Joël Plisson

ONTOLOGIES FOR COLLABORATIVE NETWORKED ORGANIZATIONS

Master's degree thesis

ONTOLOGIJE MREŽNIH ORGANIZACIJ

Magistrsko delo

Supervisor: Prof. Dr. Nada Lavrač

Co-supervisor:

June 2007

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL
Ljubljana, Slovenia

Index

Abstract ..... **V**

Povzetek ..... **VI**

Abbreviations ..... **VII**

Glossary ..... **IX**

1 Introduction **1**

2 State of the art in ontologies **3**

2.1 Ontology definition 3

2.2 Types of ontologies 3

2.3 Ontology representation and encoding 4

2.4 Methods for ontology construction 4

2.4.1 Manual construction of ontologies using Protégé 5

2.4.2 Learning ontologies from documents 6

2.4.2.1 Hierarchical clustering 6

2.4.2.2 Pattern-based extraction 6

2.4.2.3 Association rule learning 6

2.4.2.4 Conceptual clustering 6

2.4.2.5 Ontology pruning 7

2.4.2.6 Concept learning 7

2.5 Existing business related ontologies 7

2.5.1 The AIAI enterprise ontology 8

2.5.2 The TOronto Virtual Enterprise ontology (TOVE) 9

2.5.3 The Business Process Management Ontology (BPMO) 9

2.5.4 Process Specification Language Ontology (PSL) 10

2.5.5 The Yahoo! business ontology 10

3 An ontology of collaborative networked organizations **13**

3.1 Determining the goals of the ontology 13

3.2 The CNO ontology 15

3.2.1 Top-level of the CNO ontology 16

3.2.2 Organization-related ontology 17

3.2.3 VBE-role related ontology 18

3.2.3.1 VBE member role 18

3.2.3.2 Support institution-related ontology 20

3.2.4 VBE bag of assets-related ontology 22

3.3 Instantiation of the CNO ontology 22

3.4 Evaluation of the CNO ontology 23

4 Extending the CNO ontology with company specific information **27**

4.1 Methodology for semi-automated ontology construction 27

4.2 Competency structuring from the Yahoo! Business data 28

4.2.1 Description of the data 28

4.2.2 Clustering 28
4.2.3 Evaluation 30
4.3 Competency structuring of Virtuelle Fabrik 32
4.3.1 Description of the data 32
4.3.2 Clustering 33
4.3.3 Evaluation 35

5 Conclusions 37

6 Acknowledgements 39

7 References 41

APPENDIX A: Overview of existing ontologies 43
A.1. Terminological ontologies 43
A.1.1. WordNet 43
A.1.2. VerbNet 43
A.1.3. FrameNet 44
A.1.4. Sensus 44
A.2. Domain ontologies 44
A.2.1. The Gene Ontology (GO) 44
A.2.2. Process Specification Language Ontology (PSL) 45
A.3. Upper ontologies 45
A.3.1. Suggested Upper Merged Ontology (SUMO) 45
A.3.2. Mikrokosmos 46
A.3.3. Sowa's top-level ontology 46
A.4. Ontologies with common-sense knowledge 47
A.4.1. Cyc 47
A.4.2. ConceptNet 47

Index of Figures 49

Index of Tables 51

Razširjeni povzetek magistrskega dela 53

Abstract

Companies and individuals connect into networks to share their resources with the purpose of achieving a common goal, defined by a business opportunity. The field of Collaborative Networked Organizations (CNO) covers various types of organizational structures. The knowledge that is stored in such networks can be separated into two different levels. First, there is a common knowledge about the organizational structure itself, which can be used and reused in any of such networks. The second level represents the domain specific knowledge that such networks cover and use to function (e.g. companies' competencies). In this thesis we address both levels by using ontologies.

First, we propose an ontology representing the common vocabulary and identifying the actors and relationships in a specific type of network, namely a Virtual organization Breeding Environment (VBE). In this way, the thesis contributes to the formalization of the informal notions of VBEs and Virtual Organizations (VOs) in a formal ontology language. The ontology, built in Protégé, is available at a public Web site so that it can be redistributed and/or modified. The site also includes an extensive glossary of terms used/introduced in this thesis.

Second, we propose a methodology for semi automated ontology construction for the needs of VBEs, enabling the extraction of network specific knowledge related to competencies.

Both the CNO ontology and the result of the methodology for structuring companies' competencies have been checked and approved by domain experts from the European project ECOLEAD (2004-2007) in which this work was developed.

Povzetek

Podjetja in posamezniki se združujejo v mrežne organizacije z namenom doseganja skupnih ciljev, običajno v obliki realizacije poslovnih priložnosti. Področje mrežnih organizacij (Networked organizations) pokriva razne tipe organizacijskih struktur. Znanje, ki je shranjeno v takšnih mrežah, se deli na dva nivoja. Najprej je tu splošno znanje o organizacijski strukturni mreže, ki se ga da uporabiti v vsaki taki mrežni organizacij. Drugi nivo pa predstavlja specifično znanje domene, ki jo mreža pokriva in uporablja (npr. kompetence podjetij). V magisterskem delu se ukvarjamo z uporabo obeh nivojev znanja in njegovo predstavitvijo v obliki ontologij.

V delu smo razvili ontologijo mrežnih organizacij, ki vpelje terminologijo tega področja ter identificirala akterje in relacije med akterji mrežnih organizacij. Ontologija je bila razvita na širše področje kolaborativnih mrežnih organizacij (Collaborative Networked Organizations, CNO) a se osredotoča na valilnice mrežnih organizacij (Virtual organizations Breeding Environments, VBE). Ta del naloge predstavlja prispevek k formalizaciji dosedaj neformaliziranih pojmov in konceptov s tega področja ter k formalizaciji relacij med koncepti na način, ki zagotavlja konsistentnost razvite ontologije. Ontologija je implementirana v sistemu Protégé in je javno dostopna preko svetovnega spleta. Spletna stran vključuje tudi pojmovni slovar, vpeljan v uvodnem delu magistrske naloge.

Za potrebe drugega nivoja znanja je bila razvita metodologija za polavtomatsko gradnjo ontologij iz tekstovnih dokumentov, ki opisujejo kompetence podjetij, z namenom avtomatskega odkrivanja znanja o kompetencah mrežnih organizacij iz tekstovnih opisov kompetenc podjetij, ki sodelujejo v mrežni organizaciji. Medtem ko je znanje o mrežnih organizacijah, zakodirano v obliki CNO ontologije, statično in torej velja za vse mrežne organizacije, je ekstrahirano znanje o kompetencah specifično za posamezno mrežno organizacijo. Predlagana metodologija za ekstrakcijo tega znanja iz tekstovnih dokumentov je splošna in jo lahko zato uporabimo za katerokoli mrežno organizacijo, ki jo želimo modelirati, seveda pri pogoju, da imamo na voljo tekstovne opise kompetenc sodelujočih podjetij.

Razvita CNO ontologija in rezultat uporabljene metodologija za strukturiranje kompetenc podjetij sta bili pregledani in sprejeti s strani področnih strokovnjakov v okviru projekta ECOLEAD (2004-2007) integriranega projekta EU v okviru 6. okvirnega programa, ki se ukvarja z mrežnimi organizacijami, in ki je predstavljal okvir in motivacijo za pričujoče magistrsko delo.

Abbreviations

CN	=	Collaborative Network
CNO	=	Collaborative Networked Organization
VBE	=	Virtual organizations Breeding Environment
VO	=	Virtual Organization
VT	=	Virtual Team
PVC	=	Professional Virtual Community
BO	=	Business Opportunity
VF	=	Virtuelle Fabrik

Glossary

- **BO, Business Opportunity:** It is a time or occasion with favorable combination of circumstances that is suitable to start a business.

- **Capacity:** It is the aptitude to perform certain actions. The capacity related to a competency represents its availability. It is represented in terms of start time, duration, and availability rate. The availability rate of Capacity is expressed in percentage and specifies which amount of the resource can be used.

- **CNO, Collaborative Networked Organization:** A Collaborative Networked Organization is a special type of collaborative network comprising only organized collaborations while, in general, collaborative networks include both organized and non-organized collaborations. A collaborative network (Camarinha-Matos & Afsarmanesh, 2005) is constituted of a variety of entities (e.g. organizations and people) that are largely autonomous, geographically distributed, and heterogeneous in terms of their operating environment, culture, social capital and goals. These entities collaborate to better achieve common or compatible goals, and their interactions are supported by a computer network.

- **Competency:** Competencies are skills, acquired through work experience, life experience, study or training, in order to perform (business) processes and deliver products and services. (www.hstac.com.au/csassessment/infogloss.html)

- **Human resource:** An organization's human resource refers to the individuals within the organization. (en.wikipedia.org/wiki/Human_resource)

- **Ontology provider:** Ontology provider is the role performed by a participant that supports the VBE with ontologies. (Mentioned in ECOLEAD D21.1)

- **Organization:** A company, corporation, firm, enterprise or institution, or part thereof (whether incorporated or not, public or private) that has its own function(s) and administration that supplies products or services to other organizations. (www.bizmanualz.com/ISO9000-2000/ISO_definitions.html)

- **Physical resource:** Organization's physical resource refers to buildings, machines, equipment, knowledge assets within the organization. (Mentioned in ECOLEAD D21.2a)

- **Process:** A process is a set of interrelated resources and activities that transform inputs into outputs with the aim of adding value. Resources include personnel, facilities, equipment, technology, methodology and finances. (www.bizmanualz.com/ISO9000-2000/ISO_definitions.html)

- **Product-service:** A product or service is the result of activities or processes. It can be tangible or intangible, or a combination of both. (www.bizmanualz.com/ISO9000-2000/ISO_definitions.html)

- **Profile:** An organization profile comprises the information describing the organization.

- **PVC, Professional Virtual Community:** A Professional Virtual Community represents the combination of the concepts of virtual community and professional community. Virtual communities are defined as social systems of networks of individuals, who use computer technologies to mediate their relationships. Professional communities provide environments for professionals to share the body of knowledge of their professions such as similar working cultures, problem perceptions, problem-solving techniques, professional values, and behavior (Camarinha-Matos & Afsarmanesh, 2005).

- **Public entity:** Public entity is the role of a VBE participant which is not registered in the VBE. (Mentioned in ECOLEAD D21.1)

- **Resource:** A resource is anything that has identity. Familiar examples include an electronic document, an image, a service (e.g., "today's weather report for Los Angeles"), and a collection of other resources. Not all resources are network "retrievable"; e.g., human beings, corporations, and bound books in a library can also be considered resources (en.wikipedia.org/wiki/Resource_(computer_science)) In project management terminology, resources are required to carry out the project tasks. They can be people, equipment, facilities, funding, or anything else capable of definition (usually other than labour) required for the completion of a project activity. (en.wikipedia.org/wiki/Resource_(project_management))

- **Services provider:** Ontology provider is the role performed by a participant that supports the VBE with a variety of services. (Mentioned in ECOLEAD D21.1)

- **Support institution:** Support institution is the role played by participants that can provide a broad range of services, such as training, research, consulting, information services, legal and contractual services, etc. Support institutions can be used in VBE to train new partners, perform administrative tasks or even to make and keep the network running.

- **Task:** A task is part of a set of actions which accomplish a job. (en.wikipedia.org/wiki/Task)

- **Technological resource:** Organization's technological resource refers to the hardware and software within the organization. (Mentioned in ECOLEAD D21.2a)

- **Tools provider:** Ontology provider is the role performed by a participant that supports the VBE with software tools. (Mentioned in ECOLEAD D21.1)

- **VBE, Virtual Breeding Environment:** A Virtual Breeding Environment is an “an association (also known as cluster) or pool of organizations and their related supporting institutions that have both the potential and the will to cooperate with each other through the establishment of a “base” long-term cooperation agreement and interoperable infrastructure” (Camarinha-Matos & Afsarmanesh, 2005). The VBE respond to business opportunities by forming VOs. As an organization, it has also competencies but not limited to the union of the competencies of its participants. The VBE competencies are the result of combining two or more participants' competencies to realize more complex projects (e.g. building a highway, bridge, etc.).

- **VBE administrator:** VBE administrator is the role performed by a participant responsible for the VBE operation and evolution, promotion and cooperation among the VBE members. He is also filling the skills/competencies gaps in the VBE by searching and inviting new organizations in the VBE and manages the VBE general processes. (Mentioned in ECOLEAD D21.1)

- **VBE adviser:** VBE adviser is the role performed by a participant responsible for monitoring the network and suggesting recommendations to the administrator. (Mentioned in ECOLEAD D21.1)

- **VBE asset:** An asset is anything owned by a business or individual that has commercial or exchange value. Participants in the VBE have their own assets but here we refer to the assets that have been developed exclusively for the VBE only (for instance, business rules or network management tools, etc.).

- **VBE management provider:** VBE management provider is the role performed by a participant that will provide management services to the VBE.

- **VBE member:** VBE member is the role played by organizations that are registered in the VBE and are willing to participate in the VBE activities.

- **VBE participant:** A VBE participant is any organization, registered or not, within a VBE. Thus it is not to be understood as a VBE member which is a specific role that a participant can take. Each participant can play different roles in the VBE. (Mentioned in ECOLEAD D21.1)

- **VBE role:** The characteristics and expected behavior of an organization in a given position (e. g., VBE member, Support institution, etc.). Its behavior is expressed by a set of tasks that the owner of this role is supposed to perform. In addition, each role is attached to a specific VBE, since a member can have several roles in different VBEs.

- **VBE support provider:** VBE support provider is the role performed by a participant that supports the VBE with different kinds of services.

- **VO, Virtual Organization:** A VO comprises a set of (legally) independent organizations that share resources and skills to achieve its mission/goal, but that is not limited to an alliance of for profit enterprises. A Virtual Enterprise is therefore, a particular case of VO. (Camarinha-Matos & Afsarmanesh, 2005)

- **VO broker:** VO broker is the role performed by a participant that identifies and acquires new business opportunities (BO), by marketing VBE competencies and assets and negotiating with (potential) customers. (Mentioned in ECOLEAD D21.1)

- **VO coordinator:** VO coordinator is the role performed by a participant that will coordinate the VO during its life cycle in order to fulfill the goals set for the business opportunity that triggered the VO. (Mentioned in ECOLEAD D21.1)

- **VO partner:** VO partner is the role of a participant in a VO. (Mentioned in ECOLEAD D21.1)

- **VO planner:** VO planner is the role performed by a participant that will identify and select partners with appropriate competencies necessary to fulfill the goals set for the business opportunity. He will structure and plan the VO. (Mentioned in ECOLEAD D21.1)

- **VO support provider:** VO support provider is the role of a participant that supports the management of a VO. (Mentioned in ECOLEAD D21.1)

- **VT, Virtual Team:** A Virtual Team, also known as a Geographically Dispersed Team (GDT), is a group of individuals who work across time, space, and organizational boundaries with links strengthened by webs of communication technology. They have complementary skills and are committed to a common purpose, have interdependent performance goals, and share an approach to work for which they hold themselves mutually accountable. Geographically dispersed teams allow organizations to hire and retain the best people regardless of location. A virtual team does not always mean teleworker. Teleworkers are defined as individuals who work from home. Many virtual teams in today's organizations consist of employees both working at home and small groups in the office but in different geographic locations. (en.wikipedia.org/wiki/Virtual_team)

1 Introduction

It is commonly agreed that networking, as a new way of collaboration, brings benefits to its members (Camarinha-Matos & Afsarmanesh, 2005). Here the term networking refers to a type of organizational structure that takes its coordination beyond the company boundaries. A collaborative network is an association of a set of participants (profit organizations, non-profit organizations, individuals, etc.) and may include organized and non-organized collaborations. In this thesis we focus on Collaborative Networked Organizations (CNOs) which represent only organized and intentional collaborations. They are constituted by a variety of entities (e.g. organizations and people) that are largely autonomous, geographically distributed, and heterogeneous in terms of their operating environment, culture, social capital, and goals (Camarinha-Matos & Afsarmanesh, 2005).

The heterogeneity of networks appears within several dimensions. First, in terms of time scale one can distinguish between long-term associations, like strategic alliances of companies, and short-term associations that are simply trying to complete a certain task. A long-term association of organizations, called a Virtual organizations Breeding Environment (VBE), has the purpose to enable fast creation of Virtual Organizations (VOs) (Camarinha-Matos & Afsarmanesh, 2003). A Virtual Organization (VO) is a short-term alliance created in order to fulfil a common business goal. In the case of individuals, a Professional Virtual Community (PVC) is a long-term alliance, with the aim to enable dynamic creation of Virtual Teams (VTs). Furthermore, these associations can be profit or non-profit oriented. Heterogeneity can also emerge from the type of participants involved in the collaboration: they can be individuals or organizations. This diversity has lead to the identification of specific types of collaborative networks (Camarinha-Matos & Afsarmanesh, 2005).

Several types of collaborative networks exist, namely Virtual Enterprises, Professional Virtual Communities, collaborative virtual laboratories, etc. In this thesis, we focus on recently emerged Virtual organizations Breeding Environments (VBE) (Camarinha-Matos & Afsarmanesh, 2003). A VBE is an association (also known as cluster) of organizations and their related supporting institutions that have both the potential and the will to cooperate with each other through the establishment of a base long-term cooperation agreement and interoperable infrastructure (Camarinha-Matos & Afsarmanesh, 2005). VBEs respond to business opportunities by forming Virtual Organizations (VOs).

Working in a CNO is dynamic in the sense that organizations and individuals may join and leave the CNO. Therefore, an introductory and learning phase for organizations joining a VBE should be as short as possible. The learning phase and collaboration in such networks implies sharing of knowledge and communication between network participants. Since the participants might come from different fields or follow a different philosophy, it is necessary to introduce a mechanism to share common understanding of the knowledge and to agree on a controlled vocabulary used to communicate.

An ontology provides a representation of knowledge which can be used and re-used in order to facilitate the comprehension of concepts and relationships in a given domain, and the communication between different domain actors by making the domain assumptions explicit. These actors can be software agents or people that need to access or share a piece of information (Gruber, 1993). Ontologies have proven to be an unambiguous and compact way of knowledge representation enabling mutual understanding, as they provide a basis for sharing information not only among people but also among software agents. If several organizations or individuals, that join a VBE, share the same underlying concepts (for example on the Web or on their intranet), then software agents are able to extract and aggregate information and use it to gather the data and to answer queries. Such agents can also support a process of VO and VT creation (from VBE and PVC, respectively) by proposing more or less optimal VOs and VTs based on competencies of their participants. In order to share the same terminology, the participants of a VBE or PVC need to agree on the terms that they intend to use for collaboration.

The goal of this thesis is to outline existing techniques and to propose new methods and techniques appropriate for modeling VBEs. The main contributions of the thesis are the following.

First, we propose an ontology for Collaborative Networked Organizations (CNOs), representing the common vocabulary and identifying the actors and relationships. The ontology focuses especially on organizations that collaborate in a Virtual organizations Breeding Environment (VBE). In this way, the thesis contributes to the formalization of the informal notions of VBE and virtual organizations (VOs) in a formal ontology language. This formalization is not merely a codification in an ontology language since it requires a detailed elaboration of all the concepts, consideration of mutual dependencies between concepts, and the overall consistency of the developed ontology. Additionally, we have made the CNO ontology, built in Protégé, available at a public Web site so that it can be redistributed and/or modified. The site also includes an extensive glossary of terms used/introduced in this thesis. This part of the thesis was accepted as a chapter entry in the Encyclopedia of Networked and Virtual Organizations (Plisson et al., 2007a) and as a regular SCI indexed paper published in the IEEE Transactions on Systems, Man, and Cybernetics Part C (Plisson et al., 2007b).

Second, we propose a methodology for semi-automated ontology construction in order to extract network specific knowledge related to the competencies of companies that are present in the network. Whereas the knowledge that was targeted previously was static and reusable in any VBE, the knowledge that we are now interested in is specific to each network, namely its competencies. Nevertheless the methodology that we present is general and thus can be applied to any VBE having the required data (in our case textual descriptions of companies) to extract companies' competencies.

Finally, the instantiation of the proposed CNO ontology and the developed approach to semi-automated structuring of competencies have been applied to and validated on two real problem scenarios. Whereas the experiment with Virtuelle Fabrik, a Swiss-German cluster of companies in mechanical engineering, was a real-life scenario with few companies (50 companies), we further evaluated our methodology on a large-scale experiment to structure the companies' competencies present in the Yahoo! Business directory (7107 companies).

Both experiments and the methodology were presented in an international conference (Plisson et al. 2005; Ljubič et al. 2005) and published in two journals (Ljubič et al., 2006; Plisson et al., 2007b).

The structure of this thesis is as follows. First, in Chapter 2 , we start with an overview of the domain of ontologies, which is the main building block of our contribution. We give definitions and outline existing techniques for building ontologies in Section 2.4 . In Section 2.5 we analyze several existing ontologies focused on the business domain and evaluate their possible inclusion in our work.

Chapter 3 presents the first part of our contribution, namely an ontology for Collaborative Networked Organizations (CNOs). First we define in Section 3.1 the goal of the proposed ontology and then we define and explain, in Section 3.2 , the main concepts and relationships included in the ontology. It is hard to evaluate an ontology without an application, therefore we have chosen to instantiate it on two real-life case studies. In collaboration with a manager of Virtuelle Fabrik, a Swiss-German cluster of companies in mechanical engineering, we modeled two existing Virtual Organizations (VOs) that were formed to respond to two different business opportunities. We were able to model accurately both cases and received good feedback from the manager and another expert evaluator. Both instantiations and comments can be found in Sections 3.3 and 3.4 , respectively.

Chapter 4 covers the second part of our contribution: the extraction of network specific knowledge for each VBE, namely the competencies of the companies. We present and explain the methodology that we used for structuring companies' competencies in Section 4.1 . In order to evaluate the methodology, we have conducted two experiments. First, in Section 4.2 , we apply our methodology on a large-scale case study, which consists in structuring the competencies of 7107 companies present in the Yahoo! Business directory. Companies in the Yahoo! Business directory are already organized in sectors and industries, which represent also the competencies of the companies. Our goal is to re-construct this hierarchy by using only the textual descriptions of the companies. In Section 4.2.3 we evaluate the result of our methodology and show that we successfully identified and structured the companies' competencies. In a second experiment (Section 4.3), we apply the methodology on a real-life scenario: we extract and structure the competencies of companies in Virtuelle Fabrik. The results are evaluated and discussed in Section 4.3.3 .

Finally, in Chapter 5 , we give a short summary of the results achieved, and present plans for further work. APPENDIX A contains descriptions of existing ontologies.

2 State of the art in ontologies

In this chapter, we provide definitions and enumerate different types of ontologies. We also explain different possibilities of encoding and methods for manual and automated ontology construction.

2.1 Ontology definition

An ontology provides a representation of knowledge, which can be used and re-used, in order to facilitate the comprehension of concepts and relationships in a given domain, and the communication between different domain actors. These actors can be software agents or people that need to access or share a piece of information (Gruber, 1993). The most basic type of ontology is a set of terms representing a controlled vocabulary (e.g. a glossary), which are the terms that people agree to use when dealing with a common domain. By providing definitions, an ontology helps people and machines to use the same terms for expressions and thus to achieve better mutual understanding. The role of an ontology is not limited to providing information; complex ontologies can also constrain the usage of knowledge by giving axioms or micro-theories and show the relations between the different components.

2.2 Types of ontologies

The content of an ontology depends both on the amount of information and on the degree of formality that is used to express it. Generally, we distinguish two main types of ontologies: lightweight and heavyweight (Gomez-Perez et al., 2004).

A lightweight ontology is a structured representation of knowledge, which ranges from a simple enumeration of terms to a graph or taxonomy where the concepts are arranged in a hierarchy with a simple (specialization, is-a) relationship between them.

A heavyweight ontology adds more meaning to this structure by providing axioms and broader descriptions of knowledge. As a word can have several meanings, knowledge can also be interpreted in different ways, which creates ambiguity in the knowledge base. Axioms and constraints tend to reduce the ambiguity by restricting and constraining the usage of information, for instance by specifying what is possible to do with it and what is not.

The encoding of an ontology varies from informal to a highly formal representation. A lightweight ontology is usually informal and sufficient to define concepts and basic relationships between them. A formal ontology contains axioms and definitions usually stated in logic. It is also called heavyweight because it can support more complex queries and deliver comprehensive answers.

The degree of complexity of the knowledge expressed in an ontology can vary from one ontology to another. This is also true for the spectrum of the knowledge. An ontology may cover one or several domains or even focus on a specific aspect.

In every case, the construction of an ontology involves the choice of appropriate concepts that will best describe the knowledge represented in the ontology. These choices are called “ontological commitments” and are described by the ontology. “We say that an agent commits to an ontology if its observable actions are consistent with the definitions in the ontology.” (Gruber, 1993).

The number of concepts chosen and their specificity make the ontology fall into one of the following five categories (Breuker et al., 1997 ; van Heijst et al., 1996 ; Gomez-Perez et al., 2004), ordered by degree of specialization:

- Upper (top-level, generic, foundational, etc.) ontology: A foundational ontology contains very general concepts that can define the most abstract entities (object, event, physical, abstract, etc.). All other concepts specialize this top level representation.

- Core ontology: A core ontology comprises knowledge about a field or area of expertise that may include several different disciplines such as law, computer science, etc. Only the most representative (core) concepts and relations of each discipline are kept in order to represent their union. These core concepts usually contain the root of each domain-specific ontology that represents a discipline.

- Domain ontology: Knowledge encoded in a domain ontology is more specific. It presents a more specialized view of the concepts defined in a core ontology where only the most important ones are kept. A domain ontology tries to cover all the aspects of one domain (e.g. medicine, law, etc.) and the interactions between them.
- Task (application) ontology: Knowledge can be even limited to the minimum required to fulfill the needs for one task. An application ontology can be reduced to a part of one domain ontology or even be a mix of two or more different domain ontologies. Only the concepts that are suitable to the task and to the comprehension of the process are kept.
- General ontology: In a general ontology, the knowledge is reusable in different domains. The knowledge can represent different kinds of concepts such as the units of measure (time, space, etc.) or even general relations that are applicable in most of the domains, such as the is-a (hierarchical) relation or the part-of (meronymical) relations.

2.3 Ontology representation and encoding

From the representation point of view, an ontology can be presented in several ways. Generally, concepts in an ontology are first grouped into several taxonomies with the is-a and subclass-of relations. Then these taxonomies are linked together with other relations such as meronymy (part-of) or any other predefined relation. In order to be shared across the web, an ontology is encoded in a format that facilitates the interchange. The most basic language to use is XML (Beckett, 2004), but it has many disadvantages, such as the lack of description power and the lack of commitment concerning the modeling primitives. Namely, the concepts are nested without telling what kind of relations are binding them together. These deficiencies require extensions of XML.

RDF (Beckett, 2004) is the first layer on top of XML which adds semantic information to the data. RDF allows for the representation of ground binary relations in the form of triples $<subject{,}predicate{,}object>$. The encoding in RDF is usually guided by an RDF schema (RDF(S) (Brickley, 2004)) that specifies the classes and properties that are intended to be used during the encoding process. RDF(S) allows to add more meaning within the definition of classes, properties and other resources. For instance it allows to give the range and domain of the properties that are defined. RDF(S) solves some semantic problems and is appropriate for encoding lightweight ontologies. But for heavyweight ontologies it still lacks expressive power.

A step towards heavyweight ontologies encoding was made with OWL (Web Ontology Language). OWL (McGuinness, 2004) is based on the Description Logics formalism and is divided into three sublanguages: OWL Lite, OWL DL and OWL Full[1]. The main advantage of OWL is that Description Logics has been an established research field for many years and thus it benefits from all the reasoning algorithms already developed and optimized. The knowledge expressed in such a language is formally defined and contains axioms that restrain its usage to a certain context, thus removing ambiguities during the reasoning process. The choice of one of these languages has to be made according to the requirements assessed when building an ontology.

2.4 Methods for ontology construction

Most existing ontologies were constructed manually. Such process is time-intensive, error-prone, and exhibits problems in maintaining and updating ontologies. For this reason, researchers are looking for alternatives to enable generating ontologies in a more efficient and effective way. Ontology learning has emerged as a sub-area of ontology engineering due to the rapid increase of the number of web documents and advanced techniques shared by the information retrieval, machine learning, natural language

1 **See http://www.w3.org/TR/owl-features/ for details.**

processing and artificial intelligence communities. This section provides an overview of the field of manual and semi-automated ontology construction.

2.4.1 Manual construction of ontologies using Protégé

It is important to emphasize some fundamental rules in ontology design, which help making decisions during an ontology construction. First, there is no single correct way to model a domain since there are always alternative views. The best solution almost always depends on the application and the extensions to be made. Second, the construction of an ontology is an iterative process. And finally, concepts in the ontology should reflect the objects and relationships in the domain of interest. The following methodology and recommendations taken from Noy and McGuinness (2001) concern the development of an ontology using Protégé[2], which we used for the construction of the CNO ontology described in Chapter 3 .

The development of an ontology starts with the definition of its domain and scope. Important aspects are also users, who and how they will use the ontology, and which kind of questions the ontology should provide the answers to. One of the ways to determine the scope of the ontology is to sketch a list of questions that the ontology should be able to answer, also referred to as competency questions.

The next step is to consider reusing existing ontologies. Ontologies are coded in a strict manner, therefore it is worth considering whether somebody else already covered the domain, at least partially. It is worth reusing and refining such ontologies, in order to reduce construction time and not to 'reinvent the wheel'. There are many libraries of reusable ontologies in the literature and on the Web (e.g. DAML[3], Protégé Ontologies library[4]).

After reusing existing ontologies has been considered, the enumeration of all the terms we would like to explain can start. The list of important terms and their properties will also help us to define the scope of the ontology.

The definition of classes and the class hierarchy is done based on the list of terms. There are several possible approaches to developing a class hierarchy. A top-down development starts with the definition of the most general concepts in the domain and subsequent specializations of the concepts. On the other hand, a bottom-up development starts with the definition of the most specific classes, which are the leaves of the hierarchy, followed by the grouping of these classes into more general concepts. Usually, it is possible to combine both approaches. One might start with a few top-level concepts and a few specific concepts, which can then relate to the middle-level concepts. Another alternative is to start at the middle-level concepts and then specialize and generalize them to obtain a full hierarchy.

Defined classes need properties in order to provide enough information to answer the competency questions from the first step. This can also be seen as an internal structure of the concepts. Usually the properties relate to those terms, from the list of all terms, which were not defined as classes. For each property in the list it is necessary to determine to which class it belongs. These properties become slots (or attributes) attached to classes. In general, there are several types of object properties that can become slots in an ontology. These can be intrinsic properties that belong to something by its very nature (e.g. flavor of wine), extrinsic properties (e.g. name, position etc.), parts (if an object is structured), or relationship. It is important to keep in mind that properties of general classes are inherited by specialized classes.

Properties are presented as slots, and slots can have different facets which define some constraints on properties like the value type, permitted values, cardinality, and other features of the value the slot can take. For example, name is usually a slot with value type String, while age takes an integer value. If it is the age of an employee, some allowed values can be specified, for instance, it cannot be negative or higher than 100. The types that appear in ontology construction software are String, Number, Boolean, Enumerated, and Instance-type.

The last step in ontology construction is to create individual instances of classes in the hierarchy by filling in the slot values. This process is referred to as ontology instantiation, which we used to validate the CNO ontology in Section 3.3 .

---

2 **The Protégé Project: http://protege.stanford.edu**

3 **www.daml.org**

4 **http://protege.stanford.edu/plugins/owl/owl-library/**

2.4.2 Learning ontologies from documents

There are several approaches to ontology learning from text. They are based on the use of text corpora. A corpus is a set of texts that should be representative of the domain, should be complete and cover all the aspects of a certain domain, and in addition, it should be accepted by the domain experts. All methods that learn ontologies from text try to exploit certain linguistic features of terms and surrounding terms to build concepts. Therefore we can identify two levels:

- linguistic level, where the knowledge is described through linguistic terms, and
- conceptual level, where the knowledge is described using concepts and relations between them.

The different techniques of ontology learning from text are based on how linguistic level structures are projected or mapped to the conceptual level. To achieve this, different methods use a combination of natural language processing and statistics.

2.4.2.1 Hierarchical clustering

Clustering is a statistical method used to partition data points in multi-dimensional spaces so that more similar points are put into the same partition. Beside partitions the output of hierarchical clustering is also a hierarchy, which can be seen as partitioning the data points into different number of partitions. The top partition represented as root node contains all data points while nodes at lower levels contain more partitions. Beside smaller number of points in those lower partitions, another property is that the average similarity between points is higher.

In the case of ontology construction the data points are text documents. The nodes at higher level are more general, since they contain more documents. Describing each of the clusters (nodes) of documents using most descriptive words in this cluster, one can obtain a hierarchy of terms, where the terms at higher levels describe more general concepts than at the lower levels. Khan and Luo (2002) presented a method that aims at building a domain ontology from text documents using clustering techniques and WordNet (Fellbaum, 1998). The method constructs the ontology in a bottom-up fashion as follows. First a hierarchy is constructed, using some document clustering technique. Documents that have similar content are associated with the same concept in the ontology. Then, a concept is assigned to each cluster of documents relative to the same topic in the hierarchy, using a bottom up concept assignment mechanism. To achieve this goal, a topic tracking algorithm and WordNet are used.

2.4.2.2 Pattern-based extraction

Relations at the conceptual level are recognized from sequences of words in the text that follow a given pattern (Hearst, 1992). For example, in English, a pattern can be established: if a sequence of *n* names is detected, then *n-1* first names are hyponyms of the n^{th} name. According to this pattern, the term *design company* could be used to obtain hyponymy relationship between the term *design company* and *company*. This relation at the linguistic level is projected on the conceptual level as the *subclass-of* relation between the concept *design company* and the concept *company*.

2.4.2.3 Association rule learning

Association rules are defined (Agrawal et al., 1993; Maedche & Staab, 2000) as an expression of the form *X implies Y*, where *X* and *Y* are sets of items constituting a transaction. The intuitive meaning of such a rule is that transactions which contain *X* tend to contain *Y*. In the ontology area, association rules have been used to discover non-taxonomic relations between concepts, using a concept hierarchy as background knowledge, and statistics for co-occurrences of terms in texts. For example, if the word *train* frequently co-occurs with the word *travel*, then we could add to the ontology a relation between the concepts associating *train* and *travel*. Hence, association rules can be used to add relationships to an existing hierarchy of terms. Note that the name of the relation is then manually put by the user of the ontology.

2.4.2.4 Conceptual clustering

The method of conceptual clustering (Michalsky, 1980; Bisson et al., 2000) takes as input a set of concepts which are then grouped according to the semantic distance (which must be smaller than a predefined threshold). Semantic distance calculation is based on the use of the syntactical functions that the terms associated to such concepts play in the text. For example if *train* and *car* appear with the same syntactical function (e.g. *John travels by train*, *Ann travels by car*), then the concepts associated to *train* and *car* are considered semantically close, therefore they should be grouped.

2.4.2.5 Ontology pruning

The main idea of this method is to refine a domain ontology, using as its base a core ontology (e.g. WordNet) which is enriched with the learnt concepts. Those are identified using natural language analysis techniques. Kietz et al. (2000) introduced a generic method used to discover a domain ontology from given heterogeneous resources by the use of natural language analysis techniques. It is a semi-automatic process in the sense that the user takes part in the process. In their approach, they have adopted the balanced cooperative modeling, where the work of building an ontology is distributed between several learning algorithms and the user. The method is based on the assumption that most concepts and conceptual structures of the domain to be included in an ontology as well as the terminology of a given domain are described in documents. The authors propose to learn the ontology using as a base a core ontology (like SENSUS, WordNet, etc.) that is enriched with new specific domain concepts. New concepts are identified using natural language analysis techniques over the resources previously identified by the user. The resulting ontology is pruned and focused to a specific domain by the use of several approaches based on statistics. Finally, relations between concepts are learnt applying machine learning methods. Such relations are added to the resulting ontology. The process is iterative in the sense that the resulting ontology can be refined applying the method iteratively. Faatz and Steinmetz (2002) also proposed a method to enrich an existing ontology by extracting meaning from the World Wide Web. It is based on the comparison of statistical information of word usage in the corpus, and the structure of ontology itself. Each concept in the ontology should have one or more phrases or words associated with it. Using this information, the approach proposes a method to calculate the semantic similarity between words in order to enrich the concept definition, and to create clusters of words related to a new concept. The method proposed by Agirre et al. (2000) aims at enriching the concepts in an existing large ontology using text retrieved from the World Wide Web. The goal is to overcome the lack of topical links among concepts, and the proliferation of different senses for each concept. OntoLearn (Missikoff et al., 2002) is a method for ontology construction and enrichment using natural language and machine learning techniques. The method proposes to use WordNet as a source of prior knowledge to build core domain ontology, after pruning all of the unspecific domain concepts. The method follows two approaches: statistical, to determine the relevance of one term for the domain; and semantic interpretation, based on machine learning techniques, to identify the right sense of terms and the semantic relations among them.

2.4.2.6 Concept learning

This method takes as input a taxonomy, which is incrementally updated as new concepts are acquired from real-world texts. Updating the taxonomy corresponds to techniques already mentioned (*pattern-based extraction, ontology clustering*, etc.). Hahn and Schulz (2000) presented a method for the maintenance and growth of domain-specific taxonomies based on natural language text understanding. A given taxonomy is incrementally updated as new concepts are acquired from real-world texts. The acquisition process is focused around the linguistic and conceptual "quality" of various forms of evidence underlying the generation and refinement of concept hypotheses. On the basis of the quality of evidence, concept hypotheses are ranked according to credibility and the most credible ones are selected to be included into the domain ontology. In this approach, learning is achieved by the refinement of multiple hypotheses about the concept membership of an instance. New concepts are acquired by taking into account two sources of evidence: background knowledge from the domain texts, and linguistic patterns in which unknown lexical items occur. Hearst (1998) describes a procedure, called hyponymy pattern approach, for automatically learning relationships between concepts in an ontology. It consists of looking for concepts that are related in an existing ontology (e.g. WordNet) and determining whether they are associated with each other in a word pattern that expresses that relationship. For instance, Shakespeare is a hyponym of poet in WordNet. Therefore, if we find in a text the pattern "poets such as Shakespeare" we can determine that the pattern "such as" usually indicates a hyponymy relationship.

2.5 Existing business related ontologies

In the last decade, many projects aimed at creating ontologies for different purposes. Some of the most important ones and best known are presented in APPENDIX A:. They are organized into the following categories:

- terminological ontologies: Wordnet, Verbnet, FrameNet, Sensus;
- domain ontologies: The Gene ontology, PSL;
- top-level ontologies: SUMO, Mikrokosmos, Sowa's ontology; and
- ontologies with common-sense knowledge: Cyc, ConceptNet.

The following subsections focus on the most important ontologies concerning the domain of business and enterprise modeling. These ontologies will serve as a source of inspiration for our methodology, in order to build a specific ontology for CNOs in Chapter 3 .

2.5.1 The AIAI enterprise ontology

An enterprise ontology is a collection of terms and definitions used in organizations. The AIAI enterprise ontology (Uschold et al., 1998) was developed in the scope of the Enterprise Project whose goal was to provide a set of tools for enterprise modeling. The available Enterprise Tool Set contains a Procedure Builder, for capturing process models, an Agent Toolkit for supporting agent development and a Task Manager for integration and visualization. The ontology was used in order to ensure a consistent communication between agents, either human or software. The enterprise ontology built within the Enterprise Project is not meant to be a complete ontology describing the enterprise domain. It only presents the most frequent terms used in this field. Thus the ontology has to be enriched for each specific business case.

The Enterprise Ontology is divided into five top-level concepts: Activities and Processes, Organization, Strategy, Marketing and Time. Table 1 presents an exhaustive list of terms composing the enterprise ontology, divided into sections.

Table 1 The AIAI enterprise ontology: list of terms composing the ontology.

Activity & processes	Activity Specification, Execute, Executed Activity Specification, T-Begin, T-End, Pre-Conditions, Effect, Doer, Sub-Activity, Authority, Activity Owner, Event, Plan, Sub-Plan, Planning, Process Specification, Capability, Skill, Resource, Resource Allocation, Resource Substitute.
Organization	Person, Machine, Corporation, Partnership, Partner, Legal Entity, Organizational Unit, Manage, Delegate, Management Link, Legal Ownership, Non-Legal Ownership, Ownership, Owner, Asset, Stakeholder, Employment Contract, Share, Share Holder.
Strategy	Purpose, Hold Purpose, Intended Purpose, Strategic Purpose, Objective, vision, Mission, Goal, Help Achieve, Strategy, Strategic Planning, Strategic Action, Decision, Assumption, Critical Assumption, Non-Critical Assumption, Influence Factor, Critical Influence Factor, Non-Critical Influence Factor, Critical Success Factor, Risk.
Marketing	Sale, Potential Sale, For Sale, Sale Offer, Vendor, Actual Customer, Potential Customer, Customer, Reseller, Product, Asking Price, Sale Price, Market, Segmentation Variable, Market Segment, Market Research, Brand Image, Feature, Need, Market Need, Promotion, Competitor.
Time	Time Line, Time Interval, Time Point.

The Organization part contains the terms representing the actors that play a role in an enterprise. They can have legal responsibilities or not, be a human or a machine. These terms are then used to model activities and processes. The Activity part includes the concept of resources and skills that are needed and the effects of the activity. In other words, it contains the concept of input-output. The central concept of the Strategy part is Purpose. Purpose captures the idea either of something which a ‘plan’ can ‘help achieve’ or that an Organization unit can be responsible for. Finally, the Marketing part describes sales. Sale is an agreement between two Legal-Entities for the exchange of a Product for a Sale-Price.

2.5.2 The TOronto Virtual Enterprise ontology (TOVE)

The Toronto Virtual Enterprise ontology (TOVE) was developed in the scope of the TOVE project (Fox, 1992). The TOVE ontology is a formal representation of the enterprise domain. As the Enterprise Ontology, it is divided into several top-level concepts to segment the enterprise into general categories: Activity, States, Causality, Time, Resources, and Organizational structure.

In this thesis we focus primarily on the Resources and Organization parts of the ontology. In TOVE, the resource ontology comprises two sets of terms or assertions. First, the resources are defined in terms of knowledge, role, mobility and division of the resource. The role of the resource represents its nature, for instance, whether it is a product, a tool or a work area. The mobility specifies the possibility of moving the resource from one place to another or not. The divisibility of the resource specifies if the resource can be divided into several resources, without affecting its role in an activity. Each division must be able to be consumed by an activity. Once these basic terms are defined, more complex ones are introduced such as the nature of the resource or its capacity. The nature of a resource means that a resource can be continuous or discrete. The capacity of a resource represents its availability at a certain point of time.

In the organization ontology, an organization-entity can be an organization-individual or an organization-group denoting several people (e.g. board of directors, teams, etc.). Each organization has properties such as organization-role, skills, constraints, etc. The organization-role specifies the goal that the organization has to achieve. Each role has attached skills, processes, policies, etc. that are necessary to complete the goal.

The concepts encoded in the ontology are also enriched by a set of axioms that define and constrain the interpretation of these concepts. The ontology is formalized using first order logic, allowing answering questions by using the TOVE reasoning engine.

2.5.3 The Business Process Management Ontology (BPMO)

The primary goal of the Business Process Management Ontology (BPMO) (Jenz, 2003) is to provide a stable platform for the semantically rich definition of business processes, in order to better align IT with business. The Business Process Management Ontology allows to define private and public processes, business entities, business objects and services that implement process activities. It follows the UN/CEFACT Modeling Methodology (UMM) for business process and information modeling.

The BPMO uses the concept of business entities and business objects for process modeling. Their definitions rely on the UN/CEFACT glossary, which defines a business entity as “something that is accessed, inspected, manipulated, produced, and so on in the business”. Business entities are defined for all the terms that are intended to be used in business and thus form a kind of glossary. Once these entities have been defined, they are generalized under new concepts called business objects. For instance, the business entities customer and supplier may be represented by a business object named party, which is a generalization of customer and supplier (Jenz, 2003).

The BPMO also introduces the notion of Process Task Concept Type. It describes which role performs a task, the business entities and business documents it is related with, and the resources it consumes. Every task represents a defined context, which includes the following items (Jenz, 2003):

- Role: A logical abstraction of one or more physical actors, usually in terms of common responsibility or position. An actor may be a member of one or more roles. Example: Mortgage Clerk.
- Business Document: The set of information components that are interchanged as part of a task. A business document may participate in a message flow. Example: Private Mortgage Loan Application Form.
- Durable Information Entity: An information entity that a task needs to perform its function, which must be represented in a persistent storage mechanism, and whose state must exist beyond the lifetime of the service (application) that implements the task. It may be composed of multiple business objects. Example: Private Mortgage Loan Application Information.
- Resource: A real object that can be identified. Example: Flatbed Scanner.

Currently, the Business Process Management Ontology comprises approx. 650 classes. The ontology is available in the OWL format.

2.5.4 Process Specification Language Ontology (PSL)

Process Specification Language (PSL) (Schlenoff, 1999) is an attempt to create a formalism for the representation of processes that are common to all manufacturing applications. The PSL ontology is formally defined using first order logic and the KIF language (http://ksl.stanford.edu/knowledge-sharing/kif/) to encode axioms.
The ontology is organized into two main layers:

- PSL-Core that comprises concepts that are common to all manufacturing applications, and
- a set of extensions that provide the resources in order to express other concepts that are not present in PSL-Core.

PSL-Core is composed of four basic classes: Activities, Activities Occurrences, Timepoints and Objects. An Activity Occurrence is a limited, temporally extended piece of the world. A process can then be defined as: one or more Activities that occur over a period of Time in which Objects participate.

In addition to the core theories comprised in PSL-Core, the ontology can be extended with additional sets of definitions (e.g. Activity Extensions, Resource Roles, Resource Sets). The definitions can be grouped and form extensions to PSL-Core.

2.5.5 The Yahoo! business ontology

The Yahoo! business sector data (http://biz.yahoo.com) provides structured information on business. Despite the fact that this is not an ontology in the sense of providing a commonly agreed terminology and categorization of the business domain, it does provide a useful structuring of company data.

The data consists of textual descriptions of 7107 companies[5] in terms of their competencies. In Yahoo!, companies are structured into 12 sectors, which are further divided into 102 industries. This data was used in the experiment described in Section 4.2 of this thesis.

Table 2 The Yahoo! Business Ontology, the number of industries and companies (per sector).

Sector	Industry	Industries	Companies
Basic Materials	Gold&Silver, Iron&Steel, …	11	429
Capital Goods	Aerospace & Defense, …	7	361
Conglomerates	Conglomerates	1	29
Consumer Cyclical	Footwear, Tires, …	12	318
Consumer Non-Cyclical	Beverages, Crops, …	8	232
Energy	Coal, Oil & Gas, …	4	310
Financial	Insurance, S&Ls/Savings, …	10	1212
Healthcare	Facilities, Major Drugs, …	4	860
Services	Advertising, Restaurants, …	25	1486
Technology	Hardware, Software, …	11	1578
Transportation	Airline, Railroads, …	6	150
Utilities	Electric, Water, …	3	142
Total		**102**	**7107**

[5] **This is the number of companies present in the Yahoo! business sectors at the time of writing, in 2005.**

For example, the Healthcare sector is divided into four industries: Biotechnology & Drugs, Healthcare Facilities, Major Drugs, Medical Equipment & Supplies. The number of industries in each sector and the distribution of companies per sector are shown in Table 2.

3 An ontology of collaborative networked organizations

There are many new concepts and terms in the field of Collaborative Networks (CN) and Collaborative Networked Organizations (CNO) (Camarinha-Matos & Afsarmanesh, 2005). A structured way to achieve a common understanding of the newly introduced terminology is to use an ontology for this domain. Other reasons for ontology building include the ability to reuse the knowledge, making domain assumptions explicit, separating the declarative domain knowledge from the operational knowledge, and possible analysis of domain knowledge.

There are also more specific reasons due to the nature of working in collaborative network organizations (CNOs). Such work is dynamic in the sense that organizations and individuals may join and leave CNOs. This section focuses on Virtual organizations Breeding Environments[6] (VBE, Camarinha-Matos, & Afsarmanesh, 2003), whose main purpose is to enable fast virtual organization (VO) creation[7]. An introductory and learning phase for organizations joining a VBE should therefore be as short as possible, and ontologies have proven to be an unambiguous and compact way of knowledge representation. In addition, ontologies provide a basis for sharing information not only among people but also among software agents. If several organizations joining a CNO share the same underlying concepts (for example on the Web or on their intranet), then software agents are able to extract and aggregate information and use it to gather the data and to answer queries. Such agents can also support a process of VO creation by proposing VOs based on their partner's competencies needed for achieving a business goal.

3.1 Determining the goals of the ontology

For every ontology, its domain and scope should be determined first. This includes the specification of the domain, and the potential use of the ontology. A set of questions an ontology should answer can also help determining the level of detail. First, there are some general questions concerning the newly introduced terms, and later some questions concerning the real cases (instances) of those and related terms. The questions one can ask are the following:

- What is a CNO?
- What is a VBE?
- What is a VO?
- What is the difference between VBE and VO?
- What form can the following actors take (e.g. a person, a company etc.) and what are the tasks of:
  - VBE administrator
  - VBE adviser
  - Support institution
  - VBE partner
  - VO broker
  - VO planner
  - VO coordinator?
- What is a business opportunity?
- How is a business opportunity handled when it arises?

---

[6] **Existing VBEs are usually called "clusters" or "industrial clusters", while some other authors use the term Virtual Web instead of VBE (Goldman, Nagel & Preiss, 1995 ; Franke, 2000)**

[7] **Since Mowshowitz (1986) used the term virtual organization for the first time, the variety of different terms and definitions were created to describe this new form of network organizations, including virtual company (Goldman & Nagel, 1993) and virtual enterprise (Hardwick, Spooner, Rando & Morris, 1996). See Franke's work (2000) for a detailed explanation of various terms and concepts.**

- Does a VBE administrator have to be a member of the VBE?
- Who was a broker of VO1?
- How many times was X a VO broker?
- Can a VO broker be also a VO coordinator?
- Is it necessary for a VO support provider to be a VO partner?
- Is it necessary for a VO partner to be a VBE member?

These questions show what level of detail goes into the operational phase of VBEs. From the questions we can also see that different parts of the ontology emerge:

- a general part, which defines the structure and function of a VBE,
- a part describing roles of participants of VBE and VO, and
- a part of the ontology describing organizations' competencies, resources and their availability.

Defining the goals of the ontology is the first step of the methodology (described in Section 2.4.1) that we have followed. The next steps are the definition of the concepts and the relations between them and the evaluation of the ontology. They are described in the following sections.

3.2 The CNO ontology

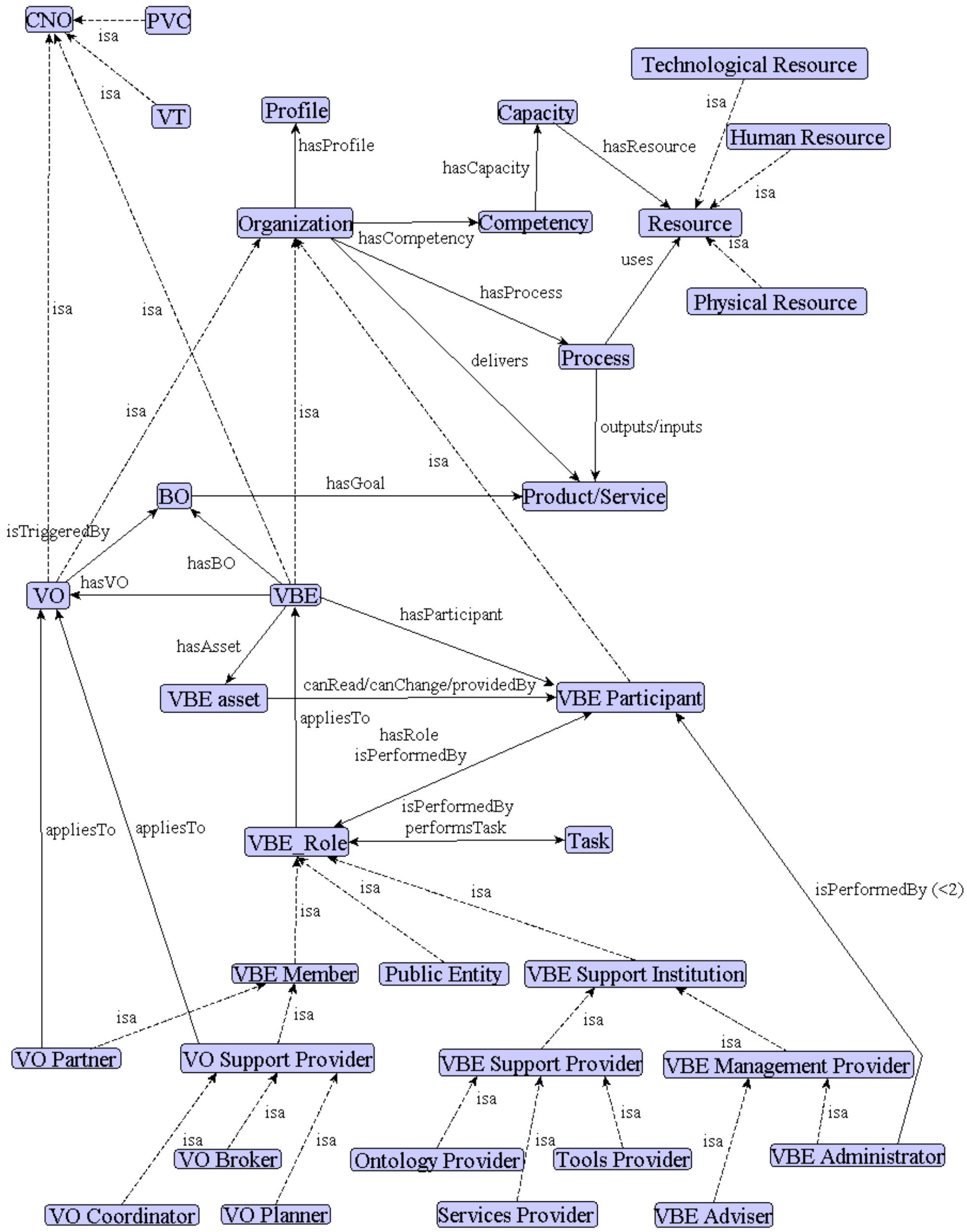

Figure 1 The CNO ontology, with a detailed elaboration of VBE and VO (without any detail on PVC and VT, which are out of the scope of this thesis).

The ontology of this thesis (shown in Figure 1) is based on the concepts developed in the ECOLEAD project (www.ecolead.org), as well as on the entity-relationships and concepts for virtual organizations developed in the SolEuNet project (Mladenić et al. 2003). The proposed CNO ontology is implemented within the Protégé framework following the methodology described in Section 2.4.1 , and is available as GPL-licensed software through a public web-site[8] so that it can be redistributed and/or modified. The site includes also an extensive glossary of terms used in this thesis. Moreover, the OWL plug-in was chosen in order to formalize the concepts and allow reasoning in the future.

[8] **http://protege.cim3.net/cgi-bin/wiki.pl?ProtegeOntologiesLibrary, pointing at http://ecolead.ijs.si/onto/cn.html**

3.2.1 Top-level of the CNO ontology

The top level concepts linking individual parts of the VBE ontology are shown in Figure 2. In further subsections other concepts of the ontology, including their instances, are introduced. The two top most concepts are Collaborative Network Organization (CNO) and Organization. A Collaborative Networked Organization is a special type of collaborative network comprising only organized collaborations while, in general, collaborative networks include both organized and non-organized collaborations. More generally, a collaborative network (Camarinha-Matos & Afsarmanesh, 2005) is constituted of a variety of entities (e.g. organizations and people) that are largely autonomous, geographically distributed, and heterogeneous in terms of their operating environment, culture, social capital and goals. These entities collaborate to better achieve common or compatible goals, and their interactions are supported by a computer network.

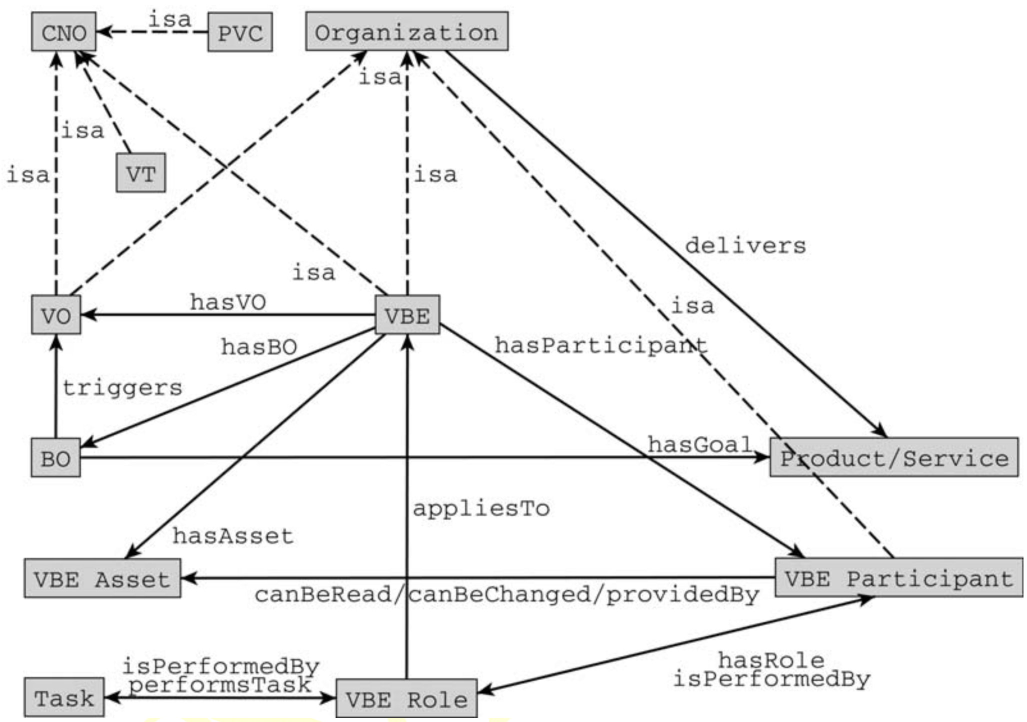

Figure 2 Top level concepts of the ontology. Specializations are shown using dashed lines, while other relations (along with their names) are represented with solid lines. Professional Virtual Communities (PVCs) and Virtual Teams (VTs) as forms of CNO are out of scope of this thesis.

Organization is a company, corporation, firm, enterprise or institution, or part thereof (whether incorporated or not, public or private) that has its own function(s) and administration that supplies products or services to other organizations[9]. All organizations have their own profiles and competencies, can perform one or more processes, and deliver some Products and/or Services. Special types of organizations are Virtual Organization (VO), Virtual Organization Breeding Environment (VBE) and VBE Participant. VO and VBE are also special types of Collaborative Networked Organizations, since they represent alliances of companies and individuals. VBE participant represents an entity collaborating with other entities in the VBE and VO. A Virtual organizations Breeding Environment (VBE) is “an association (also known as cluster) or pool of organizations and their related supporting institutions that have both the potential and the will to cooperate with each other through the establishment of a ‘base’ long-term cooperation agreement and interoperable infrastructure” (Camarinha-Matos & Afsarmanesh, 2005). A VBE responds to business opportunities by forming VOs. As an organization, a VBE has also competencies, which are not limited to the union of the competencies of its participants. The VBE competencies are the

9 **ISO definitions. Available at http://www.bizmanualz.com/ISO9000-2000/ISO definitions.html**

result of combining two or more participants' competencies to realize more complex projects (e.g. building a highway, a bridge, etc.). Each participant can take one or more different roles that are further defined in subsequent subsections. On the other hand, a VO is a shortterm association with a specific goal of being active in fulfilling a Business Opportunity (BO). A business opportunity is a time or occasion with favorable combination of circumstances that is suitable to start a business. VO represents a temporary alliance of diverse organizations that form a collaborative network, sharing knowledge, skills and resources in order to respond to a specific BO. The partners of the VO are selected from the VBE participants according to their competencies and availability to deliver products or services required to fulfill the BO. To facilitate the processes within the VBE, the participants have access to several assets, such as business rules, software tools, specifically developed for the VBE. These assets are called VBE Assets and are kept in the so-called VBE Bag of assets; they are further described in Section 3.2.4 .

3.2.2 Organization-related ontology

The organization part of the ontology introduces concepts necessary to describe organizations' ability to perform certain operations at a certain time. Concepts and their relations are shown in Figure 3.

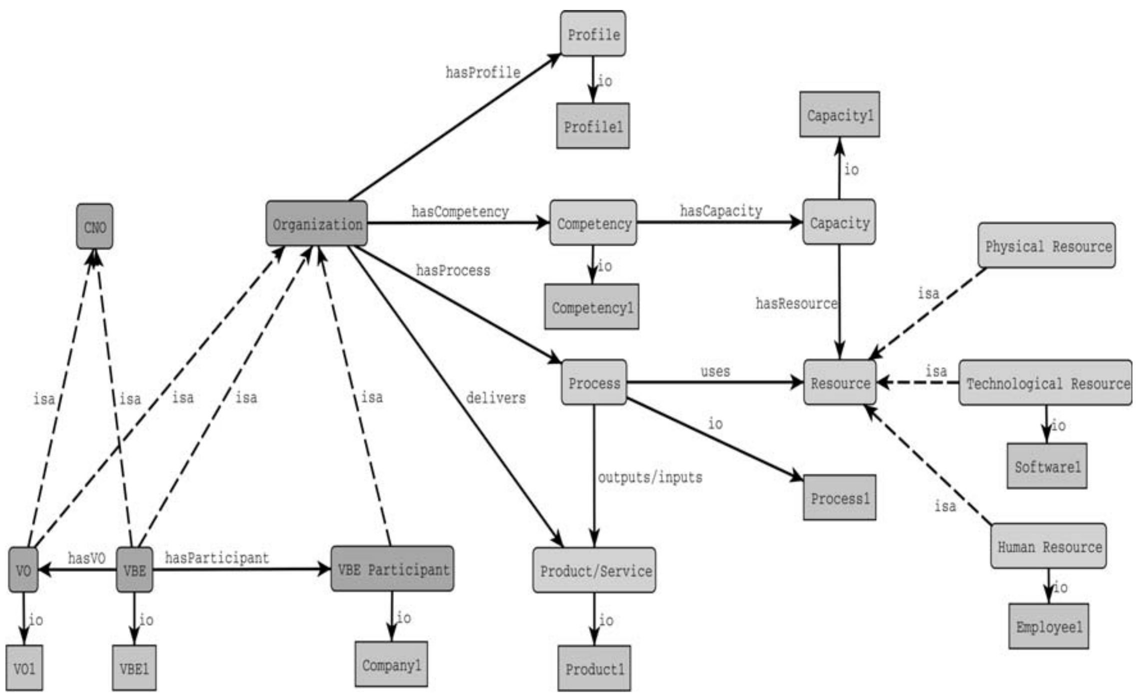

Figure 3 The organization part of the ontology, instances are linked to concepts with io (instance of) relations.

The concept Profile is a set of structured information describing the organization in general such as name, contact information, description etc. Each organization covers one or more Competencies, which define its capability to perform processes. A Process is a structured, managed and controlled set of interrelated activities that uses resources to transform inputs into specified outputs. The final aim of an organization is to offer certain products or services to the customers at a certain time. Therefore, competency is related to Capacity, describing its availability in terms of start time, duration, and availability. The attribute availability of Capacity is expressed as percentage, and specifies which amount of the resource can be used (for instance, a software engineer capable of writing 1000 lines of code per week available for one month at 50% means that he will be able to write 1000*4*0.50 = 2000 lines of code). Resource represents an element consumed in a process that performs a number of operations which can transform inputs into outputs. There are three specializations of resources, namely Physical, Technological, and Human Resource. Organization's physical resource refers to buildings, machines, equipment, etc. within the organization, technological resource refers to hardware and software within the organization, while human resource refers to the individuals within the organization. Resources are split into three categories due to different attributes used to describe them.

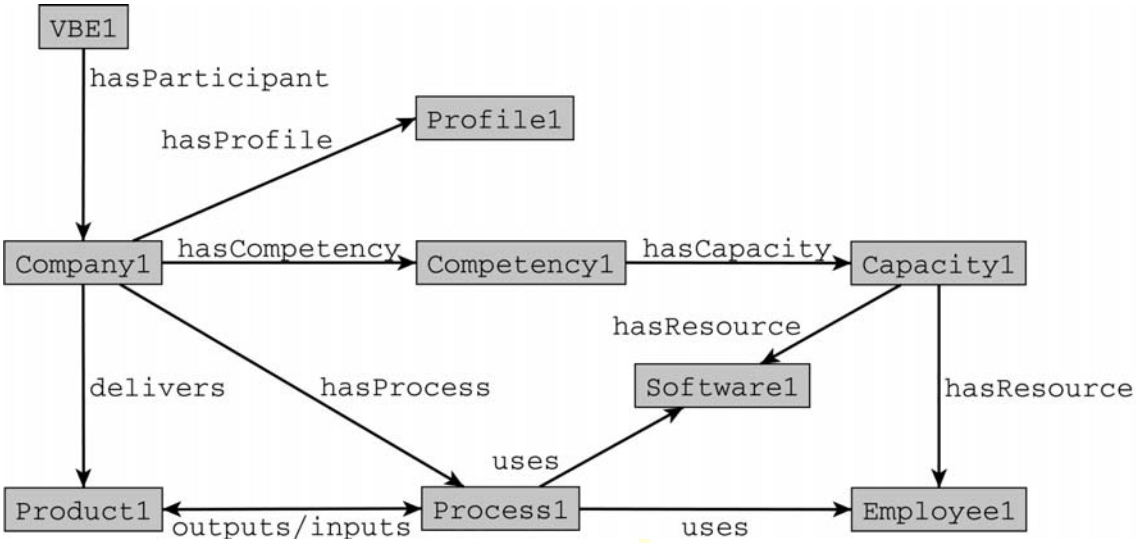

Figure 4 Example of instantiation of the organization's ontology.

Figure 4 shows instances and their relations for better understanding of the ontology. As an example there is one instance of VBE (VBE1) which has a participant named Company1. This company has a certain profile and only one competency. This competency is available at a certain time (Capacity1) with two resources, namely Software1 and Employee1. Company1 delivers its product (Product1) through Process1, that uses previously mentioned resources.

3.2.3 VBE-role related ontology

A Participant is any organization within a VBE and can participate in a different way. Each participant in the VBE can have one or more specific Roles, which are classified into three main categories:

- roles that apply to members of the VBE,
- roles that apply to support institutions, and
- roles that apply to public entities.

All these roles have in common a relation with the concept Task, which specifies all the necessary tasks to be performed within the VBE and VO. For each role certain tasks must be performed. Public entity is the role taken by a participant, which is not registered in the VBE. We do not cover it in any detail here. Member and support institution roles are defined in the following subsections since they are more complex and relevant for the ontology.

3.2.3.1 VBE member role

VBE-Member is the basic role (shown in Figure 5) of those organizations registered in the VBE and are willing to participate in the VBE's activities. The principal activity carried out by members of the VBE is to collaborate in VOs. When involved in one or more VOs, they can take different roles such as simple VO Partner or VO Support provider. The support provider role is further divided into three sub roles: VO coordinator, VO broker and VO planner. These roles can be performed by only one VBE member, whereas the role of simple VO partner can be taken by more than one member that constitute the VO. VO partner is the role of a VBE participant in a VO. Opportunity Broker is the role of a participant that identifies and acquires new collaboration opportunities (Business Opportunity (BO)), by marketing VBE competencies and assets and negotiating with potential customers. The VO planner is responsible for identifying the necessary competencies and capacities, select the appropriate partners accordingly and structure the new VO. In many cases the roles of Broker and Planner are performed by the same actor. The VO coordinator coordinates the VO during its life cycle in order to fulfill the goals set for the collaboration opportunity that has triggered the VO.

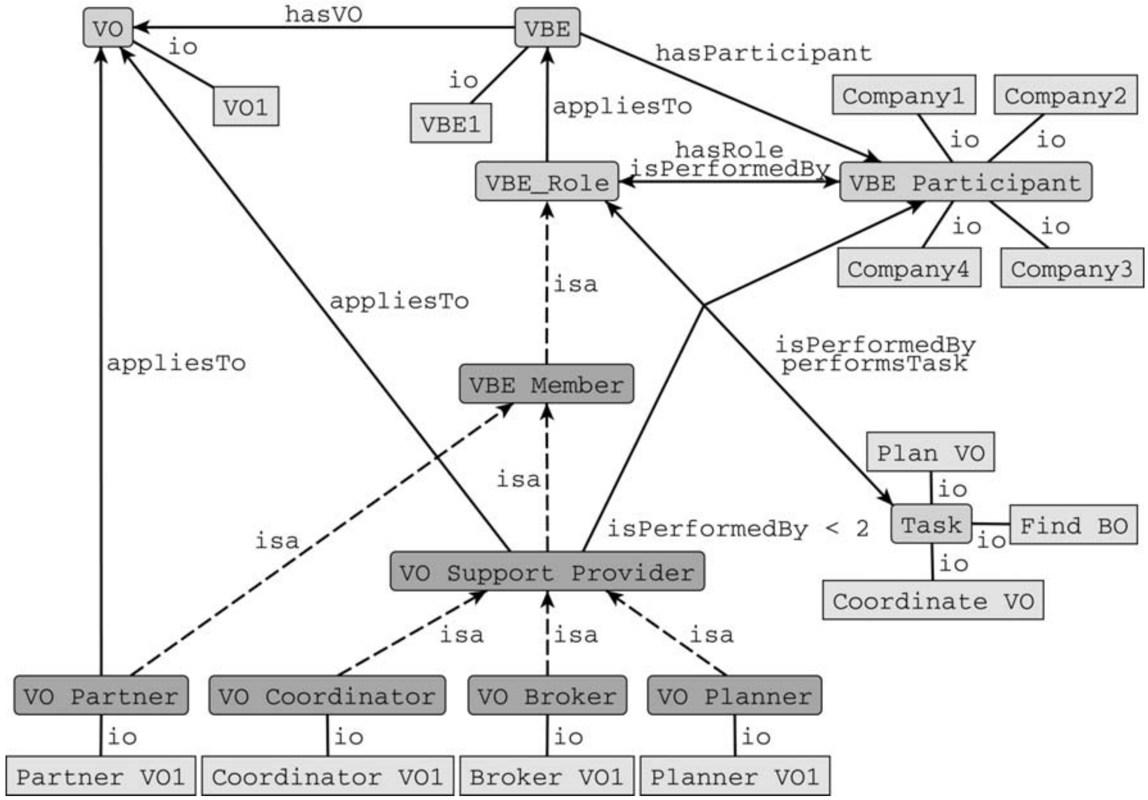

Figure 5 Roles and their structuring within a VBE.

Figure 6 shows an example using instances of the concepts previously defined. The figure illustrates the case of a VBE (VBE1) with four participants (companies 1-4). Each participant is assigned a role, VO Partner, Broker, Planner and Coordinator respectively for companies 1-4. Each role has assigned a certain number of tasks that the owner of the role has to perform. In addition, each role is attached to a specific VBE and VO, since a member can have several roles in different VBEs and VOs.

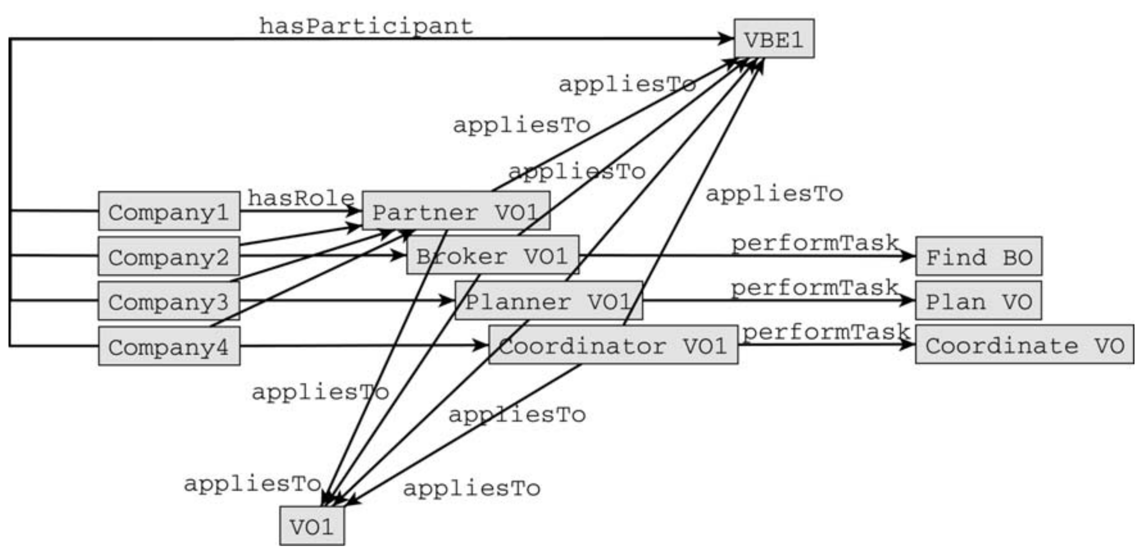

Figure 6 An example of VBE, VO, its participants, and their roles.

3.2.3.2 Support institution-related ontology

Many industry sectors (such as real estate, construction, banking, education, maintenance, etc.) collaborate with several clusters and serve as Support Institutions that provide services. The same is true within a VBE; the VBE has its own competencies to solve specific problems but may require the help of support institutions for other tasks where it does not have enough expertise or does not want to get involved. Support Institutions can provide a broad range of services, such as training, research, consulting, information services, legal and contractual services, etc. Moreover, support institutions always try to keep up with new trends and technologies in their field of expertise and thus tend to propose a solution that is up to date and optimized for the client. Usually, employees inside the support institution are also trained in order to better understand the needs of the client and try to continuously improve their methods.

Support institutions can be used in a VBE for different purposes:

- to promote entrepreneurship,
- to promote VBE capabilities,
- to help VBE members to achieve competitiveness,
- to provide an appropriate ICT infrastructure,
- to make easy the installation of enterprises that add value to the VBE,
- to make easy the application of modern theories in industry,
- to provide the best coordinated support,
- to direct funds for industry development in a VBE,
- to open markets via inter-regional and international trade agreements,
- to encourage SME's growth,
- to provide appropriate infrastructures,
- to collaborate with brokers in commercial missions,
- to collaborate between diverse governmental levels (e.g. local, regional, and national), and/or
- to support research and development activities.

For instance, building companies make use of legal support institutions to have insurance for each of their contracts. Support institutions can be even used in collaborative networked organizations to train new partners, perform administrative tasks or even to make and keep the network running. Some clusters used in the case study in Section 4.3 use a support institution to create the business development of the cluster and to implement new innovative concepts. In this case, the support institution provides services in the form of consulting or even workshops where tools are designed specifically for the needs of the cluster. A support institution also helps to develop marketing strategies and PR (Public Relations) materials such as leaflets, web-sites, etc. to promote new products. If the support institution invests sufficient effort and involvement, its help may increase the productivity of the client and thus lead to the increase of the revenues.

In the ontology of Figure 7, the Role VBE Support Institution is divided into two main sub-Roles:

- VBE Support Provider, that includes Ontology Provider, Services Provider and Tools Provider roles, and
- VBE Management Provider that includes VBE Adviser and VBE Administrator roles.

The roles under the Support Provider concept deliver services useful for the VBE, such as an ontology (Ontology Provider), software tools (Tools provider), legal services (Services Provider). The roles under the Management Provider concept deliver management services. VBE administrator is the role of a VBE participant responsible for the VBE operation and evolution, promotion and cooperation among the VBE members. He/she is also filling the skills/competencies gaps in the VBE by searching and inviting new organizations in the VBE and manages the VBE general processes. The VBE Adviser is responsible to monitor the network and suggest recommendations to the administrator.

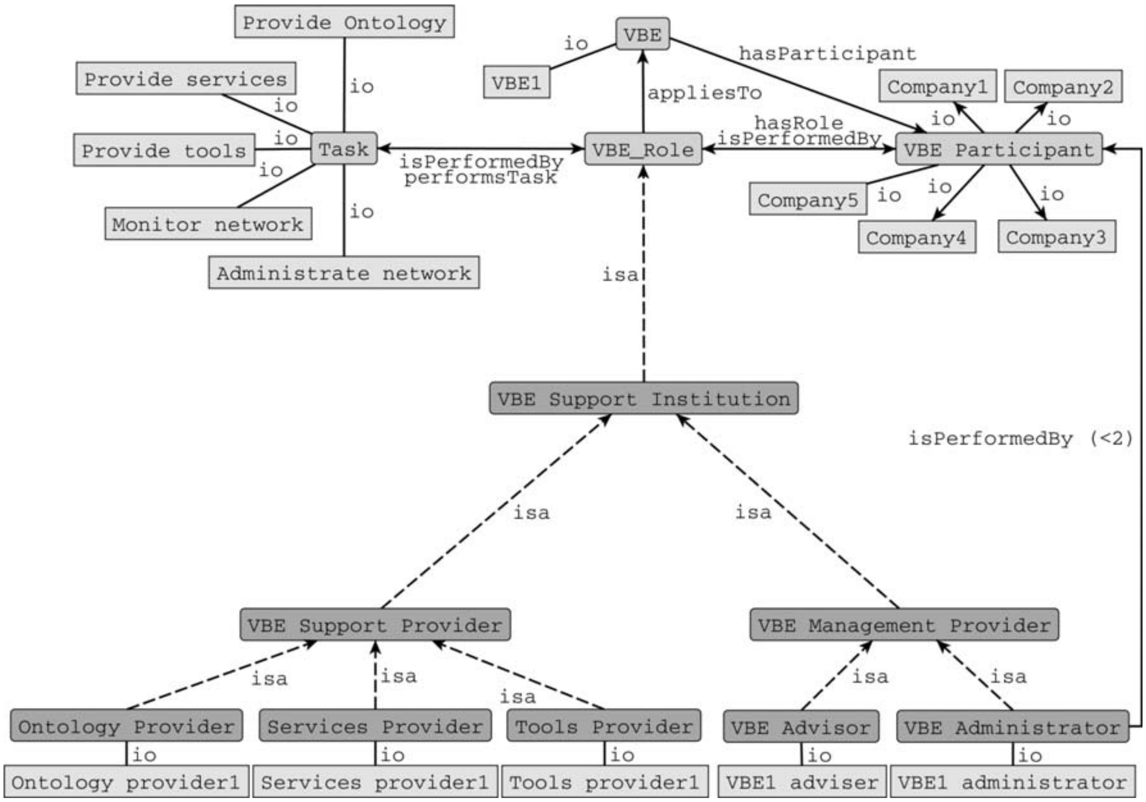

Figure 7 Support institution roles and their instantiations.

Figure 8 shows an example of a VBE with five participants (companies 1-5), which have support institution roles. Each of them is attached to a specific task with respect to the role. Each role is attached to a specific VBE, since a participant can have several roles in different VBEs.

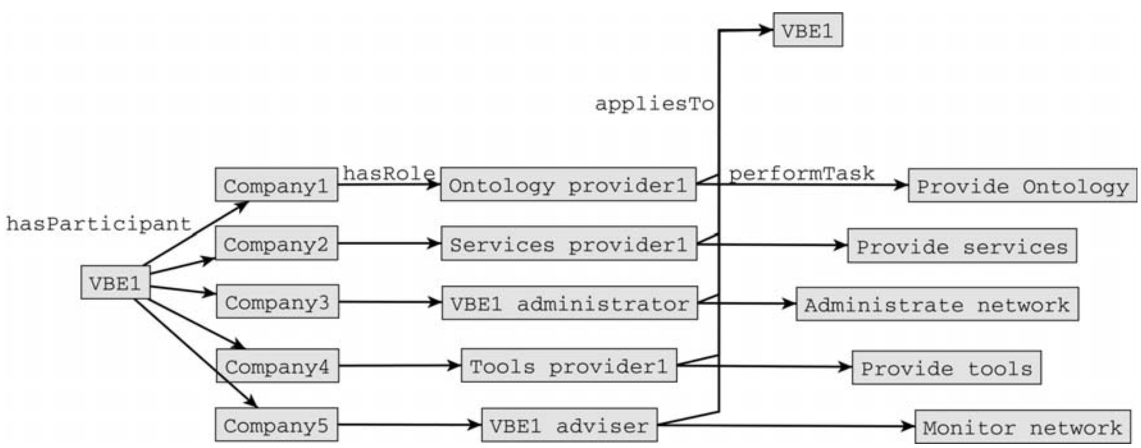

Figure 8 Example of a VBE with five participants.

3.2.4 VBE bag of assets-related ontology

In business and accounting, an asset is anything owned which can produce future economic benefit. It can be owned by a person or a group acting together, e.g. a company. Its value can be expressed in monetary terms. Assets can be classified according to the generally accepted accounting principles as follows:

- Current assets: cash and other assets expected to be converted to cash.
- Long-term investments: these are to be held for many years and are not intended to be disposed in the near future.
- Fixed assets: purchased for continued and long-term use in earning profit in a business. This group includes land, buildings, machinery, furniture, tools, wasting resources (assets which decline in value over time, e.g. gas, oil etc.).
- Intangible assets: these lack physical substance and usually are hard to evaluate. They include patents, copyrights, franchises, goodwill, trademarks, trade names, etc.
- Other assets: this type includes a high variety of assets, most commonly long-term prepaid expenses, long-term receivables, property held for sale etc.

All these types of assets can be found inside a single company. Some of them can also be a property of a VBE. The main purpose of the VBE related assets is to speed up and improve the process of a VO creation (which is the main task of a VBE). In a VBE the following potential assets have been identified:

- General policies in the form of documents, books, leaflets to help (new and old) members to easily follow the guidelines of a VBE.
- Sample contracts to speed up the contracting phase.
- General legal issues related to the sector.
- Information of interest, specific to the sector.
- Links to other sources of information.
- 'Lessons learned'. This is a database system designed to collect and make available lessons learned in the business. It enables the knowledge gained from past experience to be applied to current and future projects. Its intention is to avoid the repetition of past failures and mishaps, as well as the ability to share observations and best practices. Through this resource members seek to facilitate the early incorporation of quality into the design of their products and services.
- FAQs.

The VBE Asset concept comprises the name and the description of an asset. Each asset belongs to a VBE encoded as the property hasAsset of the concept VBE. Different rights to access an asset can be specified for reading and changing using properties canRead and canChange, respectively. The owner of an asset is identified using a property providedBy. The users and the owner of an asset must be of the type VBE Participant, as presented in Figure 2.

3.3 Instantiation of the CNO ontology

In order to verify the appropriateness of the proposed formalization of the CNO ontology we have instantiated the ontology to two existing VOs. The two actual VOs were formed by the Virtuelle Fabrik industrial cluster and represent a real-life experiment. The first VO was created for the task of construction of a maintenance machine for a nuclear power plant. The second VO was formed for the task of a gearbox construction. The two instances of the VO ontology are shown in Figure 9 and Figure 10, respectively. In accordance with the VBE ontology, each VO consists of partners which have one of the following roles: VO partner or VO support provider (VO coordinator, VO broker or VO planner). In this experiment, we were particularly interested in the VO partners and their competencies. The VO created for the task of construction of a maintenance machine for a nuclear power plant, shown as a node "maintenance machine VO" in Figure 9, consisted of five VO partners: Alwo, SMA, Innotool, Rihs, Suter&Baehler, and Beni, shown as nodes in the figure. Each partner is linked to its competencies (needed for the particular business opportunity for which the VO has been created) by the "hasCompetency" relation. Similarly, the VO created for the task of a gearbox construction, shown in Figure 10, consisted of five VO partners: Beni, Knobel, Alwo, SMA and Innotool.

3.4 Evaluation of the CNO ontology

The expert evaluation of the ontology was performed by asking the representatives of two different VBEs (Virtuelle Fabrik and IECOS) involved in the ECOLEAD project whether the proposed VBE ontology accurately describes their activities and can help them in their work. Below are the comments that we have received:

- Evaluation by the Virtuelle Fabrik expert: “The ontology describes our organization in a very accurate way. Especially Figure 9 could help us to illustrate activities within the scope of a running VO. The ontological description of competencies will also help us in further software projects and ECOLEAD activities to improve the accuracy of assumptions.”
- Evaluation by the IECOS expert: “For IECOS, the ontology you propose is accurate and appropriate for describing the VBE. We think it is complete as it is, and is also simple and easy to be understood by anyone, but could be improved with some of our feedback comments below.” The IECOS feedback comments are summarized in the following items:
    - First, there should be a link between ‘process’ and ‘competency’ (see Figure 3 and Figure 4) since “a competency allows to perform (a) process(es).”
    - Another observation that could be considered is to add the class ‘practice’ to the ontology and relate it to ‘process’ because “a process is supported by a practice(s).”
    - Finally, a ‘competency’ is related to a ‘process’ through the element ‘capability’ because “a competency has a processing capability related to a process.”

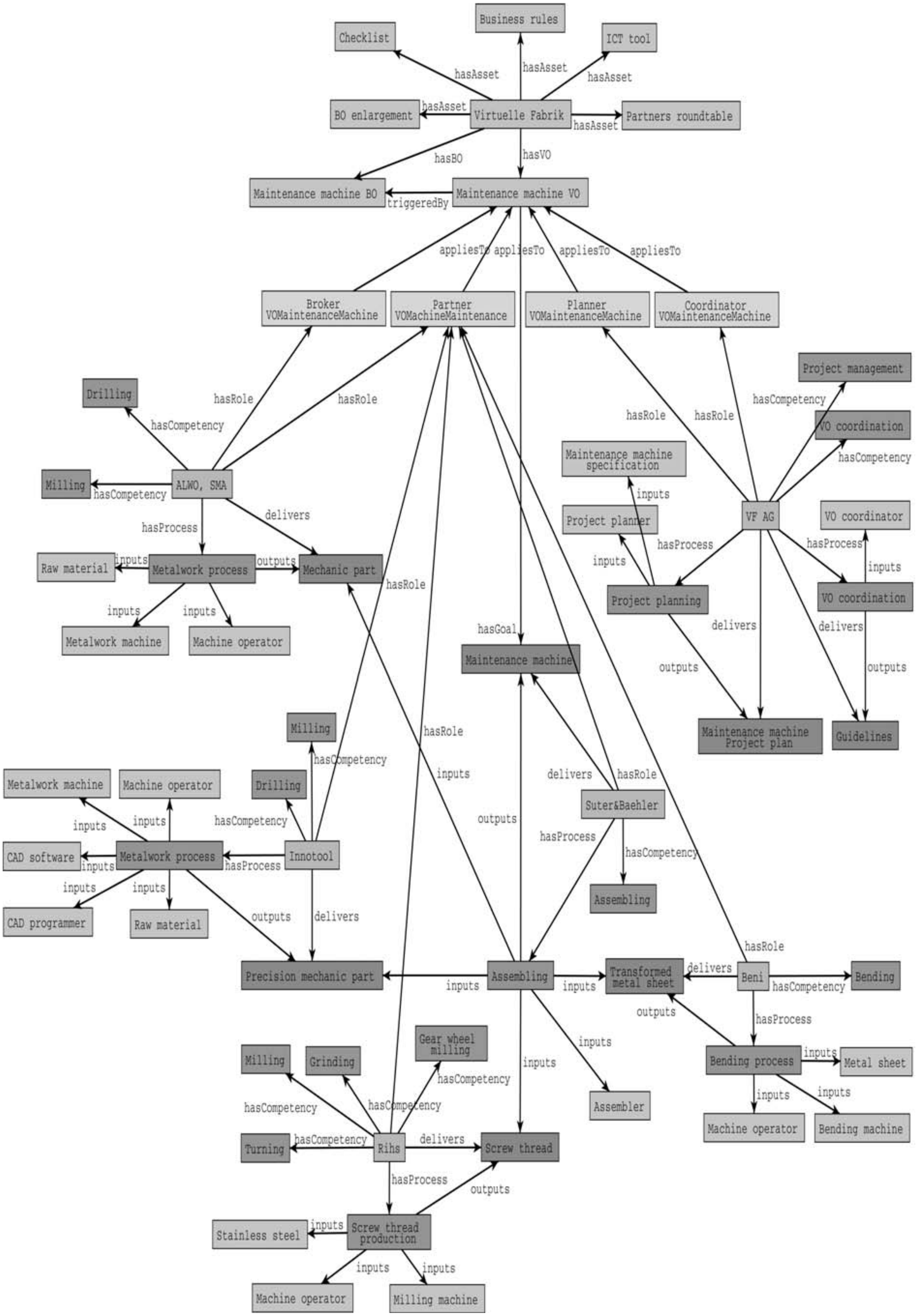

Figure 9 A VO created (within VF) for the task of construction of a maintenance machine for a nuclear power plant.

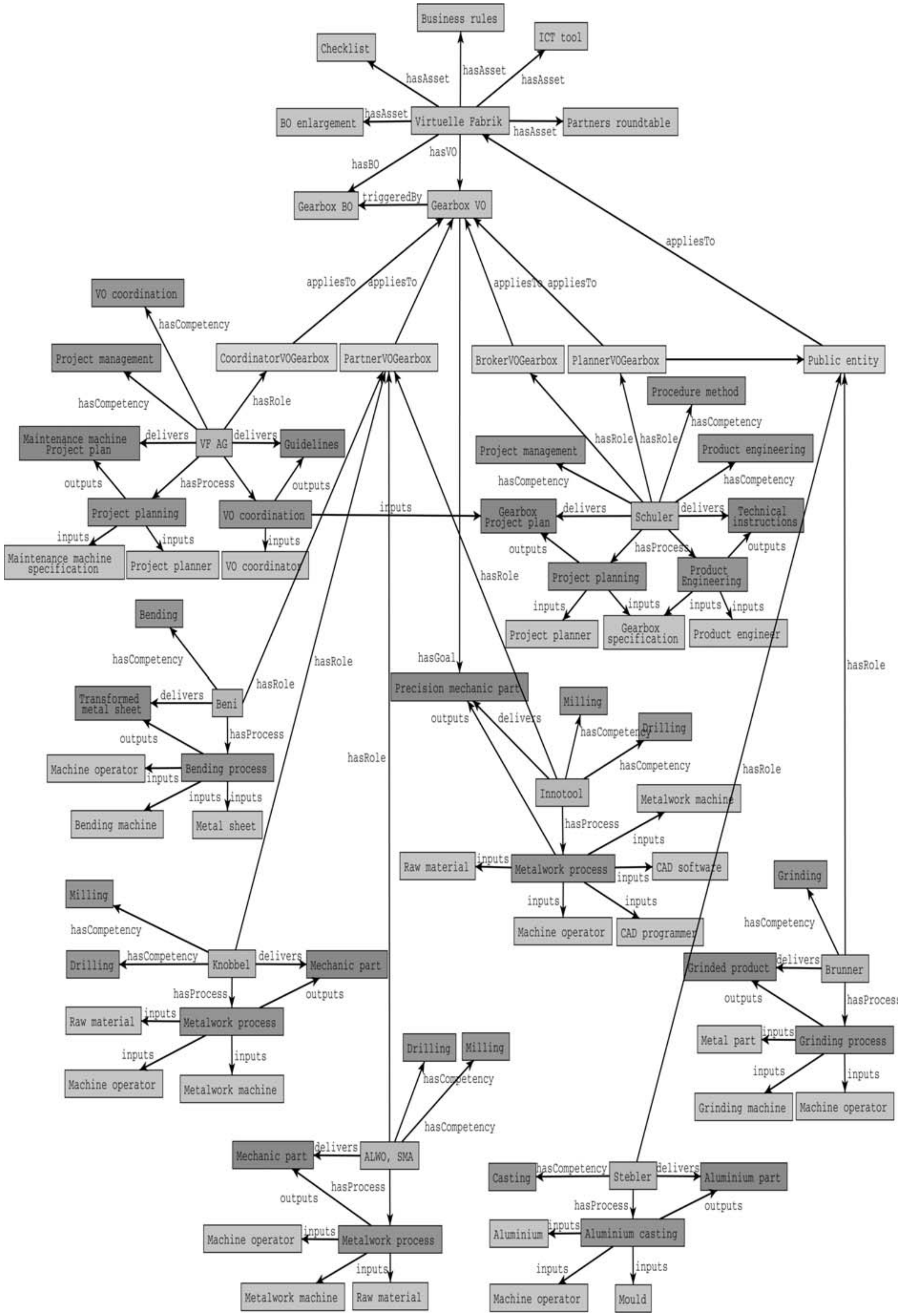

Figure 10 A VO created (within VF) for the task of a gearbox construction.

4 Extending the CNO ontology with company specific information

In this chapter, we will focus on Virtual organizations Breeding Environments (VBEs), which are a special kind of Collaborative Networked Organizations. The concept of VBE and its definition were presented in Section 3.2 . The competency part of the VBE ontology is meant to keep information about companies' competencies and their availability. This information is specific to each VBE and must be identified and structured at the creation phase of the VBE. This information is updated every time a partner joins or leaves the network or when an existing partner acquires new competencies. The number of domains and their diversity makes it unmanageable to create a universal competency ontology that can be used in all VBEs. Therefore we propose a methodology that speeds up the process of gathering and structuring companies' competencies using machine learning techniques. Identification of a hierarchy of competencies (Section 3.2.2) and linking the organizations with competencies (with hasCompetency relation) is performed using clustering on textual descriptions of companies.

4.1 Methodology for semi-automated ontology construction

The proposed methodology for structuring the competencies consists of the following steps:

1. Data gathering (yields textual data).
    a. Data can be gathered manually through questionnaires filled-in by companies.
    b. Alternatively, data is also available on the Web, including company home pages and public registers. In this case, a data gathering method employed can be focused Web crawling (Ester et al.).
2. Preprocessing (of textual data into bag-of-words):
    a. Markup tags and stop-words elimination.
    b. Stemming and/or lemmatization. (Porter, 1980).
    c. Transformation into the bag-of-words (BOW) representation where a document is encoded as a feature vector with word frequencies as elements. Elements of vectors are weighted with IDF weights (Inverse Document Frequency) (Deerwester et al. 1990). All the i-th elements are multiplied with IDFi = log(N/dfi), where N is the total number of documents and dfi is document frequency of the i-th word (the number of documents in which the i-th word appears). Such vectors are called TFIDF vectors[10].
3. Structuring (of bag-of-words into clusters). Structuring of the BOW representations is performed by document clustering (Steinbach et al. 2000). We applied document clustering to automatically build a hierarchy of companies, based on their descriptions, with a subset-of relationships between the groups of companies. In our experiments we used a kmeans hierarchical clustering system gCLUTO (Rasmussen & Karypis, 2004). The result of clustering is a simple ontology - a taxonomy - which is a tree structure with classes, subclasses, and instances.
4. Visualization (of the taxonomy). Many methods were developed for visualization of text documents or high dimensional data in general. Some examples are Themeview, Themeriver, Topic Islands (Miller et al. 1998), and Self-Organizing maps (Kohonen, 1989). In this work we have applied mountain visualization (Rasmussen & Karypis, 2004).
5. Ontology evaluation means a comparison to existing ontologies and/or manual evaluation by domain experts. Subsequently, the developed hierarchy can be manually refined and elaborated to improve the deficiencies revealed by the evaluation.

[10] **TFIDF (term frequency - inverse document frequency) is a weight often used in Information Retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document. The importance increases proportionally to the number of times a word appears in the document but is offset by how common the word is in all of the documents in the collection or corpus.**

4.2 Competency structuring from the Yahoo! Business data

In this experiment, we will evaluate quantitatively the proposed methodlogy on a large scale scenario, involving a large number of companies. We have partially implemented the proposed machine learning approach, described in Section 4.1 , through the use of two document clustering systems, both performing hierarchical k-means clustering (Steinbach et al. 2000) and providing the visualization of the generated clusters. We have implemented only steps 1 to 4 of the procedure outlined in Section 4.1 . As there was no expert involved in the experiments, we were unable to implement step 5. As an alternative, we were only able to evaluate the results of clustering by comparing the results of clustering to the existing human-labeled Yahoo! ontology, an evaluation approach which is obviously unrealistic in real-life expertise modeling scenarios. A real-life modeling scenario is described later in Section 4.3 .

4.2.1 Description of the data

We have performed the analysis of Yahoo! business data, which we have downloaded from the Yahoo! business sector (see http://biz.yahoo.com). The experimental data set consists of textual descriptions of 7107 companies. The length of the summaries varies from 180 to 1031 characters, averaging in approx. 842 characters per description. Companies are structured into 12 *sectors*, which are further divided into 102 *industries*. For example, the *Healthcare* sector is divided into four industries: *Biotechnology & Drugs, Healthcare Facilities, Major Drugs, Medical Equipment & Supplies.* The number of industries in each sector and the distribution of companies over the sectors are shown in Table 2 of Section 2.5.5 .

4.2.2 Clustering

Trying to build an ontology of 7107 company summaries manually, we would have faced the problem of not knowing the characteristics of different business areas (e.g. banking, software, healthcare etc.), which would have disabled us of producing a relevant structuring of the domain. In this experiment, our goal was thus to automatically construct a hierarchical structure of companies with distinct categories, with the potential (in step 5 of the approach presented in Chapter 3) to be interpreted as an ontology of Yahoo! companies.

We applied document clustering to automatically build a hierarchy of sets of documents, i.e., a hierarchy of company groups with a subset-of relationships between the groups of companies. In hierarchical k-means clustering, used in our approach, all companies are split into k groups; each group is further split into k subgroups, based on the similarity between company descriptions. In our experiments we used two different clustering and visualization systems.

The first system (Grobelnik & Mladenić, 2002) provides a two dimensional visual representation of document groups generated by k-means hierarchical clustering. The system performed several levels of 2-means clustering, and the stopping criterion (minimum number of companies in the clusters) was set to 1000. This resulted in a company hierarchy of 5 levels containing 11 nodes as shown in Figure 11. The main idea of tiling visualization is to split the rectangular area, representing all the companies, into sub-areas according to the size (numbers of instances) of sub-clusters. When a stopping criterion is satisfied, keywords describing the clusters are assigned to the leaves. The levels of the hierarchy are denoted by the ellipses connecting similar groups.

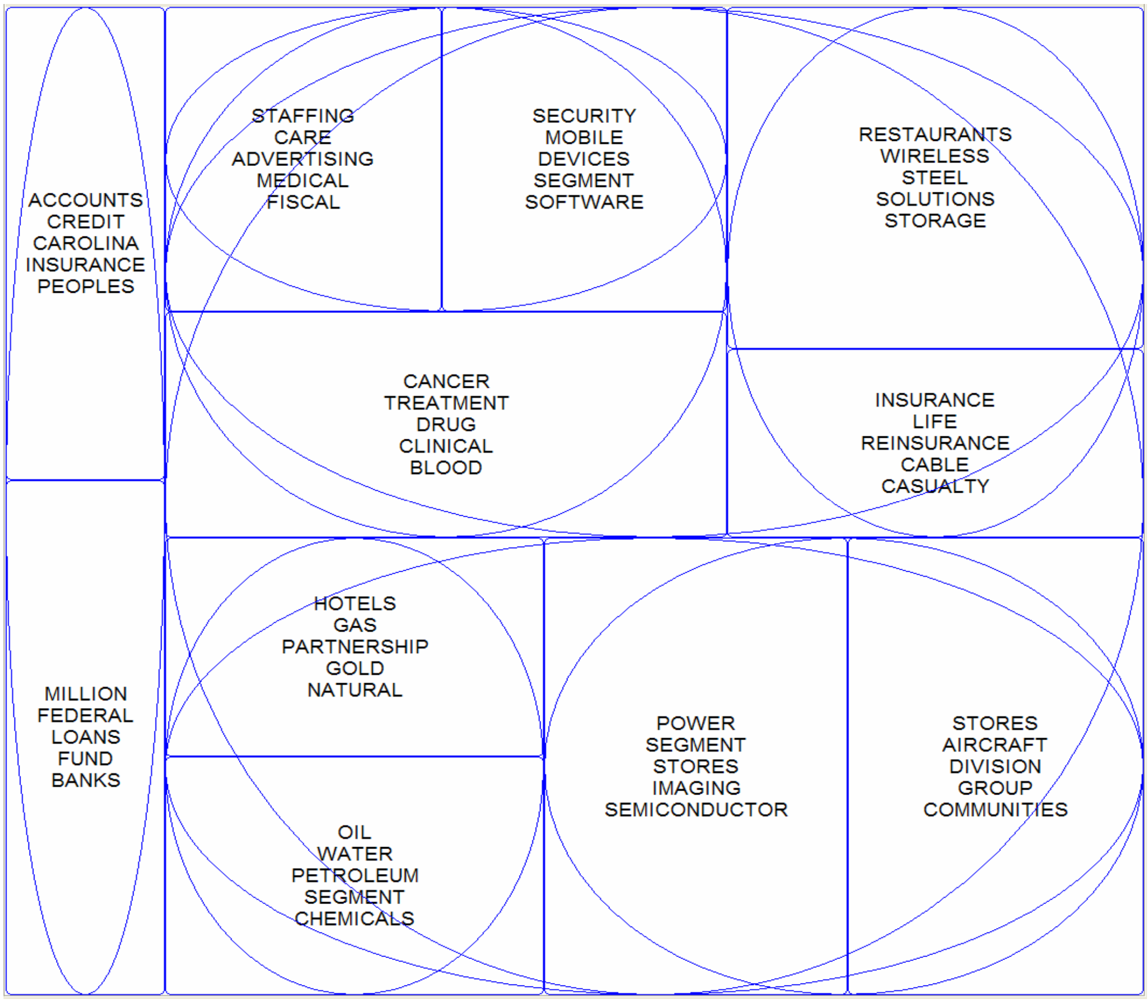

Figure 11 Tiling visualization of companies' competencies, where the companies are organized in several hierarchical levels.

The second system, gCLUTO (Rasmussen & Karypis, 2004), first performs stop-words removal and stemming in text pre-processing, followed by k-means clustering, using a predefined number of clusters of leaf-level nodes as the stopping criterion. In real-life scenarios, appropriate setting of the stopping criterion is a non-trivial task. In our experiment we have selected k equal to 12, the number of Yahoo! sectors, as one of our goals was to reconstruct the available Yahoo! business sector ontology. In gCLUTO's mountain visualization (shown in Figure 12), each peak represents an individual cluster: peak height is proportional to cluster's internal similarity, grayscale tone is proportional to cluster's internal deviation (darker tones indicate lower deviation), and peak volume is proportional to the number of elements in the cluster.

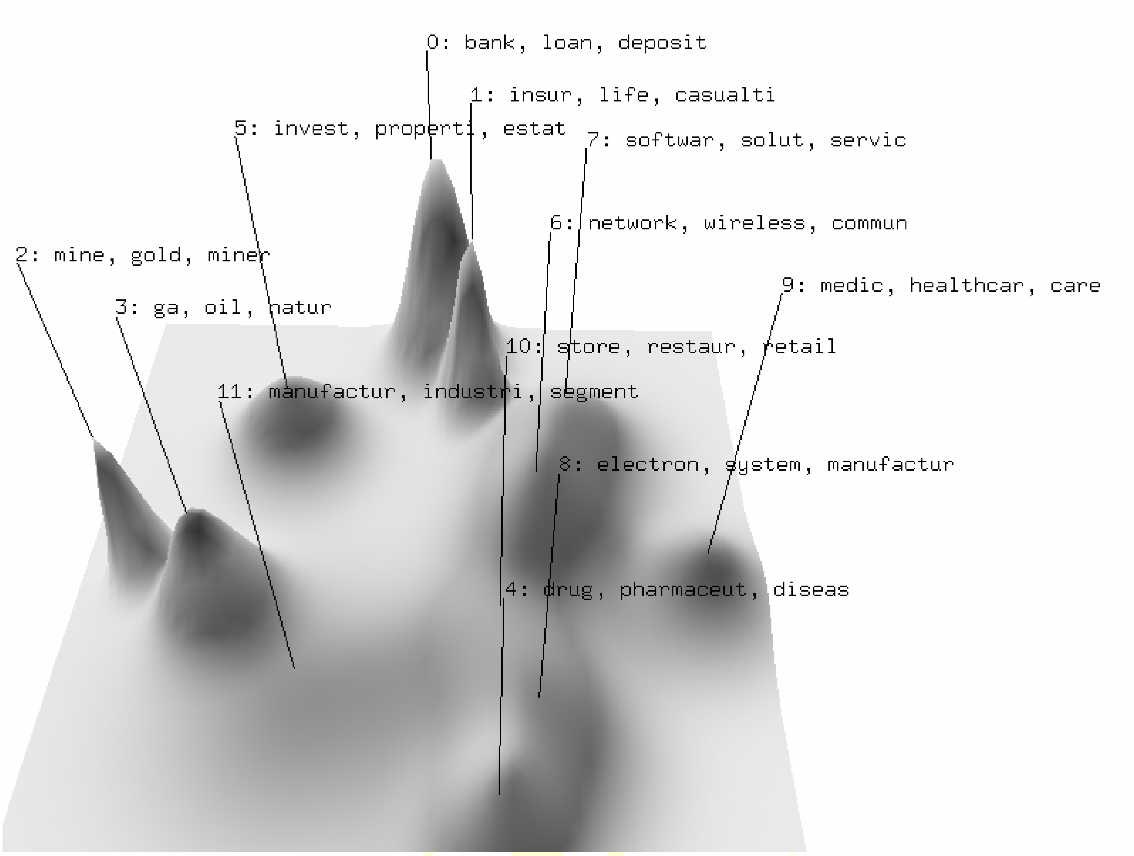

Figure 12 Mountain visualization of 12 top-level clusters with inter-cluster similarity (ISim) represented by the peak heights.

4.2.3 Evaluation

Without expert assistance we were unable to implement step 5 of the proposed semi-automated ontology construction methodology. Instead of intuitively naming the clusters by sector/industry names, we have - to the best of our capacity - manually aligned clusters to Yahoo! sectors, by comparing Yahoo! sector and industry names to the automatically assigned cluster keywords. We have evaluated the success of clustering on the scale 1 to 5, based on the number of keywords which, in our opinion, best describe the sector. The result of the evaluation is shown in Table 3.

Table 3 Clusters generated by the two clustering systems (each cluster is described by keywords and evaluated by a score) mapped to Yahoo! sectors and industries.

Yahoo! sectors and industries	**Text Garden clusters Keywords (Score)**	g**Cluto clusters Keywords (Score)**
Basic Materials Gold&Silver, Iron&Steel, …		• mine, gold, miner, exploring, property(4) • manufacturing, industry, segment, product, steel (1)
Capital Goods		
Conglomerates		
Consumer Cyclical		
Consumer Non-Cyclical		
Energy Coal, Oil & Gas, ...	• hotels, gas, partnership, gold, natural (2) • oil, water, petroleum, segment, chemicals (3)	• gas, oil, natural, energy, exploring (4)
Financial Insurance, S&Ls/Savings, …	• accounts, credit, Carolina, insurance, people (3) • million, federal, loans, fund, banks (5) • insurance, life, reinsurance, cable, casualty (4)	• bank, loan, deposit, mortgage, finance (5) • insurance, life, casualty, reinsurance, property (5) • invest, property, estate, real, trust (4)
Healthcare Facilities, Major Drugs, …	• cancer, treatment, drug, clinical, blood (5) • staffing, care, advertising, medical, fiscal (2)	• drug, pharmaceutical, disease, treatment, cancer (5) • medic, healthcare, care, health, hospital (5)
Services Advertising, Restaurants, …	• restaurants, wireless, steel, solutions, storage (1)	• store, restaurant, retail, brand, food (5)
Technology Hardware, Software, …	• security, mobile, devices, segment, software (4) • power, segment, stores, imaging, semiconductor (2)	• network, wireless, communication, internet, service (5) • software, solution, service, information, management (4) • electron, system, manufacturing, semiconductor, equipment (5)
Transportation Airline, Railroads, …	• stores, aircraft, division, group, communities (1)	
Average score	2.9	4.3

The application of Text Garden implementation of hierarchical k-means clustering resulted in a relatively weak correspondence between clusters and the Yahoo! sectors/industries (evaluated by the average score 2.9). On the other hand, the cluster keywords proposed by gCLUTO (the average score 4.3) were pertinent enough to define distinct clusters that can be relatively easily understood and interpreted. Therefore we have concentrated on the results of gCLUTO by further analyzing the distribution of companies over the Yahoo! sectors in each cluster. The companies were labeled with their respective sector, and the distribution of labels in each cluster was examined. The distribution is shown in Table 4.

Table 4 Results of gCLUTO - the distribution of 12 clusters among 12 sectors.

Id	ISim	Healthcare	Technology	Services	Basic Mat.	Financial	Cons. Cyc.	Capital Goods	Cons. Non-C.	Utilities	Transport	Energy	Conglom.
0	0,190	1	6	19	2	765	0	3	1	0	0	1	1
1	0,174	1	2	7	0	184	1	6	0	0	0	1	2
2	0,151	0	3	10	108	0	0	5	0	0	0	11	0
3	0,097	1	7	12	12	17	3	26	3	122	24	277	1
5	0,089	1	6	211	7	150	1	14	1	0	2	4	1
4	0,068	447	36	8	10	2	1	4	3	0	0	0	0
6	0,063	4	267	370	1	15	5	10	0	0	2	0	0
7	0,060	7	590	212	4	33	5	12	4	0	9	1	1
9	0,052	348	48	40	4	17	0	1	6	0	1	0	1
8	0,053	6	541	49	27	3	54	71	3	0	1	1	10
10	0,035	24	11	446	10	9	131	18	151	0	1	1	1
11	0,030	20	61	102	244	17	117	191	60	20	110	13	11

The analysis of Table 4 indicates that clusters with higher inter-cluster similarity (ISim) contain more companies with the same label. In some cases, companies are spread among two or more different sectors. For instance, the companies of cluster 6 (described by keywords *network, wireless, communications, internet, service*) are spread over sectors *Technology* and *Services*, which are closely related.

4.3 Competency structuring of Virtuelle Fabrik

The proposed methodology (described in Section 4.1) for structuring the competencies was applied also to a real-life VBE, the Virtuelle Fabrik[11] industrial cluster of mechanical engineering companies.

Company profiles of 50 partners of the Virtuelle Fabrik industrial cluster were made available for the experiment. Each company was described by its name, number of employees, products, services and their core competencies.

4.3.1 Description of the data

The procedure of translation of German company descriptions into English was as follows:

1) Each company was assigned a unique numeric identifier (see Table 5). The company identifier and company name were not used as input information for text clustering.
2) An initial document, written in German, was taken. It included descriptions of companies, their competencies, and products.
3) From the document the data on competencies and products of the companies was extracted, resulting in about 1300 words.
4) All stop-words were removed (i.e., und, mit, zu, verbs, adjectives).
5) The translation of the remaining set of German words into English was performed by a web-based translation engine.
6) Manual translation of non-translated words was performed, since the system was unable to

[11] **http://www.vfeb.ch/**

translate approximately 5% of the words.

7) Finally, the unification of the words was performed by word lemmatization (e.g., ‘system’ and ‘systems’ were transformed into ‘system’).

8) Text documents were transformed into word lists, representing simplified descriptions of the original text describing the companies. Some resulting company descriptions are shown in Table 6 below.

Table 5 The list of company names.

1	3M	11	Brueco	21	HBB	31	Mauell	41	Schuler
2	Admec	12	Brueggli	22	Heese	32	MeierundCo	42	Sika
3	AEundP	13	Buehler	23	IFTEST	33	MetallKunststoff	43	sitronic
4	Aebi	14	ccb	24	Innotool	34	MetallveredlungKopp	44	SMA
5	AESKrug	15	DCDEHNEL	25	IPG	35	LIndustrieelektronik	45	STRCNC
6	ALWO	16	ELMOTEC	26	IVM	36	OMB	46	SulzerInnotec
7	Aubry	17	EugenSeitz	27	Knobel	37	Pantec	47	SuterundBaehler
8	Baechli	18	Flube	28	KUBOTECH	38	QuadesignPartner	48	WICH
9	Beni	19	Futronic	29	MplusS	39	Rihs	49	Wiftech
10	Bichsel	20	GebrBraem	30	Marenco	40	Schaer	50	Wyser

Table 6 Simplified descriptions (nouns and noun phrases) of some companies in the Virtuelle Fabrik industry cluster.

Admec	Assembly Concepts Inspection Concepts Engineering Range Automation Development Construction Building Of Devices Plant Construction Test System Manufacturing Means Of Production Meaning Optimally Needs Plants Range Mounting Technique Handling Technology Inspection Technique Achievements Consultation Customer Project Engineering Production Plant Training Service Personnel Service Modular Transfer System
AE&P	Development Construction Range Mechanical Engineering Equipment Construction Supply Complete Plants Construction Employments Customer Project Management Mandates Total Conceptions Automation Development Hand Attachments Broad CAD Know-How Autocad Bravo Catia Me-10/30 Euklid
Aebi	Sales Engineer Services Order Manager Key Account Management Project Management Spectrum Companies Cast Parts Contact Contact Packing
AESKrug	…

4.3.2 Clustering

gCLUTO is a publicly available interactive clustering, visualization and analysis system (Rasmussen & Karypis, 2004). gCLUTO performs stopwords removal and stemming in text pre-processing, followed by k-means clustering, using a predefined number of clusters of leaf-level nodes as the stopping criterion.

The initial dendrogram induced from 50 company profiles, was obtained by the k-means hierarchical clustering method available as part of gCLUTO. Each node in the dendrogram is automatically accompanied by a list of most representative words from the document/cluster. We have decided to limit the company descriptions to only 4 most representative keywords. In k-means clustering, we selected k=2 in order to get simple hierarchical splits. We decided to limit the description of concepts at higher levels of the concept hierarchy also to 4 keywords only, in order to preserve the readability of the hierarchical structure of concepts.

The hierarchical structure of Figure 13 was produced from the initial dendrogram, by cutting the dendrogram at the level where differences between successive cluster levels are maximal. This resulted in six competency class clusters, described by automatically extracted keywords. Company names are added to the leaves of the hierarchy in order to simplify the interpretation of the obtained structure.

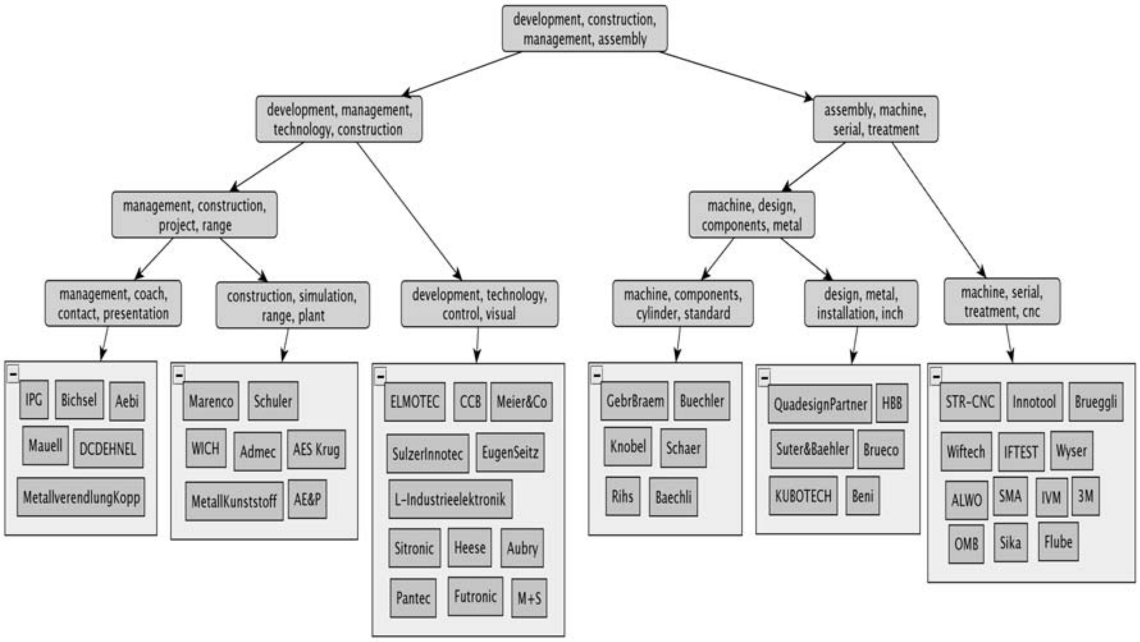

Figure 13 A hierarchy of 50 Virtuelle Fabrik company descriptions obtained by agglomerating bottom level clusters into six higher-level clusters.

The gCLUTO system (Rasmussen & Karypis, 2004) offers advanced cluster visualization tools, which we have used to visualize the results of VF clustering. The results in Figure 14 provide an overview of the clusters of competencies, their strength and their homogeneity.

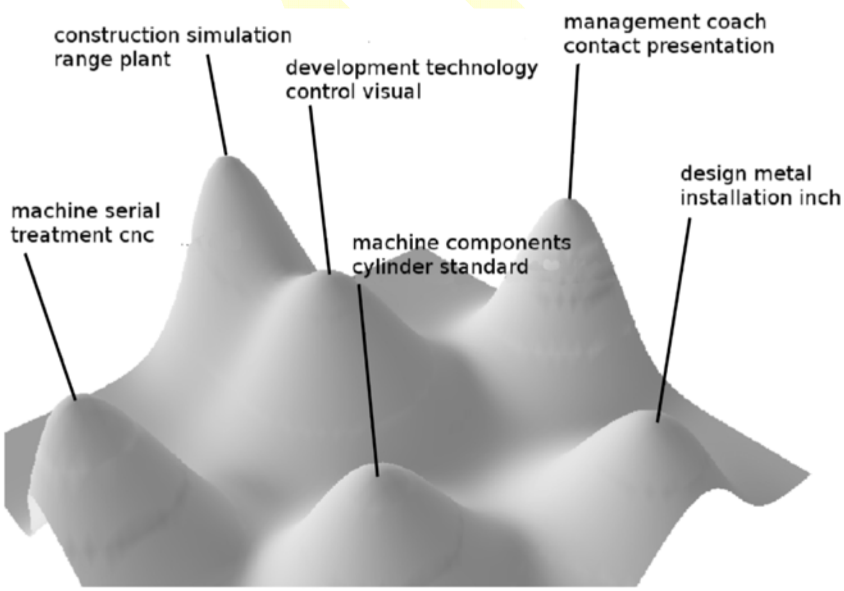

Figure 14 Mountain visualization of six clusters, described with the most descriptive words, for 50 Virtuelle Fabrik companies.

Peaks in Figure 14 represent individual clusters. The shape of each peak is a Gaussian curve, used as a rough estimate of the distribution of the data within each cluster. The height of each peak is proportional to the clusters internal similarity. The volume of a peak is proportional to the number of elements contained within the cluster. The resulting Gaussian curves are added together to form the terrain of the Mountain Visualization of gCLUTO.

4.3.3 Evaluation

In Section 3.3 we have presented two instantiations of the CNO ontology on two VOs. The first VO (Figure 9) was created for the task of the construction of a maintenance machine for a nuclear power plant and consisted of five VO partners: Alwo, SMA, Innotool, Rihs, Suter&Baehler, and Beni. Notice that these five partners all appear on the right-hand side node (super-cluster labeled by words "assembly, machine, serial, treatment") in Figure 13. This indicates that they have stronger similarity in terms of their descriptions as compared to the companies on the left-hand side node. However, they are also diverse, as they belong to three different sub-clusters of the right-hand side super-cluster, meaning that they have some complementary competencies. Similarly, the VO created for the task of a gearbox construction (Figure 10), consists of five VO partners: Beni, Knobel, Alwo, SMA and Innotool, which also fall in the same super-cluster of Figure 13. Again, these companies are sufficiently diverse, as they belong to the three different subclusters of the right-hand side super-cluster. The five partners were not sufficient for completing the task, therefore external partners (with the role "Public entity") were contracted in order to cover the missing competencies: Stebler and Brunner. In addition to the above strengths — the usefulness of the VBE ontology for formalizing the particular VOs — we can identify also some weaknesses:

First, the labeling of the cluster nodes (the 50 company hierarchy of Figure 13) is not satisfactory, since the labels were chosen automatically by the algorithm. One of the reasons for such labels is the fact that companies were described with relatively short descriptions. Instead, for text and web mining, rich and wordy descriptions would yield better results, since the techniques can extract more essential and discriminating terms from large and redundant texts. Note that, in this experiment we have performed only steps 1-4 of the proposed methodology for semi-automated ontology construction. In step 5, the human expert should actually revise the ontology and name the nodes appropriately by more meaningful cluster names.

Second, the competency structure is limited to a tree structure (a taxonomy) in contrast to a thesaurus. A thesaurus allows to link individual leaves to several nodes in the cluster which is often more desirable. Note that the above validation is qualitative only, as no quantitative measure of quality can be applied. We have, however, shown the appropriateness of the semi-automated ontology construction approach in the previous experiment (Section 4.2), in which we applied the described methodology to 7107 company descriptions from the Yahoo! business directory. In this previous experiment we have been able to quantitatively evaluate the approach by comparing the automatically constructed clusters with the original Yahoo!'s structure, created manually.

5 Conclusions

In this thesis, we introduced ontologies in the context of Collaborative Networked Organizations. After an overview of ontologies and different techniques for ontology building, we propose an ontology for Collaborative Network Organizations, with particular emphasis on Virtual organizations Breeding Environments (VBEs), which deal with long-term alliances of collaborating organizations. The knowledge encoded in the ontology was carefully chosen in order to make the ontology reusable in every kind of VBEs. The ontology has been validated through its instantiation on two case studies presented in Section 3.3 and by the feedback obtained from the representatives of two VBEs involved in the ECOLEAD project. The proposed CNO ontology is implemented in Protégé and is freely available for further refinements and improvements. This suggests to the reader to download the ontology and learn how to modify and adapt it for her/his specific needs. The intention is for the ontology to be refined and instantiated for individual types of CNOs and specially VBEs and VOs, depending on the sector and domain of their operation. Such refined and specific ontologies should be an integral part of each individual VBE information infrastructure, providing support for the whole VBE life cycle, and specifically for facilitating VO creation and management tasks.

In the second part of our contribution we provide a methodology for refining the ontology with network-specific information, namely the competencies of the members. Competencies play a major role in identifying business opportunities, acquiring projects, and forming Virtual Organizations which run and complete the projects. We present an approach to structuring the expertise of companies into a simple ontology, aimed at modeling competencies/expertise of companies from textual data. In the first experiment we have validated the methodology on a large-scale scenario of 7101 companies (The Yahoo! Business directory), where two different systems were used to implement the proposed methodology, and two different visualization tools based on hierarchical k-means clustering of documents were applied. To evaluate the results, we compared the results with the existing two-level Yahoo! ontology of companies. Despite the fact that the results are non-representative for a real-life situation in which pre-defined categories do not exist, the results of this experiment are interesting as they provide keywords representing company expertise as novel information over the human-defined Yahoo! sector categories. The results could be further improved by splitting the obtained clusters into more sub-clusters, thus achieving a complete hierarchy of companies' competencies. In addition the use of natural language processing methods could be used to provide additional information for word sense disambiguation, leading to improved clustering results and improved keyword extraction. In this experiment we were able to evaluate the approach quantitatively by comparing the automatically constructed clusters with the original Yahoo! structure, created manually. In the second experiment, we successfully extracted, using the same methodology, the competencies of companies of the Virtuelle Fabrik cluster. This validation was qualitative only, as no quantitative measure of quality could be applied. For a complete validation of the potential of the proposed approach for the Virtuelle Fabrik industry cluster, the analysis should be extended to a wider/full set of member data.

In the future we will extend the ontology to cover collaborating individuals in the form of Professional Virtual Communities (PVCs) and Virtual Teams (VTs). The generic ontology will be also incorporated into the ECOLEAD VBE management software which will support the complete VBE/VO life cycle: VBE creation and instantiation, search and identification of collaboration opportunities, as well as VO creation, management, operation and dissolution.

6 Acknowledgements

The author would like to thank his mentor Prof. Dr. Nada Lavrač, Dr. Igor Mozetič, Dr. Dunja Mladenič and Peter Ljubič for their contribution and their help in finalizing this thesis.

I am grateful to Stefan Bollhalter, the manager of Virtuelle Fabrik, for his interest and help in the application of the proposed methodology to the two real-life case studies. I am also grateful to Nathalie Galeano and David Romero from Instituto Tecnolgico y de Estudios Superiores de Monterrey (ITESM) and to Matthias Oswald from Virtuelle Fabrik for their contribution to the evaluation of the proposed VBE ontology. I also thank Jean Beeler from the Swiss Microtech enterprise network for his interest in this work.

This work was supported by the Slovenian Research Agency programme Knowledge Technologies (2004-2008) and the European 6th Framework Programme project Collaborative Networked Organizations Leadership Initiative (ECOLEAD, IST-1-506958-IP). I am grateful to the Slovenian Research Agency young researcher grant, which has founded the work on this thesis.

7 References

Agirre, E., Ansa, O., Hovy, E., & Martínez, D. (2000). *Enriching very large ontologies using the WWW*. In Proceedings of the First Workshop on Ontology Learning OL-2000, at the 14th European Conference on Artificial Intelligence.

Agrawal, R., Imielinski, T., & Swami, A. N. (1993). *Mining association rules between sets of items in large databases.* In Proceedings of the ACM SIGMOD International Conference on Management of Data, P. Buneman and S. Jajodia, Eds., Washington, D.C., 207–216.

Beckett, D. (2004). *RDF/XML Syntax Specification (Revised).* W3C recommendation published online. http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/

Bisson, G., Nedellec, C., Cañamero, D. (2000). *Designing Clustering Methods for Ontology Building - The Mo'K Workbench*. ECAI Workshop on Ontology Learning 2000.

Breuker, J., Valente, A., & Winkels, R. G. F. (1997). *Legal ontologies: A functional view*. In Proceedings of the 1st International Workshop on Legal Ontologies, P. Visser and R. G. F.Winkels, Eds. Melbourne, Australia: University of Melbourne, pp. 23–36.

Brickley, D. (2004). *RDF Vocabulary Description Language 1.0: RDF Schema*. W3C recommendation published online. http://www.w3.org/TR/rdf-schema/

Camarinha-Matos, L., & Afsarmanesh, H. (2003). Elements of a base VE infrastructure. *Journal of Computers in Industry*, *51*(2), 139-163.

Camarinha-Matos, L., & Afsarmanesh, H. (2005). Collaborative networks: A new scientific discipline. *Journal of Intelligent Manufacturing*, *16*(4), 439-452.

Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *Journal of the American Society of Information Science*, 41(6), 391–407.

Ester, M., Gross, M., & Kriegel, H.-P. *Focused web crawling: A generic framework for specifying the user interest and for adaptive crawling strategies*. Submitted to 27th Int. Conf. on Very Large Databases.

Faatz, A., & Steinmetz, R. (2002). *Ontology enrichment with texts from the WWW*. Semantic Web Mining 2nd Workshop at ECML/PKDD 2002, Helsinki, Finland.

Fellbaum, C. (1998). *WordNet, an electronic lexical database*. Cambridge, MA: MIT Press.

Fox, M. S. (1992). The TOVE Project: A common-sense model of the enterprise, industrial and engineering applications of artificial intelligence and expert systems. In Belli, F., & Radermacher, F. J. (Eds.), *Lecture Notes in Artificial Intelligence No. 604*, pp. 25-34.

Franke, U. (2000). *Knowledge Management in Virtual Organizations*. Idea group publishing. The knowledge-based view (KBV) of the virtual web, the virtual corporation, and the net-broker, 20–41.

Goldman, S., & Nagel, R. (1993). Management, technology and agility: The emergence of the new era in manufacturing. *International Journal of Technology Management*, 8(1/2), 18–38.

Goldman, S., Nagel, R., & Preiss, K. (1995). *Agile Competitors and Virtual Organizations: Strategies for Enriching the Customer*. New York: van Nostrand Reinhold.

Gomez-Perez, A., Fernandez-Lopez, M., & Corcho, O. (2004). *Ontological Engineering*. Springer-Verlag.

Grobelnik, M. & Mladenić, D. (2002). *Efficient visualization of large text corpora*. In Proceedings of the 7th TELRI seminar. Dubrovnik, Croatia.

Gruber, T. R. (1993). A translation approach to portable ontologies. *Knowledge Acquisition*, *5*(2), 199-220.

Hardwick, M., Spooner, D., Rando, T., & Morris, K. (1996). Sharing manufacturing information in virtual enterprises. *Communications of the ACM*, 39(2), 46–54.

Hearst, M. A. (1992). *Automatic acquisition of hyponyms from large text corpora.* In Proceedings of the 15th International Conference on Computational Linguistic (COLING'92), Nantes, France, 539–545.

van Heijst, G., Schreiber, A. T., & Wielinga, B. J. (1996). Using Explicit Ontologies in KBS Development. *International Journal of Human and Computer.*

Jenz, D. E. (2003). *BPMO tutorial, defining a private business process in a knowledge base*. Retrieved October 05, 2006, from http://www.bpiresearch.com/BPMO_Tutorial.pdf.

Khan, L., & Luo, F. (2002). *Ontology construction for information selection*. In Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'02), Washington, DC, 122–127.

Kietz, J.U., Maedche A., Volz R. (2000). *A Method For Semi-Automated Ontology Acquisition from a Corporate Internet*. In Proceedings of Aussenac-Gilles N., Biebow B., Szulman S. (eds) EKAW'00 Workshop on Ontologies and Texts. Juan-Les-Pins, France. Amsterdam, The Netherlands.

Kohonen, T. (1989). *Self-Organization and Associative Memory*, 3rd ed. Berlin, Germany: Springer-Verlag.

Ljubič, P., Lavrač, N., Plisson, J., Mladenić, D., Bollhalter, S., & Jermol, M. (2005). *Automated structuring of company competencies in virtual organizations*. In Proceedings of the Information Society multi-conference, 190-193.

Ljubič, P., Lavrač, N., Mladenić, D., Plisson, J., & Mozetič, I. (2006). Automated structuring of company profiles. *Metodološki zvezki*, 3(2), 369-380.

Maedche, A., & Staab, S. (2000). *Discovering Conceptual Relations from Text*. In proceedings of the 14th European Conference on Artificial Intelligence.

McGuinness, D.L., & van Harmelen, F. (2004). OWL Web Ontology LanguageOverview. W3C recommendation published online. http://www.w3.org/TR/owl-features/

Michalski, R. (1980). Knowledge acquisition through conceptual clustering: a theoretical framework and algorithm for partitioning data into conjunctive concepts. *International Journal of Policy Analysis and Information Systems No. 4*, 219–243.

Miller, N. E., Wong, P. C., Brewster, M., & Foote, H. (1998). *TOPIC ISLANDS - A wavelet-based text visualization system*. In IEEE Visualization '98, D. Ebert, H. Hagen, and H. Rushmeier, Eds., 189–196.

Missikoff, M., Navigli, R., & Velardi, P. (2002). *The Usable Ontology: An Environment for Building and Assesing a Domain Ontology*. Research Paper at International Semantic Web Conference (ISWC), Sardinia, Italy.

Mladenić, D., Lavrač, N., Bohanec, M., & Moyle, S. (Eds.). (2003). *Data Mining and Decision Support: Integration and Collaboration*. Kluwer.

Moshowitz, A. (1986). *Social dimensions of office automation*. Advances in Computers, 25, 335–404.

Noy, N.F., & McGuinness, D.L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05. March.

Plisson, J., Ljubič, P., Mozetič, I., & Lavrač, N. (2007a). Ontologies for Collaborative Networked Organizations. *Encyclopedia of Networked and Virtual Organizations*. In press.

Plisson, J., Ljubič, P., Mozetič, I., & Lavrač, N. (2007b). An ontology for Virtual organizations Breeding Environments. *IEEE Transactions on Systems, Man, and Cybernetics*. In press.

Plisson, J., Mladenić, D., Ljubič, P., Lavrač, N., & Grobelnik, M. (2005). *Using machine learning to structure the expertise of companies : analysis of the Yahoo! business data*. In Proceedings of the Information Society multi-conference, 186-189.

Porter, M. (1980). An algorithm for suffix stripping. Program, 14(3), 130–137, July.

Rasmussen, M., & Karypis, G. (2004). *gCLUTO — an interactive clustering, visualization, and analysis system*. University of Minnesota, Department of Computer Science and Engineering, Tech. Rep. TR 04-021.

Schlenoff, C., Gruninger, M., Ciocoiu, M., & Lee, J. (1999). The essence of the Process Specification Language. *Special issue on Modelling and Simulation of Manufacturing Systems in the Transactions of the Society for Computer Simulation International*.

Steinbach, M., Karypis, G., & Kumar, V. (2000). *A comparison of document clustering techniques*. In Proceedings of KDD Workshop on Text Mining, at the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, MA, 109–110.

Uschold, M., King, M,. Moralee, S., & Zorgios, Y. (1998). The enterprise ontology. *Knowledge Engineering Review, Special Issue on Putting Ontologies to Use,* 13(1), 31-89.

APPENDIX A: OVERVIEW OF EXISTING ONTOLOGIES

This appendix presents some of the most important and best known existing ontology. They are organized into the following categories:

- terminological ontologies: Wordnet, Verbnet, FrameNet, Sensus;
- domain ontologies: The Gene ontology, PSL;
- top-level ontologies: SUMO, Mikrokosmos, Sowa's ontology; and
- ontologies with common-sense knowledge: Cyc, ConceptNet.

A.1. Terminological ontologies

A.1.1. WordNet

Name	Wordnet
Goal	To find lexical information faster
Type of ontology	Terminological ontology
Encoding	Text
Degree of formalism	Informal
Characteristics	English nouns, verbs, and adjectives are organized into synonym sets, each representing one underlying lexical concept. Different relations link the synonym sets, such as antonymy, hyponymy/hypernymy, meronymy/holonymy, etc.
Reference	http://wordnet.princeton.edu/5papers.pdf

A.1.2. VerbNet

Name	VerbNet
Goal	To associate semantic components with lexical items
Type of ontology	Terminological ontology, verb lexicon
Encoding	XML
Degree of formalism	Informal
Characteristics	Implement semantic components as sets of features. Makes explicit the semantic components, argument structure, and sets of syntactic frames associated with individual lexical items.
Reference	http://www.cis.upenn.edu/~mpalmer/project_pages/VerbNet.htm

A.1.3. FrameNet

Name	FrameNet
Goal	Document the range of semantic and syntactic combinatory possibilities (valences) of each word in each of its senses.
Type of ontology	Terminological ontology, lexical database
Encoding	XML
Degree of formalism	Informal
Characteristics	An entry for each of the verbs includes a concise formula for all semantic and syntactic combinatorial possibilities, together with a collection of annotated corpus sentences in which each possibility is exemplified. FrameNet lexical database currently contains more than 8,900 lexical units, more than 6,100 of which are fully annotated, in more than 625 semantic frames, exemplified in more than 135,000 annotated sentences.
Reference	http://framenet.icsi.berkeley.edu/papers/acl98.ps

A.1.4. Sensus

Name	Sensus
Goal	Provides systems with a wide-ranging semantic thesaurus (ontology).
Type of ontology	Terminological ontology
Encoding	Loom, FrameKit, and Prolog
Degree of formalism	Formal
Characteristics	Sensus is a terminological taxonomy that contains 90000 concepts. It is the result of merging several existing ontologies or dictionaries. First, two top-level ontologies have been merged: the PENMAN Upper Model and ONTOS, forming the top-level of Sensus of approx. 400 terms. Then several branches of Wordnet have been added, together with LDOCE (semantic categories for nouns), forming the middle level of the ontology. Finally the Harper-Collins Spanish-English Bilingual Dictionary has been merged, to link Spanish words, in order to perform Spanish-English machine translation.
Reference	http://www.isi.edu/natural-language/resources/sensus.html

A.2. Domain ontologies

A.2.1. The Gene Ontology (GO)

Name	Gene ontology
Goal	The Gene Ontology project provides a controlled vocabulary to describe gene and gene product attributes in any organism.
Type of ontology	Domain ontology
Encoding	Text, fasta, xml, mysql.
Degree of formalism	Informal
Characteristics	GO terms are organized in structures called directed acyclic graphs (DAGs), which differ from hierarchies in that a child (more specialized) term can have many parent (less specialized) terms. At the time of writing, the ontology contained 18002 terms, 94.2% with definitions of 9446 biological_process, 1555 cellular_component, 7001 molecular_function. The ontology is updated every 30 mins or monthly releases are also available.
Reference	http://www.geneontology.org/GO.doc.shtml

A.2.2. Process Specification Language Ontology (PSL)

Name	PSL
Goal	The purpose of PSL is to axiomatise a set of intuitive semantic primitives that is adequate for describing the fundamental concepts of manufacturing processes
Type of ontology	Domain ontology
Encoding	KIF
Degree of formalism	Formal
Characteristics	The axioms of PSL are organized into PSL-Core and a set of extensions. PSL-Core is the set of axioms written in KIF (the Knowledge Interchange Format) and using only the nonlogical lexicon of PSL-Core. The extensions form a lattice of extensions to PSL-Core. To supplement the concepts of PSL-Core, the ontology includes a set of extensions that introduce new terminology. An PSL extension provides the logical expressiveness to express information involving concepts that are not explicitly specified in PSL-Core. The basic ontological commitments of PSL-Core are based on the following intuitions: • There are four kinds of entities required for reasoning about processes - activities, activity occurrences, timepoints, and objects. • Activities may have multiple occurrences, or there may exist activities that do not occur at all. • Timepoints are linearly ordered, forwards into the future, and backwards into the past. • Activity occurrences and objects are associated with unique timepoints that mark the beginning and the end of the occurrence of an object.
Reference	http://www.mel.nist.gov/psl/ontology.html

A.3. Upper ontologies

A.3.1. Suggested Upper Merged Ontology (SUMO)

Name	Suggested Upper Merged Ontology
Goal	To be used for research and applications in search, linguistics and reasoning.
Type of ontology	Top-level ontology, domain ontologies
Encoding	SUO-KIF, OWL
Degree of formalism	Formal
Characteristics	The Suggested Upper Merged Ontology (SUMO) and its domain ontologies form the largest formal public ontology in existence today. SUMO is the only formal ontology that has been mapped to the WordNet lexicon.
Reference	http://www.ontologyportal.org/

A.3.2. Mikrokosmos

Name	Mikrokosmos
Goal	Lexical representation of word meanings as well as text meaning representation is grounded in a broad-coverage ontology of the world. Such a language-neutral ontology has been built for the purpose of machine translation.
Type of ontology	Top-level ontology
Encoding	Frame-based, XML (Spencer notation), BNF grammar
Degree of formalism	Informal
Characteristics	All the concepts in the ontology are under three top-level concepts: event, property and object. They are encoded with frames containing slots. The value of slots are constrained with facets and fillers that define the range of values allowed in the slot. The ontology contains approximately 4500 concepts.
Reference	http://crl.nmsu.edu/Research/Projects/mikro

A.3.3. Sowa's top-level ontology

<table>
<tr><td>Name</td><td>Sowa's top-level ontology</td></tr>
<tr><td>Goal</td><td>Represent knowledge</td></tr>
<tr><td>Type of ontology</td><td>Top-level ontology</td></tr>
<tr><td>Encoding</td><td>Lattice + FCA (Formal Concept Analysis)</td></tr>
<tr><td>Degree of formalism</td><td>Formal</td></tr>
<tr><td>Characteristics</td><td>The ontology contains twelve top-level categories, derived from several primitives ("A category of an ontology that cannot be defined in terms of other categories in the same ontology"). The following table presents the twelve categories along with their primitives:
<table>
<tr><td rowspan="2">Primitives</td><td colspan="2">Physical</td><td colspan="2">Abstract</td></tr>
<tr><td>Continuant</td><td>Occurrent</td><td>Continuant</td><td>Occurrent</td></tr>
<tr><td>Independent</td><td>Object</td><td>Process</td><td>Schema</td><td>Script</td></tr>
<tr><td>Relative</td><td>Juncture</td><td>Participation</td><td>Description</td><td>History</td></tr>
<tr><td>Mediating</td><td>Structure</td><td>Situation</td><td>Reason</td><td>Purpose</td></tr>
</table></td></tr>
<tr><td>Reference</td><td>http://www.jfsowa.com/ontology/toplevel.htm</td></tr>
</table>

A.4. *Ontologies with common-sense knowledge*

A.4.1. Cyc

Name	Cyc
Goal	Provide the knowledge like in an encyclopedia, to be used for reasoning.
Type of ontology	Top-level, core ontology
Encoding	CycL
Degree of formalism	Formal
Characteristics	The Cyc Knowledge Base (Cyc KB) is divided into many (currently thousands of) microtheories, each of which is essentially a bundle of assertions that share a common set of assumptions; some microtheories are focused on a particular domain of knowledge, a particular level of detail, a particular interval in time, etc. At present, the Cyc KB contains nearly two hundred thousand terms and several dozen hand-entered assertions about/involving each term. New assertions are continually added to the KB by human knowledge enterers.
Reference	http://www.cyc.com/cyc/technology/whatiscyc_dir/whatsincyc

A.4.2. ConceptNet

Name	ConceptNet
Goal	Provide a commonsense knowledgebase and natural-language-processing toolkit, which supports many practical textual-reasoning tasks.
Type of ontology	Top-level, core ontology
Encoding	text
Degree of formalism	Informal
Characteristics	The ConceptNet knowledgebase is a semantic network presently available in two versions: concise (200,000 assertions) and full (1.6 million assertions). Commonsense knowledge in ConceptNet encompasses the spatial, physical, social, temporal, and psychological aspects of everyday life. Whereas similar large-scale semantic knowledgebases like Cyc and WordNet are carefully handcrafted, ConceptNet is generated automatically from the 700,000 sentences of the Open Mind Common Sense Project – a World Wide Web based collaboration with over 14,000 authors.
Reference	http://web.media.mit.edu/~hugo/conceptnet/

Index of Figures

Figure 1 The CNO ontology, with a detailed elaboration of VBE and VO (without any detail on PVC and VT, which are out of the scope of this thesis). 15
Figure 2 Top level concepts of the ontology. Specializations are shown using dashed lines, while other relations (along with their names) are represented with solid lines. Professional Virtual Communities (PVCs) and Virtual Teams (VTs) as forms of CNO are out of scope of this thesis. 16
Figure 3 The organization part of the ontology, instances are linked to concepts with io (instance of) relations. 17
Figure 4 Example of instantiation of the organization's ontology. 18
Figure 5 Roles and their structuring within a VBE. 19
Figure 6 An example of VBE, VO, its participants, and their roles. 19
Figure 7 Support institution roles and their instantiations. 21
Figure 8 Example of a VBE with five participants. 21
Figure 9 A VO created (within VF) for the task of construction of a maintenance machine for a nuclear power plant. 24
Figure 10 A VO created (within VF) for the task of a gearbox construction. 25
Figure 11 Tiling visualization of companies' competencies, where the companies are organized in several hierarchical levels. 29
Figure 12 Mountain visualization of 12 top-level clusters with inter-cluster similarity (ISim) represented by the peak heights. 30
Figure 13 A hierarchy of 50 Virtuelle Fabrik company descriptions obtained by agglomerating bottom level clusters into six higher-level clusters. 34
Figure 14 Mountain visualization of six clusters, described with the most descriptive words, for 50 Virtuelle Fabrik companies. 34

Index of Tables

Table 1 The AIAI enterprise ontology: list of terms composing the ontology. 8
Table 2 The Yahoo! Business Ontology, the number of industries and companies (per sector). 10
Table 3 Clusters generated by the two clustering systems (each cluster is described by keywords and evaluated by a score) mapped to Yahoo! sectors and industries. 31
Table 4 Results of gCLUTO - the distribution of 12 clusters among 12 sectors. 32
Table 5 The list of company names. 33
Table 6 Simplified descriptions (nouns and noun phrases) of some companies in the Virtuelle Fabrik industry cluster. 33

Razširjeni povzetek magistrskega dela

ONTOLOGIJE MREŽNIH ORGANIZACIJ

Podjetja in posamezniki se združujejo v mrežne organizacije z namenom doseganja skupnih ciljev, običajno v obliki realizacije poslovnih priložnosti. Področje mrežnih organizacij (Networked organizations) pokriva razne tipe organizacijskih struktur. Znanje, ki je shranjeno v takšnih mrežah, se deli na dva nivoja. Najprej je tu splošno znanje o organizacijski strukturni mreže, ki se ga da uporabiti v vsaki taki mrežni organizacij. Drugi nivo pa predstavlja specifično znanje domene, ki jo mreža pokriva in uporablja (npr. kompetence podjetij). V magisterskem delu se ukvarjamo z uporabo obeh nivojev znanja in njegovo predstavitvijo v obliki ontologij.

V delu smo razvili ontologijo mrežnih organizacij, ki vpelje terminologijo tega področja ter identificirala akterje in relacije med akterji mrežnih organizacij. Ontologija je bila razvita na širše področje kolaborativnih mrežnih organizacij (Collaborative Networked Organizations, CNO) a se osredotoča na valilnice mrežnih organizacij (Virtual organizations Breeding Environments, VBE). Ta del naloge predstavlja prispevek k formalizaciji dosedaj neformaliziranih pojmov in konceptov s tega področja ter k formalizaciji relacij med koncepti na način, ki zagotavlja konsistentnost razvite ontologije. Ontologija je implementirana v sistemu Protégé in je javno dostopna preko svetovnega spleta. Spletna stran vključuje tudi pojmovni slovar, vpeljan v uvodnem delu magistrske naloge.

Za potrebe drugega nivoja znanja je bila razvita metodologija za polavtomatsko gradnjo ontologij iz tekstovnih dokumentov, ki opisujejo kompetence podjetij, z namenom avtomatskega odkrivanja znanja o kompetencah mrežnih organizacij iz tekstovnih opisov kompetenc podjetij, ki sodelujejo v mrežni organizaciji. Medtem ko je znanje o mrežnih organizacijah, zakodirano v obliki CNO ontologije, statično in torej velja za vse mrežne organizacije, je ekstrahirano znanje o kompetencah specifično za posamezno mrežno organizacijo. Predlagana metodologija za ekstrakcijo tega znanja iz tekstovnih dokumentov je splošna in jo lahko zato uporabimo za katerokoli mrežno organizacijo, ki jo želimo modelirati, seveda pri pogoju, da imamo na voljo tekstovne opise kompetenc sodelujočih podjetij.

Razvita CNO ontologija in rezultat uporabljene metodologija za strukturiranje kompetenc podjetij sta bili pregledani in sprejeti s strani področnih strokovnjakov v okviru projekta ECOLEAD (2004-2007) integriranega projekta EU v okviru 6. okvirnega programa, ki se ukvarja z mrežnimi organizacijami, in ki je predstavljal okvir in motivacijo za pričujoče magistrsko delo.

Biography

Joël Plisson is a young researcher at the Department of Knowledge Technologies of the Jožef Stefan Institute in Ljubljana and a MSc student at the International Postgraduate School Jožef Stefan in Ljubljana, Slovenia. He received a BSc in computer science from the EPITECH School in Le Kremlin Bicêtre, France His research interests are in ontologies, semantic web, natural language processing, text mining and machine learning.

Photo

Short abstract

Companies and individuals connect into networks to share their resources with the purpose of achieving a common goal, defined by a business opportunity. The field of Collaborative Networked Organizations (CNO) covers various types of organizational structures. The knowledge that is stored in such networks can be separated into two different levels. First, there is a common knowledge about the organizational structure itself, which can be used and reused in any of such networks. The second level represents the domain specific knowledge that such networks cover and use to function (e.g. companies' competencies). In this thesis we address both levels by using ontologies. First, we propose an ontology representing the common vocabulary and identifying the actors and relationships in a specific type of network, namely a Virtual organization Breeding Environment (VBE). In this way, the thesis contributes to the formalization of the informal notions of VBEs and Virtual Organizations (VOs) in a formal ontology language. Second, we propose a methodology for semi automated ontology construction for the needs of VBEs, enabling the extraction of network specific knowledge related to competencies.