment took into account the *Believability* quality dimension, which was assessed by *Provenance*, which uses the *hasAuthor* property of the webpage. The recommendation assessment also considers the fact the user Paul is a teenager and the goal of his query is ”looksFor”. Later, in the section 4 more detail about the networked ontologies will be detailed.

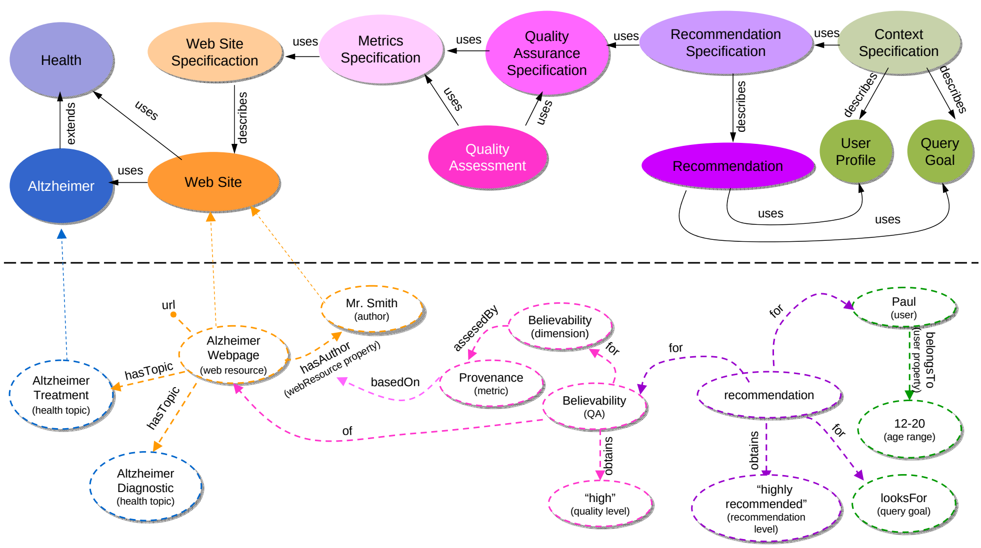

Figure 1. `Salus` ontology network

4 `Salus` Recommendation Process

The `Salus` recommendation process covers the different tasks which have to be perform in order to recommend a set of web sites to a particular user. These task are organized in three different phases. These phases correspond to the start up of the recommendation system, the quality assessment of a set web pages and the execution of recommendation assessments. The *recommendation system star-up phase* is in charge of preparing the information about web sites properties, user properties and quality and recommendation criteria, which are needed when is being performed a recommendation assessment. In the *quality assessment of web pages* phase, a set of web pages on health domain are assessed and it is determined their quality level. The *recommendation assessment phase* is in charge of generating a set of recommended web site to particular user query. In the next three subsections, they will be detailed.

The `Salus` recommendation process is characterized as an ontology-based process. Specifically, it is based-on the `Salus` ontology network described in previous section. During the execution of the `Salus` process, the `Salus` networked ontologies plays different roles: in some cases it helps to discovering knowledge domain units in the web pages (i.e. based on health ontology and the specific health ontology), while in other cases, it helps to supporting quality or recommendation assessments. In the last cases, the `Salus` ontology network can be used to both: assist in the modeling and specification of a recommendation system and check the correctness of the resulting system specification. Particularly, this section will go in deep explaining the modeling of a recommendation system based on the `Salus` ontology network during the recommendation start-up phase explanation. The ideas about the correctness checking of a recommendation system specification will be tried in details in section 5.

4.1 Recommendation Start-Up Phase

The *recommendation system star-up phase* is in charge of preparing the information about properties of web sites, user properties, quality criteria and recommendation criteria, which is needed in order to recommend web pages. This phase consists of the tasks: web site definition, quality criteria definition and recommendation criteria definition. This tasks are schematized in the figure 2 and are detailed below. In the following discussions, we will also show where, when and how the `Salus` ontology network is used.

Web Resource definition It refers to the population of the `Web site` ontology according to a a given set of webpages and their indexation based on the specific domain ontology. The `Web site` ontology is populated with `webpages` concepts (one for each given webpages) and with properties that are involved in the newly defined concepts; for example the `url` property is specified between a web page and a URL. Then, these webpages are indexed according to the specific domain ontology; in the case of the `Salus` ontology network, it corresponds to the Alzheimer ontology. In this task the `hasTopic` property is specified between webpages and Alzheimer concepts. Regarding the example of the figure 1, the `hasTopic` property is specified between the *Alzheimer webpage* and the *Alzheimer treatment* concept. There are also other properties that should be defined for a web resources, but they will specified after the Quality Criteria definition. These properties are those needed to perform a specific metric, for instance, `hasAuthor` property.

Quality criteria definition It refers to the definition of metrics and quality dimensions that will be supported by the recommender system. Based on a repertoire of metrics, the definition consists on specificating which `metric` assets each `quality dimension` and which are the possible obtained `quality level`. Fist of all, it have to be specified the repertoire of metrics. A `metric` is the main concept of the `Metric Specification` ontology and defines the metric calculus. `Metrics` are specialized in `elementary metrics`and `composite metrics`. `Elementary metrics` are specified based on `web resource properties` (concept of the `Web site` ontology). For example, when the elementary metric *provenance* is instantiated, the `basedOn` property might be also instantiated in order to link the *provenance* metric with the *hasAuthor* property. The *hasAuthor* property have be now specified as an instance of `web resource property`. On the order hand, `composite metrics` are metrics that have defined the `aggregates` property. The `aggregates` property links `composite metrics` to `metrics` (i.e elementary and/or composite metrics). Composite metrics are useful to model more complex metrics. Then, the `Quality Specification` ontology has to be populated. Quality dimension concepts have to be instantiated. These quality dimensions are those supported by the recommender system. Some examples of quality dimensions are readability, believability, completeness, accuracy. Each quality dimension concepts has defined at least once the `assesedBy` property. The `assesedBy` property links a quality dimension to the metrics that enable its assessment. Quality dimension concepts also has defined the `assesTo` property. This property links a quality dimension to its possible `quality levels`. For instance the quality dimension *Believability* has defined the `assesedBy` property which takes values in the *provenance* concept and the `assesTo` property to the set of strings: "*high*", "*medium*", "*low*".

Recommendation criteria definition It refers to the definition of recommendation criteria. Based on the quality criteria definition, a *recommendation definition* implies to indicate which quality dimensions will be assessed and which context resources will be consider for a recommendation. Context resources are mainly user properties and query resources. The output of this task is a set of *recommendation rules* which specify the *recommendation level* for each assessed web page. These rules are like:

```
if recommendation definition(thisWebPage) then
      recommendationLevel(thisWebPage)
```

where *thisWebPage* is the currently processed webpage and the `recommendation definition(thisWebPage)` is described in terms of quality assurances and context. The recommendation level for a webpage is one of the scale values of the scales of recommendation levels of the recommender system (for example, highly recommended, stronger recommended). Regarding, the example we have been followed along the paper, the below rule might be defined as follow:

if *BelievabilityQA(AlzheimerWebpage)* `assesTo` *"high"* and *Paul* `belongsTo` *12-20 age range* and *query goal* is *looksFor* then *AlzheimerWebpage* is `highly recommended`

In this rule the recommendation definition is base on the *BelievabilityQA)* and *Paul* profile and the *query goal*.

Context resource definition It is in charge of defining those context resources that have to be taken into account to make a recommendations. These context resources will be identified in the recommendation criteria definition. Mainly, they are: the user properties and the query resources. The *user properties* are those that were already relieved at the recommendation criteria definition task and will be populated at the moment of registering a user at the recommender system, for example, at the recommendation criteria definition was specified the user property *belongs*. At the moment the user is registered at the system, this property is defined between *Paul* and *12-20 range*. The query resources refers to query attributes like query goal.

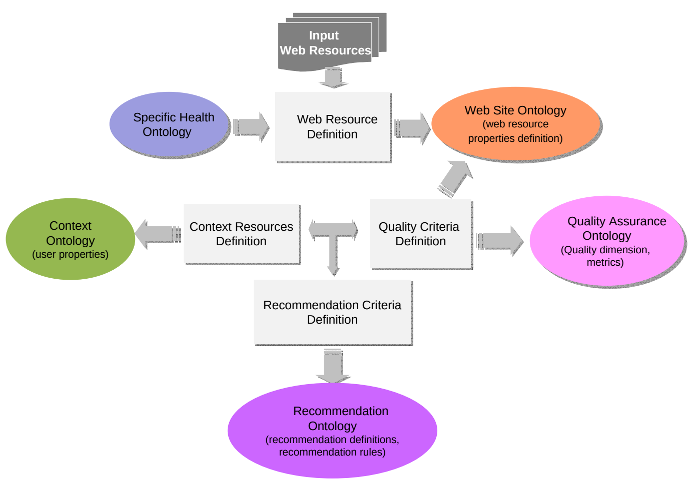

Figure 2. `Salus` Recommender start-up

4.2 Web Page Quality Assessment Phase

After the recommendation start-up phase, the quality assessment of the input web resources can be done. First of all, the web resources will be pre-processed to determine their properties and populate the web site ontology. The metrics, which are involved in a quality definition, determine the `web resource properties` to be considered. For example, in the quality definition of the dimension *Believability* is used the metric *provenance* which refers to the author of the webpage. This means that the webpage content should has associated the *hasAuthor* property. Therefore, it have to be determined which information retrieval processes have to be perform in order to discover these new `web resource properties`. The retrieved information will be used to complete the population of the `Web site` ontology. For instance, by using a specific information retrieval process that retrieve the author of a web page, the *hasAuthor* property can be defined between the *Alzheimer webpage* and *Mr. Smith* concepts. In this phase, a set of specific domain web resources (webpages) will be assessed in order to determine their quality level. The quality assessment execution involves to calculate the quality level of each web resource for each quality dimension. For that, the corresponding metric is executed and thus, it is determined the quality level of a web resources. In this phase, the quality assurance ontology is populated, mainly, by adding instances of the `quality assessment` and linking them with the `web resource` and `quality level`. Regarding the followed example, in this moment, the concept *Believability QA* is instantiated as an individual of the `Quality Assessment` class and the *obtains* property is defined between *Believability QA* and the *"high"* quality level.

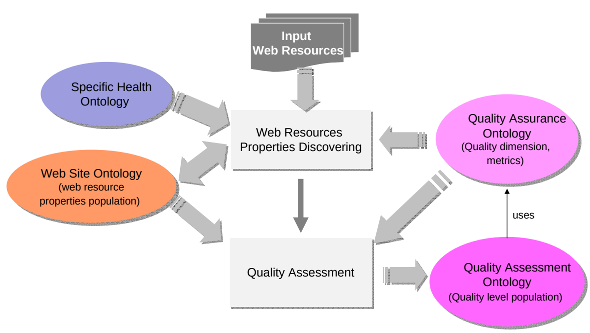

Figure 3. `Salus` Quality Assessment

4.3 Recommendation Assessment Phase

A user query is the trigger of this *recommendation assessment phase*. When a logged on user makes a query, the recommendation system evaluates the *recommendation rules* in order to determine the recommendation level. All of those web resources, which assets to an appropriated level for the considered user, will be recommended.

The evaluation of the recommendation rules is based on the user profile, the quality level of the considered web resources and the query resources. Both, the user profile and the web resource quality level, have been calculated in the previous two phases. Query resources have to be discovered at this moment. A type query resource could be a query goal.

The output of this phase is a set of recommended web resources to a particular user query. The figure 4 summarizes the recommendation assessment phase.

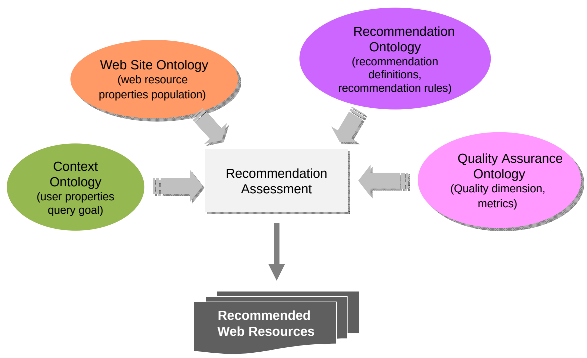

Figure 4. `Salus` Recommendation Assessment

5 Using the Ontology Networks for Validating the Recommendation System Modeling

During the configuration and execution of the recommendation process, non-expected situations can arrive. For example, someone might try to assess a quality dimension with a unappropriated metric. Having an ontology-based recommendation system is not only advantageous at the moment of configuring this kind of systems (for intance to discover knowledge domain units), but also it is helpful to validate the resulting configuration of the systems. In the last case, the `Salus` ontology network can be also used to check the validity of the resulting system and thus, to guarantee its correctness. Combining the `Salus` ontology network with a validation process will be the strategy that we will introduce in this paper.

The potential of this validation process depends on the potential of the tools which are used to express the ontology network. Ontology languages, like OWL[8] and they combination with rules likes SWRL[9], are potential tools to express the `Salus` ontology network.

[8]http://www.w3.org/TR/owl-features/
[9]W3C. SWRL Semantic Web Rule Language. URL: http://www.w3.org/Submission/SWRL/

OWL provides a powerful axiom language based on description logics [16], which allows one to constrain the model with a precise definitions of the concepts and the existing relationships among them. OWL allows implementing each networked ontology and gives facility to describe the relationships between them; for instance, the *uses* relationship can be implemented with the `owl:import`.

However, OWL does not provide facilities to draw inferences about individuals. To mitigate this drawback, OWL is combined with SWRL rules [10], due to they allow users to write Hornlike rules expressed in terms of OWL concepts to reason about OWL individuals, inferring new knowledge. Besides, the OWL query language SQWRL[11], based on SWRL, provides the chance of doing queries on OWL ontologies and also a very useful array of set operators to perform closure operations like counting and aggregation, among others. SWRL rules are used to describe the recommendation rules.

An OWL-SWRL based `Salus` ontology network, can take advantages of these two paradigms.. Both, OWL and SWRL, have an added benefit; they can help at the moment of configuring and executing the recommendation system. Following, we will illustrate the utility of OWL, SWRL and SQWRL in model checking of inconsistencies through some examples. First in section 5.1, OWL restrictions will be used to complete the specification of a class. Then, in section 5.1, SWRL and SQWRL will be used to detect specification anomalies.

5.1 Validations by Using OWL restrictions

As explained in Section 4.2, the Web Page Quality Assessment phase involves the assessment of web contents in order to determine their quality level. The *Quality Assessment Ontology* is the model that underlies that phase. It has the main concept `QualityAssessment`, that represents the assessment of a webpage for a `quality dimension` through a `metric`, categorizing the web content with a quality level. Also, the *Quality Specification Ontology* is the model that represents the relationship between the different assessed dimensions and the metrics through which the assessment is carried out. The detailed structure of *Quality Specification* and *Quality Assessment* ontologies is showed in Figure 5.

For extending the example presented in Section 4.1, we define a metric called *BasicProperties*, to measure whether a webpage has certain basic properties like *author* and *source* (site, journal). It is a composite metric that uses two elementary metrics: the *AuthorMetric*, that measures if the webpage has an author, and the *SourceMetric*, that measures if the webpage has a source. The *BasicProperties* metric can be used to assess the *Believability* dimension, i. e. in the Quality Specification Ontology, the *Believability* individual, which

[10]W3C. SWRL Semantic Web Rule Language. URL: http://www.w3.org/Submission/SWRL/

[11]http://protege.cim3.net/cgi-bin/wiki.pl?SQWRL

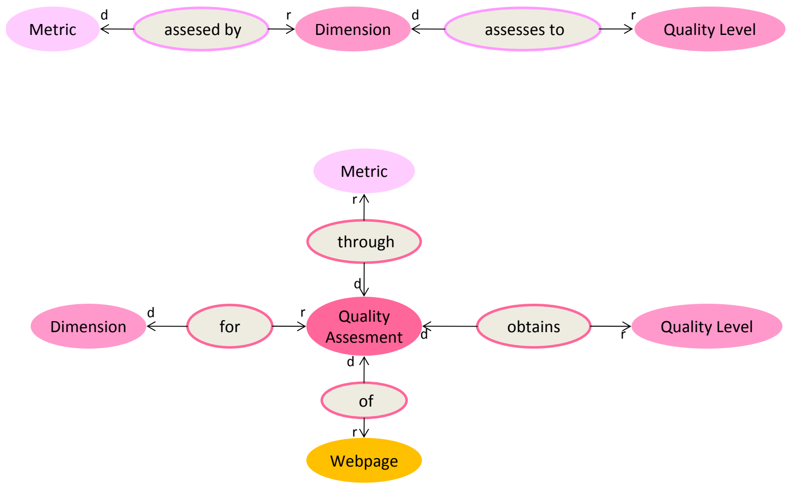

Figure 5. `Salus` Quality Specification and Quality Assessment Ontologies

is an instance of the `Dimension` class, will be related to the *BasicProperties* individual, which is an instance of the `Metric` class, through the relation `assessedBy`. If we define three quality levels as the result of the assessment of web contents, the *Believability* concept will be related to three instances of the `QualityLevel` class (”low”, ”medium”, ”high”) through the relationship `assesesTo`. As can be appreciate through the example, in order to the assessment makes sense, there must be at least two quality levels to categorize webpages. It is here, where OWL will helps us because it will allows us to constrain the model in order to satisfy this requirement. Therefore, we can attach to `Dimension` class the conjunction of the following two restrictions in order to represent this constraint (we use a description logic-style notation for simplicity):

$$\exists assessesTo.QualityLevel \tag{1}$$

$$>= 2assessTo.QualityLevel \tag{2}$$

The axioms 1 and 2 expresses an existential restriction and a cardinality restriction respectively. The first one, says: for each individual of the `Dimension` class that has defined the relationship `assessTo` must exist at least an individual of the *QualityLevel* class linked to it. The last one says: each individual of the `Dimension` class that has defined the relationship `assessTo` will be related with a minimum of two individuals of the QualityLevel class.

5.2 Validations by Using SWRL and SQWRL

In the scenario of the Web Page Quality Assessment Phase, each web content must be classified into a quality level for a dimension through a metric. In the previous example the dimension is assessed by the *BasicPropertiesMetric* metric. Then, the *BelievabilityQA* quality assessment of *webpage* for the *Believability* dimension must be carried out through the *BasicPropertiesMetric* metric. But, in the hypothetical situation where the dimension assessment was executed through another metric, for example *AuthorMetric*, would be a mistake; it is not a right metric to assess *Believability* dimension.

Therefore, before populating the `Salus` knowledge base, it is necessary to guarantee that the input of new information does not leave the knowledge base in a inconsistent state. In the example, there should be a mechanism to detect that *AuthorMetric* is not a right metric to assess *Believability* dimension. For instance, to identify invalid quality assessments, the `InvalidQualityAssessment`cass could be added as a subclass of the `QualityAssessment` class. This new class will contain inconsistent quality assessments. Then, before executing the assessment itself, a validation process can be run. This validation process will classify a quality assessment into the `InvalidQualityAssessment` class when assessment has some inconsistency. The implementation of that validation process can be done using SWRL rules with SQWRL operations and queries. The following rule implements the validation of the example:

$$QualityAssessment(?assess) \wedge Metric(?metric) \wedge through(?assess, ?metric) \tag{3}$$

$$\wedge\, Dimension(?dimension) \wedge for(?assess, ?dimension) \tag{4}$$

$$\wedge\, Metric(?metricD) \wedge assesedBy(?dimension, ?metricD) \tag{5}$$

$$\wedge\, sqwrl : makeSet(?s1, ?metric) \wedge sqwrl : groupBy(?s1, ?assess) \tag{6}$$

$$\wedge\, sqwrl : makeSet(?s2, ?metricD) \wedge sqwrl : groupBy(?s2, ?assess, ?dimension) \tag{7}$$

$$\wedge\, sqwrl : notIntersects(?s1, ?s2) \tag{8}$$

$$\Rightarrow InvalidQualityAssessment(?assess) \tag{9}$$

In this rule, for each *?assess* individual of the *QualityAssessment* class, it is retrieved:

- those *?metric* individuals related to *?assess* by the object property *through* and those *?dimension* individuals related to *?assess* by the object property *for*, by (3) and (4). They are the metric and the dimension that were used in the quality assessment.

- those *?metricD* individuals related to the *?dimension* individuals by the object property *assesedBy*, by (5) These are all metrics that can be selected to assess current dimension.
- A set *?s1* is constructed containing the *?metric* individuals, grouped by quality assessments by (6). Each set has the metric associated to the quality assessment.
- A set *?s2* is constructed containing all *?metricD* individuals, grouped by quality assessment and dimension by (7). Each set has the metrics associated to the quality assessment and the dimension.
- The built-in *notIntersects* guarantee that there is no common individuals between *?s1* and *?s2* (8).
- In case of the intersection set is empty, the instance *?assess* is classified into the InvalidQualityAssessment class, by (9).

In this way, this rule will infer that an the *BelievabilityQA* assessment for *Believability* dimension with the *AuthorMetric* metric is invalid, and then, it will be included as an individual of the `InvalidQualityAssessment` class.

In this way, this rule will infer that an the *BelievabilityQA* assessment for *Believability* dimension with the *AuthorMetric* metric is invalid, and it will be included as an individual of the `InvalidQualityAssessment` class.

6 Conclusions and future work

In this paper we have introduced a novel approach which used a ontology network (`Salus`) to assist the modeling and execution of a website recommendation system, which uses a quality-based approach to get the more adequate websites for a specific consumer context.

We have described `Salus` ontology network that models the different domain related to a recommendation system. These domain are: specific health domain, the web site domain, the quality assurance domain, the user context domain and the recommendation domain. Moreover, we showed how this ontology network can be tailored, to specific health domains and user points of views.

The main aim of this design was to obtain a flexible model that were not dependent on any particular mechanisms of websites content evaluation, such as a specific quality metric or health domain. Whenever it is required to apply a different metric for a quality factor or to consider another health domain, new extensions of quality and recommendation ontologies might be implemented, keeping up the model core intact.

In addition, a valuable feature of driving the recommendation process by ontologies is the property of checking the consistency among concepts and relationships that allow to detect inconsistencies at the design phase. Based on the intrinsic properties of ontologies, ontological model provides a high level abstraction that allows specifying in simple way relations between dimension and metrics for defining quality assurance. Besides, it also offer the possibility of defines restrictions that have to be hold in order to achieve to a consistent specifications of quality or recommendation assessments. Combination of the ontological paradigm with Horn-like rules made also possible the detection of anomalous specifications. Thus, we showed the feasibility of defining a validation process that assist the recommendation process modeling.

Starting from the presented design, good practices on Ontology Engineering lead to evaluate the model in an interaction between ontology engineers and domain experts. From this evaluation, it is expected to obtain a feedback to reach a final refinement of the structures which compose the ontology network.

7 Acknowledgment

This work was partially funded by: the SALUS/CYTED and PROSUL projects which are sponsored by the CnPq, Brasil and the CyTED, Spain. It is also supported by the PAE 37279-PICT 02203 which is sponsored by the ANPCyT, Argentina.

8 References

[1] Gunther Eysenbach, T.L.D.: Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information. BMJ **317** (1998) 1496–1502

[2] Barnes, S., Vidgen, R.: Webqual: An exploration of web-site quality. In: Proceedings of the Eighth European Conference on Information Systems, Vienna, July 3. (2000) 2000

[3] Mich, L., Franch, M., Inverardi, P.N., Marzani, P.: Choosing the"rightweight" model for web site quality evaluation. [17] 334–337

[4] Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. J. Manage. Inf. Syst. **12**(4) (1996) 5–33

[5] Wand, Y., Wang, R.Y.: Anchoring data quality dimensions in ontological foundations. Commun. ACM **39**(11) (1996) 86–95

[6] Pipino, L.L., Lee, Y.W., Wang, R.Y.: Data quality assessment. Commun. ACM **45**(4) (2002) 211–218

[7] Dubay, W.H.: The principles of readability. Costa Mesa, CA: Impact Information (2004)

[8] McLaughlin, H.G.: SMOG grading - a new readability formula. Journal of Reading (1969) 639–646

[9] Llambías, G., Motz, R., Toledo, F., de Uvarow, S.: Learning to get the value of quality from web data. [18] 1018–1025

[10] Ziegler, C.N.: Semantic web recommender systems. In: In Proceedings of the Joint ICDE/EDBT Ph.D. Workshop 2004 (Heraklion, Springer-Verlag (2004) 78–89

[11] Jimenez, J.A.A., Diaz, J.B., Gonzalez, A.C.: Logic Databases and Inconsistency Handling. In: Encyclopedia of Database Technologies and Applications. Information Science Reference (2006) 336–340

[12] Blanc, X., Mougenot, A., Mounier, I., Mens, T.: Incremental detection of model inconsistencies based on model operations. In: CAiSE '09: Proceedings of the 21st International Conference on Advanced Information Systems Engineering, Berlin, Heidelberg, Springer-Verlag (2009) 32–46

[13] Suárez-Figueroa, M.C., Dellschaft, K., Montiel-Ponsoda, E., Villazon-Terrazas, B., Yufei, Z., de Cea, G.A., García, A., Fernandez-Lopez, M., Gomez-Perez, A., Espinoza, M., Sabou, M.: Neon deliverable d5.4.1. neon methodology for building contextualized ontology networks. Technical report, NeOn Project (2008)

[14] Hartmann, J., Palma, R., Sure, Y.: Omv-ontology metadata vocabulary. In: ISWC 2005 Workshop on Ontology Patterns for the Semantic Web. (2005)

[15] Gangemi, A., Presutti, V.: C-odo: an owl meta-model for collaborative ontology design. In: Proceedings of CKC Workshop at WWW2007, Banff, Canada, Berlin, Springer (2007)

[16] Baader, F.: The Description Logic Handbook: theory, implementation, and applications. Cambridge University Press (2003)

[17] Lovelle, J.M.C., Rodríguez, B.M.G., Aguilar, L.J., Gayo, J.E.L., del Puerto Paule Ruíz, M., eds.: Web Engineering, International Conference, ICWE 2003, Oviedo, Spain, July 14-18, 2003, Proceedings. In Lovelle, J.M.C., Rodríguez, B.M.G., Aguilar, L.J., Gayo, J.E.L., del Puerto Paule Ruíz, M., eds.: ICWE. Volume 2722 of Lecture Notes in Computer Science., Springer (2003)

[18] Meersman, R., Tari, Z., Herrero, P., eds.: On the Move to Meaningful Internet Systems: OTM Workshops, OTM Confederated International Workshops and Posters, ADI,

AWeSoMe, COMBEK, EI2N, IWSSA, MONET, OnToContent + QSI, ORM, PerSys, RDDS, SEMELS, and SWWS 2008, Monterrey, Mexico,. In Meersman, R., Tari, Z., Herrero, P., eds.: OTM Workshops. Volume 5333 of Lecture Notes in Computer Science., Springer (2008)