Development of Ontologies with Minimal Set of Conceptual Relations

Natalia V. Loukachevitch and Boris V. Dobrov

Research Computing Center of M.V.Lomonosov Moscow State University;
NCO Center for Information Research
339, Research Computing Center of M.V.Lomonosov Moscow State University,
Leninskie Gory, Moscow, 119992, Russia
{louk, dobroff}@mail.cir.ru

Abstract

In the paper we describe our approach to development of ontologies with small number of relation types. Non-taxonomic relations in our ontologies are based on ontological dependence conception described in the formal ontology. This minimal relations set does not depend on a domain or a task and makes possible to begin the ontology construction at once, as soon as a task is set and a domain is determined, to receive the first version of an ontology in short time. Such an initial ontology can be used for information-retrieval applications and can serve as a structural basis for further development of the ontology.

1. Introduction

To develop an ontology for a large domain is a very serious and difficult task. It is necessary to identify main concepts of the domain, to understand what a set of relations between concepts is needed for a given task, to describe relations for the domain concepts, to test, to evaluate the ontology and to build it into the task framework.

Contemporary ontology approaches propose an arbitrary set of relations described as predicates (Niles & Pease, 2001). Properties of the relations are described as axioms. Consistent and detailed description of such rules and axioms requires a well-structured basis, good understanding of the domain structure. Such relatively simple properties as transitivity and inheritance are used only for taxonomic relations. Even the part-whole relations are not considered as a reliable basis for transitivity (Winston et.al. 1987).

We suggest to begin the construction of an ontology using a minimal set of relations and to determine the domain structure according to this set. Such a minimal set of relations does not depend on the type of a domain, on the type of a problem solved, as it is based on the fundamental properties of concepts. Besides taxonomic relations for a given concept we suppose to use relations that show how existence of a given concept depends on existence of other concepts, i.e. determine the so-called relations of ontological dependence, which are studied in the framework of the philosophical discipline «formal ontology» (Guarino, 2000; Smith, 1998) These ontological dependence relations become a basis for the second chains of transitivity besides taxonomic relations. Such an ontology can serve as a basis for explication of the domain structure and determination of a new set of relations, necessary for solving the main problem**.**

Our proposal originates from our analysis of structural principles of ontologies useful in information-retrieval applications for large heterogenous domains. We think that such ontologies can be used also as an initial level of domain structuring. An important feature of such an initial ontology that after its construction it can be evaluated and used in information retrieval tasks.

The technology has been developed in the process of creation of large and extra large ontologies and thesauri for various domains and their actual usage in multiple applications of automatic text processing such as

- Sociopolitical thesaurus for automatic text processing (29 thousand concepts) in the broad domain of public relations (Loukachevitch & Dobrov, 2002)
- Avia-Ontology for the domain, describing behavior of an operator (air crew) and board equipment in various flight operations (1200 concepts, 3400 terms) (Dobrov et.al. 2003).
- Ontology on computer security (2500 concepts) and others.

In the paper we will use examples from all these domains.

2. Specific Features of Ontology Intended for Text Processing in Information-Retrieval Applications

Real domains are associated with large collections of electronic documents different in sizes, styles, structures. An ontology of the domain has to correlate with knowledge contained in these documents and help to solve various information-retrieval tasks.

A domain ontology has to contain description of domain's concepts, conceptual relations, properties of relations described as rules: if $p(x1,\ldots xn)$ then $q(y1\ldots ym)$. The problem is that for heterogenous texts and the contemporary level of automatic text processing, it is very difficult to receive from texts that $p(x1,\ldots xn)$ is true.

The most reliable information that can be received from texts is that concept C was mentioned, that is concept C exists. Therefore the most reliable rules that can be used in information retrieval describe how existence of one concept depends on existence of other concept

The taxonomic relation is a known example of such types of relations. Non-taxonomic relations in an ontology constructed for information retrieval can be analyzed from the point of view of the philosophical theory of formal ontology, which studies existence of various entities in the world (see section 3).

In the context of automatic processing of heterogenous texts the impossibility to use complex rules of inference leads to necessity to find additional relations with such relatively simple properties as transitivity and inheritance based on knowledge of the existence of concepts.

It is obvious that problems of existence and co-existence of concepts are central for any ontology in any domain and task. Therefore such consideration is important for development of a new ontology even out of information-retrieval context. Moreover, development of such an initial ontology can be begun immideately after the project beginning, it can be finished in relatively short time and can serve as a qualitative structural basis for further development of the ontology for the given task.

3. Relations of Ontological Dependence

Basic notions of philosophical formal ontology applied to contemporary conceptual research are philosophical notions of rigidity, identity, unity and dependence (Guarino 2000).

The main question of the dependence theory is if an entity can exist by itself or it supposes the existence of something else. There are three main types of this relation:

- whether the existence of an entity supposes the existence of something else (rigid dependence), for instance, *boiling* is impossible without the existence of a certain volume of liquid which boils;
- whether existence of examples of a certain class (generic dependence) is supposed, like, the appearance of the concept *garage* is impossible without the existing concept *motor vehicle*, though a certain garage may appear without any reference to a certain motor vehicle;
- when the existence of an entity in moment *t* presumes the existence of another entity in moment *t1* before *t* (historical dependence), so, for instance, *straw* historically depends on threshing, as straw can not appear without a preliminary *threshing process*, altogether after this work has been finished, straw can continue its existence for a long time.

Ground axioms for dependence are described as (Gangemi et.al. 2001):

(D1) $D(x,x)$

(D2) $D(x,y) \wedge D(y,z) \rightarrow D(x,z)$

Among dependence relations the following sub-relations can be introduced (Gangemi et.al. 2001):

$MD(x,y) =_{def} D(x,y) \wedge D(y,x)$
(mutual dependence)

$ID(x,y) =_{def} D(x,y) \wedge P(y,x)$
(internal dependence – P(x,y) - part relation).

$ED(x,y) =_{def} D(x,y) \wedge \neg P(y,x)$
(external dependence – P(x,y) - part relation)

Let us consider several examples of description of dependence relations.

The first example is relation *TREE – FOREST*, which usually is described as PART(*TREE, FOREST*). But trees grow not only in forests but in gardens, in the streets and so on. Therefore existence of a tree does not depend on existence of a forest. At the same time a forest can not exist without trees. If all trees in a forest were eliminated, the forest does not exist. So a forest depends on a tree. We can write D(*FOREST, TREE*). If we introduce such concept as *forest tree*, we can write D(*FOREST, FOREST TREE*), D(*FOREST TREE, FOREST*), that is concepts forest and forest trees are mutually dependent – MD (*FOREST, FOREST TREE*).

Another example is a relation *WHEEL – CAR*. And again a wheel can be included in various vehicles and other machines. However if we introduce concept *CAR WHEEL*, we can again write MD (*CAR, CAR WHEEL).*

Examples of conceptual dependence relations in the Avia-Ontology are as follows:

ALTIMETER depends on *FLIGHT ALTITUDE* (generic dependence),

TANKER AIRCRAFT depends on *AIRCRAFT FUEL* (generic dependence),

AIR PATROL depends on *FIGHTER AIRCRAFT* (rigid dependence)

4. Retrieval of Documents and Relations of Ontological Dependence

It is easy to see that in case of the rigid dependence the existence of a dependent concept is very tightly connected with the existence of a main concept. It is difficult to imagine a situation (and a text) where a dependent concept participates and this situation has no relation to a main concept.

In case of the generic dependence examples of a dependent concept usually participate in situations related to a main concept, however sometimes situations, not relevant to a main concept, can arise (for example, a crime in a garage can have no relation to automobiles).

At last the historical type of dependence is the weakest type among existential situations. A main concept is necessary for appearence of a dependent concept, but then a dependent concept can exist for a long time and participate in various situations not relevant to the main concept.

There differences in subtypes of conceptual dependence relations lead to differences in behaviour of these relations in information-retrieval context and could be seen through analysis of search results of so-called simple queries.

Queries in an information-retrieval system can consist of different numbers of terms and words. From the ontology point of view the simplest query is a query consisting of a single term T of an ontology. All other queries, including several terms, words and terms have to be processed as a function from elementary queries.

We hypothesize that potential quality of query expansion based on thesaurus relations can be studied using the simplest queries. If search characteristics of expansion of elementary queries are low, then processing quality of complicated queries can not be better. If ontological relations allow effective query expansion in simple cases then it is an important step to study techniques for expansion of a complex query. The meaning of such the simplest query is "all about T" and we will denote it as SQ(T).

From this point of view we can study potential search characteristics of every ontology relation. Let us see two concepts C1 and C2, between which relation R is established. We consider a simple query consisting of a single term corresponding to concept C1 – SQ(C1), and we would like to know how relation R between C1 and C2 can be used for expansion of this query. In this process documents containing terms of C2 have to be joined to the retrieved set of documents, maybe with certain weights. Hence without any real query expansion we can take documents, containing C2, and try to estimate how many of these documents can be relevant to the query SQ(C1).

Let us study potential retrieval efficiency of simple queries, equal to main concepts C, expanded by text with ontologically dependent concept C2. We will analyse 50 best texts from retrieval set (standard tf*idf ranking) for simple query SQ(C2). The search was implemented on the full Russian collection of University Information System RUSSIA (www.cir.ru/eng), containg more than 800 thousand contemporary Russian documents. Results for several mentioned examples are presented in Table 1.

Dependent concept *D*	Type of dependence	Main concept M	*nd50*	*nm50*
FOREST	Rigid	*TREE*	49	12
SUMMIT	Rigid	*HEAD OF STATES*	49	20
PIANIST	Generic	*PIANO*	44	1 6
GARAGE	Generic	*CAR*	43	1
CAR	Historical	*CAR PLANT*	18	44

Table 1.

Here

- *nd50* - number of texts containing D, relevant to D and relevant to SQ(M),
- *nm50* - number of texts containing M, relevant to M and relevant to SQ(D).

The table demonstrates the correlation between a type of dependence and search characteristics of simple queries:

- in case of the rigid dependence for almost all texts if a text is relevant to a dependent concept, it is relevant to a main concept also;
- in case of the generic dependence the ratio is less but high enough;
- in case of the historical dependence ratio much decreases.

Search characteristics of reverse simple queries are low (that is there are a lot of texts, which are relevant to a main concept and are not relevant to a dependent concept), and this corresponds to absence of dependence. For the fifth pair a lot of texts about car plants are texts about cars at the same time, because concept *CAR PLANT* also depends on concept *CAR*. Car plants can not exist without existence of the class of cars therefore this is the generic type of dependence and again we can see correlation of search characteristics.

5. Internal and External Dependence Relations

Dependence relations can be subdivided to internal dependence relations and external dependence relations. An obvious example of an internal dependence relation is a whole dependent of a part. However it is possible to include to internal relations also relations between situations and their roles (investing – investor), entities and their properties (watercraft – nautical qualities).

So we define Internal dependence relation:

$$ID(y,x) = D(y,x) \wedge ((P(x,y) \vee participant(x,y) \vee property(x,y))$$

Then we introduce reverse internal dependence relations where we require dependence of parts, roles and properties from corresponding wholes, situations and entities:

$$(*)\ RID(x,y) = D(x,y) \wedge ((P(x,y) \vee participant(x,y) \vee property(x,y))$$

In our constructed ontologies we describe reverse internal dependence relations as a specific kind of relations. Now besides transitivity of taxonomic relations we use the following properties of the relations:

- transitivity of reverse internal dependence relations:

$$RID(x,y) \wedge RID(y,z) \rightarrow RID(x,z)$$

- **i**nheritance of dependence relations:

$$Subclass(x,y) \wedge D(y,z) \rightarrow D(x,z),$$
$$Subclass(x,y) \wedge RID(y,z) \rightarrow RID(x,z),$$
$$RID(x,y) \wedge D(y,z) \rightarrow D(x,z)$$

Let us see several examples of transitivity paths of dependence relations in Sociopolitical Thesaurus:

ACCUSED PERSON → PUBLIC PROSECUTION →
→ JUDICIAL TRIAL →
→ JUDICIAL PROCEEDINGS →
→ JUSTICE SYSTEM → LEGAL SYSTEM

MONETARY BASE → MONEY SUPPLY →
→ MONEY CIRCULATION →
→ MONETARY SYSTEM →
→ FINANCIAL SYSTEM → ECONOMY

DRUGGIST → DRUGSTORE → DRUG SUPPLY →
→ MEDICAL AID → MEDICINE (FIELD) →
→ PUBLIC HEALTH.

Examples of transitivity paths of dependence relations in Avia-Ontology:

LANDING STRIP → AIR FIELD →
→ AIRDROME → AVIATION

At present we do not use transitivity of any dependence relation as it was described in axiom D2 from section 3 and use only transitivity of RID relations (*) because we observed that through chains of not-RID dependence relations the relevancy of simple queries was lost.

6. Concept Description in Constructed Ontologies

A specific set of relations , which is used by us now besides taxonomic relations (BROADER-NARROWER relations) is the following:

- PART– WHOLE – is used to describe RID relations, that is dependent traditional parts, participants of situations, properties;
- unsymmetrical associations ASC1-ASC2 – are used for the rest of conceptual dependence relations D(x,y) – ASC1 means "is ontologically dependent of"; ASC2 – "is ontologically main concept for". At present we describe mainly generic and rigid dependence relations,
- symmetric association is used for concepts, similar by meaning.

Ontological relations described as ASC1-ASC2 are not considered as transitive relations.

Thus, two types of relations in the relations set employed by us are significantly bound with the concept of ontological dependence. Relations of these types occupy approximately half of all relations in our thesauri and ontologies.

Let us see example of relations for Computer security Ontology presented as thesaurus articles

COMPUTER VIRUS

BT	*MALWARE*
WHOLE	*VIRUS ATTACK*
NT	*BOOT VIRUS*
NT	*MACRO VIRUS*
NT	*MAIL VIRUS*
…	
PART	*VIRUS LENGTH*
PART	*VIRUS CODE*
ASC2	*ANTIVIRUS PROTECTION*
ASC2	*VIRUS ACTIVATION*
ASC2	*VIRUS WRITING*

ANTIVIRUS PROTECTION

BT	*INFORMATION PROTECTION*
ASC1	*COMPUTER VIRUS*
NT	*ANTIVIRUS SCAN*
PART	*ANTIVIRUS PROGRAM*
PART	*ANTIVIRUS COMPANY*

Conclusion

We proposed to construct ontologies mainly based on ontological dependence relations. "Minimal" relations set

- makes it possible to begin the ontology construction, as soon as a task is set and a domain is determined, to receive the first version of an ontology in short time;
- provides a conceptual basis for communicating with experts in the given domain;
- provides the initial domain structuring which may be used as a basis for singling out special relations in the domain;
- the ontology with the ontological dependence relations provides a qualitative knowledge basis for diverse information retrieval applications.

Acknowledgements

Partial support for this work is provided by the Russian Foundation for Basic Research through grant # 03-01-00472.

References

Dobrov, B., Loukachevitch, N. & Nevzorova, O. (2003). The Technology of New Domains' Ontologies Development. In Proceedings of X-th Intern. Conf. KDS 2003 "Knowledge-Dialogue-Solution". June 16-26, Varna, Bulgaria. pp.283-290.

Gangemi, A., Guarino, N., Masolo C. & Oltramari A. (2001). Understanding Top-Level Ontological Distinctions. In Proceedings of IJCAI 2001 workshop on Ontologies and Information Sharing.

Guarino, N. (2000). Some Ontological Principles for Designing Upper Level Lexical Resources. In Proceedings of First International Conference on Language Resources and Evaluation.

Loukachevitch, N.V. & Dobrov, B.V. (2002). Evaluation of Thesaurus on Sociopolitical Life as Information Retrieval Tool. In M.Gonzalez Rodriguez, C. Paz Suarez Araujo (Eds.) Proceedings of Third International Conference on Language Resources and Evaluation (LREC2002), Vol.1, Gran Canaria, Spain – pp.115-121.

Niles, I. & Pease, A. (2001). Towards a Standard Upper Ontology. In C.Welty and B.Smith, (Eds.) Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), Ogunquit, Maine, October 17-19, 2001.

Smith, B. (1998) *Basic tools of formal ontology*. In N. Guarino, (Ed.) Formal Ontology in Information Systems.

Winston, M., Chaffin, R. & Herrmann, D. (1987). A taxonomy of part-whole relationships. Cognitive Science, Vol.11, pp. 417-444.