Ontologies

B.C. Vickery

University College London, UK

Received 14 January 1997

Abstract.

The paper discusses the emergence of the term 'ontology' in knowledge engineering (and now in information science). A definition of the term as currently used is given: an explicit specification of a 'world' that is to be represented in a computer system. The background that has led to the emergence of this concept is sketched, and examples of ontologies are provided. The process of building an ontology is discussed, and the uses of such tools in knowledge engineering. There is a concluding commentary comparing ontologies with similar tools used in information science.

Introduction

The term 'ontology' has been getting into information science literature. For example, at least four authors used it in the 1996 Knowledge Organisation Conference [1]. In this paper, I wish to explore the current usage of the term, and to end with some brief comments on its relation to the work of the information scientist.

Background

Philosophically speaking, ontology is the study of what exists and what we must assume to exist in order to achieve a cogent description of reality. The term has been coming into use in artificial intelligence (AI) over the last ten years. An early example was a paper by Alexander *et al.* [2], of which the following was an abstract:

> Knowledge engineering suffers from a lack of formal tools for understanding domains of interest. Current practice relies on an intuitive, informal approach for collecting expert knowledge and formulating it into a representation scheme adequate for symbolic processing. Implicit in this process, the knowledge engineer formulates a model of the domain, and creates formal data structures (knowledge base) and procedures (inference engine) to solve the task at hand. Newell in 1982 proposed that there should be a knowledge level analysis to aid the development of AI systems in general and knowledge-based expert systems in particular. The present paper describes a methodology, called ontological analysis, which provides this level of analysis. The methodology consists of an analysis tool and its principles of use, that result in a formal specification of the knowledge elements in a task domain.

The reference back to Newell is encountered frequently in discussions of ontology in knowledge engineering. In a presidential address to the American Association for Artificial Intelligence [3], he considered computer system levels, e.g. currents and voltages at the circuit level, changing to bits at the logic level, to data structures (variables, arrays, etc) at the symbol level. He regarded the representations in knowledge bases (production rules, frames, semantic nets, etc) as being higher structures at the symbol level: above them was a knowledge level, comprising the conceptual knowledge that was embodied in the representations. One is reminded of Ranganathan's distinction [4] between the 'idea' level (corresponding to Newell's knowledge level), the 'verbal' level (the verbal expression of a concept) and the 'notational' level (the symbolic representation of the idea or concept in classificatory language).

By the end of the 1980s, it was accepted that 'domain conceptualisation' was a necessary part of knowledge acquisition for a knowledge-based system (KBS) [5]. McGraw and Harbison-Briggs considered it one of the most important and difficult of knowledge acquisition tasks, and discussed conceptual frameworks, cognitive maps, taxonomies, concept dictionaries, cluster analysis and other techniques [6].

In the early days of expert system development, a knowledge base was seen as a summarisation of knowledge that is in the mind of, or utilised by, a human

Correspondence to: Professor B.C. Vickery, 9 Clover Close, Oxford OX2 9JH, UK.

expert, and specifically relevant to the task to be solved by the expert system. However, if such knowledge was to be used effectively in problem solving, it had to reflect the real structure of the domain in which the expertise was applied. So, more recently, some workers in AI have been arguing that 'the primary concern of knowledge engineering is modelling systems in the world, not replicating how people think' [7].

In developing the representation of knowledge known as 'conceptual graphs', Sowa in 1984 used the idea of an ontology as 'a catalogue of everything that makes up a [possible] world, how it is put together, and how it works . . . a catalogue of concept and relation types', and he listed a number of examples of concept types and conceptual relations [8]. The term 'ontology' was also used by Lenat and Guha [9, 10] in their development of CYC, a very large knowledge base aimed at capturing human common-sense knowledge. They argued that 'there are deep important issues that must be addressed if we are ever to have a large intelligent knowledge-based program. What ontological categories would make up an adequate set for carving up the universe? How are they related?' They saw the need for 'a global ontology that specifies at a very high level what kinds of things exist and what their general properties are'.

These developments were all pointing to the need for task-independent conceptual analysis, which Hayes had been urging as early as 1985 [11]. Added emphasis was given to this by the need for communication among several KBSs. Clancey [7] underlined that we are now often dealing with communities of diverse intelligent 'agents' who must interact and communicate in different ways and in relation to widely different tasks. Such interaction will be facilitated if they can all draw upon a task-independent conceptual model of their joint domains. It was concluded that 'the study of ontology, as a branch of philosophy dealing with the nature of reality, can be of benefit to the knowledge-construction process in yielding high-value knowledge bases' [12]. A DIALOG search in December 1996 of the INSPEC database for 'ontology' yielded 550 references, though many – perhaps most – were not directly relevant to the problems discussed here. There are now several 'ontology pages' on the Internet, giving leads to useful World Wide Web pages [13].

Definition of an ontology

What, then, is an ontology, as the word is used in knowledge engineering? The definition developed by Gruber [14, 15, 16] in 1993 seems generally to be accepted. He uses the idea of a 'conceptualisation', meaning 'an abstract, simplified view of the world that we wish to represent for some purpose'. Every KBS, he writes, is committed to some conceptualisation, explicitly or implicitly: 'An ontology is an explicit specification of a conceptualisation'. Gruber compares an ontology with the conceptual schema that may be developed (at the knowledge level) for a database system, which provides a logical description of shared data, allowing application programs and databases to interoperate without having to share data structures (at the symbol level). In knowledge engineering, an ontology provides a vocabulary for talking about a domain, whereas a corresponding knowledge base will include the symbolically represented knowledge needed to solve specific problems or answer queries about the domain. Agents sharing a vocabulary need not share a knowledge base: each agent knows things the other does not.

Uschold [17] writes that an ontology:

> . . . is often conceived as a set of concepts (e.g. entities, attributes, processes), their definitions and their inter-relationships . . . An ontology may take a variety of forms, but will necessarily include a vocabulary of terms and some specification of their meaning (i.e. definitions). It may be:
> - highly informal, expressed in natural language,
> - semi-formal, expressed in a restricted and structured form of natural language,
> - semi-informal, expressed in an artificial formally defined language,
> - rigorously formal, with meticulously defined terms with formal semantics and theorems.

Ontologies, writes Gruber [14], are often equated with taxonomic hierarchies of classes, with class definitions and the subsumption relation, but they need not be limited to these forms.

There has been discussion of what has been called the 'granularity' or 'grain size' of an ontology – to what degree of specifity should the concept hierarchy be continued. Some authors, e.g. Poli [18], concentrate on the top-level types of concept occurring in the domain – the ontological *categories*. For him, 'an ontology is not a catalogue of the world, a taxonomy, a terminology of a list of objects, things or whatever else: it is the general framework or structure within which catalogues, taxonomies, terminologies may be given suitable organisation'. Others, e.g. Guarino [19], see the need to include all the specific concepts occurring in the domain, and urge that an ontology should be of high granularity. Mahesh [20] writes that:

Building a precise ontology where each concept is clearly distinguishable from others requires a fine-grained decomposition of concepts. Such decompositions are typically unending; but an ontology must bottom out at a reasonable level and treat many concepts as primitives. For example, the concept of a wheeled-engine-automobile can be treated as a primitive or decomposed into its structural, functional, and other components in great detail.

Another distinction among those working in this field is between those who aim to build a general ontology, taking all knowledge for its province, and those who are concentrating on an ontology for a specific domain – a distinction analogous to the makers of general and special classifications in our field.

Examples of ontologies

It is time to make these general definitions and descriptions more concrete by giving some examples of ontologies that have been developed.

CYC

Among the first of the 'generalisers' were Lenat and Guha with their CYC ontology [9, 10]. The CYC knowledge base is a formalised representation of a vast quantity of fundamental human knowledge: facts, rules of thumb, and heuristics for reasoning about the objects and events of everyday life. At the present time, the CYC knowledge base contains tens of thousands of terms and several dozen hand-entered assertions about/involving each term. Each term in the CYC ontology contains a definition and supplementary information, for example:

#Skin
DEF: A (piece of) skin serves as outer protective and tactile sensory covering for (part of) an animal's body. This is the collection of all pieces of skin. Some examples include #TheGoldenFleece (representing an entire skin of an animal) and #BodyPartFn#YulBrynner #Scalp (representing a small portion of his skin).
ISA: Physiology#AnimalBodyPartType.
SEE ALSO: #SheetOfSomeStuff.

Each entry starts with the name of the concept. Next comes an English comment, to help to clarify the meaning and intended use of the concept. In each entry, there is an ISA line, which lists one or more of the broader collections (sets) of which this term is an element. 'Skin' is itself a collection (the set of all pieces of skin, including as a subset the set of all 'whole skins'); it can therefore have supersets and subsets and elements. Some elements of this collection would be #The GoldenFleece and #BodyPartFn#YulBrynner#Scalp. Other subsets worth naming would be #SheepSkin and #Scalp. A reference to a possible superset is given under SEE ALSO.

As is implied by the ISA references, the terms are arranged into a vast (tangled) hierarchy or lattice. CYC permits (and even expects) its knowledge to be structured into microlattices that can represent different points of view, levels of granularity, cultural differences, age differences, time periods, corporate cultures, etc. In different microlattices, the same term could have different (partial) definitions, different rules which were stated about them, different assumptions that were presumed to hold true, etc.

The upper levels of the CYC ontology are believed to provide a sufficient common structure for applications of the following types:

- natural language understanding and generation;
- semantic database integration, consistency-checking, and data mining;
- semantic information retrieval;
- ontology-constrained simulation;
- building and utilising user models;
- knowledge sharing by groups working independently.

MIKROKOSMOS

Workers in the field of machine translation (MT) have also been concerned to build general ontologies as 'interlinguas': conceptual aids that can function as switching devices between languages. An example of this is MIKROKOSMOS. One of the developers of this, Nirenburg, some ten years ago [21] described an interlingua dictionary for his MT project. The concepts were arranged into an ISA hierarchy (the top concept ALL dividing into EVENT and OBJECT, the former subdividing into PROCESS and STATE, the latter into MENTAL-OBJECT and PHYSICAL-OBJECT, and so on). Each concept was represented as a frame, which contained slots in which the properties of the concept were listed. At this time, the dictionary was not called an ontology.

In the later project, MIKROKOSMOS [22], interlingual meaning representation for MT is derived from representations of word meanings in computational lexicons and from representations of world knowledge in ontologies. An ontology for MT purposes is seen as a body of knowledge about the world (or a domain) that is a repository of primitive concepts used in meaning representation, that are organised in a tangled hierarchy, and further interconnected by a rich system of semantic relations defined among the concepts. The ontology serves as a basis for:

(1) representing meanings of different languages;
(2) representing meanings of natural language texts in an interlingua; and
(3) sharing knowledge between different lexical knowledge bases.

The MIKROKOSMOS ontology has about 4,500 concepts in it and covers a wide range of domains, while focusing in particular on company mergers and acquisitions. A Spanish lexicon of about 35,000 word senses has been mapped to concepts in the ontology. The ontology and the lexicon are currently being used to analyse Spanish news articles and to study a variety of phenomena in the literal and non-literal semantics of natural language expressions.

The top levels of the MIKROKOSMOS hierarchy are as follows:

```
ALL
        OBJECT
                PHYSICAL
                MENTAL
                SOCIAL
        EVENT
                PHYSICAL
                MENTAL
                SOCIAL
        PROPERTY
                ATTRIBUTE (a property of an
                        OBJECT or EVENT)
                RELATION (a relation between
                        OBJECTs and/or EVENTs)
```

PHYSICAL OBJECTS can be MATERIAL, SEPARABLE ENTITY or PLACE. MATERIALS are categorised according to various dimensions, thus:

FLUID/SOLID
ELEMENT/CHEMICAL COMPOUND
METAL/PLANT-DERIVED/ANIMAL-DERIVED
AGRICULTURAL/PHARMACUETICAL/CLEANSING/
    POISONOUS/ and other material groups
    classed according to origin, use, properties

SEPARABLE ENTITIES are categorised an ANIMATE or INANIMATE. One chain of the hierarchy in the ANIMATE class is:

ANIMAL-VERTEBRATE-MAMMAL-PRIMATE-
    HUMAN-COMPUTER USER

The INANIMATE class includes a big tree of ARTIFACTS.

MENTAL OBJECTS are ABSTRACT, a class that includes FIELD-OF-STUDY, a classification of the fields of human learning; or REPRESENTATIONAL, which includes NAME, MEDIA OBJECT, MATHEMATICAL OBJECT, MEASURE UNIT, LEGAL OBJECT, FINANCIAL OBJECT (e.g. ASSETS, MONEY, TAXES), etc.

SOCIAL OBJECTS are divided into ORGANISATIONS, which is divided according to the Standard Industrial Classification, and SOCIAL ROLES such as FAMILY, SERVICE, POLITICAL, RELIGIOUS, GOVERNMENTAL, FINANCIAL, and the professions generally.

PHYSICAL EVENTS include CHANGE LOCATION, CHANGE STATE, LIVING EVENT, APPLY FORCE, PERCEPTUAL EVENT, DISASTER, etc. MENTAL EVENTS can be EMOTIONAL, COGNITIVE, COMMUNICATIVE or PERCEPTUAL. SOCIAL EVENTS are divided into RELIGIOUS, POLITICAL, FINANCIAL, WORK ACTIVITY (e.g. MILITARY, MANUFACTURING, SERVICE, FARMING, ACADEMIC), and so on.

A given concept can occur at more than one point in the hierarchy, e.g. note that PERCEPTUAL EVENTS are regarded as both physical and mental. There is significant parallelism between several hierarchies in the ontology. The most visible parallelism in structure is between the set of three hierarchies in the social world: SOCIAL ROLE, WORK ACTIVITY and FIELD-OF-STUDY. Some of these three hierarchies are also in parallel with other SOCIAL OBJECT and SOCIAL EVENT structures.

In the MIKROKOSMOS material inspected, there is no exposition of the information provided for each concept in the ontology. Since it serves as an interlingua in MT, one may expect that the information is similar to that indicated earlier by Nirenburg [21]. For example, the entry for the animate object COMPUTER USER might contain pointers to its immediate superclass (HUMAN), to PROPERTIES of humans, to the related work activity (COMPUTING) and to 'case roles' in which the concept might figure, thus:

agent of: OPERATE, TAKE, PUT, FIND, SPEECH
    PROCESS, MENTAL PROCESS, EAT,
    INGEST, DRINK, MOVE, ATTACK . . .
experiencer of: FIND, MENTAL PROCESS, SPEECH
    PROCESS, ATTACK, TAKE, PUT . . .
destination of: SPEECH PROCESS . . .

The possible ‘case roles’ listed in MIKROKOSMOS are: agent, theme, instrument, experiencer, beneficiary, accompanier, purpose, location, source, destination, path. They serve as categories defining relations between concepts, as do the facts familiar in bibliographic classification.

Poli’s ontology

As already noted, Poli [18] concentrates on the categories of concept to be included in an ontology. He postulates at least five ontological ‘levels’:

- the inanimate physical world;
- the animate physical world;
- the psychological world;
- the social world;
- the world of ideas.

General categories apply to all levels, e.g. object, event, substratum, substance, form, relation, determination, dependence, structure, part, whole, unity, multiplicity, dimension, continuum, discrete, internal, external, identity, diversity, possibility, actuality, necessity, change.

Other categories apply to particular levels. Examples, given only for illustration, are:

- Inanimate: space, time, cause, situation, reciprocal action, dynamic structure, dynamic equilibrium, becoming;
- Animate: organic structure, adaptation, end-directedness, material exchange, self-regulation, species life, species degeneration;
- Psychological: act, content, consciousness, unconsciousness, pleasure, displeasure;
- Social: social system, family, community, conflict, class, institution, integration;
- Ideal: activities and products of knowledge, art, faith.

It is not clear from the paper cited how Poli would use these categories in the further development of an ontology or dictionary.

GALEN

Clinical medicine is inherently large and complex. As the demand has grown for wider coverage and new uses for terminologies is this field, some feel that traditional techniques of coding and classification are inadequate; they tend to ‘explode’ in size and become unwieldy, inconsistent and unmanageable. Advanced clinical systems need more than just terminologies, they need computer systems which can provide a sophisticated and appropriate set of terminological services, allowing applications to be developed to use whatever coding system or natural language local circumstances demand. Clinical application developers can therefore concentrate on the clinical tasks they must support, knowing that not only are the details of coding and classification abstracted for them but that they have access to a powerful model of clinical information to support their dialogue with clinical users.

To address the problems of clinical terminologies, project GALEN (Generalised Architecture for Languages, Encyclopaedias and Nomenclatures in Medicine) [23] is constructing a ‘semantically sound’ model of clinical terminology – the GALEN Coding Reference (CORE) model. This model comprises elementary clinical concepts such as ‘fracture’, ‘bone’, ‘left’, and ‘humerus’; relationships such as ‘fractures can occur in bones’, that control how elementary concepts may be combined; and complex concepts such as ‘fracture of the left humerus’ composed from simpler ones. This compositional approach allows for detailed descriptions while preserving the structure provided by the individual components. GALEN separates the model of the concepts from the natural language phrases used to refer to them (terms). The CORE model is intended to be language-independent and hence information entered in one language can be displayed in another. The natural language phrases for concepts are generated by a Multilingual Module within a Terminology Server, using the structure of the concept together with appropriate lexicons and grammar rules associated with the CORE model.

Existing medical coding schemes are important to GALEN. These schemes are widely used (and frequently mandatory) in current information systems and represent a large investment in expertise. Many schemes are detailed and aim for extensive clinical coverage. However, they typically lack the structure and formal basis that is necessary to meet the needs of advanced systems. GALEN relates to existing schemes by drawing on them to help to construct the CORE model; by mapping concepts in those schemes to structured concepts in the CORE model; by acting as an interlingua between schemes and thus supporting sophisticated code conversion; and by enhancing existing schemes by using the structure of the CORE model to derive new relationships and verify or correct existing ones.

ENTERPRISE

The ENTERPRISE ontology [24] is a collection of terms and definitions relevant to business enterprises. It is intended as a communication medium between people, between people and computer systems, and between systems. It contains about 100 defined terms, with added non-preferred synonyms and borderline terms. The defined terms are listed in five groups, of which four are shown in Table 1 (the fifth relates to Time).

Each term has an extended definition as to how it is to be used in the ontology. For example:

Activity: something that is done over a particular time period. It has Preconditions and Effects, is performed by one or more Doers, may have Subactivities, uses Resources, has Authority requirements, and has an Activity owner.

Superimposed on this ontology is a 'meta-ontology', a series of categories that serve to interrelate terms (comparable to the case roles of MIKROKOSMOS). In ENTERPRISE, an ontology is composed of a set of ENTITIES and a set of RELATIONS between ENTITIES, which have ROLES within the RELATIONS. An ATTRIBUTE is a special kind of RELATION, an ACTOR a special kind of ROLE. A STATE-OF-AFFAIRS is a situation in which any combination of ENTITIES is in any number of RELATIONS to each other. Examples of the categories are:

ENTITY: Plan
RELATION: Have-capability is a RELATION between Person and Activity; Sale is a RELATION between two legal-entities to exchange a Product for a Sale-price.
ROLE: Vendor is a ROLE played by a Legal-entity in a Sale RELATION.
ATTRIBUTE: Date-of-birth is an ATTRIBUTE associating only one date with a given Person.
ACTOR: a ROLE involving doing or cognition, e.g. the RELATION Perform-activity involves the ACTOR role Performer.

Building ontologies

Guidance as to the methods of building ontologies is given by some authors. The most detailed that I have come across is the discussion by Uschold and Gruninger [25]. Clearly, the domain to be covered by the ontology must first be decided, and a specification document may be drawn up. Concept terms are collected by scanning the literature of the domain, and by consulting domain experts. A 'brainstorming' session with experts is recommended, to produce significant terms and their relative importance.

Table 1
Four groups of defined terms in the ENTERPRISE ontology

ACTIVITIES AND PROCESSES	ORGANISATION
Activity	Person
Activity-specification	Machine
Execute	Corporation
Task-begin	Partnership
Task-end	Partner
Precondition	Legal entity
Effect	Organisational-unit
Doer	Manage
Subactivity	Delegate
Authority	Management-link
Activity-owner	Legal-ownership
Event	Nonlegal-ownership
Plan	Owner
Subplan	Asset
Planning	Stakeholder
Problem-specification	Employment-contract
Capability	Share
Skill	Shareholder
Resource	
Resource-allocation	
Resource-substitute	
STRATEGY	MARKETING
Purpose	Sale
Hold-purpose	Potential-sale
Intended-purpose	For-sale
Purpose-holder	Sale-offer
Strategic-purpose	Vendor
Objective	Actual-customer
Vision	Potential-customer
Mission	Retailer
Goal	Product
Help-achieve	Asking-price
Strategy	Sale-price
Strategic-planning	Market
Strategic-action	Market-segment
Decision	Segmentation-variable
Assumption	Market-research
Critical-assumption	Brand
Noncritical-assumption	Image
Influence-factor	Feature
Critical-success-factor	Need
Risk	Market-need
	Promotion
	Competitor

The collected terms may then be grouped, perhaps according to work areas in the domain, 'such that terms are more related to other terms within a group than they are to terms in other groups'. During this process, potential synonyms may be noted, and semantic cross-references, i.e. 'concepts that are likely to be referred to or be referred to by concepts in other groups'.

Then begins the work of producing definitions. In discussing this, the authors cited [25] introduce the idea of a 'meta-ontology' – a statement of the categories of term required in the ontology, as already illustrated in the ENTERPRISE ontology. They recommend 'initially, do not commit to any particular meta-ontology . . . Instead, let the careful consideration of the concepts and their interrelations determine the requirements for the meta-ontology'.

In creating definitions, 'start with the work areas that have the most semantic overlap with other areas . . . Define the most fundamental [or basic] terms in each area before moving on to more abstract [or general] and more specific terms'. The authors call this the 'middle-out' approach, and give an example: 'dog' may be regarded as a basic term, 'mammal' is then a generalisation, 'spaniel' is a specialisation. They comment: 'A bottom-up approach results in a high level of detail . . . and makes it difficult to spot commonality between related concepts . . . A top-down approach . . . can result in choosing and imposing arbitrary high-level meta-ontological categories', that later prove restrictive.

The following guidelines for developing definitions are proposed [25]:

- produce a natural-language text definition, as precise as possible;
- ensure consistency with terms already in use, making ample use of dictionaries, thesauri and technical glossaries;
- indicate the relationship with other similar terms (synonyms and variants);
- avoid circularity in definitions.

Mahesh [20] suggests a number of principles for 'ontological engineering', some of which are summarised as follows.

(1) Not unique: there is no unique ontology of the world (or even of a narrow domain). Ontologies are not natural entities to be discovered, they are artifacts to be constructed (and their structure will depend on their intended use).
(2) Task specific: an ontology is built for a specific task. An ontology built for natural language processing may not be suitable for some other task such as planning, design, or other reasoning. The types of knowledge that an ontology must contain differ from task to task.
(3) Usability: it must be easy to browse the ontology, find the right concept to map a word, see the definition/description of that concept in its entirety, and visualise the organisation (taxonomic or otherwise) of that concept relative to other concepts in the ontology.
(4) Modularity: it must be possible to add new concepts and conceptual relations incrementally.
(5) Grain size: building a precise ontology where each concept is clearly distinguishable from others requires a fine-grained decomposition of concepts. The choice of granularity must be determined on the basis of utility to the task involved.
(6) Redundancy: the classification of concepts in an ontology is necessarily redundant. Multiple dimensions of classifying a set of concepts are often overlapping. It is futile to try to eliminate such ontological redundancy entirely from an ontology of non-trivial size.

Guarino writes [19]:

> One of the main motivations of the recent emphasis on ontology building is the possibility of knowledge sharing and re-use across different applications: as soon as a particular domain (such as medicine or automotive manufacturing) is fixed, it seems reasonable to expect a large part of domain knowledge to be the same for a variety of applications, so that the high costs of knowledge acquisition can be better justified.

He defends the thesis of the independence of domain knowledge:

> This thesis should not be intended in a rigid sense, since it is clear that – more or less – ontological commitments always reflect particular points of view (for instance, the same physical phenomenon may be described in different ways by an engineer, by a physicist or by a chemist). Rather, what I would like to stress is the fact that reusability across multiple tasks or methods can and should be systematically pursued even when one is modelling knowledge related to a single task or method: the closer we get to the intrinsic, task-independent aspects of a given piece of reality . . . the more this knowledge can be re-used for different tasks.

He links granularity to generality:

> What is dependent on the particular task at hand is the granularity of the domain knowledge used. Consider for example the case of tasks as different as car driving, car maintenance, car troubleshooting and car repairing: it is evident that, although some specialized bodies of knowledge (such as traffic regulation rules or good maintenance practices) are only relevant for a single task, all of the tasks above may make use of the same basic knowledge related to the car structure, provided its granularity is fine enough. This means that, when modelling domain knowledge with a single task in mind, reusability may be pursued by paying the cost of higher granularity and generality.

Some uses of ontologies

An ontology may be regarded as a database with information about what categories and/or concepts exist in the world/domain, what properties they have, and how they relate to one another.

In interlingual MT, the principal reasons for using an ontology are to provide a grounding for representing text meaning in an interlingua; to enable lexicons for different languages to share knowledge; to enable source language analysers and target language generators to share knowledge; and to resolve semantic ambiguities and interpret non-literal language by making inferences using the topology of the ontology to measure the semantic affinity between meanings [20].

An ontology can also be of great value in a variety of other tasks such as database merging or integration of software or business enterprise models. Essentially, an ontology is invaluable wherever a 'semantic wall' is to be scaled, i.e. a situation where there are two or more systems that overlap conceptually, but have differing knowledge representations. Thus, it may be used to translate between a pair of natural languages or a pair of database schemas, or to integrate different models of the same domain or similar phenomena in the world.

An example of ontology use is in spacecraft mission operation [26]. Various KBSs were developed to assist in different aspects of this, each with its own structured knowledge base. To integrate the systems, an ontology as a unifying framework was constructed, and a software 'agent' embodying the ontology acts as an interlingua to make knowledge in each system available to the others.

The KATUS project [27] is investigating the notion of ontologies as a mechanism to specify explicitly the concepts, relations, axioms and constraints present in an electrical network, and that are shared by many applications in the domain. This research involves the development of an ontology for the electrical network and its use in the development of two applications: diagnosis of faults and service recovery planning. The long-term objective of this research is to obtain a standard representation of the network that can be shared and reused across many applications and electrical companies.

The PLINIUS project [28] aims at semi-automatic extraction of knowledge from natural language texts, particularly in the field of the mechanical properties of ceramic materials. A lexicon is used to map natural language expressions into representations using the conceptual terms of the ontology.

Uschold and Gruninger [25] suggest 'the need to develop more expressive ontologies for activities/processes, resources, products, services and organisations'. This, they write, would widen the scope for application of ontologies in such domains as materials science and engineering, petrochemical and plastics industries, and medicine. There is a need to construct libraries of ontologies which can be reused, customised, and adapted to different general classes of problems and environments.

Commentary

Summing up what has been set out above, we can describe an ontology as a schedule, in some form that may involve the use of semantic categories, of concepts significant in a particular domain (that may be as wide as the universe of knowledge), together with a definition or scope note for each concept, and mechanisms for displaying its relationships to other concepts.

The analogy with bibliographic classifications and thesauri is obvious, although there are equally obvious differences because the uses intended for ontologies are not the same as for classifications and thesauri. Despite the differences, it is to be regretted that the 'ontological engineers' make little or no reference to work in information science. As a consequence, they do not appear to draw at all on the rich experience of constructing knowledge schedules that is summarised in such texts as *Theory of Subject Analysis* [29] and *Vocabulary Control for Information Retrieval* [30], or on the experience of developing concept lexicons in intelligent online search aids [31].

The linkage of ontologies to lexicography is also obvious – particularly as several ontologies are designed for use in MT – but, except in Guarino [12], there is not much reference to such work as *Relational Models of the Lexicon* [32], and only limited reference to the long history of the study of semantic relations [33].

Indeed, except for Sowa, who looks briefly at the history of categories in philosophy [34], there is not much appreciation that constructing an ontology is not such a new activity. For example, in 1668, John Wilkins published 'a regular enumeration and description of all the things and notions, to which marks or names ought to be assigned', set out in hierarchical 'Philosophical Tables' [35]. This directly influenced the well-known thesaurus produced in 1852 by Roget. Much valuable detail of this kind of development can be found in McArthur's historical review of 'the long effort involved in knowing, and struggling to retain what we think we know' [36].

Let us consider some of the issues that have been raised by those working in ontological engineering.

There is first the conflict between generalisation and specialisation. The considerable emphasis on the potential reuse of ontologies, and their value as communication links between specialist agents (links similar to the 'switching systems' often discussed in our field), urges that they be task-independent, standard inventories of concepts. However, as in our field, the desire to mould an ontology to the needs of a particular user group urges that it be tailor-made.

Guarino [19] stresses the importance of achieving as much task-independence as possible (for the purpose of reuse), implying that it is possible to arrive at a standard 'unique' ontology of a certain subset of knowledge, that can be used in several or all domains. On the other hand, Mahesh [20] stresses the importance of making an ontology task-specific, and urges that 'there is no unique ontology'. This is a long-standing issue in classification theory – within biological taxonomy as well as within information science. As Guarino's argument (already quoted) suggests, classifiers seek to construct a standard schedule (say, for plant types) wherever commonality of purpose in use can be assumed, but provide varying schedules to meet differing purposes (say, for cultivated plants or medicinal plants). This relates to what Mahesh calls 'redundancy' and CYC [10] calls differing 'microlattices': the need to include sub-schedules containing the same concept terms, but in different contexts.

The general ontologies commonly imply the existence of 'levels' of conceptual entity. Poli [18] does this explicitly, and MIKROKOSMOS [22], by dividing objects into physical, mental and social, does it implicitly. This is reminiscent of the 'integrative levels' that have influenced some bibliographic classifiers [37].

Lastly, I will mention interrelations between concepts, as expressed by the case roles of MIKROKOSMOS [22], the meta-ontology of ENTERPRISE [24] and the relations of GALEN [23]. Such relations are built into the very structure of a faceted classification, which may group its terms into such categories as thing, part, substance, process, property, operation. Uschold and Gruninger [25] urge that, initially, one should 'not commit to any particular meta-ontology . . . Instead, let the careful consideration of the concepts and their interrelations determine the requirements for the meta-ontology'. There is much analogous discussion of the techniques of facet analysis and formulation of facet categories in information science literature [4, 38].

In conclusion: this new development of tools at the 'knowledge level' shows once again the growing understanding of the importance of semantic analysis in information processing. The problems with which information scientists have for so long been struggling are now being faced by a wider community of knowledge engineers. It is to be hoped that all involved will continue to learn from each other's experience. Soergel's proposed SEMWEB – 'an open, multifunctional, multilingual system for integrated access to knowledge about concepts and terminology' [39] – could provide a helpful integrating background to work at the knowledge level.

References

[1] R. Green (ed.), *Knowledge Organization and Change: Proceedings of the Fourth International ISKO Conference, 15–18 July 1996, Washington* (Indeks Verlag, Frankfurt, 1996).

[2] J.H. Alexander *et al.*, Knowledge level engineering: ontological analysis, *Proceedings of AAAI* 86 (1986) 963–968.

[3] A. Newell, The knowledge level, *Artificial Intelligence* 18 (1982) 87–127.

[4] S.R. Ranganathan, *Prolegomena to Library Classification* 3rd ed. (Asia Publishing House, London, 1967).

[5] A.C. Scott *et al.*, *Practical Guide to Knowledge Acquisition* (Addison-Wesley, Reading, MA, 1991).

[6] K.L. McGraw and K. Harbison-Briggs, *Knowledge Acquisition* (Prentice-Hall, Englewood Cliffs, NJ, 1989).

[7] W.J. Clancey, The knowledge level reinterpreted, *International Journal of Intelligent Systems* 8 (1993) 33–49.

[8] J.F. Sowa, *Conceptual Structures* (Addison-Wesley, Reading, MA, 1983).

[9] D. Lenat and R.V. Guha, *Building Large Knowledge-based Systems* (Addison-Wesley, Reading, MA, 1990).

[10] Information on CYC is available at the Internet site: http://www.cyc.com/

[11] P. Hayes, The second naive physics manifesto. In: J.R. Hobbs and R.C. Moore (eds), *Formal Theories of the Commonsense World* (Ablex, Norwood, 1985), pp. 1–36.

[12] N. Guarino, Formal ontology, conceptual analysis and knowledge representation, *International Journal of Human-Computer Studies* 43 (1995) 625–640.

[13] A good starting point is the site maintained by E. Franconi at: http://mnemosyne.itc.it:1024/

[14] T.R. Gruber, *What is an Ontology?* Available at Internet site: http://www-ksl.stanford.edu/people/gruber/

[15] T.R. Gruber, A translation approach to portable ontology specifications, *Knowledge Acquisition* 5 (1993) 199–220.

[16] T.R. Gruber, Toward principles for the design of ontologies used for knowledge sharing, *International Journal of Human-Computer Studies* 43 (1995) 907–928.

[17] M. Uschold, *Building Ontologies: Towards a Unified Methodology*. Presented at Expert Systems '96 Conference and available from Internet site: ftp://ftp.aiai.ed.ac.uk/pub/documents/1996/

[18] R. Poli, Ontology for knowledge organisation. In: R. Green (ed.), *Knowledge Organization and Change: Proceedings of the Fourth International ISKO Conference, 15–18 July 1996, Washington* (Indeks Verlag, Frankfurt, 1996), pp. 313–319.

[19] N. Guarino, *Understanding, Building, and Using Ontologies*. Available from Internet site: http://www.ladseb.pd.cnr.it/infor/Ontology/

[20] K. Mahesh, *Ontologies for Natural Language Processing*. Available from Internet site: http://cr1/nmsu.edu/users/mahesh/

[21] S. Nirenburg *et al.,* The structure of interlingua in TRANSLATOR. In: S. Nirenburg (ed.), *Machine Translation* (Cambridge University Press, 1987), pp. 90–113.

[22] Material on MIKROKOSMOS is available at Internet site: http://crl.nmsu.edu/users/mahesh/

[23] Information on GALEN is available at Internet site: http://www.cs.man.ac.uk/mig/galen/

[24] Information about ENTERPRISE is available from the Internet site: http://www.aiai.ed.ac.uk/~entprise/enterprise/

[25] M. Uschold and M. Gruninger, Ontologies: principles, methods and applications, *Knowledge Engineering Review* 11(2) (1996) 93–136.

[26] M.W. Jones *et al.*, An agent-based approach to spacecraft mission operation. In: N.J. Mars (ed.), *Towards Very Large Knowledge Bases* (IOS Press, Amsterdam, 1995), pp. 259–269.

[27] Information on KAKTUS is available from Internet site: http://www.swi.psy.uva.nl/projects/Kactus/

[28] P.E. van der Vet and N.J. Mars, Ontologies for very large knowledge bases in materials science. In: N.J. Mars (ed.), *Towards Very Large Knowledge Bases* (IOS Press, Amsterdam, 1995), pp. 73–83.

[29] L.M. Chan *et al.* (eds), *Theory of Subject Analysis* (Libraries Unlimited, Littleton, CO, 1985).

[30] F.W. Lancaster 2nd ed. (Information Resources Press, Arlington, VA, 1986), *Vocabulary Control for Information Retrieval.*

[31] B.C. Vickery and A. Vickery, Online search interface design, *Journal of Documentation* 49(2) (1993) 103–187.

[32] M.W. Evens (ed.), *Relational Models of the Lexicon* (Cambridge University Press, 1988).

[33] B.C. Vickery, Conceptual relations in information systems, *Journal of Documentation* 52 (1996) 198–200.

[34] J.F. Sowa, Top-level ontological categories, *International Journal of Human-Computer Studies* 43 (1995) 669–685.

[35] B.C. Vickery, The significance of John Wilkins in the history of bibliographic classification, *Libri* 2 (1953) 326–343.

[36] T. McArthur, *Worlds of Reference* (Cambridge University Press, 1986).

[37] D.J. Foskett, Classification and integrative levels. In: L.M. Chan *et al.* (eds), *Theory of Subject Analysis* (Libraries Unlimited, Littleton, CO, 1985), pp. 212–220.

[38] B.C. Vickery, *Classification and Indexing in Science* 3rd ed. (Butterworths, London, 1975).

[39] D. Soergel, SemWeb: proposal for an open, multifunctional, multilingual system for integrated access to knowledge about concepts and terminology. In: R. Green (ed.), *Knowledge Organization and Change: Proceedings of the Fourth International ISKO Conference, 15–18 July 1996, Washington* (Indeks Verlag, Frankfurt, 1996), pp. 165–173.