Modeling a Web Site Quality-based Recommendation System

Edelweis Rohrer
Fac. Ingenieria, UdelaR
Reissig 565
Montevideo, Uruguay
erohrer@fing.edu.uy

Regina Motz
Fac. Ingenieria, UdelaR
Reissig 565
Montevideo, Uruguay
rmotz@fing.edu.uy

Alicia Diaz
LIFIA, Fac. Informatica, UNLP
CC 11
La Plata, Argentina
alicia@lifia.info.unlp.edu.ar

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 concepts, among others. At the moment of the specification of a recommendation system it must be guaranteed a right interrelation of all of these features. In this paper, we propose an ontology network based process for web site recommendation modeling. The 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.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous: Recommendation Systems

General Terms

Documentation, Design, Verification

Keywords

Quality Assessments, recommendation system, Ontology Network

1. INTRODUCTION

Web site recommendation systems must help to get high quality information. In order to explicitly conceptualize this property, the modeling of recommendation system involves the combination of many features: metrics of quality, quality criteria, recommendation criteria, user profile, specific domain concepts, among others. At the moment of the specification of a recommendation system it must be guaranteed a right interrelation of all of them.

The use of the web by common people, as a repository where to find information, increases drastically day by day. This is a very worrying reality because many of websites do not contain data of good quality: precise, believable, relevant to the userÂĕs profile. There are several characteristics of websites which make attention to quality issues necessity. Particularly, the lack of quality controls (i.e. editorial boards) at the stage of production. Then, quality-based recommendation systems are a help to get high quality web sites for users needs.

It is quite common to find recommendation systems based on ontologies that model the user profile and the domain of resources to be recommended [17, 7, 14]. However, it is uncommon to find models that explicitly represent the criteria used by the recommendation systems or that express the quality dimensions of resources on which criteria are applied.

In this paper we model a web site quality-based recommendation system by an ontology network. Moreover, we show how the ontology network is useful to check the correctness of the recommendation system specification.

Quality in websites is determined by several diverse factors, some of which are enough general, and therefore, they can be 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 operation of the site, its conformity with standards of the used language or of accessibility like those described in normative such as the Web Content Accessibility Guidelines of the W3C 11 . There are quality models that take these features into consideration, some of them are for example WebQual [4] and WebQEM [10]. Particularly, in this work, we focus on the quality that arises of the information value that the site provides and its adequacy to the consumerÂĕs context.

Regarding the correctness of the specification, the ontology network helps to reach this goal, since it allows the ontologist to asses restrictions and rules to classify invalid quality and recommendation criteria. Particularly, in this paper we present an application on the health domain, however, it can be extended to any other domains.

The remain of the paper is organized as follow: Section 2

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
iiWAS2010 , 8-10 November, 2010, Paris, France
Copyright 2010 ACM 978-1-4503-0421-4/10/11 ...$10.00.

presents an introduction to web site quality assurance from webÂěs content point of view. Then, in Section 4, the `Salus` ontology network is introduced. Following, in Section 5 we argue through the utility of the ontology network in validating the correctness of the recommendation system specification. Finally, in Section 6 we present a brief conclusion and future work.

2. QUALITY ASSURANCE

The general term "data quality" is a combination of quality factors, where each factor focuses on a specific point of interest. There are several data quality factors: relevance, accuracy, reliability, accessibility, freshness and syntactic correctness are some of them. There are several works that describe each one of these factors [3, 16, 11]. Which factor is the most relevant, depends on the specific domain of the information system and on the intended use of the data. In this section we analyze some quality dimensions specially tailored to the case when data is provided for web pages.

2.1 Timeliness

Timeliness is widely identified as one of the most important factors of data quality to consumers. Hence, giving information to the systems about the timeliness degree on the web data that they are consuming is a major challenge in the developing of applications. However, there are several definitions about data freshness, in general it is accepted that timeliness measures how much updated are data for a specific task. In Pipino et al. [12], the following definition can be found: *Timeliness is the extent to which data is sufficiently up-to-date for the task at hand.*

Considering web data, one approach is to focus on the concepts of age and *volatility* as defined in [1, 16, 11]. The age suggests how old are data, captures the time interval between the creation or updating of data and the time at what user receives data. On the other hand, volatility measures the frequency with which data change over time.

Note that it is possible to have updated data and yet they are useless, since their usage time expired. In this sense, the volatility of data is a relevant element in the composition of the timeliness degree. Therefore, timeliness degree is obtained by the metric defined by the following formula, which relates the concepts of age and volatility:

$$Timeliness = Max(0, 1 - (age/volatility)) \quad (1)$$

This metric results in a value between 0 (the data are not up-to-date) and 1 (data are extremely up-to-date).

However, depending on the specific web site contentÂěs domain, users can find useful other metrics or different definitions. The quality dimension ontology that we have developed is flexible to model different points of view [11].

2.2 Readability

Readability dimension concerns what makes some texts easier to read than others. In Dubay et al. [6] were introduced different readability metrics created for different domains and user profiles. *Readability is what makes some texts easier to read than others.* There are a lot of readability formulas created for different authors, like FOG and SMOG [9] grade levels, that reached good results when they were tested [6]. The FOG and SMOG grade levels are defined by the next two formulas:

$$FOG = 0.4(averageSentenceLength + hardWords) \quad (2)$$

$$SMOG = 3 + ?polysyllableCount \quad (3)$$

Here also, the decision on which formula, 2 or 3, to use must be taken for a domain expert.

Syntactic correctness is a factor that is involved in some readability metrics of web data quality. The syntactic correctness concerns that data are free of syntactic errors such as typing errors or format. The data are considered syntactically correct if they comply with user-defined rules or restrictions. Examples of rules are: *classrooms are identified by three digits numbers* or the one that says that a date is represented by the format *mm/dd/yyyy*, where *mm*, *dd* and *yyyy* are integers such that $00 < dd <= 31$, $00 < mm < 13$, considering also the different number of days according to each month and year (31, 30 or 28). In this example, syntactic correctness verifies that the date is a valid date, without verifying the relationship that the date may have with the reality; for example, that a given date is really the date of my birthday. The latter kind of correctness is what is called semantic correctness.

The value of the outcome of the metric that measures the syntactic correctness takes values true or false. A useful idea may be, the use of a range of values more significant than simply Boolean to represent the seriousness of the error. However, the measure of the seriousness of the error varies among different situations depending on the use will be given to the data, the domain of interest and the viewpoints of the user; making it impossible to establish this range of values generically. Here also, the decision on which formula to use must be taken for a domain expert.

2.3 Believability

Regarding believability, in Pipino et al. [12] 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 for example, it is important to take into account the existence of sites with certified quality labels, such as HON [1], WIS [2] and WMA [3], which means that documents linked by these sites will be evaluated with a higher level of quality than those that have no certification.

Metrics to measure believability regarding as true and credible data may involve different factors, for example, semantic correctness and consistency.

The semantic correctness refers to the degree to which data represent real-world. To measure this factor, it is necessary to make a comparison of data with the real world that may be represented by a trusted reference, called *oracle*, considered always as correct. Considering, for example a webpage that has information on health clinics and their

[1] http://www.hon.ch/
[2] http://www.portalesmedicos.com/web_interes_sanitario/index.htm
[3] http://wma.comb.es/

directors, in order to verify whether in the real world each clinic was directed by the director that is indicated on the page, one should access an external trusted reference in order to evaluate some questions. This reference may be, for instance, a Public Health Database. Possible questions that this oracle should answer are:

- Is it correct that <DirectorName> is a doctor?
- Is it correct that <ClinicName> is a health clinic?
- Is it correct that <DirectorName> directed the <ClinicName>?

This example shows that semantic correctness of data from webpages must be evaluated considering the relationships among all the data. The difficulty lies in that semi-structured data, such as data from a webpage, does not explicitly express these relationships, as they do, for instance, functional dependencies in relational databases. Therefore, metrics to evaluate this factor depend on the ability to the information extractor to recognize relevant relationships among web data.

Assuming that these relevant relationships among data are known, one possible metrics for this quality factor may be this:

- If establishing a single question to check its correctness, and this is passed successfully, the page is considered absolutely correct in terms of semantics, evaluated to 1
- If establishing only two questions and only one was successfully approved, the semantic correctness of the web page is evaluated to 0.5
- and so on

Ultimately, this is a correct response rate compared to the number of replies.

The *consistency factor* measures the fulfillment of integrity restrictions. This factor is strongly related to the semantic correctness factor, if we have data semantically correct, these must be consistent, because they correspond with reality which is always consistent. Anyway, the consistency factor measures complementary aspects from semantic correctness. While the semantics correctness verifies the correctness of the values that take place in relationships within concepts, the consistency factor measures the correctness of relationships itself. For example, whereas the clinic name and the director name in a webpage that really corresponds to the clinic directed by this doctor is evaluated as semantically correct, if the data is about a clinic directed by an animal this is evaluated as an inconsistency.

The detection of this inconsistency is achieved because there exists the restriction that the director of a health clinic must be human. The extractor of data from the webpage must be able to validate this restriction from an ontology that describes the domain relations. Hence, as has been shown in [8], the metrics of the consistency factor depends directly on the quality of the data extractor.

3. QUALITY ASSURANCE ONTOLOGY

In order to get the values of quality from web data in a flexible and consistent way, the first step is to specify a

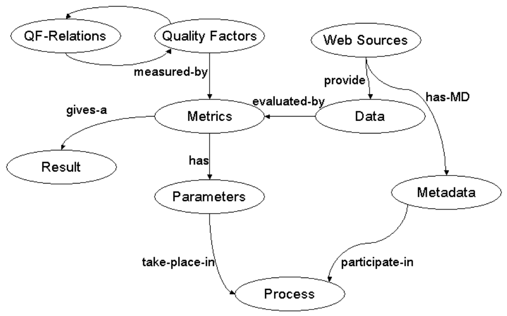

Figure 1: The Quality Factor Ontology (QF-Ontology)

formal model that represents the factors 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 the works of *Qurator* project on biological data quality [7], and also from work in the area of QoS [5]. However, we differentiate from these projects in the sense that our proposal is to model a generic ontology to quality dimension (Q-Ontology), independent from the specific domain and from the different factors and metrics applied.

In addition to the valuable property of checking the consistency among concepts and relationships, the ontological model provides a high level abstraction that allows specifying in simple way relations between factors and metrics. It also offers a rule language, SWRL, which in a declarative way can be used to specify the mechanisms to measure the quality of factors and dimensions.

Figure 1 depicts, in a simplified way, our proposed ontology for representing the context to assessing quality dimensions of web data. As described in Section 2, data quality is a composition of quality dimensions such as timeliness, semantic correctness, consistency, among others which holds some relationships. For each quality dimension can be defined factors with different metrics. Metrics are functions that have input parameters and produce a result. Web data belongs to a web source, which provides metadata. Web source metadata and metricsÂě parameters participate in the process of acquisition of the quality dimensions values for a given data or relationship between data.

In the QF-Ontology, the `Freshness` factor for example, is a subclass of `QualityFactors` and it can be measured using different `Metrics`. If we use the metric proposed in Section 2.1, the input parameters for movies domain are *age* and *volatility*. The user must give the value of volatility according to the features of the domain, but the value of age must be calculated according to the features of each source. In this sense, web sites have different metadata that can be used to get the value of the age of their data (i.e. headers http, RDF, RSS, ATOM). Therefore, QF-Ontology should be developed according to the profile of a user-specific domain and by defining the metrics that should be used to get values of each quality factor in terms of the data from specific sources on the web. An OWL version of complete Quality Factor Ontology (QF-Ontology) can be accessed through the home page of `Salus` Project [4].

[4] http://www.fing.edu.uy/inco/grupos/csi/Proyectos/SALUS/

The problem of how getting the values of quality factors from web data is tailored in paper [8]. In this paper, we mainly focus on exploiting the quality ontology as a component of an ontology network in order to check the whole correctness of the modeling recommendation system.

The quality assessment of web contents and their recommendation 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 system administrator determines the quality dimensions that are going to be considered to evaluate web contents, and what metrics are used to assess each factor. Moreover, it is important to remark that the combination of user profile, context features and quality levels associated to web contents is a key aspect to reach a recommendation that suggests the most suitable contents to each user.

A key issue of this approach is that the validation of the relationships between these different ontologies are exploited by the use of OWL language and SWRL rules (explained in Section5). In the next section, we will explain the concept of ontology network by a motivated example in the context of the `Salus` Project.

4. THE SALUS ONTOLOGY NETWORK

The `Salus` ontology helps to recommend health-related web pages 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 [15, 2, 13]. These domains 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:

The *Specific Health Ontology* conceptualizes a health domain. This ontology may be an already existing ontology like UMLS[5] which conceptualizes, for instance, risk factors, impact, treatment, diagnostic, effects, phases of a disease. This ontology can be refined in terms of a specific disease; for example "Alzheimer treatment", ÃňDiabetes diagnosticÃő, etc.

The *Web Site Ontology* conceptualizes the domain of web resources that will be consider in a quality assessment. The main concepts of this ontology are `web resource` and `web resource property`. A `web resource` is any resource which is identified by a URL; for instance a webpage. `Web resource properties` models the properties that can be attached to a `web resource`. For instance, possible properties of a webpage could be the author, the amount of words, etc. Among these properties there is a particular one, the *hasTopic* property that relates `web resources` with concepts at the Specific Health Ontology. The *hasTopic* property describes what a `web resource` is talking about. This property is always presented in webpages.

The *Quality Assurance Ontology* conceptualizes metrics, quality assurance specifications and quality assessments. `Metrics` are specific calculus based 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` associates to each `quality dimension` a corresponding `metric`. A `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`.

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

The *Context Ontology* describes the user profile and query situation. The `user profile` describes `user properties` as *user age range*, *role*, *academic level*, *health domain expertise*, among others. The *query situation* represents the context of the query. A characteristic query situation is the *query goal* (i.e. educational, commercial, academic, etc.).

Finally, the *Recommendation Ontology* describes the different criteria of recommendation for a particular context (user and query situation), quality assessment and the obtained recommendation level.

Particularly, `Salus` ontology network is specific tailored to the health domain, but it could be easily adapted to another domain, just by changing the health ontology by another specific domain ontology.

All of these ontologies, each one built for different purposes, may be used together in complex applications. However, the information about how is the relationship among them, is usually hidden in the application code. The explicit use of an ontology network does not only describe relationships between ontologies but also serves as the conceptualization of the main challenges of software development process.

The challenge is to get relationships for each networked ontology in order to help us to specify the ontology network and to detect inconsistencies among them. Then, our approach is to consider, in an integrated way, the specific health domain of interest (i.e. Alzheimer diagnosis, Diabetes treatment,etc.), the query goal (i.e. educational, commercial, etc.), the user profile (i.e. user age, language, genre, etc.), the dimensions of quality and the criteria to assure that some information is in accordance to the goal of "fitness for use" for a consumer.

`Salus` networked ontologies are interrelated by three different relationships. They are: *uses*, *extends* and *describes* relationships. The semantic of these relationships is described below:

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 A*lzheimer 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 Specificatio*n 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.

The Figure 2 shows all of these networked ontologies; each one by a different color. In this figure, it is possible to appreciate the different interrelationships hold among the networked ontologies.

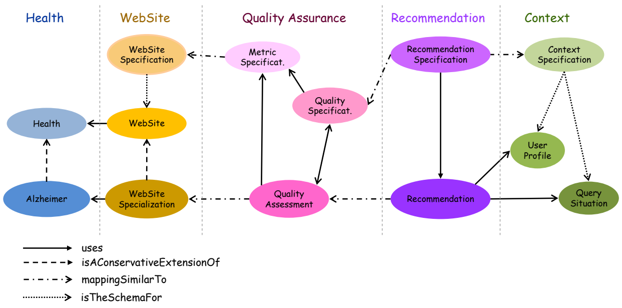

Figure 2: `Salus` ontology network

During the execution of the process to get a set of websites to be recommended, the `Salus` ontology network 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 and recommendation assessments. In the last cases, the `Salus` ontology can be used to both: assisting in the modeling and specification of a recommendation system and checking the correctness of the resulting system specification.

The modeling and specification of a recommendation system based on the `Salus` ontology network involves the population of the `Salus` ontology network. The population of the `Salus` ontology is done in three different moments: at the start up of the recommendation system, when performing the quality assessment of a set of given webpages and finally, at the execution of recommendation assessments. The Figure 3 represents 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".

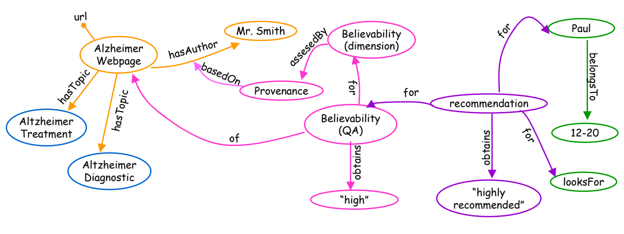

Figure 3: A population example of the `Salus` ontology

The ideas about the correctness checking of a recommendation system specification will be developed following in the next section.

5. SPECIFYING A CONSISTENT RECOMMENDATION SYSTEM

Having an ontology-based recommendation system is not only advantageous at the moment of configuring this kind of systems (for instance to discover knowledge domain units), but also it is helpful to validate the resulting configuration of the systems. For example, someone might try to assess a quality dimension with a unappropriated metric. In this section, we show how by using web semantic technologies like: OWL, SWRL and SQRWL, it is possible to maintain consistent the Recommendation System specification.

Ontology languages, like OWL[6] and their combination with rules likes SWRL[7], are potential tools to express restrictions on the ontology network. 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, allowing users to write Hornlike rules expressed in terms of OWL concepts. Besides, the OWL query language SQWRL[8], based on SWRL, provides the chance of doing queries on OWL ontologies and enables to perform closure operations like counting and aggregation, among others. SWRL is used to describe the recommendation rules.

At the following, we illustrate by examples the utility of OWL, SWRL and SQWRL in order to preserve the consistency of the recommendation system model. First we depict how OWL restrictions are used to complete the specification of a class. Then, we show how SWRL and SQWRL are used to detect specification anomalies.

5.1 Consistency by using OWL

Within the Quality Assurance ontology, a `quality dimension` assesses to a `quality level`. Considering *believability* as an individual of the *Dimension* class, in order to the assessment make sense, there must be at least two instances of the *QualityLevel* class related to the believability individual, for example "high believability" and "low believability". Here, we can use OWL in order to define a constraint that satisfies this requirement. Thus, the *Dimension* class has attached the conjunction of the following two restrictions to represent this constraint (we use a description logic-style notation for simplicity):

$$\exists assessesTo.QualityLevel \tag{4}$$

$$>= 2 assessTo.QualityLevel \tag{5}$$

Axioms 4 and 5 express an existential restriction and a cardinality restriction respectively. According to 4 for each individual of the Dimension class that has the relationship *assessTo* must exist at least an individual of the *QualityLevel* class linked to it. According to 5 for each individual of

[6] http://www.w3.org/TR/owl-features/
[7] W3C. SWRL Semantic Web Rule Language. URL: http://www.w3.org/Submission/SWRL/
[8] http://protege.cim3.net/cgi-bin/wiki.pl?SQWRL

the Dimension class that has the relationship *assessTo* it must be related with a minimum of two individuals of the *QualityLevel* class.

5.2 Consistency by using SWRL and SQWRL

Considering an scenario to model a *Quality Criteria Definition*, we can define for example a metric called *BasicPropertiesMetric*, to measure whether a webpage has certain basic properties like *author* and *source*. It may be 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. An expert working in the recommendation system can define that to assess the *Believability* dimension the *BasicPropertiesMetric* metric must be applied. Then, the *Believability* individual, which is an instance of the *Dimension* class, will be related to the *BasicPropertiesMetric* individual, which is an instance of the *Metric* class, through the relation *assessedBy*. At the moment of carrying out the quality assessment of a webpage for the *Believability* dimension, the *BasicPropertiesMetric* metric must be used. But, 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 the 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 this hypothetical scenario, 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* class could be added as a subclass of the QualityAssessment. 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:

$$QualityAssessment(?assess) \tag{6}$$
$$\wedge\, Metric(?metric) \tag{7}$$
$$\wedge\, through(?assess, ?metric) \tag{8}$$
$$\wedge\, Dimension(?dimension) \tag{9}$$
$$\wedge\, for(?assess, ?dimension) \tag{10}$$
$$\wedge\, Metric(?metricD) \tag{11}$$
$$\wedge\, assesedBy(?dimension, ?metricD) \tag{12}$$
$$\wedge\, sqwrl : makeSet(?s1, ?metric) \tag{13}$$
$$\wedge\, sqwrl : groupBy(?s1, ?assess) \tag{14}$$
$$\wedge\, sqwrl : makeSet(?s2, ?metricD) \tag{15}$$
$$\wedge\, sqwrl : groupBy(?s2, ?assess, ?dimension) \tag{16}$$
$$\wedge\, sqwrl : notIntersects(?s1, ?s2) \tag{17}$$
$$\Rightarrow InvalidQualityAssessment(?assess) \tag{18}$$

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*, by 7 and 8 and those *?dimension* individuals related to *?assess* by the object property *for*, by 9 and 10. 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*, by11 and 12. 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 13 and 14. 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 15 and 16. 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*, by 17.
- In case of the intersection set is empty, the instance *?assess* is classified into the InvalidQualityAssessment class, by 18.

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.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we present a quality-based approach to get the more adequate websites for a specific consumer context. We show the Quality Assurance Ontology as a networked ontology for a Web Site Recommendation System. Moreover, we show how OWL ontology and SWRL and SQWRL may be very useful in order to predict inconsistencies among different definitions of the Recommendation System.

We have described the `Salus` Ontology Network that models the different domains related to a recommendation system These domain are: the specific health domain, the web site domain, the quality assurance domain, the user context domain and the recommendation domain.

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.

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. ACKNOWLEDGMENTS

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

8. REFERENCES

[1] J. Akoka, L. Berti-Equille, O. Boucelma, M. Bouzeghoub, I. Comyn-Wattiau, M. Cosquer, V. Goasdoué-Thion, Z. Kedad, S. Nugier, V. Peralta, and S. S.-S. Cherfi. A framework for quality evaluation in data integration systems. In J. Cardoso, J. Cordeiro, and J. Filipe, editors, *ICEIS (3)*, pages 170–175, 2007.

[2] C. Allocca, M. d'Aquin, and E. Motta. Door - towards a formalization of ontology relations. In J. L. G. Dietz, editor, *KEOD*, pages 13–20. INSTICC Press, 2009.

[3] D. Ballou, R. Wang, H. Pazer, and G. K. Tayi. Modeling information manufacturing systems to determine information product quality. *Manage. Sci.*, 44(4):462–484, 1998.

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

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

[6] W. H. Dubay. The principles of readability. *Costa Mesa, CA: Impact Information*, 2004.

[7] G. EPSRC Programme Fundamental Computer Science for e Science: GRS67593. Quarator project: Describing the quality of curated e-science information resources. http://www.csd.abdn.ac.uk/research/qurator/, [last visited september 2010.

[8] G. Llambías, R. Motz, F. Toledo, and S. de Uvarow. Learning to get the value of quality from web data. In R. Meersman, Z. Tari, and P. Herrero, editors, *OTM Workshops*, volume 5333 of *Lecture Notes in Computer Science*, pages 1018–1025. Springer, 2008.

[9] H. G. McLaughlin. SMOG grading - a new readability formula. *Journal of Reading*, pages 639–646, May 1969.

[10] L. Mich, M. Franch, P. N. Inverardi, and P. Marzani. Choosing the"rightweight" model for web site quality evaluation. In J. M. C. Lovelle, B. M. G. Rodríguez, L. J. Aguilar, J. E. L. Gayo, and M. del Puerto Paule Ruíz, editors, *ICWE*, volume 2722 of *Lecture Notes in Computer Science*, pages 334–337. Springer, 2003.

[11] V. Peralta. *Data Quality Evaluation in Data Integration Systems.* PhD thesis, Universidad de Versalles, France - Universidad de la Republica,Uruguay, November 2006.

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

[13] E. Rohrer, R. Motz, and A. Diaz. Modelling and use of an ontology network for website recommendation systems. In R. Meersman and P. Herrero, editors, *OTM 2010 Conference Posters*, LNCS. Springer Velag, October 2010.

[14] R. Sena, Oscar; Motz. Hacia un modelo genérico para la calidad de los servicios web. In *Proceedings of II Congreso Espanol de Informatica*, September 2007.

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

[16] D. Wang, R; Strong. Beyond accuracy: What data quality means to data consumers? *Journal on Management of Information Systems*, 12(4):5–34, 1996.

[17] C.-N. Ziegler. Semantic web recommender systems. In W. Lindner, M. Mesiti, C. Türker, Y. Tzitzikas, and A. Vakali, editors, *Current Trends in Database Technology - EDBT 2004 Workshops*, volume 3268 of *Lecture Notes in Computer Science*, pages 521–521. Springer Berlin / Heidelberg, 2005. 10.1007/978-3-540-30192-9-8.