An Approach to Ontological Modeling and Establishing Intercontext Correlation in the Semistructured Environment

N. A. Skvortsov, L. A. Kalinichenko
Institute for Problems of Informatics
Russian Academy of Sciences
nskv@mail.ru, leonidk@synth.ipi.ac.ru

Abstract

Presented research is intended for ontological identification of relevant specifications for semantic context integration of heterogeneous semistructured sources. Metainformation model is defined which includes uniform features for ontology, thesaurus and classifier modeling. Special technique for integration and mapping of different ontologies in this model is defined. The method for identification of specification element correlations in different contexts is considered.

KEYWORDS

Ontology, thesaurus, classifier, heterogeneous context integration, semistructured environment

1 INTRODUCTION

The approach reported in the paper has been developed as a part of subject mediating environment aiming at semantic interoperability of heterogeneous digital library collections[1] [15]. To provide for interoperability of heterogeneous information sources it is required to establish a global, uniform view of the underlying digital collections and services. It is assumed that specific, intermediary layer is formed by mediators providing a uniform query interface to the multiple data sources to free the user from having to locate the relevant collections, query each one in isolation, and combine manually the information from the different collections.

Here we are focused on ontological modeling and establishing intercontext correlation between heterogeneous information sources registered at the mediator. Ontologies are used to explicitly represent application semantics of the mediator's subject domain and of various information sources connected to the mediator.

An important issue of heterogeneous information integration consists in establishing correlations between various elements of schemas of heterogeneous information sources. Such correlations can be established relying on semantic references of such elements to ontological concepts and reasoning in a formal ontology [12, 19]. Regretfully, usually ontological modeling and reasoning is considered separately of the mediation context. For instance, the On-To-Knowledge-Project [2] is focused on sharable and reusable knowledge ontologies considering them as an independent entities for various tasks of information integration.

In our research from the very beginning we consider ontological modeling as a part of the more general information modeling in the mediator. The important consequence of such approach is that ontological integration and modeling and heterogeneous information integration and access in the mediator should be based on one and the same canonical model. There is no possibility to consider simplified models to make them just tractable for ontological reasoning (as it is done in [2]). If tractability is an important issue, a subset of the canonical model mentioned can be considered. At the same time, the metainformation repository and ontological modeling facilities can be considered separately of the mediator for different applications.

Another distinguishing feature of this research is that taking into account that most of the information in the Internet media is textual, visual, audial and rather weakly structured and attempting to provide quality controllable access to such information from the mediator we consider thesauri and classifiers to be an important part of a subject domain definition. It means that we consider ontological and terminological modeling in an integrated way. This leads to a combination of ontological methods with that of information retrieval.

Ontologies together with thesaurus definitions are used for semantic integration of information contexts. Establishing of context correlation includes procedures of mapping or integration of ontologies and thesauri themselves, storing of statistics about unstructured information, and identification of related specifications of structure schema of data. Therefore, one of the objectives of ontologies and thesauri joint use in this work is to form a basis for establishing semantic corre-

[1] This research has been supported by the RFBR 98-07-91061 and INTAS 97-11109 grants

lations of schema elements and ontological classes from various contexts for mediator specification and heterogeneous collections registration as well as for compositional design of information systems [5]. To use ontologies from different contexts for this objective, they should be mapped into a common ontological context. The process of integration and mapping of ontologies of different contexts expressed in the canonical model as well as methods for correlated schema elements identification with the help of a common ontological context are considered in the paper.

The proposed approach to metainformation modeling is based on a uniform model of representation for thesauri, classifiers and ontologies, and uniform methods for manipulation of these sorts of information. Additionally to a high semantic level of structural properties description of semistructured data we need also to take into account the statistics about unstructured or textual data to apply information retrieval methods to metainformation and source data.

We emphasize that the approach that has been developed is applicable to semistructured environment (as a more general one comparing to conventional structured databases) at least due to the following:

- an approach is applicable to schemas discovered from semistructured data as to conventional database schemas though schemas of semistructured data are characterized by non-strict typing of information fragments (defining variants of possible types, a posteriori type definition), often changing definitions, using implicit schema in data, etc. We assume in this paper that if a schema of a source is known or discovered, it is represented in the unified canonical model SYNTHESIS [14] and stored in the metainformation repository supporting this model. The same relates also to ontologies and thesauri. Taking into account that schemas of semistructured data may be contained implicitly in data and changed often, the process of related schema elements identification may become quite frequent operation. Ontological approach to this operation simplifies it and increases reliability of identification. Unified approach for ontologies and thesauri helps to find related elements during data analysis in case of absence of explicit schema or absence of an ontology related to considered sources;

- an approach is applicable for provision of semistructured data with application semantics directly as it is being done in XML applications (e.g., OpenMath [7] using Content Dictionaries) or in RDF[2] [3, 4].

- schema specifications that become instances of ontological classes are semistructured data by their nature.

We obtained such possibilities due to the flexibility of the canonical model used and the flexibility of the unified ontological/thesaurus model where we can freely use frames (unstructured data) and objects (typed data) as instances of ontological classes.

The paper is structured as follows. In the next section the canonical model for ontologies and thesauri is represented. Section 3 contains description of methods used for integration and mapping ontologies in the canonical model. In section 4 an approach to detection or intercontext schema elements correlation is proposed.

2 CANONICAL MODEL FOR THESAURI AND ONTOLOGIES

Ontological descriptions of subject domains contain specifications of ontological concepts, relations between them and constraints over these concepts. For ontological context modeling the SYNTHESIS model [14], the canonical model of the mediator environment is used. This model is sufficient for construction of an ontological model, model of thesauri and classifiers.

There exist mappings of metadata to this canonical model from well-known models used for ontology representation, such as Ontolingua [12], OKBC [8], various description logics and others.

The canonical model of ontologies and thesauri builds on basic notions of categories, concepts (unified for ontological and lexical ones), their properties, relations and assertions.

Various information collections are developed in different subject domains with specific terminology and interpretation of object structure in a respective domain. Ontological context is a collection of ontological information providing for a correct interpretation of concepts in a subject domain of a collection.

Any name may be provided as a lexical concept (lexical unit) in schemata or thesauri (vocabularies) of metainformation. Natural language definitions of all names are assumed. More formal ontological definitions related to the names can also be introduced.

The ontological model considered as well as most other ontological models is based on principles of knowledge representation systems. The main constituent of the model is an ontological concept. Ontological concept is an entity of knowledge representation, artifact that reflects characteristics of all similar objects of real world which could exist for agents in a given subject domain. Since ontological concepts are usually knowledge base entities, their structural and logical properties could be specified in terms of abstract data types (ADT). Each concept may have also a respective ontological class. The extension of this class contains metaobjects (other concepts, classes, elements of the object schema descriptions) semantically related to this concept.

Thesauri are represented as collections of lexical units and relationships between them. The model of description of lexical concepts is a subset of the ontological model. The model complies with standard requirements to multilingual thesauri [1]. Thus, lexical units can be considered as weakly formalized ontological concepts.

The hierarchy of classes is formed to categorize the subject domain. Each class defines certain subject category. Instances of a category can represent different artifacts including:

[2] Values of properties in RDF-descriptions have no pure typing. Semantics of properties in RDF-descriptions is provided with the help of the namespace dictionaries where data is semistructured too. They can be considered as own ontologies of contexts

- lexical units of a thesaurus;
- ontological specification of a concept;
- type specifications in various structure definition modules.

On the Figure 1 the fragment of the metainformation repository schema is presented, indicating how the ontological and thesaurus model are implemented in the metainformation repository.

Concepts are represented as metaobjects of `Concept` type which is an immediate subtype of the metainformation repository type `ADT`. Ontological concept type may have associated class, such that the concept is the type of this class as an object. This class may contain schema elements and lexical units related to the given concept. For each concept its frequency in each local collection and its weight over each local collection must be given. It is represented by `ConceptWeight` with `weight`, `frequency` attributes and `collection` association to metaobject `class` as a collection of information.

Let normalized weight of a term in one document of collection reflects a frequency of a concept in the collection and inverse document frequency (number of documents of collection in which the concept occurs at least once) [17]:

$$W_{dk} = \frac{f_{dk} \cdot \log \frac{N}{n_k}}{\sqrt{\sum_{i \in V_d} (f_{di} \cdot \log \frac{N}{n_i})^2}} \qquad (1)$$

where f_{dk} is frequency of term k in document d, N is number of documents in the collection, n_k is number of documents containing at least one occurrence of term k, V_d is the vector of all concepts in the document.

The first factor in the product increases the significance of terms that are frequently mentioned in the document. The second factor increases the significance of terms that occur in a smaller number of documents in the collection. The more the frequency and the less the number of documents containing given term the more its significance and, thus, its weight. Weights W_{dk} are normalized by the denominator to eliminate the dependence on the difference in length of vector V_d of different documents.

We use frequency of the concept over collection (class) that equals to number of documents or objects of this collection containing this concept (n_k). Weight of the concept in collection is sum of weights of this concept in all documents of collection:

$$W_{ck} = \sum_{d \in c} W_{dk} \qquad (2)$$

Frequency and weight of concepts in collections are used for evaluation of collections to be relevant to concepts of interest [11].

Ontological specifications of a given subject domain is defined within the ontological module that represents the respective ontological context. Analogously, thesaurus is defined in the module of thesaurus. Inside such modules their submodules may be defined. A module with the concept specifications can be imported into a module of the information source schema whose subject domain is described in this ontological module. Thus concepts can be related to schema specification elements. Names that are not included into the thesaurus but exist in collections become members of special module of an auxiliary lexicon.

A concept specification can contain the following definitions:

- identifier (word or phrase);
- verbal description;
- descriptor list;
- relationships to foreign equivalents (translations);
- linguistic semantic relationships;
- properties (attributes and associations);
- constraints.

Concept identifier (`name`) is one or several words in one of languages of thesaurus. Equivalent concepts in different languages must be linked by `foreign` association with each other. To define belonging of the concept to a given language we make this concept an instance of a special class (for example `russian` class for all Russian terms). Names of ontological concepts may be not in natural language. A kind of name can be assigned to a concept in `wordClass` attribute. We indicate here if the name is a noun, an adjective, a phrase or not a natural language identifier.

Verbal description of a lexical unit or an ontological concept defined in the `definition` attribute is a natural language description of the concept needed for a human understanding, for application of information retrieval methods and for preliminary mapping of concepts from one ontological context to another. These descriptions are assumed to have a form similar to one in Webster dictionary.

Using lexical and morphological analysis of verbal definition and name of the concept, the list of descriptors is generated. Normalized words are detected for Russian descriptions, for English descriptions word stems can be used as descriptors. Descriptor list (`descriptors` association) consists of lexical units of thesaurus defining a given concept. It can represent keywords list of a concept or terms related to a category (constituting its terminological portrait). Usually a list of descriptors is retrieved applying a lexical analysis of verbal descriptions. Each descriptor has its weight in the concept calculated with equation (1) like normalized weight of a term in a document. In the formula the descriptor list is treated as a document and a set of concepts in the context is treated as a collection.

Four kinds of relationships can be defined between concepts. They can be fuzzy, i. e. have strength in the interval [0.0,1.0], and default value of the strength is 1.0. These kinds of relationships are:

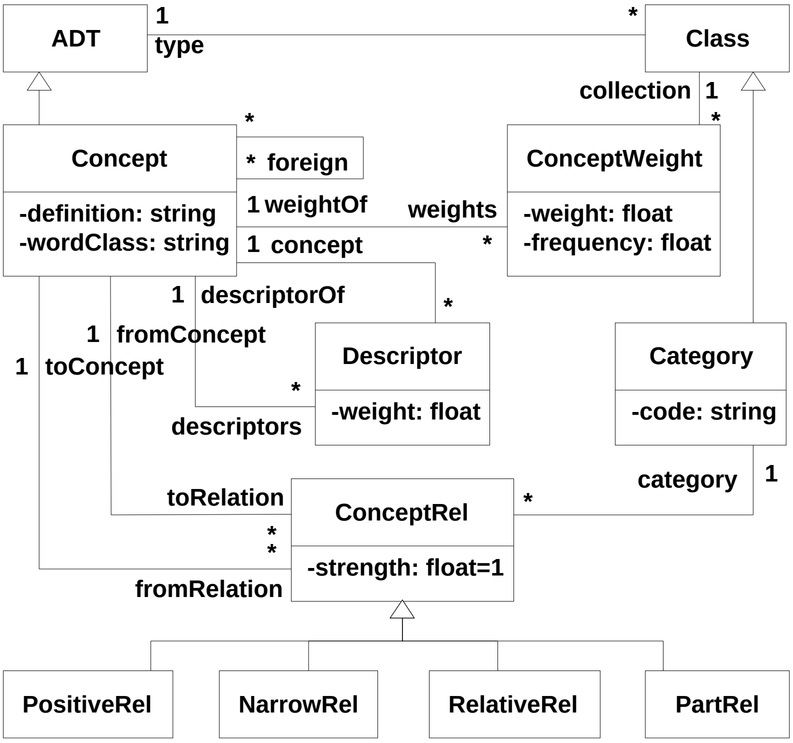

Figure 1: Representation of ontologies and thesauri in the metainformation repository

- positive (synonymous terms),
- hypernym/hyponym (generic term hierarchy),
- part/whole,
- associative (related terms).

All these relationships are represented in the metainformation repository by metaobjects of respective subtype of ConceptRel type. Positive association (PositiveRel) defines that two concepts are positively related in a given context. Hypernym/hyponym (supercon cept/subconcept) association (NarrowRel) binds more specific concept with more general one. It can express also such relations as “is a”, “form of”, “belong to” and so on. Hypernym/hyponym relationship is duplicated for ontological concepts, it is modeled by su pertype/subtype associations in ADT. Part/whole relationships are metaobjects of PartRel type. Associative relationship (RelatedRel) is established if concepts are related by another kind of concept relationship.

The logical structure of ontological concepts is described by properties: attributes and associations. Attributes can be typed by other concepts of the given context or by primitive data types. Relationships are represented by attributes and association metaclasses [14] defining these attributes if required.

Functional attributes may be useful for knowledge base manipulations. Functions are defined in predicative form. Invariants establish logical relations and constraints in concepts. All these definitions use facilities of Concept supertypes.

Elements of concept specifications can contain the information about concepts themselves or about instances of ontological classes related to these concepts. Depending on that they are defined in specification of types defining concepts or in specifications of instance types of those classes.

Lexical units themselves and relationships between them of the kinds mentioned may be related to some categories of classifier. This is usual practice for many existing thesauri. For this purpose lexical units can become instances of category classes, in ConceptRel the category attribute is defined for relating concept relationship to a category.

Every category is represented by a metaobject of Category type which is an immediate subtype of Class. Categories as classes can form subclass/superclass hierarchy. Since categories are classes, they contain concepts and type specifications as instances. Every category has a respective concept related to it. It is represented by using type relationship between ADT and Class.

Terminological portrait of category characterizing it by terms related to the category actually is a descriptor list of a respective concept. Category must have weights of concepts over each information collection, they are stored as weights of respective concepts in ConceptWeight.

Every concept metaobject becomes an instance of preexisting `concept` class. There are two subclasses of this class: `lexUnit` and `ontoUnit` to store lexical and ontological concepts respectively. These classes are not necessarily disjoint. Every category metaobject is also an instance of `category` class.

Since every concept is a type in our metamodel and lexical units from the thesaurus carry only small part of information, every attribute in metainformation repository schema is implemented as a binary relation.

3 ONTOLOGY INTEGRATION AND MAPPING

Depending on an objective, the task of ontological integration of various contexts may be formulated differently. To use ontologies in the task of context integration, first of all it is necessary to map them to a common ontological context. There are two alternative scenarios:

- integration of different ontological contexts in a common ontology,
- mapping them into an existing common ontology.

Mapping differs from integration so that in the case of integration a common ontology can be changed or extended, but during mapping it remains unchanged.

Sometimes it is necessary to integrate ontologies without involvement of the common one. For several participating ontologies the same approach is applied. Each next ontological context is integrated or mapped independently on the other contexts to a common ontology, to selected most representative ontology or to a fragment of such ontology. In each case we consider two ontologies, one of them is assumed to be a common ontology and another one is integrated in the common one.

Before integration, ontological concepts are described in terms of the canonical SYNTHESIS model. It should be stressed that the context of the common ontology must be described in the canonical model too. For this purpose, models used for representation of ontologies should be mapped to the canonical SYNTHESIS model. Ontological concepts from an ontological context in the canonical model are integrated into the common ontology using following criteria/techniques:

- integration by names and relationships,
- integration by lists of descriptors,
- structural integration and construction of views.

The first two criteria are based on the analysis of linguistic information related to concepts. These techniques are also used in the task of thesauri integration. Some issues of integration by names and relationships are considered in [13, 18].

Tools for lexical and morphological analysis are used in the process of name parsing. Words in a name are normalized for the Russian language or word stems are selected in names for the English language.

All names of an integrated ontology are compared with names of the common ontology. If the name of a concept is a phrase then minimal phrases (consisting of two words) are detected from this name. If names of different contexts or at least their minimal phrases are coincident then concepts assumed to be equivalent.

We link such concepts by intercontext positive relationships, sort them by average weight of the concept (see equation (2)) over all collections, and in this order displace them to an expert for equivalence confirmation. If the name of one concept includes the name of another concept in different context then it is assumed to be linked to it by hy pernym/hyponym relationship.

In case of different names of concepts in integrated ontology such concepts are stored in an auxiliary lexicon. They also are sorted by weights. By this order those of them whose average weights are greater then ℓ_1 are advised to be included into the common ontology.

Similarity and moreover partial coincidence of names doesn't guarantee the same semantics of concepts. To determine such situations it is useful to evaluate the sum of distances from similar concepts to one that is a common superconcept of them and to check if this value is sufficiently small. Analogously, if categories the concepts belong to are far from each other then those concepts are not similar.

To identify positively related and hypernymous/hyponymous concepts, an estimation of their proximity by lists of descriptors is used. Such estimation does not depend on name differences of the compared concepts. This kind of concept correlation is aimed at mapping concepts of one ontological context into the other.

For this purpose, the degree of proximity of concepts taken from two ontological contexts is calculated. It is based on the vector-space retrieval approach with normalized weights [6, 17]. We have already mentioned how to use equation (1) to calculate weights of descriptors.

Let X and Y be concepts of different ontological contexts (of the integrated and common ontologies), V_X and V_Y be vectors consisting of descriptors that define the corresponding concepts X and Y. W_{Xk} and W_{Yk} are weights of descriptors k that participate in the descriptor lists X and Y, respectively. The functions for estimating the correlation between ontological concepts are defined as follows:

$$sim(X,Y) = \frac{\sum_{k \in V_X \cup V_Y} (W_{Xk} \cdot W_{Yk})}{\sqrt{\sum_{k \in V_X} (W_{Xk})^2 \cdot \sum_{k \in V_Y} (W_{Yk})^2}} \tag{3}$$

$$r(X,Y) = \frac{\sum_{k \in V_X \cup V_Y} \min(W_{Xk}, W_{Yk})}{\sqrt{\sum_{k \in V_X} (W_{Xk})^2}} \tag{4}$$

$$r(Y,X) = \frac{\sum_{k \in V_X \cup V_Y} \min(W_{Xk}, W_{Yk})}{\sqrt{\sum_{k \in V_Y} (W_{Yk})^2}} \tag{5}$$

If lists of concept descriptors are disjoint then the function (3) [17] returns minimal value 0.0. If concepts have identical lists of descriptors then the value 1.0 is returned.

The concept X is assumed to be positively related to the concept Y if this function is greater than a certain threshold value ℓ_2. In this case, intercontext positive relationship with the strength equal to the returned value is established and advised for confirmation.

Functions (4) and (5) are used to establish hyper nym/hyponym relationships between different contexts. If $r(X, Y)$ and $r(Y, X)$ are less than a certain threshold value ℓ_3, the concepts X and Y are not referred to each other. If both the values of $r(X, Y)$ and $r(Y, X)$ are greater than ℓ_3, the concepts X and Y are positively associated with each other, and the correlation strength is the minimum of these values. If $r(X, Y)$ is greater than ℓ_3 while $r(Y, X)$ is less than ℓ_3, then X is a hypernymous concept of Y. In this case, the correlation coefficient is equal to $r(X, Y)$. On the other hand, if $r(X, Y)$ is less than ℓ_3 and $r(Y, X)$ is greater than ℓ_3 then X is a hyponymous concept of Y. In this case, the correlation coefficient is equal to $r(Y, X)$. If necessary, the results of the automatic mapping of one context into another can later be refined manually by an expert.

Since we need to include new concepts into common ontology and to integrate concepts for which intercontext relationships have been established, we may need to integrate semantic relationships between concepts. For this purpose concepts of integrated ontology linked by at least one semantic relationship are grouped into three vocabularies containing:

- completely coincident names (identical to name of common ontology);
- partially coincident names (not all words coincide);
- completely different names.

First of all it is advised to experts to add into the common ontology those concepts from an integrated ontology, which have relationships with at least one concept in vocabulary of completely coincident names. Their relationships are added into the common ontology too or some relationships may be changed by experts if there are contradictions. After this procedure the concepts in two last vocabularies are sorted by weights and those whose weights are greater then the threshold ℓ_1 are considered to be added into common ontology. For partially coincident names it is decided if these concepts are semantically different and if it is required to change a respective name of a common ontology concept.

After any manipulations with semantic relationships of the concepts the following integrity constraints should to be satisfied:

- Hypernym/hyponym graph should be acyclic.
- Hypernym/hyponym relationship is transitive. Thus relationships are redundant if they are included into the transitive closure of the hypernym/hyponym relationships.
- Hyponym relationship should be inverse to hypernym relationship.
- Part relationship should be inverse to whole relationship.
- Associative relationship is inverse to itself.
- There can be the only one relationship between two concepts because it is not possible to establish semantic relationships of different kinds simultaneously between two given concepts.
- A concept cannot be related to itself by a relationship of any kind.

We have not yet considered the internal structure of concepts and the respective logical constraints. Thus, a deeper and more accurate structural integration may be performed (if required) with respect to the internal structure of the concepts.

During structural integration and construction of views we consider concepts as types. The process of structured concept integration consists of construction of type reducts and composition of concretizing types using operations over types [5]: reduct, meet and join. Reduct operation chooses a fragment of a type specification. Meet operation gives most common supertype of operand types. Join operation returns least common subtype of operand types.

Key role in the process of structural integration belongs to construction of views over ontological classes corresponding to concepts in the ontologies being integrated. It aims at producing in common ontology an extent of artifacts contained in those classes.

The process of type specification integration is discussed in [5, 16]. However structural integration of ontologies has certain difference from the technology of object schema integration.

To apply type operations to concept specifications we need to have information about relevant elements of specifications. It is possible since intercontext relationships between concepts are established. So attributes of concept specifications in different contexts are assumed to be relevant if there exists a path beginning at mutually relevant concepts of those contexts and ending at mutually relevant types of considered attributes. Some missing attributes could be added to concepts during the integration.

Specifications of ontological classes corresponding to concept types do not usually include types of their instances because artifacts contained in these classes may be very different. So in the task of ontology integration only concept specifications themselves (which are own types of those classes) are involved into type operations.

Possible composition of concepts and view definitions above ontological classes are suggested by linguistic integration technique described above. If we know which concepts of ontologies being integrated are relevant to the concept of the common ontology (hyponymous/hypernymous or positively related), then most probably we will form total extent of respective classes in the view related to the concept of the common ontology. In this case internal structure of the concept is formed as composition of specifications of relevant concepts from ontologies being integrated. More difficult manipulations are possible.

To complete structural integration we must check and reconcile constraints. Correctness of changes may be checked by mapping of specifications to a description logic [9] definitions and verifying their satisfiability and subsumption between concept definitions.

4 ESTABLISHING A CORRELATION BETWEEN SCHEMA SPECIFICATION ELEMENTS WITH THE USE OF ONTOLOGY

After integration of ontologies, relationships of some kinds and common views over ontological classes are established between ontological contexts. We pass to ontology-based procedure for identifying relevant schema elements. It uses both relationships and views established so far.

At this procedure two schemata from different resource contexts participate. Each schema has own ontological context or relates to some existing ontology. These ontologies have been integrated into the common one. Specification elements of these schemata must be included as class instances of some concept in their respective ontological contexts.

Figure 2 presents part of metainformation repository schema that reflects relevance information to be stored. As shown, elements of specifications are frames and their slots, types, attributes, functions, parameters and invariants. Two elements of the same kind specified in different contexts can become ontologically relevant. Relationship of relevance is characterized by similarity value in the interval [0.0,1.0]. And it can be accepted or rejected by experts.

The semantic relationships between ontologies are used for evaluating of relevance between various specification elements of the same kind. The notion of weak ontological relevance of specification elements is based on such relationships.

Definition 1. The element I_1 of one information source specification is called ontologically weakly relevant to the element I_2 of the same kind (type, class, function, attribute, and so on) of another source specifications if I_1 is related to a concept C_1, I_2 is related to a concept C_2, and there exists a positive relationship between C_1 and C_2, or C_1 is a subconcept of C_2.

To identify relationships between two concepts, the correspondence paths must be analyzed and the concept graph must be complemented with the missing relationships. It is necessary to use completion algorithm to evaluate auxiliary relationships between every two concepts from different contexts. The search for new relationships is performed with the help of the algorithm that is close to one described in [10]. This algorithm uses transitivity of positive and hypernym/hyponym relationships. Auxiliary relationships are marked to differ them from initial ones, they required only for ontological identification of schema elements, but not for adding to ontological models. The results of the algorithm are used to find correspondences between the specification elements. For this purpose, the concept of the weak ontological relevance of specification elements is used. Identified relevance may be evaluated by expert to accept or reject it.

If the process of integrating ontological contexts is performed only using lexical properties and thesaurus features, then the subsequent integration of data relevant to the context concepts can use only the concepts themselves and their relationships, without regard to internal structure of concepts and respective classes.

Views are used for more reliable identification of specification interrelations. The notion of tight ontological relevance is applied for this purpose.

Definition 2. The element I_1 of one information resource specification is called ontologically tightly relevant to the element I_2 of the same kind (type, class, function, attribute, and so on) of the second resource specifications if I_1 is ontologically weakly relevant to I_2, and I_1 is an instance of at least one ontological class corresponding to a concept C_1 that is a specialization (subconcept) of an ontological concept C_2 that has ontological class with I_2 as its instance, or I_1 and I_2 should belong to the same class of an ontological concept.

So when ontologies of sources are mapped to or integrated in the common one, then resource specification elements are relevant if they belong to the same ontological classes. If we search elements relevant to a given one, they may belong not only to the same class, but to its subclasses too.

Similarity value for tight ontological relevance is borrowed from weak relevance, but any tight relevance is considered to be more probable than any weak one. That is why tight relevances must be displaced first in lists of probable relevances to experts for evaluation.

5 CONCLUSION

In this paper, a uniform metainformation model for ontologies and thesauri in the semistructured data environment is presented, methods of integration and mapping of ontologies are proposed for this model and an approach proposed for ontology-based identification of relevant specification elements of different contexts in case of semistructured information.

References

[1] *ISO 5964. Documentation - Guidelines for the Development and Establishment of Multilingual Thesauri.* ISO, 1985

[2] *On-To-Knowledge: Content-driven Knowledge-Management through Evolving Ontologies.* [http://www.ontoknowledge.org/]

[3] *Resource Description Framework (RDF) Model and Syntax Specification.* [http://www.w3.org/TR/REC-rdf-syntax]

[4] *Resource Description Framework (RDF) Schema Specification 1.0.* [http://www.w3.org/TR/rdf-schema]

[5] D. O. Briukhov, L. A. Kalinichenko. *Component-Based Information Systems Development Tool Supporting the*

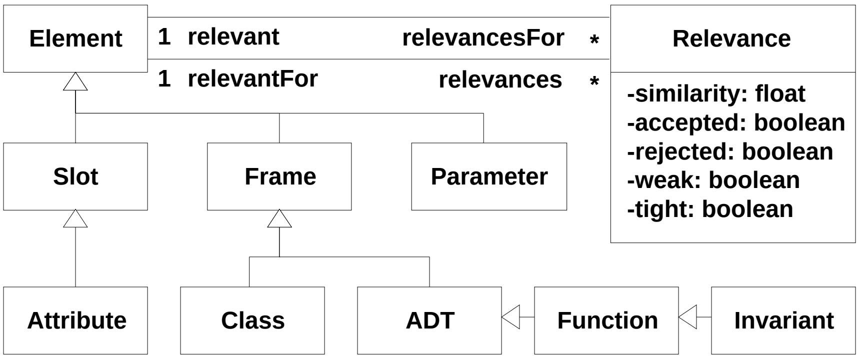

Figure 2: Relevance of schema elements in the metainformation repository structure

SYNTHESIS Design Method. Proc. of the East European Symposium on "Advances in Databases and Information Systems", September 1998, Poland, Springer, LNCS N 1475, 1998

[6] D. O. Briukhov, S. S. Shumilov. *Ontology Specification and Integration Facilities in a Semantic Interoperation Framework,* Proc. of the International Workshop ADBIS'95, Springer, 1995

[7] O. Caprotti, D. P. Carlisle, A. M. Cohen. *The OpenMath Standard. Version: 1.0.* The OpenMath Esprit Consortium, 2000

[8] V. K. Chaudhri, A. Farquhar, R. Fikes, P. D. Karp, J. P. Rice. *Open Knowledge Base Connectivity 2.0.2.* Artificial Intelligence Center of SRI International, KSL, Stanford University, Feb 1998

[9] F. M. Donini, M. Lenzerini, D. Nardi. *The complexity of concept languages.* Information and Computation 134(1), 1997

[10] P. Fankhauser, E. J. Neuhold. *Knowledge Based Integration of Heterogeneous Databases.* Integrated Publication and Information Systems Institute (GMD-IPSI), Darmstadt, 1993

[11] L. Gravano. *Querying Multiple Document Collections Across the Internet. Ph.D. Thesis.* Stanford University, 1997

[12] Gruber T.R. *Ontolingua: a Mechanism to Support Portable Ontologies.* Stanford University, June 1992

[13] E. N. Kuznetsov. *Creation and Keeping of Thesaurus within the Mediator between Users and Net of Digital Libraries.* 1st Russian scientific conference "Digital Libraries: Perspective Methods and Technologies, digital collections", Saint-Petersburg, 1999

[14] L. A. Kalinichenko. *SYNTHESIS: the Language of Definition, Design and Programming of Interoperable Environments of Heterogeneous Information Resources.* Institute for Problems of Informatics RAS, Russian Academy of Sciences, Moscow, 1993

[15] L. A. Kalinichenko, D. O. Briukhov, N. A. Skvortov, V. N. Zakharov. *Infrastucture of the Subject Mediating Environment Aiming at Semantic Interoperability of Heterogeneous Digital Library Collections.* 2st Russian scientific conference "Digital Libraries: Perspective Methods and Technologies, Digital Collections", Protvino, 2000

[16] L. A. Kalinichenko, N. A. Skvortsov, D. O. Briukhov, D. V. Kravchenko, I. A. Chaban. *Design of Personalized Digital Libraries over Web-Sites with Semistructured Data.* 1st Russian scientific conference "Digital Libraries: Perspective Methods and Technologies, Digital Collections", Saint-Petersburg, 1999

[17] G. Salton, C. Buckley. *Term-Weighting Approaches in Automatic Text Retrieval.* Readings in Information Retrieval under edition of K. S. Jones and P. Willett, Morgan Kaufmann Publishers Inc., 1997

[18] M. Sintichakis, P. Constantopoulos. *A Method for Monolingual Thesauri Merging.* SIGIR'97, Philadelphia, 1997

[19] M. Uschold. *Converting an Informal Ontology into Ontolingua: Some Experiences.* Proc. of the Workshop on Ontological Engineering, Budapest, 1996