Concepts, Ontologies, and Knowledge Representation

SANIDA OMEROVIC
University of Ljubljana
VELJKO MILUTINOVIC
University of Belgrade
and
SASO TOMAZIC
University of Ljubljana

Recording knowledge in a common framework that would make it possible to seamlessly share global knowledge remains a central challenge for researchers. This annotated survey of the literature examines ideas about concepts, ontologies, and knowledge representation that address this challenge. A separate section to each of the above topics is dedicated following a uniform outline: definition, organization, and use.

General Terms: Concepts, Knowledge, Meaning, Modeling, Ontology, Semantics
Additional Key Words and Phrases: data, relations, representation

---

1. INTRODUCTION

The information world that we live in today presents us with a vast amount of data stored separately in books, newspapers, radio, TV, Internet, etc., all of them increasingly digitized. Moreover, there is an exponential increase in these data day to day so that the ability of an average computer-educated person to find a specific data element or subject-related useful piece of information is decreasing rapidly. As an example of an inadequate response, most text-search engines only find matches based on keywords without regard to their various meanings [GAUCH 2002]. This poses several central questions: How can one efficiently extract the desired data from a huge data source? How can a searcher find a necessary and potentially available, but unknown, piece of knowledge that represents the answer to some question or that helps to resolve some problem? How complex must the underlying records be? It has been asserted that if the structure and function of all organisms that live or have lived on earth can be coded by triplet sequences of four nitrogenous base pairs A, G, T and C, there is no reason for a knowledge record to be more complex than this [NOVAK 2007]. (However, pursuing the analogy, such records must be imbedded in an appropriate highly reliable processing environment such as the cell.)

This research was supported by the Slovenian Research Agency under the program "Algorithms and optimization methods in telecommunications".
Authors' addresses: Sanida Omerovic, Department of Telecommunications, The University of .Ljubljana, Slovenia; Veljko Milutinovic, Department of Computer Science, The University of Belgrade, Serbia; Saso Tomazic, Department of Telecommunication, The University of Ljubljana, Slovenia.

To answer such key questions, it is useful to proceed from well known problems. For example, most knowledge retrieval systems in the text domain suffer from the following shortcomings:

- An inability to recognize ambiguities in the terms used in their object descriptors.
- An inability to query effectively in an uncontrolled text environment [CHUA94].

Here we follow a widespread general agreement, among most of the authors referred to in this survey that uniform knowledge representation should be achievable by the use of ontologies, populated with concepts, as indicated in Figure 1. We extend this hypothesis to propose that the answer to efficient knowledge retrieval lies in the use of such imbedded concepts and ontologies as keys to discriminate processing.

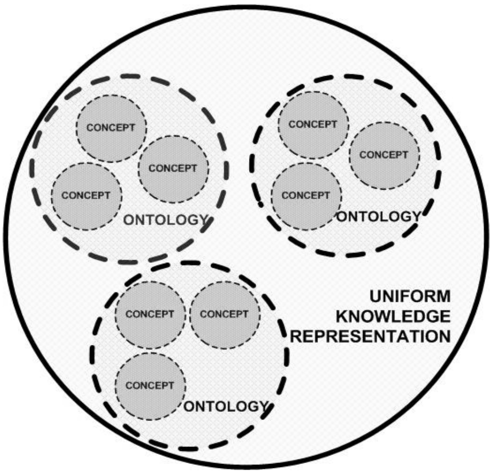

Figure 1. Uniform knowledge representation model, consisting of ontologies that are populated by concepts.

The brain remains the first and best human knowledge storage locality. Actually, one can observe the brain as "hardware" and our thoughts as "software" operating on such "hardware" [QUINZ 2004]. But there is still no machine that can simulate the efficient way that the human brain thinks. However, some research efforts are made in that direction. [HAWKINS 2007] created a software platform that simulates the work of brain's neocortex (neocortex represents a thin sheet of cells which is responsible for almost all high-level thoughts and perception in human brain). On the other hand, [DEACON 1998] analyses the essence of a thinking process in language. Language is not merely a mode of communication; it is also the outward expression of an unusual mode of thought – symbolic representation. Without symbolization, the entire virtual world (telling stories about our real experiences, inventing stories about imagined ones, pondering what it will be like not to be) is out of reach – inconceivable. So species that have not acquired the ability to communicate symbolically cannot have acquired ability to think this way either. Of course, we are far away of understanding in total the algorithm of thinking embedded in the human brain; however, certain components are recognizable:

People use concepts every day to express their thoughts, although there is no unique definition of concept or a commonly accepted agreement of what a concept is. Nevertheless, we understand by observation how they are used in human communication to carry a circumscribed meaning (for example: a house, a dog, a car, or some more abstract idea). Still, one does not know how concepts are derived from everyday perception or learned knowledge. For every person, concept derivation appears to be unique. Because of the importance of expressing specific delimited meanings in knowledge representation, the first section of this survey is dedicated to various approaches to concepts.

Concepts alone are not enough. Grouping related concepts into ontologies has proved to be a very efficient way to capture and structure meaning within natural languages [DAML 2007]. It is a convenient means of uniting a subject, a relationship, and an object to talk about. For example, one is able to present an abstract concept of a ***person*** by means of ontologies using the Ontology Web Language OWL [OWL 2004] with datatype properties such as: firstName, lastName, gender, birthday, homeAddress, officeAddress, email, cellPhone, fax, pager, homepage, etc. Because ontologies thus enable meaning to be captured in a uniform manner, they become the essence of successful knowledge representation. Therefore, the second section of this survey is dedicated to different views of ontologies.

Knowledge as usually presented arrives unstructured in a non-uniform manner making it unsuitable for further joint processing (see, for example, the numerous related but incompatible computerized record systems present within every government department or business organization). The main goal of introducing appropriate formal schemes for concept and ontology is to so structure knowledge as to make it shareable among both computers and people. As a consequence, this survey concludes with a section specifically dedicated to work on various forms of knowledge representation that are under-girded by concepts and ontologies.

2. CONCEPTS

A concept is an entity of consciousness. We know a concept when we see one in action because it exceeds its stand-in descriptive label as a word, phrase, sentence or paragraph. It might be directly conceived or an intuited object of thought. In general, every object, issue, idea, person, process, place, etc. can generate a concept. Embedded in language, concepts can migrate to incorporate new phenomena as they arise – leading to an evolution in their meaning over time. This malleability is both the strength and weakness of language, which lacks the precision of mathematics. (It is very apparent, for example, in every attempt to write an unambiguous law or contract.) Concept malleability is one of the underlying issues in creating a universal framework for the exchange of knowledge. Thus the goal of information computing is to unite the flexibility of language with the strict definitions of – for example – mathematics or description logics, through the application of specific measures such as statistical weighting, neural nets, and various approaches to fuzzy logic.

2.1 Concept definition

The substance of a concept may be abstract or concrete, elementary or composite, real or fictitious. A concept can be anything about which something is thought, for example, a neural excitation induced by an object linked with an object name [SOWA 2000] (this neural excitation is illustrated in Figure 2). Concepts can describe a task, function, action, strategy, reasoning process etc [PEREZ 2002], or to be expressed in terms of other concepts [VOSS 1999]. In order to manage all such concept types they must be assigned common formal properties.

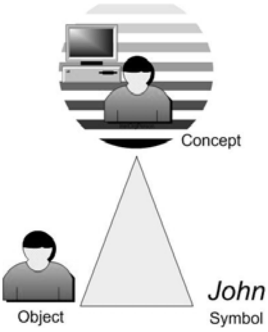

Figure 2. Concept presentation. On the lower left is an icon that resembles a person named John. On the lower right is a printed symbol that represents a person's name. The cloud on the top designates the neural excitation induced by John working at his office. This excitation is called a *concept*.

There are many view points from where one can define concepts. We have chosen to present possible concept definitions on the basis of criterion whether the concepts are implicit or explicit.

2.1.1. Implicit concept definition

If concepts can be recognized, but cannot be defined explicitly, how can a machine, for example, distinguish which words in a text are concepts, and which are not? What terms carry more "meaning" than the others? How can one make concepts recognizable, so that they are automatically extractable from any type of texts?

A Vector Space Model (VSM) [SALTON 1975] presents a possible answer to the above questions. Each document processed in VSM is in the form of a vector with its coordinates representing the values of occurrences of the index terms in that document (Figure 3a). Similarity level is measured (it might be, for example, the angle between two vectors) when the new index term is assigned to a document collection (Figure 3b.). If the similarity level decreases, a new assigned index term has a "good" discriminating property. The reverse holds for a "bad" index term.

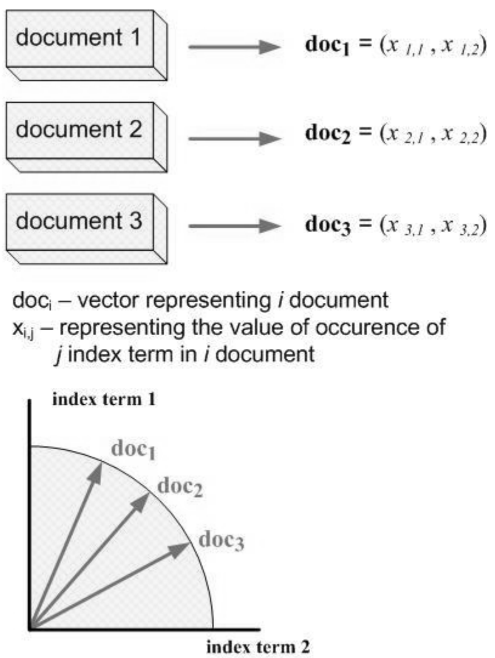

Figure 3a. Each indexed document in the Vector Space Model is represented by a j-dimensional vector (j - number of different index terms). Documents are indexed by index terms and reside within the planes defined by index term axes. Depending on j (the number of index terms), vectors can reside in j-dimensional space within the sphere.

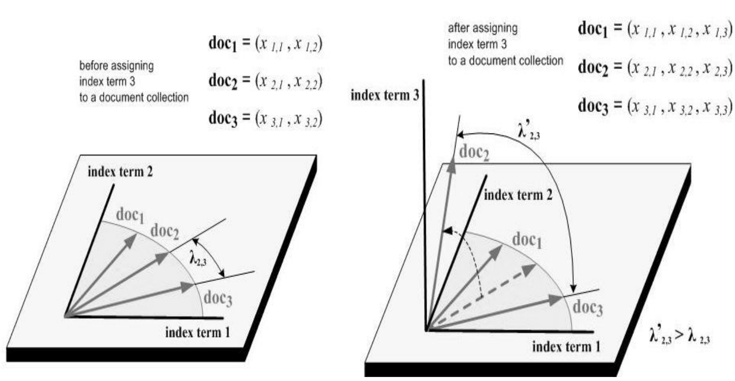

Figure 3b. Operation on a “good” discriminating term in Vector Space Model. The similarity measure is taken as the inverse function of the angle between two corresponding vector pairs (when the angle between two vectors is zero, the similarity function is at a maximum; and vice versa). Before assigning index term 3, three vector documents reside on one plain, formed by the axes of index term 1 and index term 2. After including an index term 3 to a collection of documents, a new dimension is added to a vector space. Therefore, the coordinates of all three vectors are changed, the corresponding angles have increased, and the similarity measure is decreased.

Therefore, “good” index terms can be recognized as concepts, since they represent the smallest units of knowledge that carry the most of meaning. (enough “meaning” to decrease the similarity level between the documents when assigned to a document collection).

2.1.2. Explicit concept definition

The explicit definition of concepts can be made to depend upon the way concepts are used in such areas as simulated knowledge, description logic, and concept maps.

In [HALLADAY 2004] the atoms of Simulated Knowledge are its concepts. For a potential concept to be an atom, it must be dynamically capable of abstraction into higher entity forms, relationships and/or processes. In addition its meaning must be both syntactically and semantically interdependent, but language independent.

Description logic (a formalism for representing logic-based knowledge) is based on concepts (classes) and roles. Concepts are interpreted as sets of objects and roles as binary relations between objects [NAKABASAMI 2002].

Concept maps are mostly used for representing already organized knowledge. Here, concepts are defined as a perceived regularity among events or objects, or records of events or objects, designated by a label [NOVAK 2005]. The label for most concepts is a word or symbol (a similar idea is presented in [SOWA 2000]).

This subsection has presented several possible concept definitions with respect to simulated knowledge, description logic, and concept maps. Once so defined, the next subsection discusses issues of how concepts can be organized and thus be made predictably available for use.

2.2 Concept organization

[ZELLWEGER 2003] presents concept organization in a database as a data item with its cross-data relationships. Some data modeling techniques that aim in that direction are:

- An entity-relationship model [CHEN 2007],
- A Unified Modeling Language [UML 2007], and
- An object-role modeling [HALPINO 2007].

The above three techniques share the same basic concept structure:data items exhibit connections among different neighboring data. The essence of grasping meaning lies in the ability to assign a verbal explanation to these connections. Such a verbal explanation of the relationships between data items provides a conceptual model for the linked data items. The importance of identifying relationships between data is given in an illustrative example in [HALLADAY 2004]:

> "The statement "John has an IQ of 150" explicitly describes only a very simple relationship (i.e, that John has some attribute named IQ that equals 150). However, the statement assumes a set of other implicit relationships (like IQ being an acronym for Intelligence Quotient, or that 150 is a value that precedes 151 and is proceeded by 149, or that John is a common human male name, or that an IQ equal to 150 indicates a person of above-average intelligence, etc). However, without the context of all these relationships, the statement loses some of its meaning. In fact, meaning is the sum-total of relationships."

As discussed in [HALLADAY 2004], among the major issues in concept organization are relations. Relations can be treated as interactions between the concepts of a domain and their attributes [PEREZ 2002] (Figure 4).

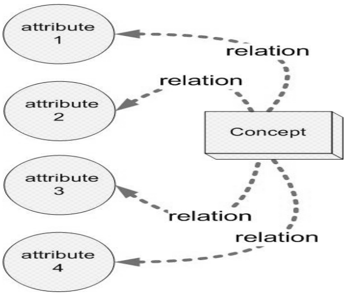

Figure 4. Relations are defined as mediators between concepts and their attributes. Attributes are classified in the following four groups, depending on their relation to a specific concept: Instance attributes - its value might be different for each instance of the concept; Class attributes - its value is attached to the concept, meaning that its value will be the same for all instances of the concept; Local attributes - same-name attributes that attach to different concepts; Global attributes - its domain is not specified and can be applied to any concept in the ontology.

Another viewpoint presenting the importance of attributes in concept organization is [Han 1996]. The Knowledge Rich Data Base KRDB consists of concepts that are presented as a set of entities and relationships. Entities together with their attributes could be defined by:

- physical data (i.e.,data relations),
- virtual data (i.e., deduction rules), or
- a mixture of physical and virtual data.

A fragment of KRDB Schema organization is presented in Figure 9.

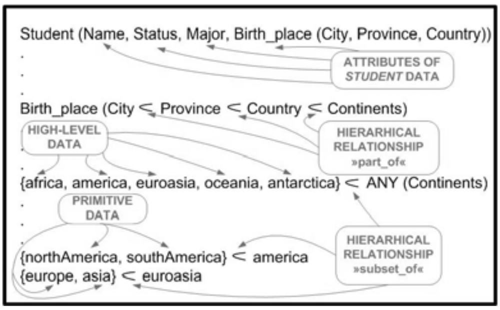

Figure 9. A fragment of a Knowledge-Rich Database Schema for *Student* as high-level data. Primitive data represents the original input data. High-level data represent a

superset of primitive data and in contrast to primitive data, can contain attributes.
Possible types of hierarchical relationships between data are: *part_of*, *is_a*, *subset_of*, etc.
Such semantical relationships enable query intent analysis and intelligent query
answering, which are suitable further for knowledge retrieval processing.

A great deal of interface design research has been devoted to determining mechanisms for making productivity tools (e.g., word processors and drawing tools) easy to use and intuitive so that users can perform a given task more smoothly and efficiently. Graphical, textual, and visual presentation are some of the possible mechanisms that require different concept organizations.

2.1.1. Graphical presentation

Two possible solutions to the graphical interface design are conceptual graphs and concept maps. A Conceptual Graph (CG) [SOWA 1999] contains only two kinds of nodes: concepts and conceptual relations, as presented in Figure 5. [CHEIN 1992] defined mathematically the Conceptual Graph operations.

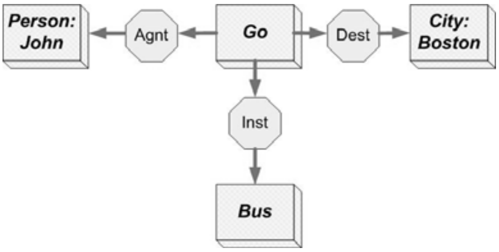

Figure 5. Conceptual Graph (CG) representing the propositional content of the English sentence: *John is going to Boston by bus.* Concepts are presented by 3-D boxes, and conceptual relations are presented by octagons. Every arc in CG must link a conceptual relation to a concept: *Go* has an *agent* (Agnt) which is a person *John*; *Go* has a *destination* (Dest) which is a city *Boston*; *Go* has an *instrument* (Inst) which is a *bus*.

Concept maps present graphical tools for organizing and representing knowledge [NOVAK 2005]. They include concepts, usually enclosed in circles or boxes of some type, and relationships between concepts indicated by a connecting line linking two concepts. Words on the line – referred to as linking words or linking phrases – specify the relationship between two concepts, as shown in Figure 6.

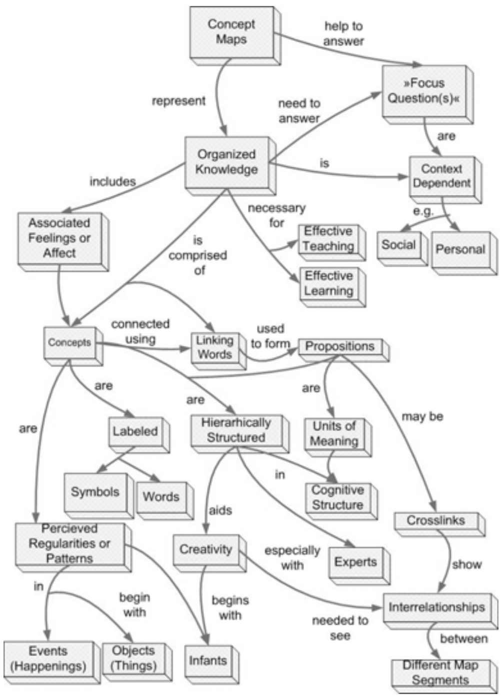

Figure 6. A concept map graphical presentation. Concepts are enclosed in 3-D boxes, and relationships between concepts are presented by arcs linking two concepts. Words on the arcs are referred to as linking words or linking phrases, specifying the relationship between the two concepts. Propositions are statements about some object or event in the universe, either naturally occurring or constructed. Propositions usually contain two or more concepts connected using linking words or phrases to form a meaningful statement usually called semantic unit, or unit of meaning. Concept maps tend to be read proceeding from the top downwards.

An example of a software package intended to search for a graph in a database of graphs is GraphGrep [Giugno 2007]. Within given collection of graphs and a pattern graph as the query input, GraphGrep is able to find occurrences of the input pattern in each database graph. The input pattern represents a sub graph and it can also be a tree, a path, or a node. Graphical presentation can serve as a foundation for knowledge discovery systems. SUBDUE [COOK 2007] system represents data using labeled, directed graphs. It finds structural, relational patterns in the observed data.

2.2.2. Textual presentation

Conceptual indexing [WOODS 1997] presents one possible form of textual presentation, based on concepts. In conceptual indexing, instead of alphabetical indexing, phrases are indexed by their meaning. This is done by automatically parsing each phrase into one or more conceptual structures. Each conceptual structure represents the way the elements of the phrase are assembled to construct its meaning(s). An example of a conceptually parsed document is presented in Figure 7.

```
brokers
        automobile brokers
        truck brokers
cleaning
        automobile cleaning
            automobile steam cleaning
            automobile upholstery cleaning
            automobile washing
                    car washing
        industrial cleaning
            industrial steam cleaning
        steam cleaning
            automobile steam cleaning
            industrial steam cleaning
        upholstery cleaning
            automobile upholstery cleaning
        washing
            automobile washing
                    car washing
```

Figure 7. A fragment of Conceptual Indexing Taxonomy in a text document related to automobiles. The system first automatically parses each phrase into one or more conceptual structures. Then automatically determines when the meaning of one phrase is more general than another, given that it knows about the generality relationships among the individual elements that make up the phrase. For example, a system can automatically determine that *car washing* is a kind of *automobile cleaning* if it has the information that a *car* is a kind of *automobile* and that *washing* is a type of *cleaning*.

One practical implementation of a conceptual indexing textual presentation is thesauri. To illustrate, the International Nuclear Information System thesaurus [INIS 1981] contains information about applications in nuclear science and technology. The record for each chunk of literature consists of three main components:

- A bibliographic description - identifying authorship, publishing, and similar ideas.
- A set of descriptors - identifying the subject content in a piece of literature.
- An abstract - summarizing the information contained in the piece of literature.

In textual presentation as discussed in [VOSS 1999] concepts are created by a user marking pieces of text in documents. Therefore, concepts are not formally defined, but one must interpret the concepts in the context of their occurrences and their use in the documents. Two simple relations in concept organization are supported by the users:

- A “comprise” relation - grouping several concepts into new concept.
- An “associated” relation - two concepts are simply seen to be closely associated, but not necessarily grouped into another concept.

2.2.2. Visual presentation

The image retrieval system based on concepts is presented in [CHUA 1994]. Every image is described with concepts, where each concept contains its descriptor and the relationships with other concepts, as shown in Figure 8.

```
SP-Chinese :==(chinese, hokkien, hakka, catonese,…);
SP-SkillJob :==(carpenter, tailor, barber, blacksmith, …);
SP-UnskillJob :==(hawker, assistant, labourer, servant);
SP-Job :== (occupation, job, SP-SkilJob, SP-UnskillJob);
SP-Occupation :==(occupation, rule, conquer, SP-Japanese);
…

Concept Occupation-of-Chinese
  Description: occupation of early Chinese immigrants in Singapore
  Component: SP-Chinese, SP-Job,…;
  Parent: Chinese-Immigrant
  Child: Rickshawmen, Merchant, Labourer,…;
  Synonym: null
  QueryProfile: null;
End of Concept;
…
```

Figure 8. The concept structure for image descriptors consists of two nodes: Semantic Primitives (SP) and Concept Nodes. SP represents the sense and use of a term or phrase in a domain. SPs are used to reduce the variability of vocabulary used by the users when issuing queries. Most important terms or phrases in the domain have their corresponding SPs to map to. Since a term may have different meaning in different context, the mapping of some terms to their corresponding SPs may be conditional. One or more SPs constitute CN.

In summary, this subsection has presented various uniform concept organizations applied to the areas like graphical, textual, and visual presentation. The following subsection is dedicated to how concepts when organized in a uniform manner can be are in practice.

2.3 Concept use

Conceptualization – as a process to structure concepts for practical use – involves an understanding of both semantics and contexts [FUJIHARA 1997]. In text analysis, concepts can be used to ensure that only meaningful and non-ambiguous terms become text descriptors. Standard techniques based on keywords text analysis use two basic measures: the frequency of words and the distance between words. A weighted terms approach [ROBERTSON 1994] can also be introduced, where each term-document combination is given a certain weighted value (document retrieval then depends upon the corresponding input query, word frequency, word distance, and the weight values). A concept-based search is a step ahead of keyword search, and will be the main focus of this subsection related to concept use.

2.3.1. Concept-based search

In concept-based searches all comparable keywords converge to the same concept. Progress over time in achieving this [SCHATZ 1997] is traced in Figure 10. In a typical situation the user enters a query using keywords that are recognizable in some, but not all, domains of the document collection (because different domains describe similar concepts using different terminologies). For example, if one says: “I have a doctorate” and “I am a PhD” these are two different semantic statements, but they map to the same concept. Obviously, for efficient concept-based retrieval a system must perform vocabulary switching to automatically translate terms across different domains.

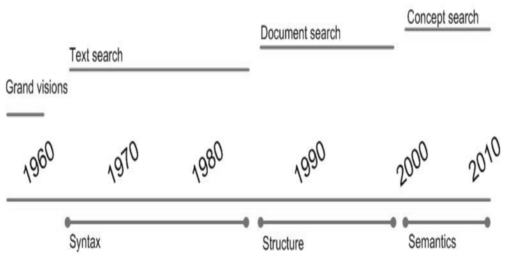

Figure 10. Rough timeline vision of the concept-searching evolution from 1960 till 2010.

Three types of concept-base search systems are presented next: KeyConcept, the Automated Generated Thesaurus Approach, and the Semantic Web.

2.3.1.1. KeyConcept

An example of a concept-based retrieval engine is KeyConcept [GAUCH 2002]. There, concepts are first processed by a traditional indexer using training documents under a ***tf.idf*** indexing method (Figure 11). The outputs of the ***tf.idf*** method are concepts where each concept is presented as a centroid of the training set of documents for the observed concept.

$$tf.idf_{i,d} = tf_{i,d} \times idf_i$$

$$idf_i = \log(n/df_i)$$

$tf_{i,d}$ (term frequency) = some measure of term ***i*** density in document ***j***

idf_i (inverse document frequency) = some measure of informativeness of a term ***i*** in the collection of documents

df_i = the number of document that contain term ***i***

n = total number of documents

Figure 11. The *tf.itf* indexing method.

The conceptual indexer of KeyConcept processes new (non training) documents by using a Vector Space Model (VSM) [SALTON 1975]. Concept-based retrieval is done by processing similarity level values obtained from VSM, words/concepts and an L-factor (entered by the user). The KeyConcept system architecture is presented in Figure 12.

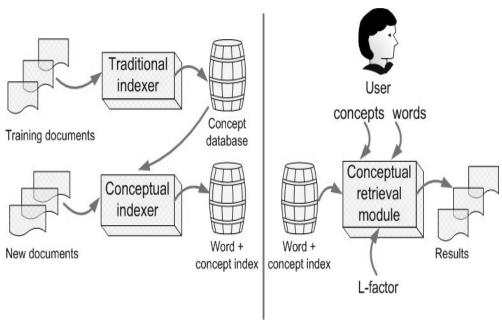

Figure 12. KeyConcept system architecture. Indexing is done by a fixed number of sample documents which are collected and processed through a Traditional indexer for each concept. The output of the indexer is a set of concepts in the Concept Database (CD) which is the essence of the Conceptual Indexer (CI). Each new document is processed through CI and the output of CI is a Word plus a Concept Index (WCI). The L-factor specifies the relative importance of concept matches to word matches and is provided by the user in a scale from 0 to 1. If L is 1, only concept matches are considered. If L is 0, only word matches are considered. When L is 0.5, concept and word matches contribute equally.

2.3.1.2. The Automated Generated Thesaurus Approach

Thesauri (listing of words with similar, related, or opposite meanings) can be an efficient tool for concept-based retrieval in the text domain. The Automated Generated Thesaurus Approach (AGTA) [CHEN 1999] provides an ability to fine tune what keywords a user had in mind (or should have had in mind) when proposing a particular query (the initial step in knowledge retrieval). AGTA carries this out in several phases:

- Document collection – specifying the set of documents in a specific subject domain(s) that is to serve as the thesaurus base.
- Automatic indexing of the collection of terms in the document set using an automatic indexing technique [SALTON 1975].
- Co-occurrence analysis of term frequency, inverse document frequency, and cluster analysis to assign weights to each term in a document in order to represent the term’s level of the importance.
- Associative retrieval treats each term in this network-like thesaurus as an active node or neuron and the asymmetric weight between any two terms is taken as the unidirectional, weighted connection between these terms. In order to consolidate related terms and to cast out confusing outliers, the Hopfield algorithm [HOPFIELD 1982] is introduced. It first takes user-supplied terms as input patterns and then activates term neighbors (i.e., strongly associated terms), combines weights from all associated neighbors by adding collective association strengths and repeats this process until term convergence is achieved. The Hopfield algorithm also causes a damping effect in which terms farther away from the initial terms receive gradually decreasing activation weights and eventually are excluded from further processing altogether.

2.3.1.3. The Semantic Web

The Semantic Web [LEE 2001] represents an envisioned future evolution of the World Wide Web, where information will be human and computer understandable. It gathers together both use concepts and ontologies. As presented in Figure 13, the essence of the Semantic Web are Uniform Resource Identifier’s (URI’s) [W3C 2007], which can be formulated as concepts. Ontology vocabulary puts meaning into these concepts, and through XML [W3C 2007] and RDF [W3C 2007] enables a concept-based Web search. Examples of concept-based retrieval web-search engines may be found in [LINKS01].

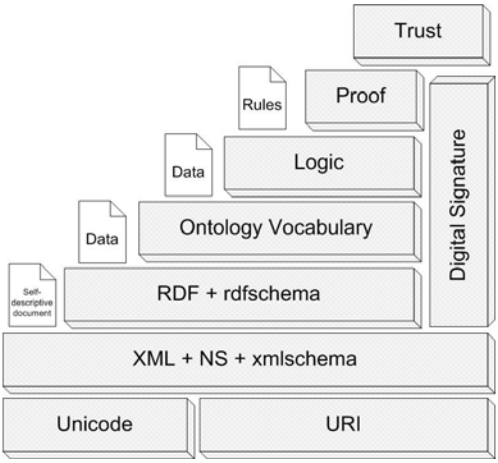

Figure 13. The Semantic Web architecture. Unicode and URI provide means for identifying objects in the SemanticWeb.eXtensible Markup Language (XML) together with the namespaces and XML schema provide syntax without semantic constraints for objects (URI's). A Resource Description Framework (RDF) triple contains three components: the subject, which is an RDF URI reference or a blank node; the predicate, which is an RDF URI reference; the object, which is an RDF URI reference, a literal or a blank node. Therefore, at this level, statements about the subject, object, and predicate are made. An ontology vocabulary defines properties and possible classes for statements built in RDF Layer. A digital signature represents small bits of code that one can use to unambiguously verify that some party wrote a certain document. The Logic Layer contains a logical reasoning mechanism in which it is possible to define logic rules. The Proof Layer executes rules defined in the Logic Layer and the Trust Layer processes security issues. Generally, Trust Layer contains a decision making mechanism to differentiate whether to trust or not to trust the given proof from the bottom layers.

Although concepts and ontologies are presented as separate layers in the Semantic Web, we are fully aware that in practice there is no clear way of distinguishing where the use of concepts stops and use of ontologies begins. Therefore, our format of "definition"; "organization", and "use" represents more the authors wish to give a formal structure to the paper, than that the separation of concepts from ontologies really allows. Still, we tried to present concepts in this form, and in the next section, ontologies will be presented using the same framework.

3. ONTOLOGY

The term ontology is taken from philosophy, where it means the study of being or existence (“What exists?”, “What is?”, “What am I?”). Questions about being exemplify and highlight the most basic problems in ontology: how to find a subject, a relationship, and an object to talk about. Within the more limited scope of the works cited in this paper, an ontology is a concept that groups together other (in some sense “like”) concepts as shown in Figure 1. This grouping of concepts is brought under a common specification in order to facilitate knowledge sharing.

3.1 Ontology definition

Even within the limited scope of information sharing, the term ontology has been defined from many different view points and with different degrees of formality. Ontologies mostly include metadata such as concepts, relations, axioms, instances, etc [NAVIGLI 2003]. Ontologies can be viewed as mediators in the acquisition of knowledge from concepts. Therefore, ontologies lie between the concepts (that they subsume) on one hand and the overarching knowledge domain (within which they are embedded) on the other. Here are some ontology definitions from the viewpoint of concepts:

- A “specification of a shared conceptualization” [GRUBER 1993].
- An arrangement of concepts that represents a view of the world that can be used to structure information [CHAFFEE 2000].
- A conceptual model shared between autonomous agents in a specific domain [MOTIK 2002].

And here are some ontology definitions from the viewpoint of knowledge:

- An organized enumeration of all entities of which a knowledge-relation system is aware [HALLADAY 2004].
- A description of the most useful, or at least most well-trodden, organization of knowledge in a given domain [CHAN 2004].

Given ontology definitions from both of these concept and knowledge perspectives, the next issue is how ontologies can be organized.

3.2 Ontology organization

Modeling is used to achieve a consistent organization among and within ontologies; moreover finding (or inventing) consistent descriptive metadata for ontology modeling purposes is cited as the main obstacle to the introduction of ontology-based knowledge management applications into commercial environments [WARREN 2006]. Various approaches are suggested to address these problems.

Created from subsumed components, ontologies can unite classes, relationships, and entities that are equivalent but differently expressed. Ontologies themselves can be combined to model a certain knowledge domain [CHAN 2004] by organizing them as a set of terms and constraints in some form of ontology vocabulary. Alternatively, [MOTIK 2002] presents the organization of an ontology as a set consisting of: a relation, a sub-concept, an instance, a property, and a concept. In order to achieve interoperability between information presented in different ontologies, an application can consolidate ontologies into one, through a process of ontology mapping (Figure 14).

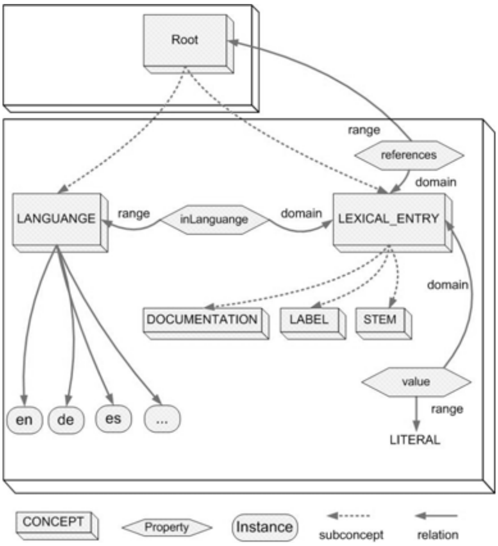

Figure 14. A Lexical Ontology-Instance-Model Structure. Each instance of a ROOT concept may have a lexical entry which reflects various lexical properties of an ontology entity, such as a label, stem, or textual documentation. Before interpreting a model, the interpreter must filter out a particular view of the model (whether a particular model can be observed as a concept, a property, or an instance) – it is not possible to consider multiple interpretations simultaneously. However, it is possible to move from one interpretation to another - if something is viewed as a concept in one case, in another case

it is possible to interpret the same thing as an instance. *This is similar to the vocabulary switching process proposed in [SCHATZ 1997].*

Ontologies can be organized as a set of hypercubes, where each hypercube represents a single concept [SCHLOOSSER 2002]. A hypercube composed of peers supporting a ***TravelService*** concept, in a peer to peer network, is presented in Figure 15.

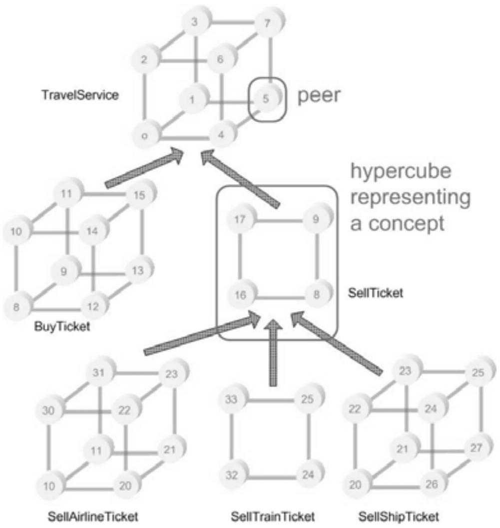

Figure 15. A peer to peer (P2P) ontology-structured network topology representing the process of buying and selling airline, train, and ship tickets. The internal peer organization of a hypercube is structured so that the network can support queries that are logical combinations of ontology and concepts . Every peer should be able to become a root of a tree spanning all nodes in the network. Also, to maintain the network symmetry, crucial for P2P networks, any node in the network should be allowed to accept and integrate new nodes in the network. Querying the network works in two routing steps. The first step is to propagate a query to those concept clusters that contain peers that the query is aiming at. In the second step, a broadcast is carried out within each one of these concept clusters, optimally forwarding the query to all peers in the clusters. This involves shortest-path routing as well as restricted broadcast in the concept coordinate system.

Some software tools for grouping and organizing ontologies are:

- Ontology library systems [DING 2001],

- Automatic ontology derivation [GAUCH 2002] from hierarchical collections of documents like Open Directory Project [OPEN 2002],
- Protégé [PROTEGE 2006],
- KAON [KAON 2001], etc.

One sees that the primary purpose of all of this effort, with an eye towards enterprise applications, is to so define and organize ontologies as to facilitate information sharing among originally incompatible data elements – possibly with the assistance of software tools that have been developed to automate this effort. This leads directly to a consideration of the ways that ontologies so constructed are used in applications to support information sharing.

3.3 Ontology use

Ontologies cover a broad range of knowledge. They are variously presented in this section as applied to information systems, software agents, automatic translation process, photo descriptors, and text mining. Many different applications use ontologies to explicitly declare the knowledge embedded in them [PEREZ 2002]. As an illustration, a DARPA Agent Markup Language DAML [DAML 2007] ontology record for the concept of ***address*** is presented in Figure 16.

```
<?xml version="1.0" ?>
<!DOCTYPE rdf:RDF (View Source for full doctype...)>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:address="http://
daml.umbc.edu/ontologies/ittalks/address#" xmlns="http://
daml.umbc.edu/ontologies/ittalks/address#" xml:base="http://
daml.umbc.edu/ontologies/ittalks/address">
- <owl:Ontology rdf:about="">
<rdfs:comment>This file describe the postal address. We are more
consider about the delivery aspect rather than accurate geographical
localtion, such as latitude and longitude.It is created by Li Ding -- http://
www.csee.umbc.edu/~dingli1/, Harry Chen -- http://www.csee.umbc.edu/
~hchen4/, Lalana Kagal -- http://www.cs.umbc.edu/~lkagal1/, Tim Finin -
- http://www.csee.umbc.edu/~finin/.</rdfs:comment>
<owl:versionInfo>$Revision: 1.1 $</owl:versionInfo>
</owl:Ontology>
- <owl:Class rdf:ID="Address">
<rdfs:label>Address</rdfs:label>
<rdfs:comment>Address</rdfs:comment>
</owl:Class>
- <owl:DatatypeProperty rdf:ID="roomNumber">
<rdfs:domain rdf:resource="#Address" />
<rdfs:range rdf:resource="http://www.w3.org/2001/
XMLSchema#string" />
</owl:DatatypeProperty>
- <owl:DatatypeProperty rdf:ID="streetAddress">
<rdfs:domain rdf:resource="#Address" />
<rdfs:range rdf:resource="http://www.w3.org/2001/
XMLSchema#string" />
</owl:DatatypeProperty>
- <owl:DatatypeProperty rdf:ID="city">
<rdfs:domain rdf:resource="#Address" />
<rdfs:range rdf:resource="http://www.w3.org/2001/
XMLSchema#string" />
</owl:DatatypeProperty>
- <owl:DatatypeProperty rdf:ID="state">
<rdfs:domain rdf:resource="#Address" />
<rdfs:range rdf:resource="http://www.w3.org/2001/
XMLSchema#string" />
</owl:DatatypeProperty>
- <owl:DatatypeProperty rdf:ID="zip">
<rdfs:domain rdf:resource="#Address" />
<rdfs:range rdf:resource="http://www.w3.org/2001/
XMLSchema#string" />
</owl:DatatypeProperty>
- <owl:DatatypeProperty rdf:ID="country">
<rdfs:domain rdf:resource="#Address" />
<rdfs:range rdf:resource="http://www.w3.org/2001/
XMLSchema#string" />
</owl:DatatypeProperty>
</rdf:RDF>
```

Figure 16. DAML ontology for the concept of *address* in OWL [OWL04]. The concept *address* is observed as a class, with the following subclasses: roomNumber, streetAddress, city, state, zip, and country.

3.3.1. Information systems

Ontologies can serve as a basis for an information system as is the case in Ontology-Driven Information Systems (ODIS) [GUARIN 1998]. ODIS system illustrates how four distinct general ontology types can be involved in building an information system in terms of Top-level, Domain, Task and Application ontologies. This unified hierarchical organization is presented in Figure 17.

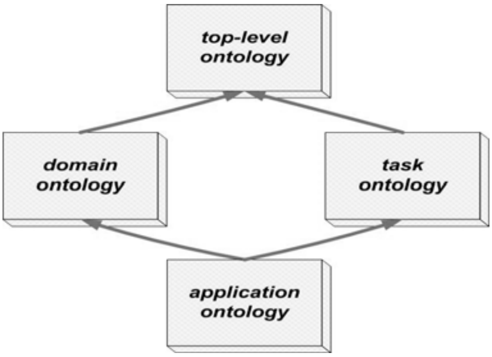

Figure 17. Organization of the ontologies in the Ontology-Driven Information System. Top-level ontologies describe very general concepts like space, time, matter, object, event, action etc., which are independent of a particular problem or domain. Domain ontologies and task ontologies describe, respectively, the vocabulary related to a generic domain (like medicine, or automobiles) or a generic task or activity (like diagnosing or selling), by specializing the terms introduced in the top-level ontology. Application ontologies describe concepts depending both on a particular domain and tasks related to a specific application.

3.3.2. Software agents

Ontologies can also be applied to agent-based computing environments. One approach in that direction is presented in [SMIRNOV 2001] where ontologies act as multi-agents in three forms:

- An application oriented ontology - a conceptual model that describes a real-world application domain depending on the particular domain and problem.
- A resource ontology – a knowledge source description in application ontology terms.
- A request ontology – a user request description in application ontology terms.

In agent-based computing environments, devices, software agents and services are expected to seamlessly integrate and cooperate with each other in support of human objectives – i.e. anticipating needs, negotiating for the service, acting on our behalf, and delivering services in an anywhere, anytime fashion. To serve as the core for such an environment, the authors [CHEN 2003] propose an intelligent Context Broker (CB). CB has the ability to integrate and reason over retrieved information in order to maintain a coherent model of the space, the devices, the agents, and the people in it. An ontology graph based on OWL [OWL 2004] supporting the work of an intelligent CB, is presented in Figure 18.

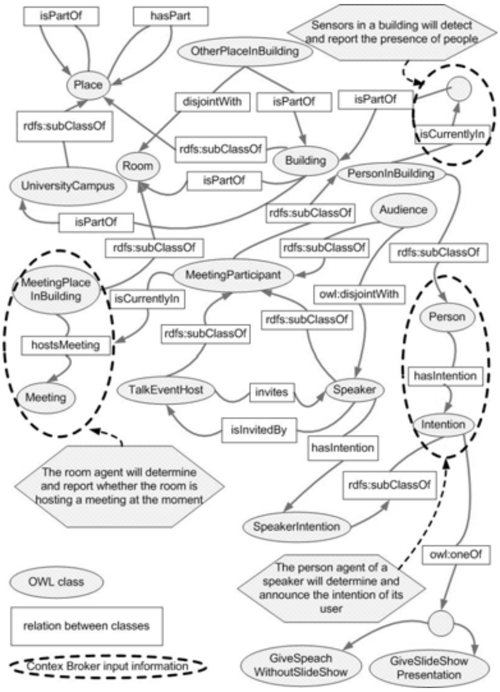

Figure 18. Ontology graph for context broker processing support. It consists of 17 classes and 32 property definitions. Each one of the classes and properties are used to describe "Person", "Place", and "Intention" from retrieved data. The "Person" class defines the most general properties about a person in an intelligent space (i.e., conference room, office room, and living room). The "Place" class defines the containment relationship properties (i.e., isPartOf, and hasPartOf) and naming properties of a place (like fullAddressName). The "Intention" class defines the notion of user intentions; for example, a speaker's intention to give a presentation and an audience's intention to receive a copy of the presentation slides and handouts. Each oval with a solid line represents an OWL [OWL 2004] class. Each oval with a broken line indicates the kind of information that CB will receive from other agents and sensors in the environment.

“Semantic interoperability” represents a type of communication between two software agents that work in overlapping domains, whether they use the same or different notations and vocabularies. Ontology-based agents such as given by OntoMerge and OntoEngine [DOU 2004] are offered as one possible solution to implement such “semantic interoperability”. Like its name, OntoMerge is a tool for ontology merging, and OntoEngine is a tool to automate reasoning over merged ontologies. Here, the merger of two related ontologies is obtained by taking the union of the terms and the axioms defining them. Reasoning is automated by means of an inference mechanism that uses a dataset of several ontologies as input and automatically projects the conclusions into one or several target ontologies, as output. Another example of automated reasoning mechanism on top of ontologies is Ontology Inference Layer. It uses the Fast Classification of Terminologies [HORROC 2003] as a system to provide reasoning support for ontology design, integration, and verification [FENSEL 2001].

3.3.3. Natural language automatic translation

OntoLearn [NAVIGLO 2003], a system that automatically associates multiword English and Italian terms, is a practical example of the use of ontologies in automatic translation. This automated learning system extracts relevant domain terms from a corpus of text, relates them to appropriate concepts in a general-purpose ontology, and detects taxonomic and other semantic relations among concepts. The main features of semantic interpretation in OntoLearn are:

- A determination of the right concept (finding the sense) behind each component of a complex term (semantic disambiguation).
- An identification of the semantic relations holding among concepts to build a complex concept.

An example of semantic net in OntoLearn is presented in Figure 19.

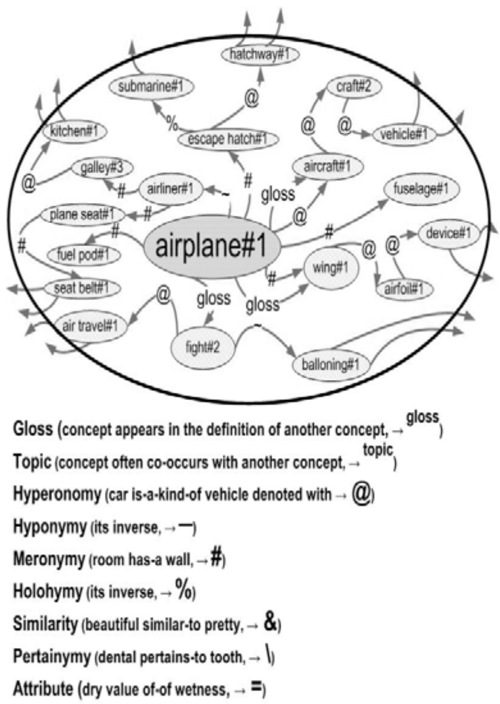

Figure 19. OntoLearn semantic net for the concept *airplane* (sense number 1, airplane#1). The system automatically builds semantic nets by using the following lexicosemantic relations: Hyperonomy, Hyponymy, Meronymy, Holohymy, Pertainymy, Attribute, Similarity, Gloss, and Topic.

3.3.4. Photo descriptors

Ontologies can also be used as a tool for describing photos, in order to help in the photo retrieval process. In [SCHREIBER 2001], the ontology-based photo annotation tool consists of the following two features:

- A photo annotation ontology and
- A subject matter vocabulary.

The photo annotation tool provides the description template for annotation construction and consists of the following features:

- A subject matter feature - what does the photo depict?
- A photograph feature - how, when, and why was the photo made?
- A medium feature - how is the photo stored?

The subject matter vocabulary is a domain-specific ontology for the animal domain (basically describing photo's subject matter). It consists of the following four elements:

- An agent (for example: "an ape"),
- An action (for example: "eating"),
- An object (for example: "a banana"), and
- A setting (for example: "in a forest at down").

3.3.5. Text mining

Text mining usually reefers to a process of automatically obtaining information from texts. An example of a text mining software tool based on ontologies is Artequakt [ALANI 2003]. It has implemented the Knowledge Extraction Tool (KET), which searches online documents and extracts knowledge that matches the given classification structure. Artequakt links KET with ontologies in order to achieve efficient information extraction from Web pages. An example of the Artequakt knowledge extraction process is presented in Figures 20a and 20b.

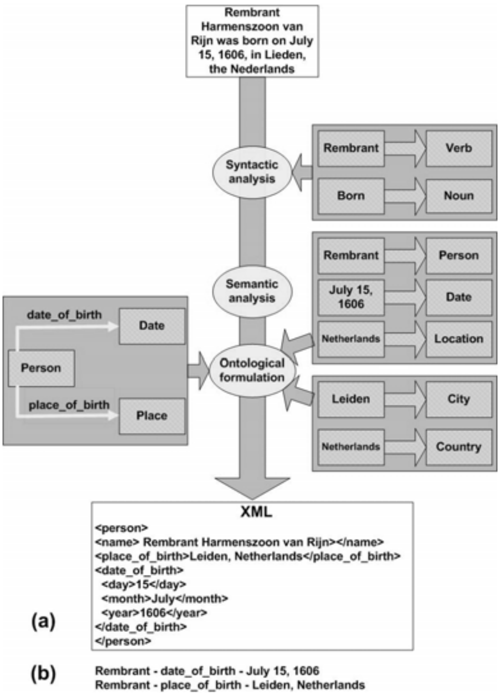

Figure 20a. An example of Artequakt knowledge extraction from a Web page. When a Web page is recognized to match an input query, it is further processed in a form of syntactic analysis, semantic analysis, and ontological formulation. Outputs are extracted knowledge triplets from the web page in XML syntax, as shown in example (a). After the web page extracted information is presented in a form of XML, it is further processed in a form of ontology, with corresponding instances and relationships, as shown in example (b).

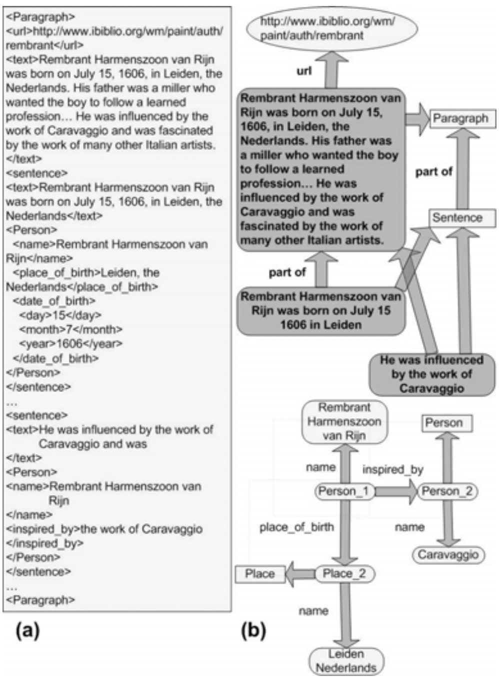

Figure 20b. Automatic Artequakt ontology population process. Based on the XML file of extracted information from the web page (a), the corresponding instances and relations are made (b), supported by Protégé [PROTEGE 2006].

This section has presented various scientific contributions related to ontology definition, organization, and use. All seek to provide a common means of knowledge representation suitable for further processing. However, the thing that most of the ontologies lack (generally) is atomicity. They usually start in the middle of the problem, not at the beginning and, as a result, tend to be domain specific. They can find concepts in mammography, but not in seismological reports. As such, they can be considered to be content theories about the types of objects, properties of objects and relations between objects that are possible in a specific knowledge domain [CHANDRASEKARAN 1999]. Working up from this basis, the final section focuses on knowledge representation as bringing together both concepts and ontologies

4. KNOWLEDGE REPRESENTATION

A person can experience knowledge as information at its best. Loosely stated, it is information in support of or in conflict with some hypothesis or it serves to resolve a problem or to answer some specific question. This kind of knowledge that may be expected as the outcome of information processing – or it may be something new and surprising. Information as initially gathered is often fragmented and unstructured and in that form is not suitable for further exchange and processing across different systems. Moreover, a priori one does not usually have a firm grip on what the atoms of knowledge are, how they are connected, how populated, and how one can retrieve or deduce new knowledge from them. In order to answer some of these important questions, the next section begins by examining different definitions of knowledge followed by a discussion of knowledge organization and concluding with practical applications of how knowledge representation uses both concepts and ontologies.

4.1 Knowledge definition

Knowledge and *concept* are among most abstract terms in human vocabulary. Just as stated with the term of *concept*, all of the characteristics of knowledge cannot be captured within a single definition. Therefore, we start with some abstract definitions of knowledge:

- The content of all cognitive subject matter [MERRILL 2000].
- A critical resource for any activity [SMIRNOV 2001]: enterprise activity [YOON 2002], intelligent systems [GUO 2005] etc.
- A net made of entities and relationships [MILLIGAN 2003] where relationships between entities provide meaning, and entities derive their meaning from their relationships.

Some more concrete definitions of knowledge related to both concepts and ontology are:

- Conceptual models of information items or systems, including principles that can lead a decision system to resolution or action [HALLADAY 2005].
- In scale-free networks only two types of nodes exist: a few nodes with many connections, and many nodes with very few connections. Concept organization in a scale-free network can be considered as knowledge [HALLADAY 2004].

Because knowledge based on entities and relationships (upgraded in the form of concepts and ontologies) represent the foundation for many intelligent systems, this introduces the problem of how to organize the knowledge in a uniform manner to make it suitable for further processing. In order to provide answers to this important question we next discuss knowledge organization issues.

4.2 Knowledge organization

Generally speaking, knowledge organization is directly related to a particular way of thinking [YOON 2002]. There are many ways to characterize this. [MERRIL 2000] describes process of thinking consisting of knowledge objects. These knowledge objects variously describe the subject matter, the content or what is to be taught. From this perspective, four types of knowledge objects are essential for knowledge organization:

- Entities - things or objects.
- Actions - procedures that can be preformed by a learner on/to/with entities or their parts.
- Processes - events that occur often as a result of some action.
- Properties - qualitative or quantitative descriptors for entities, actions, or processes.

[HALLADAY 2004] simulates the acquisition of knowledge that has been previously organized for education purposes by introducing the concept of clusters in knowledge objects. As the subject matter of an area is learned, the relevant clusters undergo a phase transition among the connections that make up the way the cluster was originally formed.

Starting from another point of view, [PEREZ02] specifies knowledge organization using five components: concepts, relations, functions, axioms, and instances.

In [LAND01] the organization of knowledge is distinguished by the level of formality and by the level of individuality. Formality levels can be expressed as:

- Implicit knowledge, i.e., not well structured, and cannot be easily articulated, or
- Explicit knowledge, i.e., formally represented using a precise and sufficiently formal knowledge representation scheme.

(While it is possible to conceptually distinguish between explicit and implicit knowledge, in practice these are seldom separated, because new knowledge is created through the dynamic interaction and combination of both types.)

Knowledge can also be organized according to the level of individuality as:

- Individual knowledge - as resides in the brains of the individual, and is owned by the individual, or
- Collective knowledge - distributed and shared among members within same team, different teams, and organizations.

Database organization is a common form of explicit knowledge representation that facilitates both mathematical analysis and computer processing. To establish a database organization [ZELLWEGER 2003] uses navigation structures like a network of topic lists, topic data and data relationships (such as one-to-one, one-to-many, many-to-many and many-to-one). Such a database structure is presented in Figure 21a. A structure of semantic relationships database nodes is illustrated in 21b.

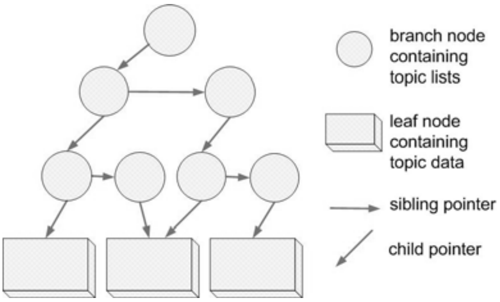

Figure 21a. Database network graph structure. Data flow starts from the root node and progresses downwards through one or more branch nodes to form paths that link to the leaf nodes. Each branch node has a sibling pointer and a child pointer. The sibling pointer creates a list of topics and the child pointer connects each list to a successor node (either another branch or leaf node). The advance is that any number of topic lists can link to the same topic data. *It is analogous to the situation where different words can link to the same concept.*

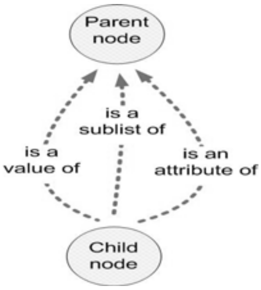

Figure 21b. Relationships between parent-child nodes within a database organization. As a knowledge structure, each parent-child node pair represents a semantic relationship like "is a sublist of", "is an attribute of", and "is a value of".

Now that several proposed ways of characterizing knowledge organization are presented, the following section focuses on various possibilities in knowledge use which combine with both concepts and ontologies use.

4.3 Knowledge use

Clearly, a main use of knowledge is in problem-solving [MERRIL 2000]. A successful knowledge structure incorporates the information required for an interested party to solve a particular problem. If the required knowledge components and their relationships are incomplete, then the party will not be able to use the available information efficiently. Problem solving by computer requires not only an appropriate knowledge representation, but also proper algorithms or heuristics to manipulate the knowledge components. A successful problem-solving sequence passes through the following phases:

- knowledge integration,
- knowledge modeling,
- knowledge storage, and
- knowledge retrieval.

4.3.1 Knowledge integration

To begin the information at hand must be cleaned up to remove redundancy, subsumption and contradiction between different knowledge entities – which is the task of knowledge integration [GUO 2005]. [MEDSKER 1995] cites the benefits of knowledge integration in one expert system:

- Existing knowledge can be reused.
- Knowledge acquired from different sources usually has a better validity and is more comprehensive than from only one source.
- Knowledge integration by computer can build a knowledge-based system faster and less expensively than can human experts.

In order to integrate different knowledge sources, the relationships between these different sources must be made explicit. In the process of integrating knowledge hidden relationships may be uncovered that reveal new knowledge. To assure that all such relations remain consistent both before and after knowledge integration one requires a knowledge modeling process – which is the next step.

4.3.2 Knowledge modeling

Knowledge modeling takes the way one thinks about data, information and knowledge from the real world (a human cognition process) and combines it with knowledge models from the information world [WEIQI 2004]. As a consequence, knowledge models incorporate the set of information entities such as data, ontology, rules, logic, and propositions. An example of such a knowledge modeling process is presented in Figure 22.

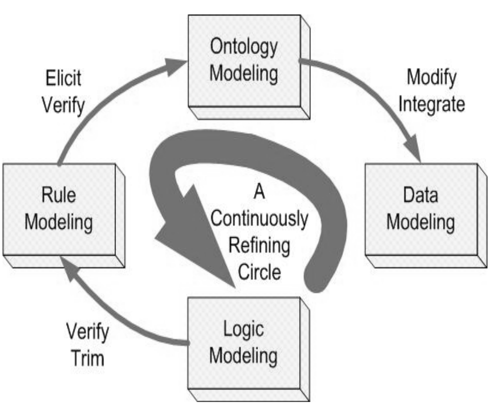

Figure 22. A Unified Knowledge Modeling process consisting of knowledge models: data, ontology, rule, and logic, forming an inner and outer circle. In the inner circle processes are carried out as follows: data can be used to build ontologies, rules can be formed on the top of these ontologies, and logic can be inferred from these rules. Each

knowledge model forms the underlying base for the next model, in contrast to the outer cycle. In the outer cycle each newly built model can be useful to the previously built model: the ontology model can be used in modifying and integrating a data model, a rule model can be used in eliciting and verifying an ontology model, and a logic model can be used in verifying and trimming a rule model.

[CHAN 2004] presents another possible way to model knowledge by using Knowledge Modeling Systems (KMS) based on an Inferential Modeling Technique (IMT). IMT first models the domain objects and relations before deciding what tasks are involved and what problem-solving method to adopt. Thus KMS consists of two primary modules:

- A class module – gathers user knowledge on classes of objects, the attributes and values associated with each class, and the relationships between the classes, all related to the problem-solving domain.
- A task module – represents an organized structure or a sequence of activities that is performed to accomplish some objective in the problem-solving process.

The main benefit of building a KMS is to gain a shared and reusable knowledge base. The shared and reusable knowledge base paradigm leads us into the next section where our paper discusses how to store knowledge in such a knowledge base, once it is modeled in a uniform manner.

4.3.3 Knowledge storage

There is still no machine that can simulate the efficient way that the human brain stores its data and thinks about them, but generally a person does not even have many static records in his or her head. Over time, mankind has invented increasingly sophisticated means to store knowledge – by writing on stone, papyrus, and paper, later adding recorded speech and film. Now, all kinds of knowledge records are stored digitally in machines. With advances in computer science, knowledge stored in knowledge basis has begun to serve as the foundation for intelligent systems [GUO 2005]. But the lack of consistency in the vast amount of implicit knowledge poses a particular storage problem. [LAND 2001] has designed a specialized conceptual framework to first capture, organize and finally store implicit knowledge in the domain of software engineering. The two phases of this proposed conceptual framework are:

- Knowledge Capture (KC) and

* Knowledge Organization (KO)

as indicated in Figure 23.

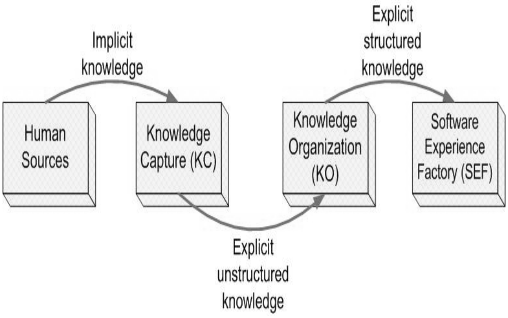

Figure 23. The process of storing implicit knowledge. Knowledge Capture (KC) extracts implicit knowledge (related to software development) residing in the minds of the parties involved, with other mechanisms such as anecdotes, case studies, lessons learned, best practices, failures, successes, etc. The knowledge retrieved with KC is explicit, but it lacks structure and organization, thus Knowledge Organization (KO) is necessary. KO usually includes transcription (translation from voice or video formats to written form), summarization (production of the main points from transcribed data), and coding (assigning symbols to transcribed data). The output of KO is an explicitly structured knowledge, suitable for further exchange and comparison in a computer system; and serves to populate the Software Experience Factory (SEF). SEF represents the storage of explicit and structured knowledge related to software development.

After knowledge from different sources has been integrated, modeled in a uniform manner, and stored in a knowledge base, the next step is the purpose of it all: the extraction of knowledge, or knowledge retrieval.

4.3.3 Knowledge retrieval

Knowledge discovery in databases or data mining refers to the nontrivial extraction of implicit, previously unknown, and potentially useful information from the data stored in Databases [FRAWLEY 1992]. Two types of queries and answers for efficient knowledge retrieval in the database domain are cited by [HAN 1996]:

* A simple data query – to find a stored data item in the database (which corresponds to a basic retrieval statement in a database system).

- A knowledge query – to find a certain rules and other kinds of complex knowledge in observed database.

The answering to a database query can take two forms:

- Direct answers that are simple examples of data or knowledge from a database.
- An answer to a query using intelligence – by first analyzing the intent of the query and then providing generalized, neighborhood, or associated information relevant to the query (by means of data summarization, concept clustering, rule discovery, etc).

One possible way to increase efficiency in a Web pages domain knowledge retrieval process is to collect user feedback from the pages visited (so that in future iterations, user searches can better refine and match the system searches). [TSAI 2003] presents a multi-agent framework that iteratively collects user's feedback and updates the user Web page profile. Its task cycle is presented in Figure 24.

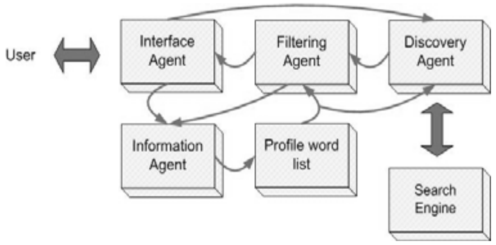

Figure 24. A Multi-agent based framework for efficient knowledge retrieval from World Wide Web. The framework consisting of agents involves the agents' task cycle as follows: An interference agent receives a user's query and redirects it to existing search engines. Then, an Information agent analyzes the Web pages chosen by the user and derives a temporary user profile. A Discovery agent, based on a user profile, performs query expansion and modification. A Filtering agent ranks the retrieved Web pages correspondingly to the user profile and recommends the most relevant web pages in future queries. The user labels useful pages which are then further processed in profile updating. The knowledge retrieval procedure continues iteratively until a user terminates the search.

Knowledge conceptualization is a special form of knowledge retrieval processing. The knowledge conceptualizing tool [FUJIHARA 1997] uses concept clustering and ranking techniques by applying a Vector Space Model [SALTON 1975] and a Probability Ranking Principle [ROBERTSON 1997]. An interview transcript, containing several question and answer pairs and consisting basically of unstructured conversational sentences, represents the system input. After processing, the outputs are a set of rules and facts extracted from the input data, thereby forming a new knowledge representation.

This section has presented a list of papers related to knowledge definition, organization, and use. Problem-solving in some sense is the final goal of every knowledge use. Therefore, most of the papers presented focused on how to get there through knowledge representation as a stepwise layered process consisting of knowledge integration, knowledge modeling, knowledge storage, and knowledge retrieval. In this way it is possible to combine different knowledge representations and merge them in order to answer a particular question or, some more general, problem-solving issue.

5. CONCLUSION

The research efforts presented here are focused on knowledge representation by ontologies populated with concepts. Concepts, ontologies, and knowledge representation are almost impossible to separate in practice, since there is no clear distinction where the use of concepts stops and use of ontologies begins in knowledge representation. Therefore, most of the research efforts presented are a combination of all tree topics. Thus the survey can be viewed as an annotated guide to this literature.

This paper sheds more light on a selected number of different avenues leading to the same future goal of knowledge retrieval based on conceptual queries, as opposed to the current state of the art based on semantic queries. As indicated in this paper, statements “I am a PhD” and “I have a doctorate,” are two different semantical entities, but they both represent the same concept. Therefore, a semantic query (e.g., focused on only one of the above two statements) will be able to retrieve only a subset of relevant knowledge, while a conceptual query (focused on both statements above, as well as all other statements supporting the same concept) would retrieve the full set of relevant knowledge. A trivial solution to the problem is, for each relevant concept, to create a case structure that includes all statements supporting that particular concept. This solution is based on exhaustive approaches, and has no practical value. Practical value lies in the many sophisticated approaches discussed in this survey paper.

The authors believe that this survey will benefit both those who want to enter the field of knowledge retrieval quickly, and those who would like to extend the state-of-art. To the best of our ability, all of the relevant work up to the present has been cited and discussed. For those who are concerned with implementation, there are examples of numerous working systems. Quite clearly there is no overarching "Killer Ap"; the results achieved so far in this domain remain both tentative and incomplete. Much work remains to be done.

ACKNOWLEDGMENTS

The authors would like to thank Charles Milligan of Sun Microsystems, USA, and Gerald O'Nions of StorageTek, France, who initiated research interest in this exciting field. Also, to Dr. Tom Lincoln, University of Southern California, USA, Dr. Roger Shannon, Duke University, USA, Dr. William Robertson, Dalhousie University, Canada, Ognjen Šćekic, B.Sc., University of Belgrade, Serbia, and Djordje Popović, B.Sc., IPSI Belgrade, Serbia, for constructive feedback on this paper.

REFERENCES

[ALANI 2003] Alani, H., Kim, S., Millard, D., Weal, M., Hall, W., Lewis, P., Shadbolt, N., "Automatic Ontology-Based Knowledge Extraction from Web Documents," IEEE, Intelligent Systems, Jan-Feb 2003, pp. 14- 21 Vol. 18, Issue 1.

[CHAFFEE 2000] Chaffee, J., Gauch, S., "Personal Ontologies form Web navigation," ACM Press New York, USA, 2000

[CHAN 2004] Chan, C., "The Knowledge Modelling System and its Application," Canadian Conference on Electrical and Computer Engineering, 2-5 May 2004, pp. 1353 - 1356 Vol.3.

[CHANDRASEKARAN 1999] Chandrasekaran, B., Josephson , J., Benjamins, V., "What are Ontologies, and why do we need them?," IEEE, Intelligent Systems and Their Applications, Jan-Feb 1999, pp. 20-26 Vol. 14, Issue 1.

[CHEIN 1992] Chein, M., Mugnies, M., « Conceptual Graphs : fondamental notions, » Revue d'Intelligence Artificielle, 1992, pp. 365 – 406, Vol. 6, no 4.

[CHEN 2003] Chen, H., Finin, T., "An Ontology for Context Aware Pervasive Computing Environments," Cambridge University Press, September 2003, Vol. 18, Issue 03.

[CHEN 2007] Chen, P., "The Entity-Relationship Model - Toward a Unified View of Data," ACM Press, ACM Transactions on Database Systems, pp. 9-36 Vol. 1, Issue 1.

[CHEN99] Chen, H., Houston, A., Sewell, R., Schatz, B., "Internet Browsing and Searching: User Evaluations of Category and Concept Space Techniques," Journal of the American Society for Information Science, 6 Jan 1999, pp.582 – 603 Vol. 49, Issue 7.

[CHUA 1994] Chua, T., Pung, H., Lu, G., Jong, H., "A Concept-based Image Retrieval System," Proceedings of the Twenty- Seventh Hawaii International Conference on System Sciences,. Vol. III: Information Systems: Decision Support and Knowledge-Based Systems, Wailea, Hawaii, USA, 4-7 Jan. 1994, pp. 590-598.

[COMP 1988] "Principles of Database and Knowledge Base Systems," Computer Science Press, Rockville, Maryland, USA, 1988, Vol. 1.

[COOK 2007] Cook, D., "SUBDUE Graph Based Knowledge Discovery," http://www.uta.edu/ra/real/editprofile.php?pid=176&onlyview=1

[DAML 2007] DAML ontologies, DARPA, USA, www.daml.org/ontologies

[DEACON 1998] Deacon, T., "The Symbolic Species – the Co-evolution of Language and the Brain," W.W.Northon & Company, USA, 1998.

[DING 2001] Ding, Y., Fensel, D., "Ontology Library Systems: The key to successful Ontology Re-use," Proceedings of the 1st Semantic Web working symposium, Stanford, CA, USA, 30 July – 1 August 2001.

[DOU 2004] Dou, D., McDermott, D., Qi, P., "Ontology Translation by Ontology Merging and Automated Reasoning," Yale University, New Haven, USA, 2004.

[FENSEL 2001] Fensel, D., Harmelen, F., Horrocks, I., McGuinness, D., Patel-Schneider, P., "OIL: An Ontology Infrastructure for the Semantic Web," IEEE, Intelligent systems, Mar-Apr 2001, pp. 38- 45, Vol.16, Issue 2.

[FRAWLEY 1992] Frawley, W., Piatetsky-Shapiro, G., Matheus, J., "Knowledge Discovery in Databases: An Overview," The American Association for Artificial Intelligence, USA, 1992.

[FUJIHARA 1997] Fujihara, H., Simmons, D., "Knowledge Conceptualization Tool," IEEE Transactions on Knowledge and Data Engineering archive, March 1997, pp. 209 – 220 Vol. 9, Issue 2.

[GAUCH 2002] Gauch, S., Madrid, J., Induri, S., Ravindran, D., Chadalavada, S., "KeyConcept: A conceptual Search Engine,"Information and Telecommunication Technology Center, Technical Report: ITTC-FY2004-TR-8646-37, University of Kansas, USA, 2002.

[GIUGNO 2007] Giugno, R., Sasha, D., "GraphGrep", http://alpha.dmi.unict.it/~graphgrep/index.html

Field Code Changed

[GRUBER 1993] Gruber, T., "A Translation Approach to Portable Ontologies," Knowledge Acquisition, Nol. 5, No. 2, 1993, pp. 199–220.

[GUARIN98] Guarino, N., "Formal Ontology and Information Systems," Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998.

[GUO 2005] Guo, P., Fan, L., Ye, L., Cao, J., "An Algorithm for Knowledge Integration and Refinement," Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005.

[HALLADAY 04] Halladay, S., Milligan, C., "The Application of Network Science Principles to Knowledge Simulation," Proceedings of the 37th Annual Hawaii International Conference on System Sciences, Hawaii, 5-8 Jan. 2004.

[HALLADAY 05] Halladay, S., Milligan, C., "Knowledge VS. Intelligence, IPSI Belgrade, Proceedings of the IPSI-2005 Montengro conference, Sveti Stefan, Montenegro, 2005.

[HALPIN 2007] Halpin, T., Object Role Modeling, Neumont University, USA, http://www.orm.net/

[HAN 1996] Han, J., Huang, Y., Cercone, N., Fu, Y., "Intelligent Query Answering by Knowledge Discovery Techniques," IEEE Transactions on Knowledge and Data Engineering, June 1996, pp. 373-390 Vol. 8, No. 3.

[HAWKINS 2007] Hawkins, J., "Learn Like a Human," IEEE Spectrum on-line, April 2007

[HOPFIELD 1982] Hopfield, J., "Neural networks and physical systems with emergent collective computational abilities," Proceedings of the National Academy of Sciences of the United States of America, USA, 15 April 1982, Vol. 79, No. 8.

[HORROC 2003] Horroc, I., "Fast Classification of Terminologies," School of Computer Science, University of Manchester, UK, 2003.

[INIS 1981] INIS Thesauri, International Atomic Energy Agency (IAEA), Vienna, Austria, Januar 1981

[KAON 2001] KAON, University of Karlsruhe, Germany, 2001, http://kaon.semanticweb.org.

[LAND 2001] Land, L., Aurum, A., Handzic, M., "Capturing Implicit Software Engineering Knowledge," IEEE Computer Society, 13th Australian Software Engineering Conference, 2001, pp. 108-114.
[LEE 2001] Berners-Lee, T., Hendler, J., Lassila, O., "The Semantic Web," Scientific American, USA, May 2001, pp. 28-37, Vol. 284, No. 5.

[LINKS 01] www.vivismo.com, www.grokker.com, www.clusty.com, etc.

[MEDSKER 95] Medesker, L., Tan, E. Turban "Knowledge Acquisition from Multiple Experts: Problems and Issues," Expert Systems with Applications, 1995, pp.35-40 Vol.9, Issue 1.

[MERRIL 2000] Merrill, M., "Knowledge Objects and Mental Models," Proceedings of the International Workshop on Advanced Learning Technologies, Palmerston North, New Zealand 12 Apr. - 12. June 2000, pp. 244-246.

[MILLIGAN 2003] Milligan, C., Halladay, S., "The Realities and Facilities Related to Knowledge Representation," IPSI Belgrade, Proceedings of the IPSI-2003 Montengro conference, Sveti Stefan, Montenegro, 2003.

[MOTIK 2002] Motik, B., Maedche, A., Volz, R., "A Conceptual Modeling Approach for Semantic-driven Enterprise Applications," Springer Berlin / Heidelberg, Book on the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE: Confederated International Conferences CoopIS, DOA, and ODBASE 2002. Proceedings 2002 Vol. 2519.

[NAKABASAMI 2002] Nakabasami, C., "An Inductive Approach to Assertional Minning for Web Ontology Revision," International Semantic Web Conference (ISWC2002), Sardinia, Italy, 9-12 June 2002.

[NAVIGLI 2003] Navigli, R., Velardi, P., Gangemi, A., "Ontology Learning and Its Application to Automated Terminology Translation," IEEE, Intelligent Systems, 2003, pp. 22-31.

[NOVAK 2005] Novak, J., Cañas, A., "The Theory of Underlying Concept Maps and How to Construct them," Technical ReportFlorida Institute for Human and Machine Cognition CmapTools 2006-01, USA, 2005.

[NOVAK 2007] Novak, J., private communication, 23.1.2007

[OPEN 2002] "The Open Directory Project," 2002, http://www.dmoz.org.

[OWL 2004] OWL, Web Ontology Working Group, 2004, http://www.w3.org/2004/OWL/

[PEREZ 2002] Gomez-Perez, A., Corcho, O., "Ontology Languages for the Semantic Web," IEEE, Intelligent Systems, Jan-Feb 2002, pp. 54-60 Vol.17, Issue 1.

[PROTEGE 2006] Protégé, Stanford University, USA, 2006, http://protege.cim3.net/cgi-bin/wiki.pl?ProtegeOntologiesLibrary.

[QUINZ 2004] Quinz, D., Acacetus Inc, Canada, private communication, 20.9.2004

[ROBERTSON 1994] Robertson, S., Jones, K., "Simple, Proven Approaches to Text Retrieval," Technical report TR356, Cambridge University Computer Laboratory, UK, 1994.

[ROBERTSON 1997] Robertson, S., "The Probability Ranking Principle in Information Retrieval," Morgan Kaufmann Publishers, Readings in information retrieval archive 1997 pp. 281 – 286.

[SALTON 1975] Salton, G., Wong, A., "A Vector Space Model for Automatic Indexing," Communications of the ACM, 1975, pp. 613 - 620 Vol. 18, Issue 11.

[SCHATZ 1997] Schatz, B., "Information Retrieval in Digital Libraries: Bringing Search to the Net," Science, 17 January 1997 pp. 327-334 Vol. 275.

[SCHLOSSER 2002] Schlosser, M., Sintek, M., Decker, S., Nejdl, W., "HyperCup – Hupercubes, Ontologies and Efficient Search on P2P Networks," International Workshop on Agents and Peer-to-Peer Computing, Bologna, Italy, July 2002.

[SCHREIBER 2001] Schreiber, A., Dubbeldam, B., Wielemaker, J., Wielinga, B.,"Ontology-Based Photo Annotation," IEEE, Intelligent Systems, May-Jun 2001, pp. 66- 74 Vol. 16, Issue 3.

[SMIRNOV 2001] Smirnov, A., Pashkin, M., Chilov, N., Levashova, T., "Ontology Management in Multi-agent System for Knowledge Logistics," Proceedings of the International Conferences on Info-tech and Info-net, Beijing, China 2001, pp. 231-236 Vol.3.

[SOWA 2000] Sowa, J., "Ontology, Metadata, and Semiotics," Proceedings of the Linguistic on Conceptual Structures: Logical Linguistic and Computational Issues table of contents, 2000, pp. 55 – 81.

[SOWA 1999] Sowa, J., Tepfenhart, W., Cyre, W. "Conceptual graphs: draft proposed American National Standard,". Springer-Verlag, Berlin, Germany, Lecture Notes in Computer Science, 1999, pp. 1-65.

[TSAI 2003] Tsai, T., Lee, W., "An Interactive Agent-based System for Concept-based Web Search," Expert Systems with Applications, May 2003, pp. 365-373 Vol. 24, Issue 4.

[UML 2007] Unified Modeling Language, Object Management Group, http://www.uml.org/

[VOSS 1999] Voss, A., Nakata, K., Juhnke, M., "Concepts as Knowledge Handles in Collaborative Document Management," International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, Stanford, USA, 16-18 June 1999, pp. 245-252.

[WEIQI 2004] WeiQi, C., JuanZi, L., KeHong, W., "CAKE: The Intelligent Knowledge Modeling Web Services for Semantic Web," The 8th International Conference on Computer Supported Cooperative Work in Design Proceedings 26-28 May 2004 Xiamen, China, pp. 209-216.

[WOODS 1997] Woods, W., "Conceptual Indexing: A better Way to Organize Knowledge," Sun Microsystems, USA, Technical Report: TR-97-61, 1997.

[W3C 2007] World Wide Web consortium, http://www.w3.org

[YOON 2002] Yoon, T., Fujisue, K., Matsushima, K., "The progressive Knowledge Reconstruction and its Value Chain Management," Engineering Management Conference, 2002. IEMC '02. 2002 IEEE International, 2002, pp. 298- 303 Vol.1.

[ZELLWEGER 2003] Zellweger, P., "A Knowledge –based Model to Database Retrieval," Proceedings of the International Conference on Integration of Knowledge Intensive Multi-Agent Systems, 30 Sept.- 4 Oct. 2003, pp. 747- 753.