Context dependent semantic granularity

by Riccardo Albertoni, Elena Camossi, Monica De Martino, Franca Giannini, Marina Monti

Int. J. of Data Mining, Modelling and Management, 2011 Vol.3, No.2, pp.189 - 215

DOI: 10.1504/IJDMMM.2011.041495

Context Dependent Semantic Granularity

Riccardo Albertoni[1]*, Elena Camossi, Monica De Martino[1], Franca Giannini [1], Marina Monti[1]

[1]IMATI-GE,
Consiglio Nazionale delle Ricerche,
Via De Marini, 6 - Torre di Francia - 16149 Genova, Italy
E-mail: albertoni@ge.imati.cnr.it
E-mail:demartino@ge.imati.cnr.it
E-mail: giannini@ge.imati.cnr.it
E-mail: monti@ge.imati.cnr.it
*Corresponding author

Abstract: A fundamental issue to improve the accessibility to information resources is how to efficiently deal with huge amount(s) of data. In this respect, ontology driven techniques are expected to improve the overlap between the Cognitive Space applied by the user and the Information Space, which is defined by the information providers. In this paper we describe a powerful method to extract *semantic granularities*, which enable the navigation of a repository according to different levels of abstraction. In the formalization we present, granularities are explicitly parameterized according to criteria induced by the *context*, which improves the method flexibility. Furthermore, the parameterization assists the user allowing to formulate and refine the browsing criteria. Case studies are described to demonstrate how granularities ease the information sources browsing and to illustrate how they may vary according to the context. A validation of the cognitive principles behind the method is presented, together with the analysis of the results obtained by the experimentation.

Keywords: semantic granularity; ontology; semantic browsing; application context.

Reference to this paper should be made as follows: Albertoni, R., Camossi, E., De Martino, M., Giannini, F. and Monti, M. (xxxx) Context Dependent Semantic Granularity, *Int. J. Data Mining, Modelling and Management*, Vol. x, No. x, pp.xxxxxx.

Biographical Notes: Riccardo Albertoni received his Ph.D. in Electronic and Computer Engineering from the University of Genoa, Italy in April 2007. Since April 2002 he has carried out his research activity at the Institute of Applied Mathematics and Information Technology of the National Research Council in Genoa (CNR-IMATI-GE). His research interests are in the fields of Knowledge Management: Data Visualization, Metadata Analysis, Ontologies, Semantic Similarity and Granularity. He has actively worked in different EU-funded Projects pertaining to the analysis of metadata and the exploitation of information resource semantics. He has been programme committee member of various

conferences and co-author of more than 50 publications including scientific papers and technical reports.

Elena Camossi holds a Ph.D. (2005) and an advanced Degree (2001) in Computer Science awarded from the University of Milan, in Italy. She has been post doctoral researcher at the Institute for Applied Mathematics and Informatics of the Italian Research Council (IMATI-CNR) and at the School of Computer Science and Informatics, University College Dublin. Her research activity focuses on modeling, consistency checking and analysis of multi-granular spatio-temporal data in Spatial and Geographic Information Systems. She has been also investigating the multi-granular modelling and querying of three-dimensional geographic data in databases, and the knowledge based representation and extraction of multiple granularities.

Monica De Martino is a researcher at CNR-IMATI-GE where she is leading the research activity related to the knowledge technology for Geographic Information Management. She graduated from the Department of Mathematics, University of Genova in 1992. Since then she has been working at the IMATI-CNR Genova. She has been involved in several National and International Projects on Spatial Data Processing and Analysis and their specific application in Geographic Information System. Successively she has been extended her research expertise to the Knowledge Management field on Metatada Analysis, Ontology knowledge exploitation and Semantics Analysis. Most of her research results has been carried on and validated within European project: where she has been scientific responsible for her Institute (e.g TAUTEM, INVISIP, IDE-UNIVERS, Nature-SDIplus). The results have been published in high relevant journals and conferences.

Marina Monti is involved in academic research since 1998 when she joined the Institute for Applied Mathematics and Information Technologies of the National Research Council in Genoa, Italy, after 14 years of experience in research and development departments of IT companies. Her main scientific activities focused mainly on 3D modelling, product model representation in industrial product development and knowledge formalization and processing. Main industrial application fields are mechanical design, industrial design and process plant. Working in research departments both industrial and public, she actively participated to several international funded projects. She has published extensively in high profile journals and conferences.

Franca Giannini is senior researcher at the Institute for Applied Mathematics and Information Technologies of the National Research Council in Genoa, Italy. She has been working on different issues for the specification of tools and methodologies for the representation and analysis of 3D objects and the related knowledge in industrial product development. She has participated and been responsible for IMATI of several national and international projects. The results of her research activity have been published in more than 90 reviewed papers presented in international conferences and journals. Her current research interests include multidimensional media modelling and understanding and knowledge formalisation and processing.

1 Introduction

Semantic web is rising as an extension of the current web to provide sophisticated and powerful inferences improving the accessibility to web content. Indeed, in the traditional web, the huge amount of results returned by a search activity completely overwhelms the user capability to exploit the available information. The difficulties pertaining to information access are mainly due to a poor overlap between the information model employed by the user (i.e., the Cognitive Space) and the model defined by information providers (i.e., the Information Space) (Newby (2001)). As a consequence search activity is a highly interactive process: the seeker refines the selection criteria according to the results obtained, alternating querying and browsing activities.

Metadata represented in ontologies to characterize information resources, namely ontology-driven metadata (Sicilia (2006)), may formally describe portions of both the Cognitive and the Information Spaces, improving the accessibility to the repository content. The available ontology technology provides reasoning facilities that are very useful in supporting querying activities as well as in checking the ontology consistency, but lacks effective tools for fully exploiting its content. Therefore, ontology driven methods that consider both the implicit (i.e., encoded in the resources) and the explicit (i.e., formalized in the ontology) semantics of information resources are a primary research issue. Among existing ontology driven methods, *semantic granularity* (Albertoni et al. (2006)) enables the browsing of information resources according to different levels of abstraction, i.e., granularities.

Moreover, to fill the gap between Cognitive and Information Spaces, it is mandatory to take into account the influence of the context. For example, to access the ontology content by using granularities, it is important to consider that ontology entities concur differently in the granularity assessment according to the application context.

In this paper, we propose a context dependent formalization of semantic granularity. It relies on an ontology model and a layered framework originally inspired by Ehrig et al. (2005) and originated from the research results presented by Albertoni & De Martino (2008), where the *application context* has been formalized in order to parameterize the semantic similarity among ontology instances. The application context models the importance of ontology entities (i.e., classes, attributes and relations) as well as the different operations used to analyse them.

The resulting instrument is a powerful ontology driven method that eases the browsing of a repository of information resources where the user may formulate and modify the granularity criteria induced by the context. The main advantages of the method are:

- the exploitation of implicit and explicit semantics of information resources organised by an ontology;
- the tailoring of the resource organization according to requirements arising from a specific application context;
- the improvement of the decision making of the user in the selection of browsing criteria according with his/her needs.

The paper is an extension of the ongoing research presented in Albertoni et al. (2008). With respect to our previous work, we provide more details useful for the method exploitation: in particular the granularity computation is fully described and the principles behind the method are justified through a validation process.

The paper is organized as follows. Section 2 provides an overview of related works. Section 3 illustrates the application scenario in which the semantic granularity is usually applied. Section 4 gives an overview of the semantic granularity presented in Albertoni et al. (2006), and the basic assumptions upon which the method relies. Section 5 describes the extension of the method parameterized according to the context: it illustrates the formalization of the application context and the computation of the semantic granularity. Section 6 summarizes and discusses the results of the first stage of evaluation. Finally, Section 7 concludes the paper outlining future research directions.

2 Related Work

Semantic Granularities compared with Granularities in Information System
Granularities allow to explore data according to different levels of detail, enhancing the flexibility in the information representation and retrieval. In the area of Information Systems granularities have been widely studied for the spatio-temporal domain (Khatri et al. (2002), Bertino et al. (2009)). The recent research has focused mainly on cognitive issues pertaining to the perception of vagueness, indeterminacy, imperfection, roughness, etc. (see for instance Bittner & Stell (2003)). Moreover, some attempts to define semantic granularities have been made with respect to terminologies by Fonseca et al. (2002). They introduced the term *semantic granularity* exploiting object oriented cast to represent ontology instances at different levels of detail.

Differently from previous work on granularities, the semantic granularity method we propose in (Albertoni et al. (2006)) is aimed at repository browsing. It dynamically generates the granules and the granularities to use for browsing a collection of information resources, and creates the structure encoding the relationships among the different levels of abstraction (namely, the *granularity lattice*). Granularity generation encompasses both the explicit semantics of data, given by the ontology, and their implicit semantics, which is encoded in the repository. Only the relevant granularities for the repository under evaluation are generated, according to the distribution of its population, that greatly affects the capability of abstracting information sources. Therefore, two repositories described by the same ontology, but with different instances, may generate different sets of granularity. By contrast, granularities applied in databases are static, i.e., pre-arranged by the database designer and embedded in the database schema, independently of the database population.

Semantic Granularities compared with techniques to organize repositories
Automatic techniques such as clustering and classification can be employed to organize a repository and to ease its browsing (see as example, Xu & Wunsch (2005), Krowne & Halbert (2005)). Clustering only relies on the model emerging from the Information Space, whereas the classification relies on a set of classes, belonging to the user's Cognitive Space, that are expected to be meaningful. On the contrary, semantic granularity as discussed in this paper takes into account both the spaces. It considers part of the Cognitive Space represented in the ontology as well as the Information Space by balancing the sources at a given granularity according to their occurrence. In this work semantic granularities are defined dynamically, according to the data model (represented by an ontology schema) as well as to data (represented as ontology instances).

Context in Information Systems
The representation of the context is discussed in several fields like Cognitive and Computer Science. In the early 90's, context has been applied to accomplish interoperability in heterogeneous and multidatabase systems or federated databases (Sheth & Larson (1990)). In particular, Sciore et al. (1992) represent context through object metadata, given by object attributes and their values; Ouksel & Naiman (1993) propose context building to represent the knowledge required for information exchange and dynamic schema integration of

different Information Systems against a given set of queries, encompassing both semantic and structural components of data. Kashyap & Sheth (1996, 1998) propose an explicit representation of context, which expresses the *real world semantics* of objects in a database given in term of their intensional description. Context is represented as a collection of *contextual coordinates*, formalized in a domain specific ontology, and their values. Context is specified also by the database in which the objects are stored, by their relationships, and by a subset of properties of interest. Furthermore, the definition of context implies the choice of a vocabulary. In this approach, relationships among different contexts may be handled within a semi-lattice structure. In all these works, the representation of context is fundamental for expressing the similarity, or proximity, among objects in a database and to give the functional transition among heterogeneous information systems. By contrast, in our work context is applied to express user needs, which affect the semantic granularity definition.

Context as representation of user needs
User's context is discussed for example in the Geo-spatial field for service personalization (see for instance Weakliam et al. (2008)) and to define collaborative systems (see Petit et al. (2008)). In Information Retrieval and related areas, contexts are employed to represent information user needs (see for example Hernandez et al. (2007)), often through the use of ontologies. In the field of Knowledge Representation for the Semantic Web, proposals to support contextual representation of the background knowledge have been presented by Bouquet et al. (2004) and by Segev & Gal (2005). Aleman-Meza et al. (2003) propose an explicit context formalization for semantic association ranking to express users' interests in RDF (the ontology languages Resource Description Framework by Klyne et al. (2004)), where the context defines a sub-graph of the RDF schema, and the semantic associations are given as weighted *paths of interest*, including properties and RDF classes. This approach is similar to what proposed by Guha (1990), that defines the context as a partition of an ontology (namely, a microtheory). More generally, in Computer Science the concept of context is often related to the notion of view (see Noy & Musen (2004), Volz et al. (2003)).

The definition of context proposed in this paper differs from the aforementioned formalizations because it is intended as an explicit parameterization of semantic granularity, and beside to the relevant features, it also includes the operations that should be applied to these features.

3 The Application Scenario

In this section we introduce the typical scenario in which the semantic granularity is worked out. Two main actors are assumed in this scenario (see Figure 1): the *user* and the *ontology engineer*. The user of the method is the domain expert and owns the knowledge to give the informal browsing criteria that will drive the semantic granularity evaluation. The ontology engineer acts as communication channel for the user requests to the system; he/she is in charge of translating the user needs into a context formalization that indicates which ontology entities are relevant and how they may be used during the granularity evaluation. The scenario is characterized by an interactive exchange of information between the user and the ontology engineer.

In the first step, the user provides the ontology engineer with the elements to identify (i) the set of resources to be explored (namely *Res*); (ii) the set of nominal values (namely

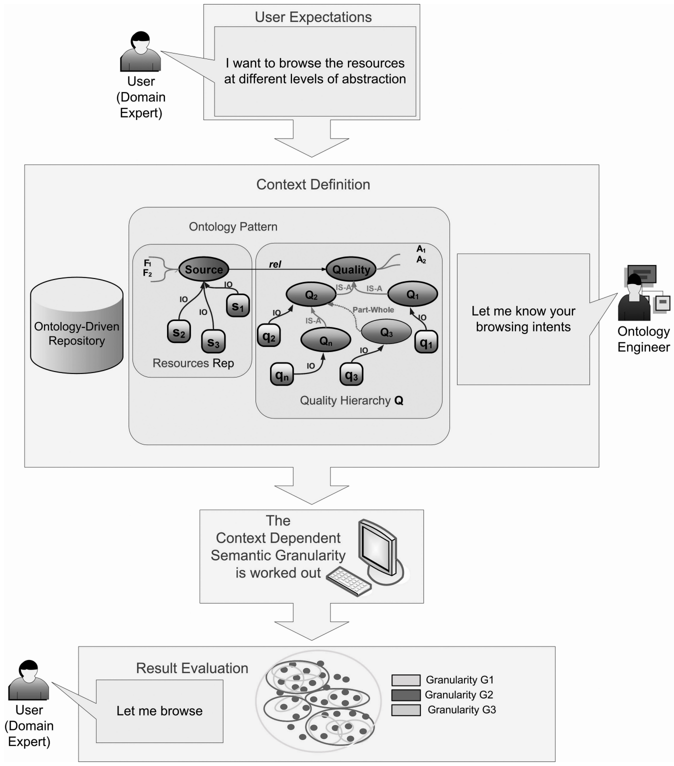


Figure 1 Application scenario.

qualities $\mathcal{Q}$) according to which the resources have to be organized; (iii) the aims of the browsing activity. For example, hereafter, scientific papers will be considered as resources; the features characterizing them will be the qualities (e.g., topic, date of publication, journal, authors); finally, the aim of the browsing is the identification of the hottest scientific topics.

In the second step, the ontology engineer analyzes the ontology and the information the user has provided to work out a context formalization to parameterize the semantic granularity. From a logical point of view, the method assumes that the repository is organized according to an ontology pattern where the resources and the qualities satisfy some structural constraints, such as they are instances of classes connected by some relations.

Moreover, specific criteria are formalized to tailor the granularity assessment to the user's aims. An application context is defined at this step to trace a generic repository to the ontology pattern and to customize the computation of the semantic granularity. Therefore, the semantic granularity can be applied to repositories organized according to any ontology schema and can be customized for the specific user aims.

In the third step the granularities are generated: they are defined dynamically according to the data model represented by the ontology schema; the data given by the ontology instances; and the context formalization.

The method results in a sequence of granularities $G_1, G_2, G_3, \cdots, G_n$, which group the resources at different levels of abstraction enabling a user oriented browsing. The scenario can be iterated refining the granularity criteria to further increase the adherence to the user expectations.

4 Semantic Granularity

This section provides an overview of the semantic granularity method, which is discussed in details in Albertoni et al. (2006) and which is the starting point of the work presented in this paper. Semantic granularity structures a repository at different levels of abstraction taking into account both its conceptual structure and its content. The approach we propose assumes that the repository is organized according to an ontology pattern which recalls the structure depicted in Figure 2, where:

- **Resource** is a class grouping the set of resources *Res* the user is going to browse;
- **Quality** is a class grouping the set of qualities $\mathcal{Q}$ according to which the repository has to be structured;
- F_1,F_2, A_1, A_2 are other class properties, which might be employed to characterize the resources, the qualities and the context;
- Q_1,Q_2, ..., Q_n are classes representing the qualities. Their names correspond to the labels to be exploited during the browsing;
- *Is-A* and *Part-Whole* are partial order relations among classes representing the qualities;
- s_1, s_2, s_3, q_1, q_2, .., q_n are instances;
- **IO** is the relation *Instance-Of* between classes and instances;
- rel is a relation, which joins the resources s_1, s_2, s_3, etc. to the instances q_1, q_2, .., q_n representing the quality values;

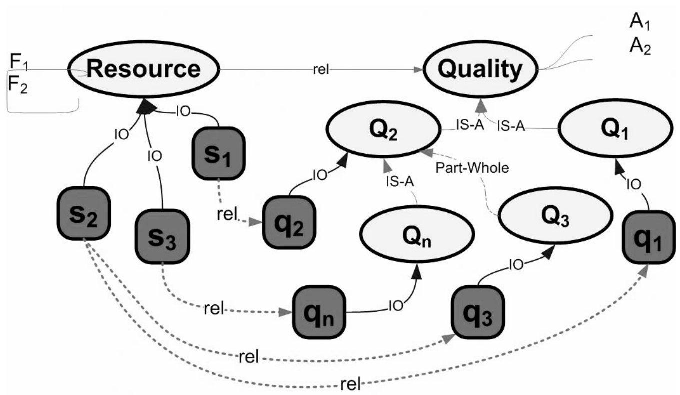


Figure 2 The ontology reference pattern for semantic granularity.

Qualities are represented by classes organised in a hierarchy $\prec_Q$ induced by the relations Is-A and Part-Whole. Qualities are the values of the rel relation; therefore they are represented both as classes and instances to preserve the ontology pattern as specified by OWL-DL (McGuinness & Van Harmelen (2004)). We suppose that a top hierarchy class Q^T exists such that, for each quality Q, $Q \prec_Q Q^T$. Distinct parthood relations can be identified, depending on how the parts differently contribute to the structure of the whole. We restrict the possible interpretations of the relation Part-Whole assuming the parthood among the qualities in $\mathcal{Q}$ adheres to the following properties, as defined by Winston et al. (1987):

1. *transitivity*, i.e., parts of parts are parts of the whole;
2. *reflexivity*, i.e., every part is part of itself;
3. *antisymmetricy*, i.e., nothing is part of own parts;
4. *homeomericity*, i.e., parts are of the same type as their wholes.

Properties 1–3 induce the partial order needed to preserve the hierarchical structure of the qualities. Property 4 instead ensures the parts in the hierarchy are still qualities of the same type as their whole (e.g., a paper topic can be part of another topic). Since the granule labels correspond to qualities in Q, we may define a relationship between labels matching $\prec_Q$ occurrences. For convenience, we use the same notation to denote both relationships.

The method follows a two-step process. The first step, namely *quality filtering*, evaluates each quality with respect to its capability of abstracting information resources. The evaluation of the *abstraction capability of a quality* Q takes into account the attributes and the relations that characterize the resources in *Res* as well as the attributes and the relations of their related instances. It is defined by a ratio which measures the abstraction capability of a quality with respect to the qualities included in its subtree. The higher is the ratio, the higher is the abstraction capability of the quality. The quality filtering returns the qualities with a good value of abstraction capability: the qualities which have the aforementioned ratio higher than a given threshold are promoted to be granules of some granularity (Lin (1995) and Albertoni et al. (2006)).

The second step, namely the *granularity building*, distributes the granules among different granularities according to $\prec_Q$. It returns the set of granularities $\mathcal{G} = < G_1, G_2, ..., G_n >$ to assist the repository browsing. Each granularity G_i groups the resources at increasing levels of detail, i.e., G_{i+1} gives a finer view on the resources in *Res* than G_i (G_{i+1} is said to be finer than G_i). A granularity G_i is extensionally defined by the set of granules that provide a discrete view of the qualities in $\mathcal{Q}$. A generic granule in G_i is a unique textual label that represents a quality and that is semantically meaningful for the user. Each granule identifies a set of resources which share the features identified by the quality. According to this method, all the qualities in $\mathcal{Q}$ resulting from quality filtering become granule labels.

Example 1 *Suppose the user is a researcher who wants to browse a repository of scientific papers organized according to the ontology schema in Figure 3. The schema describes the main entities represented in the repository by the classes Paper, Researcher and Topic; the relation between a paper and its topics as the relation isAbout; the relation between a paper and its authors as the relation hasAuthor; the type of paper (e.g., if it is a conference paper or a journal paper) by the attribute type; its publication date by the attribute date; the authors's affiliation and name by the attributes affiliation and name.*

Figure 4 shows a topic taxonomy excerpt obtained applying the semantic granularity on the repository of scientific papers represented by the schema. The first value in brackets is the abstraction capability of the topic resulting by the quality filtering phase: the lower is the value, the better the topic abstracts its subtopics in the hierarchy. Setting an abstraction threshold of 0,20, Multi-Agent Systems and Ontology Languages are considered as good granules for the browsing, while Semantic Web and Ontology are discarded. By contrast, setting the threshold equals to 0,31, just the quality topic Ontology is discarded.

The second element in the brackets represents the outcome of the granularity building phase: the granularities G_1, G_2, G_3, which correspond to distinct levels of abstraction, are

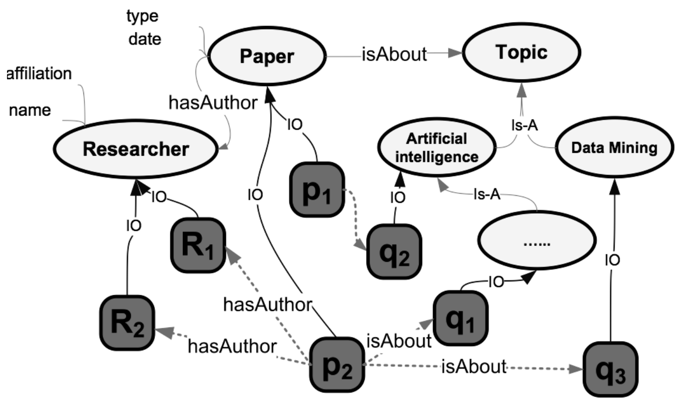


Figure 3 An ontology schema organizing a repository of scientific papers. R_1 and R_2 are two examples of instances of Researcher; p_1 and p_2 are examples of instances of Paper; q_1, q_2 and q_3 are examples of instances of Topic .

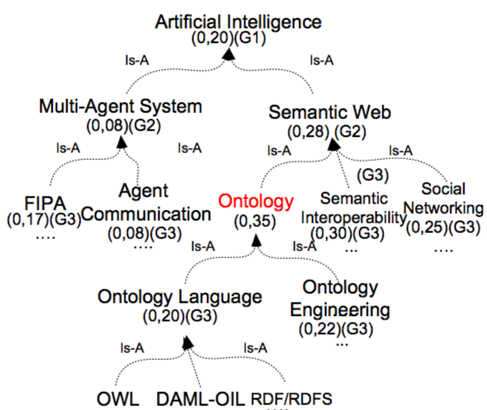


Figure 4 Topic taxonomy and semantic granularity results. For each quality, the values in brackets indicate its abstraction capability and the granularity (G_1,G_2, or G_3) to which it is assigned. The topic *Ontology* does not belong to any granularity because it is discarded by the quality filtering setting the abstraction threshold equals to 0,31.

identified and the granules are associated with them. For example, increasing the level of detail during the browsing activity, Artificial Intelligence belonging to G_1 is converted into Multi-Agent System and Semantic Web, which belong to G_2. Furthermore, Semantic Web is converted into Ontology Language, Ontology Engineering, Semantic Interoperability, and Social Networking belonging to G_3.

5 Context Dependent Parameterization of Semantic Granularity

In this section we formally describe a context dependent parameterization of the semantic granularity method. The method refers to a layered semantic framework inspired by the work of Ehrig et al. (2005), which is structured in terms of *data*, *ontology* and *context* layers plus the *domain knowledge* layer which spans all the others (see Albertoni & De Martino

(2008)). The *data layer* provides the *functions* onto the data type values (e.g., functions which filter the values of simple or complex data types, statistical and user defined functions). The *ontology layer* provides the mechanism for processing semantic granularity by considering the way ontology's entities are related. It provides the implementation of the semantic granularity and of the *operations* (e.g., intersection, count) which may be recalled by the semantic granularity in a given application context. The *context layer* provides the *application contexts*, i.e., the criteria for the computation of semantic granularity considering how ontology entities are used in the given application domain. The *knowledge layer* represents special shared ontology domains, which have their own additional vocabulary. For example, it contains information about the relations of equivalence among terms used in specific knowledge domain.

In the following, first we describe the ontology model we adopted, secondly we formalize the application context and the computation made at the ontology layer.

5.1 Ontology Model

The ontology model formally describes the expressiveness of the ontologies defined according to the semantic framework. Herein, an ontology model is equivalent to an ontology with data types, which is defined as a structure $O := (C, T, \leq_C, R, A, \sigma_R, \sigma_A, \leq_R, \leq_A, I, V, l_C, l_R, l_A)$, where the C, T, R, A, I, V are disjointed sets of classes, data types, binary relations, attributes, instances and data values, respectively, and the following relations and functions are defined:

- $\leq_C$ the partial order on C, which defines the classes hierarchy;
- $\leq_R$ the partial order on R, which defines the relation hierarchy;
- $\leq_A$ the partial order on A, which defines the attribute hierarchy;
- $\sigma_R : R \to C \times C$ the function that provides the signature for each relation;
- $\sigma_A : A \to C \times T$ the function that provides the signature for each attribute;
- $l_C : C \to 2^I$ the function called class instantiation;
- $l_T : T \to 2^V$ the function called data type instantiation;
- $l_R : R \to 2^{I \times I}$ the function called relation instantiation;
- $l_A : A \to 2^{I \times V}$ the function called attribute instantiation.

The ontology model corresponds to a common subset of what is supported by the ontology languages Resource Description Framework (RDF) (Klyne et al. (2004)) and Ontology Web Language (OWL) (McGuinness & Van Harmelen (2004)), which are designed by the World Wide Web Consortium (W3C) as standards for expressing the ontologies in the web. Additionally, we consider the following ontology model functions that retrieve the attributes, the relations and the concepts reachable by a given concept or relation, that are extensively employed in the formalization of the application context and in the granularity parameterization:

- $\delta_a : C \cup R \to 2^A$ where $\delta_a(c) = \{a : A |\ \exists t \in T, \sigma_A(a) = (c, t)\}$ is the set of attributes of c; and $\delta_a(r) = \{a : A |\ \exists c, c' \in C \exists t \in T, \sigma_R(r) = (c, c') \wedge \sigma_A(a) = (c', t)\}$ is the set of attributes of the classes reachable by the relation $r \in R$;

- $\delta_r : C \cup R \rightarrow 2^R$ where $\delta_r(c) = \{r : R | \exists c' \in C, \sigma_R(r) = (c, c')\}$ is the set of relations of c; and $\delta_r(r) = \{r' : R | \exists c \in C, \exists c' \in \delta_c(r); \sigma_R(r') = (c', c)\}$ is the set of relations of the concepts reachable through the relation $r \in R$;
- $\delta_c : C \cup R \rightarrow 2^C$ where $\delta_c(c) = \{c' : C | \exists r \in \delta_r(c); \sigma_R(r) = (c, c')\}$ is the set of concepts related to $c \in C$ through a relation in R; and $\delta_c(r) = \{c' : C | \exists c \in C, \sigma_R(r) = (c, c')\}$ is the set of concepts reachable by the relation $r \in R$;
- $\delta_{r^{-1}} : C \cup R \rightarrow 2^R$ where $\delta_{r^{-1}}(c) = \{r : R | \exists c' \in C, \sigma_R(r) = (c', c)\}$ is the set of relations that reach $c \in C$; and $\delta_{r^{-1}}(r) = \{r' : R | r' \neq r, \exists c \in C, \exists c' \in \delta_c(r), \sigma_R(r') = (c, c')\}$ is the set of relations which differ from r and reaches the concepts reachable through the relation $r \in R$.

5.2 Application Context

In this section, we describe the formalization for the application context we use to parameterize the semantic granularity. The application context is defined by an ontology engineer, according to specific application needs. Alternatively interactive tools supporting the semi-automatic context extraction might be designed. In this paper the context is intended as a specialization of the context to represent information user needs tailored to explicitly parameterize the semantic granularity. Each application context specifies the attributes and the relations to consider as well as the operations and functions to apply on them. Such a definition of context is enough general to parameterize different semantic methods, for example it has been already applied to parameterize the semantic similarity proposed in Albertoni & De Martino (2008) and Albertoni & De Martino (2006). It is not difficult to extend it in order to consider further operations or different ontologies at the same time. Its formalization relies on the concepts of *sequence of elements* and *path of recursion*.

Definition 1 (Sequence of Elements) *Given a set X, a* sequence s of elements *in X with length $n \in N^+$ is defined as a function $s : [1, n] \rightarrow X$. A sequence may be represented as a list of functional values $[s(1).s(2)....s(n)]$.* $\triangle$

$S_X^n = \{s \mid s : [1, n] \rightarrow X\}$ is the set of sequences of X having length n, and $\circ : S_X^n \times S_Y^m \rightarrow S_{X \cup Y}^{n+m}$ is the operator to concatenate two sequences. A *path of recursion* tracks the recursion during the assessment of the semantic granularity, and represents the navigation path in the ontology to collect the information of interest.

Definition 2 (Path of Recursion of length n) *A* path of recursion p *of length n is a sequence of elements with length n whose elements are classes in C and relations in R (i.e., $p \in S_{C \cup R}^n$), such that p starts from a class c and whose other elements are relations either starting from or ending in c or c', where c' is a class involved in some relation in p, that is $p(1) \in C \wedge \forall j \in [2, n]\ p(j) \in R \wedge (p(j) \in \delta_r(p(j-1)) \vee p(j) \in \delta_{r^{-1}}(p(j-1)))$.* $\triangle$

P^n denotes the set of all paths of recursion with length n, whereas P denotes the set of all paths of recursion $P = \bigcup_{n \in N} P^n$.

The *application context (AC)* function is defined inductively according to the length of the path of recursion. It yields the set of attributes and relations to consider and the operations to apply when computing the semantic granularities, e.g., sum, average, minimum, maximum, which could indirectly recall the functions in the data layer, and different forms of count operations: $Count$, which evaluates the cardinality of a set of instances; $WCount$, which evaluates a weighted count of instances according to the cardinality of related attributes or relations; $InvCount$, which evaluates the inverse cardinality of a set of instances, (i.e., a set with less instances has more importance than a set with greater cardinality). The application context is formally defined as follows.

Definition 3 (Application Context AC) *Given the set P of paths of recursion, L the set of operations provided by the ontology layer (i.e. Count, WCount and InvCount for the semantic granularity), G the set of datatype functions available in the data layer, the application context for the semantic granularity is defined by the partial function AC $: P \rightarrow 2^{A \times (L \cup G)} \times 2^{R \times L}$.* $\triangle$

Note that each application context AC is characterized by the operators AC_A: $P \rightarrow 2^{A \times (L \cup G)}$ and AC_R: $P \rightarrow 2^{R \times L}$, which yield respectively the context AC related to the attributes and to the relations.

For each path of recursion p, $AC(p) = (AC_A(p), AC_R(p))$ represents the portion of application context related to the path of recursion p. In particular, $AC_A(p) = \{(a_1, opg_1), (a_2, opg_2), ..., (a_l, opg_l)\}$ and $AC_R(p) = \{(r_1, op_1), (r_2, op_2), ..., (r_k, op_k)\}$, with $j \in N^+$, $opg_j \in L \cup G$, and $op_j \in L$, are, respectively, the attributes and the relations with the corresponding operations to use when computing the semantic granularity against the path of recursion p.

Example 2 *Considering the ontology schema in Figure 3, two examples of application contexts AC_1 and AC_2 are defined.*

AC_1 corresponds to the hard coded context implicitly used in Example 1. It starts from the path of recursion $[Topic]$ and considers the instances of Paper associated with each Topic to calculate the capability of abstraction. It is formalised as follows:

$[Topic] \overset{AC_1}{\rightarrow} \{\{\phi\}, \{(isAbout^{-1}, Count)\}\}$

AC_2 considers the date of publication, the number of authors, the type (i.e., journal, conference proceedings, or book) of papers. It is formalised as follows:

$[Topic] \overset{AC_2}{\rightarrow} \{\{\phi\}, \{(isAbout^{-1}, WCount)\}\}$

$$[Topic.isAbout^{-1}] \overset{AC_2}{\rightarrow} \{\{(type, i(Paper, Book))(date, g(today))\} \\ \{(hasAuthor, InvCount)\}\}$$

AC_2 starts from the path of recursion $[Topic]$ and moves along to the inverse of relation isAbout to focus on the attributes and relations of Paper. The change of focus is tracked by the path of recursion $[Topic.isAbout^{-1}]$. AC_2, when applied to the new path of recursion, returns the attributes type and date and the relation hasAuthor. Type and date are processed respectively by the data layer functions i and g: $i(Paper, Book)$ returns the inverse of the cardinality of the papers associated with a given topic that are published in a book or a journal, whereas $g(today)$ counts only the papers published in the last three years. Finally, the inverse cardinality of the relation hasAuthor is considered. $\square$

5.3 Semantic Granularity Computation

The semantic granularity computation is a two-step process: namely *quality filtering* and *granularity building*. We first illustrate the parameterization of the *quality filtering* phase and, secondly, of the *granularity building*. Finally an illustrative example of the method application is provided.

5.3.1 Context Dependent Quality Filtering

As we discussed in Section 4, the *quality filtering* phase selects the quality values that are good candidates to be adopted as granule labels. In the proposed approach, a quality value is considered as a granule label whenever it ensures a good *abstraction capability* or it is involved in a Part-Whole. The level of abstraction of a resource quality is worked out according to four cognitive principles, which express the propension of a concept in a taxonomy to subsume other concepts:

1. *Quality children balance:* the more the tree rooted in the quality is balanced, the bigger is its abstraction capability.

2. *Quality cardinality:* the higher the quality cardinality is with respect to its sub-qualities, the bigger is its abstraction capability.

3. *Number of balanced children:* The higher is the number of the quality's children in a balanced tree, the bigger is the quality's abstraction capability.

4. *Non-locality:* The cardinality of non immediate children of a quality (i.e., grand children, great grandchildren and so on) affects its abstraction capability.

The approach we adopt to evaluate the abstraction capability of qualities originates from existing techniques applied in the area of Natural Language Processing for topic identification and generalization by Lin (1995). Lin introduced the notion of *degree of informativeness and summarization* of a concept C in a lexical taxonomy as a measure of the capability of C to generalize its specializations, i.e., the children in the taxonomy, according to the terms occurrence in a corpus. According to Lin, the more the children of C have a closer number of occurrences, the more the concept C is a good generalization. The Lin's algorithm is based on the assumption that the occurrences of the corpus are associated only to the leaves in the noun taxonomy. Pike & Gahegan (2003) extend the Lin's approach to identify and to abstract the arguments of a discourse allowing the intermediate concepts of the taxonomy to have their own occurrences associated.

In Albertoni et al. (2006) we extend such work in order to consider full ontologies instead of lexical taxonomies; therefore, complex relationships among qualities and among resources and qualities may be considered to evaluate their degree of informativeness. In Albertoni et al. (2006) we consider also quality aggregations, i.e., Part-Whole structures, differently from Lin (1995) and Pike & Gahegan (2003) that consider only the relation Is-A for structuring concepts. Furthermore, we modify the filtering algorithm defined by Pike & Gahegan (2003) in order to take into consideration also the influence of the occurrences of a quality in the evaluation of its abstraction capability. In this work, we further exploit general ontology relationships, as well as the application of data layer functions, in order to parameterize the quality filtering phase according to the user preferences and the given application.

Definition 4 expresses the capability of a quality Q to abstract its direct sub-qualities defined with respect to $\prec_Q$ (including, as we specified above, both the qualities reachable through Is-A and Part-Whole). Specifically, Q has a good abstraction capability if the ratio between the maximum number of resources in the repository associated to Q or to one of its sub-qualities is less than a given a threshold R_t. The ratio value ranges in [0,1]. We assume that the leafs in the quality hierarchy have ratio equal to 0.

Definition 4 (Abstraction capability of a quality Q**)** *Let* $Q \in \mathcal{Q}$ *be an information resource quality, such that* $\prec_Q$ *holds on* $\mathcal{Q}$*, and let* Q^T *be the most generic quality in* $\mathcal{Q}$ *according to* $\prec_Q$*. Then, given* p *a starting path of recursion initialized as* $p = [Q^T]$*, the abstraction capability of* Q *with respect to the application context* AC *and the path of recursion* p*, denoted by* $R_Q^{AC,p}$*, is defined as follows:*

$$R_Q^{AC,p} = \frac{\sum_{x \in \{AC_A(p) \cup AC_R(p)\}} R_Q^{AC,p,x}}{|AC_A(p)| + |AC_R(p)|}. \tag{1}$$

△

$R_Q^{AC,p,x}$ in Equation (1) is the abstraction capability of Q according to the relation or attribute x that appears in the application context AC for the path of recursion p. It is defined as follows:

$$R_Q^{AC,p,x} = \begin{cases} R_Q^{AC,p\circ x} & \text{if (x,}WCount\text{)} \in AC_R(p) \\ \frac{\max_{\{Q'|Q' \prec_Q Q\}} s_{AC,p,x}^{Q'*}}{\sum_{\{Q'|Q' \prec_Q Q\}} s_{AC,p,x}^{Q'*} + s_{AC,p,x}^{Q}} & \text{otherwise.} \end{cases} \tag{2}$$

In Equation (2), for each relation x in the context whose associated operation is $WCount$, $R_Q^{AC,p}$ is defined recursively considering the instances related to Q by the path of recursion $p \circ x$. Therefore, the abstraction capability of Q is evaluated considering also the relationships and attributes belonging to instances that are not directly associated to Q. Otherwise, when the context does not prescribe a recursive assessment, the abstraction capability is parameterized according to the context defining $s_{AC,p,x}^{Q}$ and $s_{AC,p,x}^{Q^*}$ as follows:

$$s_{AC,p,x}^{Q} = \sum_{q \in Q} \sum_{\iota \in I(q,p)} f_{AC}^{p,x}(\iota) \qquad s_{AC,p,x}^{Q^*} = \sum_{\{Q'|Q' \prec_Q Q\}} s_{AC,x}^{Q'}.$$

$f_{AC}^{p,x}(\iota)$ measures the *weight* of the instance q of Q according to the relation or attribute x belonging to the set of instances $I(q,p)$, which are reachable through the recursion path p, and considering the operations indicated in AC.

Assuming: (i) X a placeholder that works as a metasymbol to replace with R or A, whether x is respectively a relation or an attribute; (ii) $i_A(\iota, a) = \{v \in V \mid (\iota, v) \in l_A(a), \exists y \in C \; s.t. \; \sigma_A(a) = (y, T) \wedge l_T(T) = 2^V\}$ the set of values assumed by the instance ι for attribute a; (iii) $i_R(\iota, r) = \{\iota' \in l_c(c') \mid \exists c \; \iota \in l_c(c) \; \exists \; c' \; s.t. \; \sigma_R(r) \in (c, c') \wedge (\iota, \iota') \in l_R(r)\}$ the set of instances related to the instance ι by relation r; (iiii) g a function provided by the data layer, w the metasymbol that works as placeholder for the function parameters that have been already fixed in the application contexts; $f_{AC}^{p,x}(\iota)$ is defined as follows:

$$f_{AC}^{p,x}(\iota) = \begin{cases} g(w) & \text{if} (x, g(w)) \in AC_A(p), v \in i_A(\iota, x) \\ |i_X(\iota, x)| & \text{if } (x, \text{Count}) \in AC_X(p) \\ \frac{1}{|i_X(\iota, x)| + 1} & \text{if } (x, InvCount) \in AC_X(p) \end{cases} \tag{3}$$

5.3.2 Granularity Building

The *granularity building* phase distributes the granules selected by the *quality filtering* phase among the semantic granularities in $\mathcal{G}$. The partial order $\prec_Q$ induced by the relations Is-A and Part-Whole onto $\mathcal{Q}$ leads such a distribution. In particular, considering two distinct granule labels a and b, the principles applied are:

1. *Partial order exploitation:* if $b \prec_Q a$ then the two granules have to belong to distinct granularities;

2. *Part-Whole exception:* if $b \prec_Q a$, b Part-Whole a holds and the whole granule with label a belongs to the granularity G_i, then the part granule with label b has to belong to the granularity G_{i+1}, such that G_{i+1} is finer than G_i.

The granularity building phase is performed according to the algorithm in Figure 5. It returns the sequences of granularities $\mathcal{G}$, starting from a context AC defined according to Definition 3, being Q^T the top hierarchy class of qualities, R_t is the ratio threshold for the evaluation of the degree of informativeness of qualities. It performs a breath first visit of $\prec_Q$, inserting the granules in distinct granularities according to the aforementioned principles. It terminates whenever the visit has reached all the $\prec_Q$ leaves. In Figure 5, ds is the starting level (from the root) in $\prec_Q$; next(Q) returns the children of Q in $\prec_Q$; node(l) returns the qualities laying at the level l in the hierarchy; $+$ and $-$ are the set operators for union and difference.

```
i = 0; G_i ={};
  NodeToConsider= Node(ds);
  While (!  empty(NodeToConsider)) {
   For each Q belonging to NodeToConsider {
    If (R_Q^{AC,Q^T} <R_t) G_i += { Q };
    else If Q is not leaf NodeToConsider += next(Q);
    NodeToConsider -= { Q };
   }
   i++; G_i = {};
   For each Q belonging to G_i - 1 {
    For j = 1 to i - 1 {  check for multi-inheritance
     For each Q_1 belonging to G_j
      If ((Q_1 part-whole Q) or (Q_1 is-a Q)){
       G_j -= { Q_1 }; G_i+= { Q_1 };
      }
      If Q is not leaf
       NodeToConsider += next(Q);
      }
     }
    }
```

Figure 5 Granularity Building Algorithm

5.4 Context Enabled Semantic Granularity: Application Example

Suppose a user needs to browse a repository of scientific papers with two different purposes: (1) to get an overview of the repository content and (2) to identify the hottest topics in the Computer Science research. These two aims correspond to two distinct contexts. In the former, the user is interested in browsing the repository considering the principal scientific topics, thus it is reasonable he considers the number of papers associated with each topic. In the latter, he wants to find out the most promising topics (e.g., to plan his upcoming research activity). In this context he could focus only on the topics on which few authors have been working on, namely papers with recent publications, not yet mature to appear in scientific journals and books.

The test is based on real data extracted from Faceted Dblp, a repository of Computer Science papers[a], and organized in the ontology of Figure 4. The two contexts have been formalized respectively by the functions AC_1 and AC_2 given in the Example 2.

The semantic granularity has been applied and a fragment of the result is illustrated in Figure 6. Different granularities are obtained considering the contexts AC_1 and AC_2: comparing Figure 6(a) and Figure 6(b) granularities G2 and G3 contain different granules, i.e., topics. Indeed, the filtering phase results in different abstraction capabilities for the topics, according the two contexts. For example, *Semantic Web* disappears in Figure 6(b), as it has been discarded by the filtering, whereas *Ontology* increases its importance: moving from AC_1 to AC_2, *Semantic Web* decreases its abstraction capability as R_Q increases from 0.28 to 0.35, whereas *Ontology* increases its importance as R_Q decreases from 0.35 to 0.31. Adopting the threshold 0.31, the granularity building phase returns the granularities depicted in Fig 6.

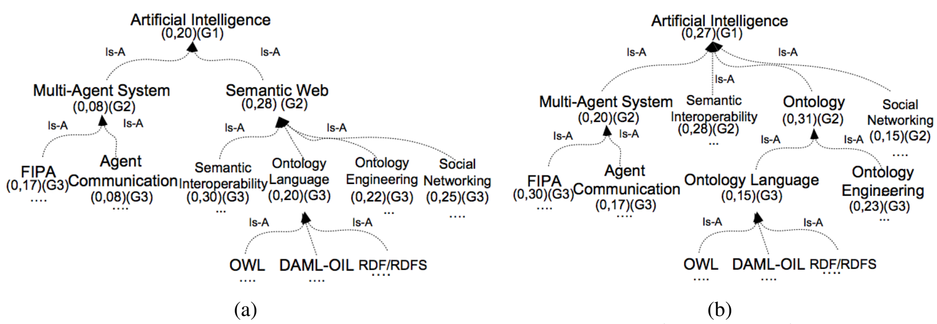


Figure 6 Results of semantic granularity: according to contexts AC_1 (a) and AC_2 (b).

6 Validation and discussion

In the following we illustrate the results of a preliminary evaluation we undertook to validate the cognitive principles at the base of the semantic granularity method. In particular, we consider: (i) the characterization of good abstractor we adopted to select the qualities during the quality filtering (see the four cognitive principles in section 5.3.1); (ii) the Partial order exploitation and the Part-whole exception to assigns the granules to the proper granularities during granularity building (see section 5.3.2).

During the experiment, each participant was asked to play the role of a user that browses for the first time different repositories of scientific books. We provided the user

[a]Faced Dblp is available at http://dblp.l3s.de/.

with graphic representations of these repositories, where the books are grouped according to their topics organized in simple taxonomies, as represented in Figures 7-11, where:

- the labels in the circles represent the book topics;

- the number beside each topic represents the number of books in the repository which are classified with that topic: e.g., considering Figure 7, Question 1 and the diagram on the left, 33 books are about "Maths", 33 books are about "Physic", 33 books are about "Computer science" and 1 is about "Science". The total number of available books in the repository is the sum of these numbers;

- the arrows represent topic specializations, i.e., the IS-A relation that identifies subtopics (e.g., "Maths" is considered as a subtopic of "Science").

Different diagrams, representing the same taxonomy but containing different resources, simulate different resource organizations. We have defined a questionnaire focusing on situations that are significant for testing the aforementioned principles.

The questionnaire includes 11 questions to explicit the user perception: some questions address the validation of a specific principle with a direct approach; others assess the relative relevance of a principle against to another one. The user was invited to compare diagrams representing two repositories organized according to the same taxonomy but containing different books, then to indicate in which of the two repositories a specific topic (i.e., "Science") better/worst represents the repository content. Figures 7 to 11 illustrate the diagram comparisons proposed in the questionnaires.

The questionnaire is realized with the SurveyMonkey (http://www.surveymonkey.com), which allows to design and distribute questionnaires through the web collecting and analyzing the answers. The questionnaire has been distributed to a set of potential users with mathematic and computer science background; 22 persons participated actively answering to most of the questions. The outcome of the answer analysis are very encouraging and show that the assumptions we made in the quality filtering are coherent with the way the majority of the users perceives an abstractor as good.

In the following we illustrate each proposed question, highlighting its aim and the gathered results. The abstraction capability (AC) is calculated applying the formula (2) without considering the context to avoid perceptual misleads concerning the context during the validation of semantic granularity principles. Disregarding the context, the abstraction capability of a root node in a tree (formula (2)) corresponds to the ratio between the maximum cardinality of its children and the overall tree cardinality.

- Question 1 is aimed at the validation of the assumption concerning the *Quality's children balance*. 86,4% of the responders has answered Rep 1: it indicates that "science" is perceived as a better abstractor when its children are balanced and it conforms the AC we calculate (i.e. $33/100$ for Rep 1 $<$ $97/100$ for Rep 2).

- Question 2 is intended to validate the assumption concerning the *Quality cardinality*. 63,6% of responders has answered Rep 1: it indicates that when the subtopics are equally balanced the level of abstraction provided by a topic also depends on its cardinality: the more are the book referring to the "science" the more this topic is perceived as a good abstractor. These results conform the AC we calculate (i.e. $33/109$ for Rep 1 $>$ $33/199$ for Rep 2).

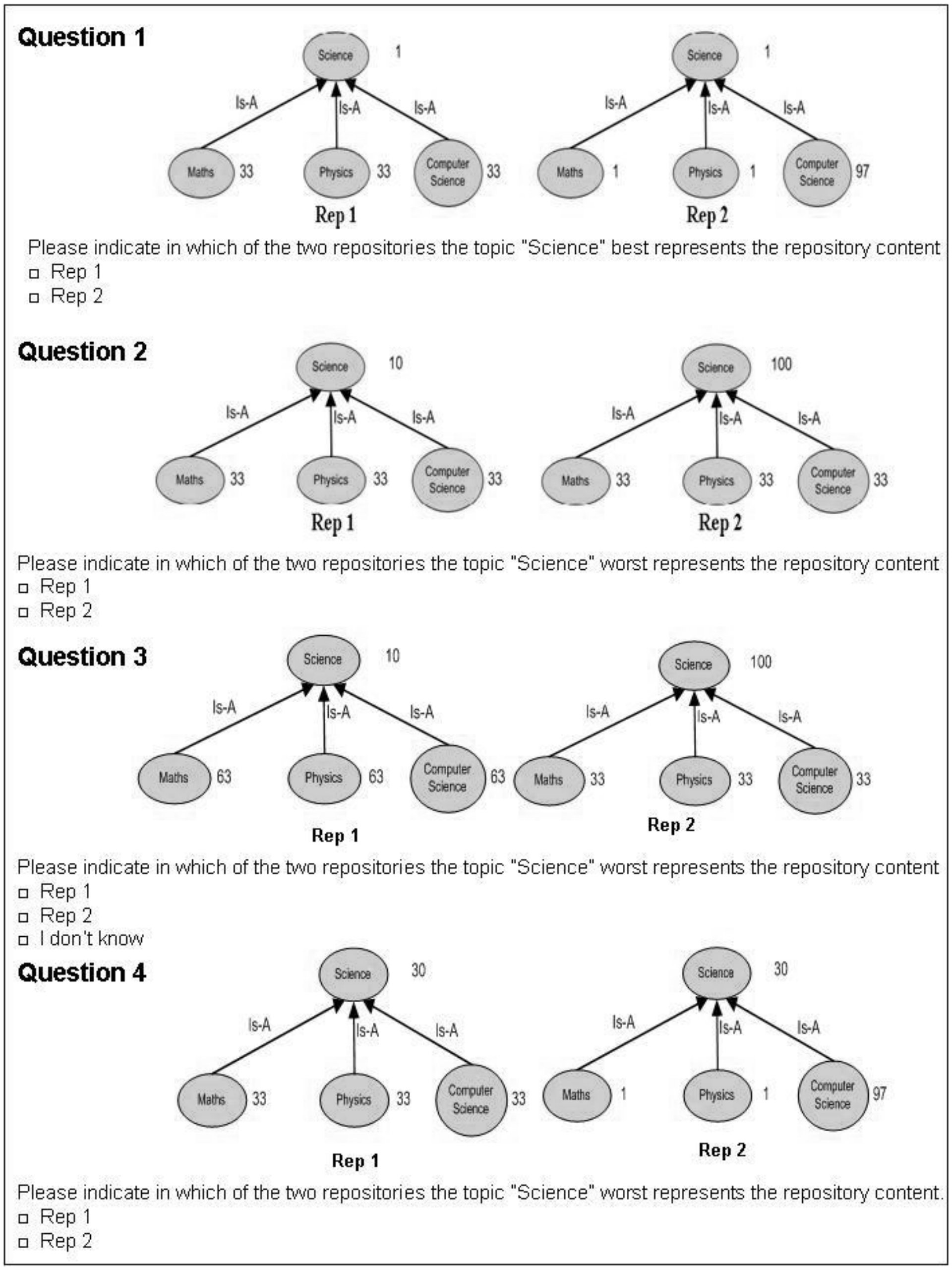


Figure 7 Questions 1- 4 of the questionnaire used for the validation of the cognitive principles.

- Question 3, analogously to Question 2, is aimed at the validation of the assumption concerning the *Quality cardinality*. 59,1% of the responders has answered Rep 1; it indicates that when the subtopics are equally balanced, the level of abstraction provided by a topic also depends on its cardinality even if the repositories have the same number of books. These results conform the AC we calculate (i.e. 63/199 for Rep 1 > 33/199 for Rep 2).

- Question 4, exactly as Question 1, is aimed at the validation of the assumption concerning the *Quality's children balance*. 85,5% of the responders has indicated that Rep 2 is a worst abstractor, thus confirming that children balance highly influences the ancestor abstractor capability. These results conform the AC we calculate (i.e. 33/129 for Rep 1 < 97/129 for Rep 2).

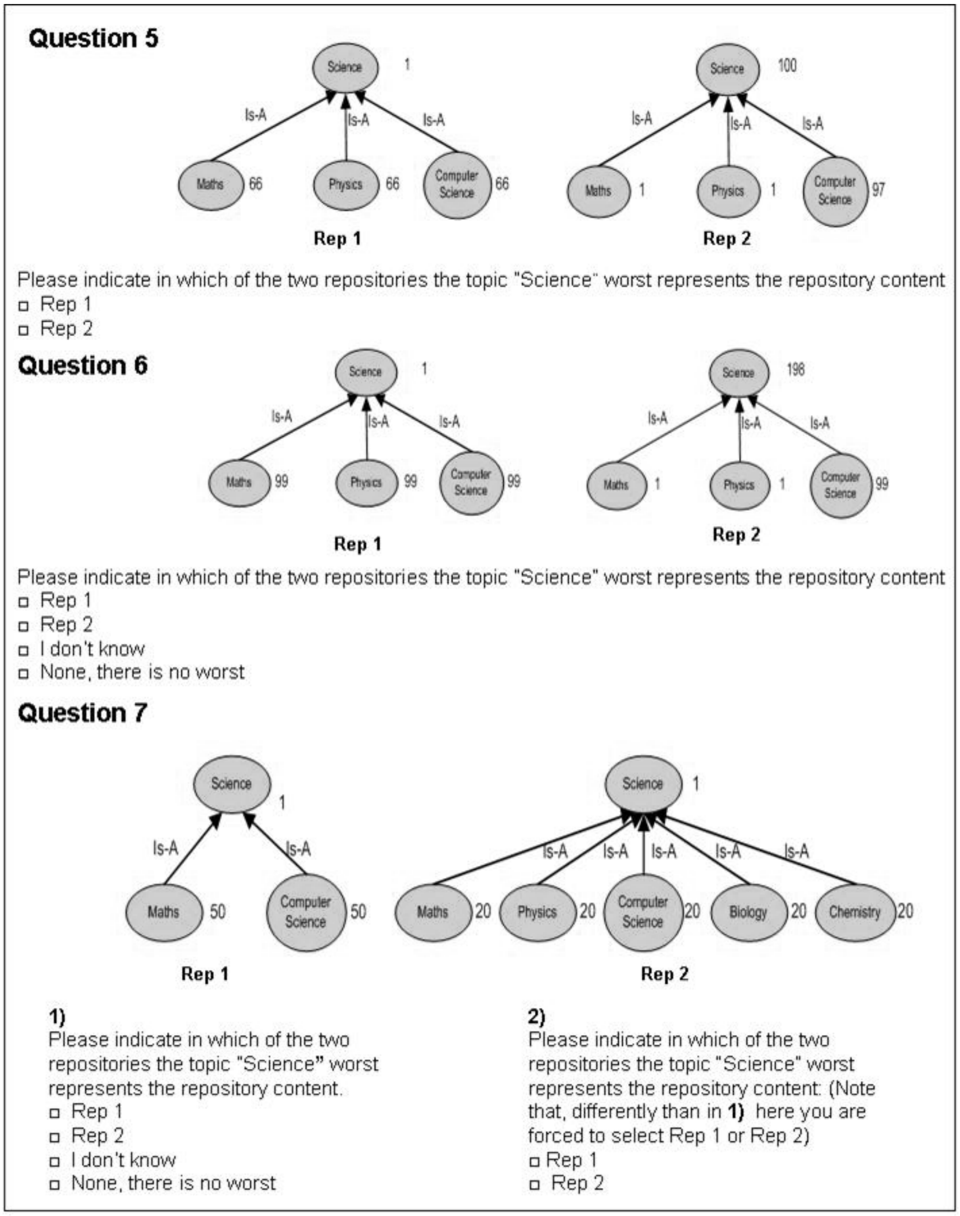


Figure 8 Questions 5- 7 of the questionnaire used for the validation of the cognitive principles.

- Question 5 and 6 are both intended to understand if the ancestor abstraction capability is most influenced by the *children balance* or by the *abstractor cardinality*: the difference is that for Question 6 the distribution in the second repository is much more unbalanced. To Question 5 59,1% of responders has indicated that Rep 2 is the worst abstractor, thus suggesting the major weight of the children balance and it conforms the AC we calculate (i.e. 66/199 for Rep 1 $<$ 97/199 for Rep 2). In Question 6 the answer distribution (27,3% indicated Rep1, 40,9% Rep 2 and 31,8% answered that there's no worst representative) indicate that in this extreme situation there's no meaningful difference in the user perception, conforming the AC we calculate(i.e. 99/299 for Rep 1 = 99/299 for Rep 2).

- Question 7 is aimed at the validation of the assumption concerning the *Number of children*; The user is asked to answer the question twice, but the second time he/she is forced to indicate one of the two repositories. In the first case the answer distribution (40,9% indicated, as worst abstractor, Rep 1 and 45,5 answered that there was no worst abstractor) indicates that Rep 2 is a better abstractor. In this case there's no a clear majority. However in the second case, 68,4% of the responders has answered Rep 1 and 31,6% answered Rep 2: it more strongly suggests that when the topic has more children it is perceived as a better abstractor. Results conform the AC we calculate (i.e. 50/101 for Rep 1 $>$ 20/101 for Rep 2).

- Question 8 and 9 are both aimed at the validation of the assumption concerning the *Non-locality*. To the Question 8 50% of responders answered Rep 2, conforming the AC we calculate (i.e. 75/145 for Rep 1 $<$ 137/145 for Rep 2). For Question 9, 66,7% of the responders indicated Rep 2. This confirms that the majority of the users perceive the balance not as a local property, on the contrary they consider the balance of the whole sub-tree. It conforms the AC we calculate (i.e. 75/145 for Rep 1 $<$ 137/145 for Rep 2).

- Finally Question 10 is aimed at addressing the two assumptions made in Section 5.3.2: the principle of *Partial order exploitation* makes use of the relationship $\prec_Q$ to distribute the granules resulting after the quality filtering; thus it is justified by the fact that IS-A is defined as a partial order. The principle of *Part-Whole exception* sets the granules involved in a Part-Whole as granules belonging to distinct and contiguous granularities. Unfortunately the answers distribution to this question did not provide sound indications.

Summarizing, considering the responses to the questions concerning the validation of the *Quality's children balance*, we can say that about 90% of responders indicate that a topic is perceived as a better abstractor when its children are balanced. Concerning the *Quality cardinality*, about 60% of responders indicate that when the subtopics are equally balanced the level of abstraction provided by the topic also depends on its cardinality: e.g. the more are the books referring to the topic, the more this topic is perceived as a good abstractor. Concerning the *Number of balanced children*, about 75% of responders indicate that when a topic has more children it is perceived as a better abstractor. Analogously for the *Non-locality principle*, the questionnaire results indicate that the majority of the users (about 70%) does not perceive the balance as a local property; in other words, it does not depend only on the immediate children of the topic, but also on the subnodes. The principle of *Partial order exploitation* is mathematically justified by the assumption of partial order

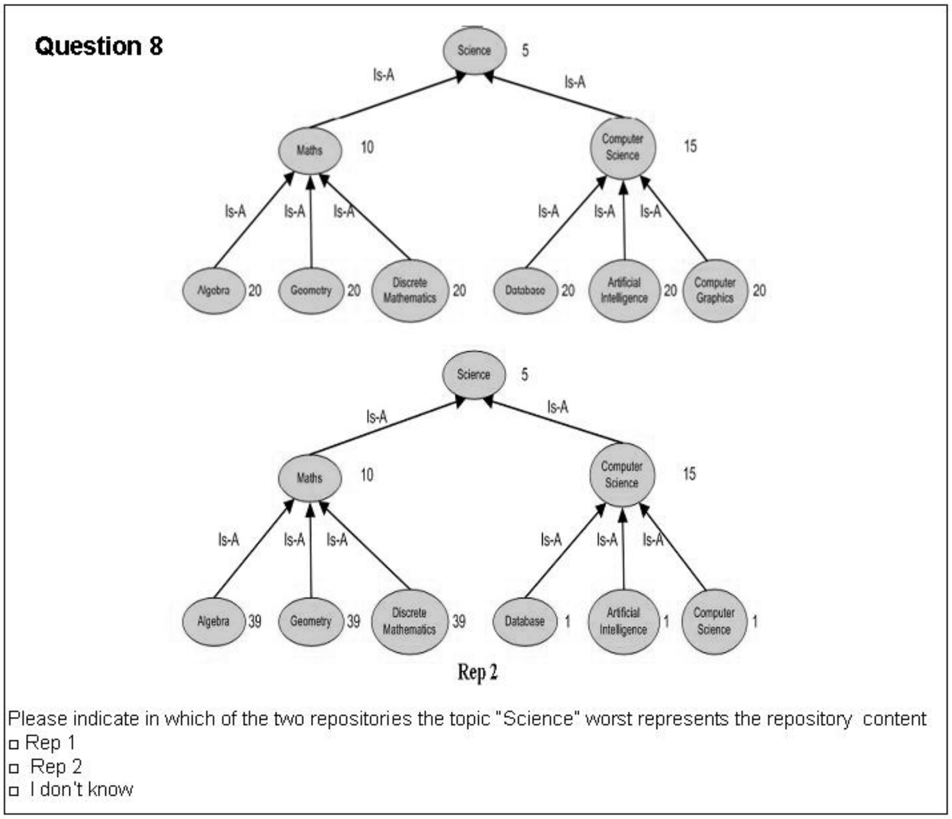


Figure 9 Question 8 of the questionnaire used for the validation of the cognitive principles.

we did on IS-A and PART-WHOLE. On the contrary for the *Part-Whole exception* the experimentation cannot be considered satisfying and further investigation are needed.

The entire questionnaire is available at http://purl.oclc.org/NET/SGQuestionnaire.pdf and the summary of the results at http://purl.oclc.org/NET/SGSurveySummary.pdf.

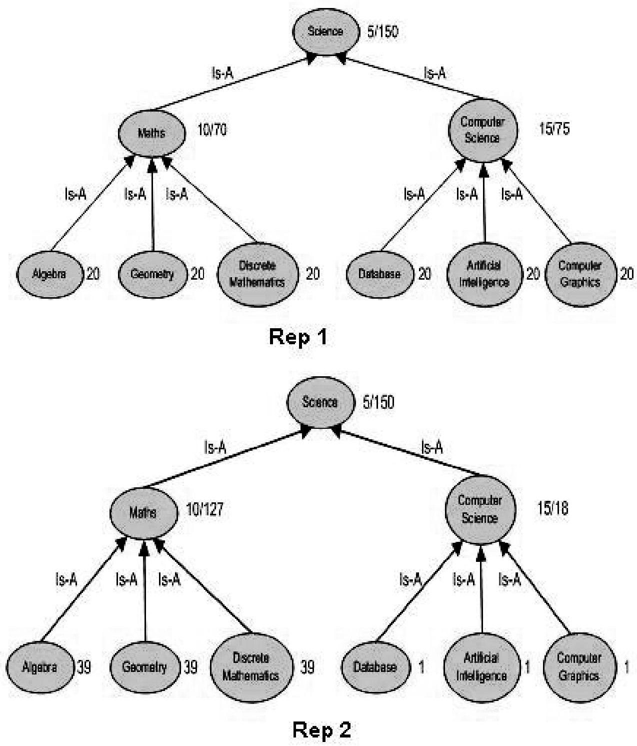
7 Conclusions and Future Works

In this paper we have described a context dependent parameterization of semantic granularity to browse any kind of information resources represented in ontologies. The method extends our previous research on semantic granularity (Albertoni et al. (2006) and Albertoni et al. (2008)) and originated from the application context formalization for semantic similarity presented by Albertoni & De Martino (2008). Here we provide a complete description of context dependent semantic granularity method, in particular:

- an application context formalism for the granularity: we define new operations and functions to be adopted for the analysis of ontology entities;

- a flexible computation of the semantic granularity throughout its context dependent parameterization;

Question 9



Please indicate, considering the effect of the IS-A, in which of the two repositories the topic "Science" worst represents the repository content:

- ☐ Rep 1
- ☐ Rep 2
- ☐ I can't answer to this question since I don't understand the effect of the formal IS-A

Please indicate if you would modify the answer you gave in question 8?

- ☐ Yes
- ☐ No

Figure 10 Question 9 of the questionnaire used for the validation of the cognitive principles

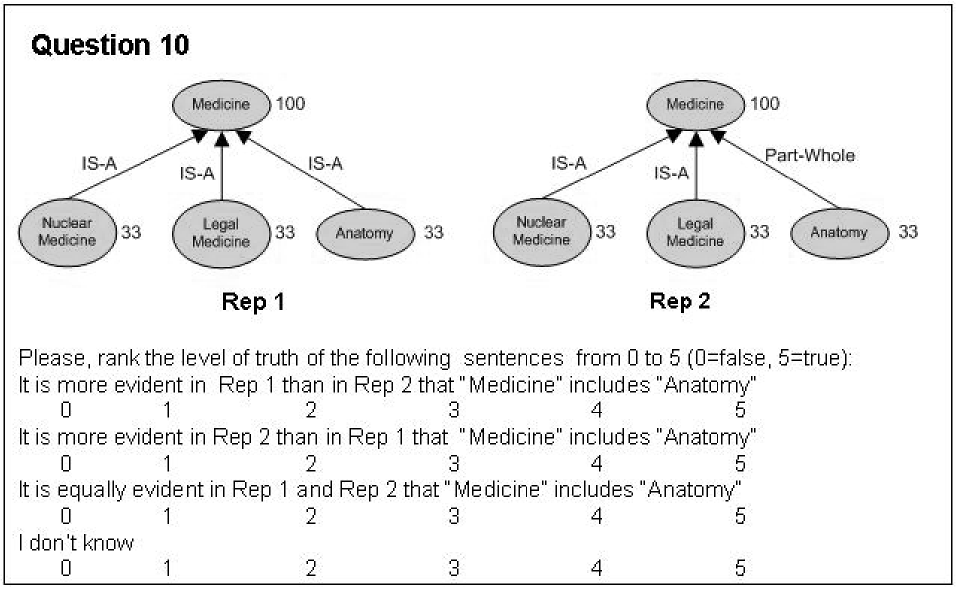


Figure 11 Question 10 of the questionnaire used for the validation of the cognitive principles

- a validation of the cognitive principles which are behind the defined method.

The achieved results indicate the validity of the undertaken approach to the definition of a powerful ontology driven method that eases the browsing of a repository of information resources, allowing a user-oriented formulation of the granularity criteria induced by the context.
We have provided an example of application of the semantic granularity according to two different contexts, which demonstrates the benefits of explicitly adopting contexts to manage different application scenario. We have validated the main cognitive assumptions which are behind the quality filtering and the granularity building phases. However, we did not succeed in the validation of the principle which set the granules involved in a Part-Whole as granules belonging to distinct and contiguous granularities. We think that this principle holds only in specific application contexts; in the other cases, the Part-Whole might be handled like the IS-A, applying the quality filtering also to the granules involved in the Part-Whole relations. However further investigations are needed.

In the future we are going to investigate how to exploit linked data technology to empower this method applicability. Moreover we will deeper analyse the relationships between similarity and granularity in order to get an integrated method which fully exploit semantics to improve data accessibility.

References

Albertoni, R. & De Martino, M. (2006), Semantic similarity of ontology instances tailored on the application context, *in* R. Meersman & Z. Tari, eds, '5^{th} International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 06)', Vol. 4275 of *LNCS*, Springer-Verlag Berlin Heidelberg, pp. 1020–1038.

Albertoni, R. & De Martino, M. (2008), ‘Asymmetric and context-dependent semantic similarity among ontology instances’, *J. Data Semantics* **10**, 1–30.

Albertoni, R., Camossi, E., De Martino, M., Giannini, F. & Monti, M. (2006), Semantic granularity for the semantic web, *in* R. Meersman, Z. Tari & P. Herrero, eds, ‘2^{nd} International IFIP Workshop On Semantic Web & Web Semantics (SWWS ’06)’, Vol. 4278 of *LNCS*, Springer-Verlag Berlin Heidelberg, pp. 1863–1872.

Albertoni, R., Camossi, E., De Martino, M., Giannini, F. & Monti, M. (2008), Context enabled semantic granularity, *in* I. Lovrek, R. J. Howlett & L. C. Jain, eds, ‘KES (2)’, Vol. 5178 of *Lecture Notes in Computer Science*, Springer, pp. 682–688.

Aleman-Meza, B., Halaschek, C., Arpinar, I. B. & Sheth, A. (2003), Context-aware semantic association ranking, *in* ‘International Conference on Semantic Web and Databases’.

Bertino, E., Camossi, E. & Bertolotto, M. (2009), Multi-granular spatio-temporal object models: Concepts and research directions, *in* ‘2^{nd} International Conference on Object Databases (OODB09), 1-3 July 2009 ETH Zurich’, ETH Zurich, to be published by Springer-Verlag Berlin Heidelberg, pp. 131–148.

Bittner, T. & Stell, J. G. (2003), Stratified rough sets and vagueness, *in* Kuhn et al. (2003), pp. 270–286.

Bouquet, P., Giunchiglia, F., Harmelen, F. V., Serafini, L. & Stuckenschmidt, H. (2004), ‘Contextualizing ontologies’, *Journal of Web Semantics* **1**, 325–343.

Ehrig, M., Haase, P., Hefke, M. & Stojanovic, N. (2005), Similarity for ontologies - a comprehensive framework., *in* ‘Proceedings of the 13th European Conference on Information Systems, Information Systems in a Rapidly Changing Economy, ECIS 2005,’.

Fonseca, F. T., Egenhofer, M. J., Davis, C. A. & Camara, G. (2002), ‘Semantic granularity in ontology-driven geographic information systems’, *Annals of Mathematics and Artificial Intelligence, Special Issue on Spatial and Temporal Granularity* **36**(1-2), 121–151.

Guha, R. V. (1990), Micro-theories and contexts in cyc part i: Basic issues., Technical Report ACT-CYC-129-90, Microelectronics and Computer Technology Corporation, Austin, TX.

Hernandez, N., Mothe, J., Chrisment, C. & Egret, D. (2007), ‘Modeling context through domain ontologies’, *Information Retrieval* **10**(2), 143–172.

Kashyap, V. & Sheth, A. (1996), ‘Semantic and schematic similarities between database objects: a context-based approach’, *The VLDB Journal* **5**, 276304.

Kashyap, V. & Sheth, A. (1998), Semantic similarity between objects in multiple databases, *in* ‘Management of heterogeneous and autonomous database systems, Morgan Kaufmann Series In Data Management Systems’, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, pp. 57–89.

Khatri, V., Ram, S., Snodgrass, R. T. & O’Brien, G. M. (2002), ‘Supporting user-defined granularities in a spatiotemporal conceptual model’, *Annals of Mathematics and Artificial Intelligence* **36**(1-2), 195–232.

Klyne, G., Carroll, J. J. & McBride, B. (2004), 'Resource description framework (RDF): Concepts and abstract syntax', *W3C recommendation*.

Krowne, A. & Halbert, M. (2005), An initial evaluation of automated organization for digital library browsing, *in* 'JCDL '05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries', ACM, New York, NY, USA, pp. 246–255.

Kuhn, W., Worboys, M. F. & Timpf, S., eds (2003), *Spatial Information Theory. Foundations of Geographic Information Science, International Conference, COSIT 2003, Ittingen, Switzerland, September 24-28, 2003, Proceedings*, Vol. 2825 of *Lecture Notes in Computer Science*, Springer.

Lin, C.-Y. (1995), 'Knowledge-based automatic topic identification', *Proc. of the 33rd Annual Meeting on Association for Computational Linguistics* pp. 308–310.

McGuinness, D. L. & Van Harmelen, F. (2004), 'OWL web ontology language overview', *W3C recommendation* **10**, 2004–03.

Newby, G. B. (2001), 'Cognitive space and information space', *J. Am. Soc. Inf. Sci. Technol.* **52**(12), 1026–1048.

Noy, N. F. & Musen, M. A. (2004), Specifying ontology views by traversal, *in* '3rd International Semantic Web Conference', Vol. 3298 of *LNCS*, Springer-Verlag Berlin Heidelberg, pp. 713–725.

Ouksel, A. M. & Naiman, C. F. (1993), 'Coordinating context build-ing in heterogeneous information systems', *Journal of Intelligent Information Systems* **3**, 151183.

Petit, M., Ray, C. & Claramunt, C. (2008), 'An adaptive interaction architecture for collaborative gis', *Cartographic & Geographic Inf. Science* **35**, 91–102.

Pike, W. & Gahegan, M. (2003), Constructing semantically scalable cognitive spaces, *in* Kuhn et al. (2003), pp. 332–348.

Sciore, E., Siegel, M. & Rosenthal, A. (1992), Context interchange using meta-attributes, *in* 'Proceedings of the CIKM', ACM.

Segev, A. & Gal, A. (2005), Putting things in context: A topological approach to mapping contexts and ontologies, *in* '20th National Conference on Artificial Intelligence (AAAI-05), 1st International Workshop on Context and Ontologies: Theory, Practice and Applications'.

Sheth, A. & Larson, J. (1990), 'Federated database systems for managing distributed, heterogeneous and autonomous databases', *ACM Computing Survey* **22**, 183236.

Sicilia, M. A. (2006), 'Metadata, semantics, and ontology: providing meaning to information resources', *International Journal of Metadata, Semantics and Ontologies*.

Volz, R., Oberle, D. & Studer, R. (2003), Implementing views for light-weight web ontologies, *in* 'IEEE Database Engineering and Application Symposium (IDEAS)'.

Weakliam, J., Wilson, D. & Bertolotto, M. (2008), Personalising map feature content for mobile map users, *in* 'Map-based Mobile Services Design, Interactivity and Usability', Geo-Information and Cartography, Springer-Verlag Berlin Heidelberg.

Winston, M. E., Chaffin, R. & Herrmann, D. (1987), 'A taxonomy of part-whole relations', *Cognitive Science* **11**(4), 417–444.

Xu, R. & Wunsch, D., I. (2005), 'Survey of clustering algorithms', *Neural Networks, IEEE Transactions on* **16**(3), 645–678.