Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

Ontological Conceptualization Based on the Simple Knowledge Organization System (SKOS)

Farhad Ameri
Department of Engineering Technology
Texas State University
Tel: 512-245-1984 Email: ameri@txstate.edu

Boonserm Kulvatunyou
Systems Integration Division
National Institute of Standards and Technology
Tel: 301-975-2720 Email: serm@nist.gov

Nenad Ivezic
Systems Integration Division
National Institute of Standards and Technology
Tel: 301-975-3536 Email: nenad.ivezic@nist.gov

Khosrow Kaikhah
Department of Computer Science
Texas State University
Tel: 512-245-3666 Email: khosrow@txstate.edu

Abstract:
Ontological conceptualization refers to the process of creating an abstract view of the domain of interest through a set of interconnected concepts. In this paper, a thesaurus-based methodology is proposed for systematic ontological conceptualization in the manufacturing domain. The methodology has three main phases, namely, thesaurus development, thesaurus evaluation, and thesaurus conversion and it uses Simple Knowledge Organization System (SKOS) as the thesaurus representation formalism. The concept-based nature of a SKOS thesaurus makes it suitable for identifying important concepts in the domain. To that end, novel thesaurus evaluation and thesaurus conversion metrics that exploit this characteristic are presented. The ontology conceptualization methodology is demonstrated through the development of a manufacturing thesaurus, referred to as ManuTerms. The concepts in ManuTerms can be converted into ontological classes. The whole conceptualization process is the stepping stone to developing more axiomatic ontologies. Although the proposed methodology is developed in the context of manufacturing ontology development, the underlying methods, tools, and metrics can be applied to development of any domain ontology. The developed thesaurus can serve as a standalone lightweight ontology and its concepts can be reused by other ontologies or thesauri.

Keywords: ontology development, ontological conceptualization, SKOS, thesaurus,

1 Introduction

Ontologies are the backbone of knowledge-based systems as they facilitate sharing and reuse of knowledge through providing explicit representations of shared conceptualizations. In the engineering domain, development of ontologies has emerged as one of the fast growing research topics and several ontologies have been developed over the past decade [1-4] with the objective of enhancing semantics for interoperability [1] throughout the product value chains and promoting structured knowledge capture, synthesis, and dissemination. The focus of the existing research in this area has been primarily on the applications of ontologies in various knowledge-intensive activities such as creating manufacturing process plans, configuration of distributed supply chains, or searching product design repositories. However, the ontology development process itself has not been investigated adequately particularly in the engineering domain and the existing approaches are often unstructured and informal [5, 6]. Ontologies are engineering artifacts and their development should follow a structured process with appropriate metrics to determine the sufficiency and correctness of the ontology.

The need for a structured approach to ontology development becomes more pronounced when developing complex ontologies that require involvement of multiple Domain Experts (DE) and Knowledge Engineers (KE). A methodical approach to ontology development is beneficial as it makes the decision-making process explicit, facilitates collaboration and communication among knowledge engineers and domain experts, and enforces systematic documentation and validation. Several methods have been proposed for the development of ontologies to guide knowledge engineers through the general activities of the development life cycle including requirements planning, knowledge acquisition, conceptualization, formalization, evaluation, and documentation [5, 7-10]. However, the existing methods mainly address the development phases from a generic and high-level perspective without providing sufficient details with respect to the mechanics of ontology design and the analysis process. In particular, *ontological conceptualization* (conceptualization for short) still suffers from a lack of structure and it is often conducted in an ad hoc manner without following any systematic approach.

Ontological conceptualization is the process of creating an abstract view of the domain of interest represented through a set of interconnected concepts [7]. Once the concepts are identified and defined, they are implemented in an ontological language, linked to one another through relationships, and constrained logically to create a formal ontology. The importance of the conceptualization phase in

[1] The capability of two or more networks, systems, devices, applications, or components to interwork, and to exchange and readily use information, securely, effectively, and with little or no inconvenience to the user (Definition form Smart Grid 2.0)

Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

ontology development is comparable to that of the concept development phase in product design. A poor conceptual design will result in a low quality product that doesn't meet customer needs; likewise, a poor ontological conceptualization will lead to an eventual ontology that does not meet the requirements with respect to expressivity and completeness regardless of the level of formality incorporated in coding the ontology. By systematically exploring domain knowledge sources and identifying, organizing, and representing engineering concepts in a neutral and platform-independent format, the process of ontological conceptualization can be streamlined. Decades of information modeling and ontology research have proven the merits of applying systems engineering principles and rigor to the development of information engineering artifacts and tools.

The objective of this paper is to propose a methodology for systematic and incremental ontological conceptualization in the manufacturing domain. The proposed methodology is intended to support the social process of ontology development in a multi-developer and multi-user environment. A thesaurus-based approach is adopted in this research that incorporates the Simple Knowledge Organization System (SKOS) standard [11] as the thesaurus representation formalism. The proposed conceptualization methodology has three main phases, namely, thesaurus development, thesaurus evaluation, and concept conversion. It is not the intention of this research to develop a comprehensive thesaurus composed of an exhaustive collection of terms in the manufacturing domain, as that would require active involvement of multiple parties from academia, government, and industry. The thesaurus, developed with the proposed methodology provides the stepping stone for a more comprehensive ontology. In other words, the thesaurus can be evolved into a more axiomatic ontology, such as an OWL ontology, by incremental enrichment of the formal semantics represented through relationships, logical constraints, and rules.

The developed thesaurus in this work is referred to as *ManuTerms* and it is specifically geared toward organizing and classifying the concepts used for describing the technological capabilities of manufacturing suppliers in the contract manufacturing industry. At the time of preparing this document, ManuTerms has more than 2100 concepts that are organized in eighteen concept schemes. Approximately, 2800 alternative terms have been identified for those concepts. ManuTerms is developed as the conceptualization medium for MSDL (Manufacturing Service Description Language) [12], an axiomatic OWL-based ontology developed for formal description of manufacturing services and their capabilities in the Semantic Web. MSDL has grown at a steadier rate and in a more structured fashion since adoption of the proposed methodology for conceptualization. Although ManuTerms is intended to serve as a supporting controlled vocabulary to be used "behind the scene" while extending MSDL, it is envisioned that ManuTerms, as a SKOS-compliant thesaurus, can be used as a standalone lightweight ontology due to its simple semantics and also its open, standard format. The focus of this paper is on conceptualization while other ontology development activities, such as implementation and validation, will not be addressed in this paper.

The remainder of this paper is organized as follows. The existing works in ontology development methodologies are discussed in Section 2. Section 0 provides background information about formal thesauri. Section **Error! Reference source not found.** presents the rationale for selecting Simple Knowledge Organization System as the thesaurus representation language. In Section 5, the proposed

Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

ontological conceptualization methodology is introduced. Finally, Section 6 concludes the outcome of this research.

2 Related Works

The early works in methodical approaches to ontology development date back to mid-1990's when ontologies emerged as essential components of knowledge management systems. Pinto and Martines [5] provided an overview of the existing ontology development methodologies and described the research trends and challenges. They concluded that ontology development is still more of a craft than engineering and new methodologies with more formal and structured approach to knowledge acquisition and conceptualization need to be developed. Also, they emphasized the fact that any ontology development methodology should support a social development process in which distributed knowledge engineers and domain experts can develop the ontology in collaborative fashion. This feature has not been adequately addressed by the existing methodologies and tools. Below an overview of prominent ontology development methodologies is provided. This review focuses on the methods used for collecting domain concepts early in the ontology development process.

Gruninger and Fox [8] introduced a three-step methodology in the context of TOVE (Toronto Virtual Enterprise) ontology development project. The first step in their methodology deals with defining the ontology requirements through identifying the competency questions (CQ) that the ontology should be able to answer. The second step includes conceptualization, formalization, and implementation and the third step is about evaluating the competency and completeness of the ontology. TOVE uses First-order Logic (FOL) as the knowledge representation formalism. There is no specific recommendation or structured approach to knowledge acquisition and conceptualization in TOVE.

Gruber [7] proposed a method for development of portable ontologies that defines terms at the knowledge level, independent of any specific representation language. Such a representation-neutral knowledge model can provide the base concept model for formal ontologies, however, since the expressivity of the ontology modeling language is not known upfront during knowledge capture, the initial knowledge model may become over-constrained or under-constrained depending on the selected modeling language.

Ushold and King [13] proposed a generic framework for ontology development composed of four activities, namely, identifying the purpose, building the ontology through knowledge capturing and coding, evaluation and documentation. The proposed framework, known as ENTERPRISE, uses brainstorming as the main technique for collecting key concepts within the domain of interest. One important aspect of this methodology is the adoption of the middle-out approach to produce the conceptual model. The middle-out approach is enforced by focusing on the concepts that have more connections. ENTERPRISE does not mandate a priori commitment to any particular ontology representation language for formalizing the ontology.

METHONTOLOGY is another structured method introduced by Fernandes et al. [14] based on the experience acquired in the chemicals' domain. The major steps in METHONTOLOGY are specification, knowledge acquisition, conceptualization, integration, evaluation, and documentation. Due to its systematic and evolutionary nature, METHONTOLOGY has been widely adopted in

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

software engineering practice. METHONTOLOGY promotes ontology reuse and incremental development. For knowledge acquisition, METHONTOLOGY provides guidelines for structured and unstructured interviews with the domain experts. The conceptualization step in METHONTOLOGY is based on creating a Glossary of Terms (GT) and organizing the terms into concept trees. However, METHONTOLOGY does not provide any guideline as for how to select and organize the relevant terms for a given application and how to convert them into ontology classes.

De Nicola et al. [10] introduced a methodology called UPON for building large-scale domain ontologies based on Unified Process (UP) that has been widely used in software engineering. The proposed methodology is composed of five major steps, namely, requirements, analysis, design, implementation, and test. In the first three steps, the terminology of the domain is collected from domain expert interviews and enhanced semantically to create a domain lexicon and eventually a reference glossary. This methodology uses UML for preparing the blueprint of the ontology. UPON is use-case driven and is more appropriate for creation of application-dependent ontologies (also called task ontologies). Similar to METHONTOLOGY, UPON uses glossary development as a prelude to ontology development. However, no structured approach is proposed for creating the reference glossary in either methodology. In the engineering domain, there are few works that address methodical development of engineering ontologies.

Ahmed et al. [9] proposed a methodology for creating an engineering design ontology. The resulting ontology, called EDIT, is mainly a taxonomy of engineering design concepts with *design*, *process*, *function*, *product*, and *issue* as the root concepts. Interviewing domain experts is the main technique used for the knowledge acquisition phase, which includes conceptualization. A validation method is provided for each stage of the proposed methodology. The methodology also recommends that a thesaurus be developed at the later stages of the proposed methodology to provide lexical support for the developed ontology.

Li *et al.* [15] proposed a methodology for developing engineering ontologies to be used for indexing unstructured engineering documents and facilitating information retrieval in engineering design. In their methodology, concepts are collected form engineering knowledge resources such as engineering handbooks, textbooks, and online catalogs but concept acquisition does not follow a formal procedure. The other notable ontology development methodologies include On-To-Knowledge [16], DILIGENT [17], and NeOn [18]. NeOn methodology has a particular emphasis on reuse of ontologies and suggests a variety of pathways for developing ontologies under different scenarios. One of the scenarios in NeOn deals with reusing the existing non-ontological resources such as classification schemes, thesauri, and lexicons.

This study revealed that the existing ontology development methodologies do not provide adequate guidelines for the conceptualization stage. The use of a thesaurus has been proposed in some methodologies such as NeOn but no guidelines have been provided for creating the thesauri that can serve as the baselines for ontologies. Furthermore, most of the proposed methodologies are related to the software engineering domain. This paper supplements the existing ontology development methodologies through proposing a systematic and evolutionally method for the early stages of the development process. Although the proposed approach is illustrated using the examples from the manufacturing domain, it can be generalized for other domains.

Corresponding Author: Farhad Ameri

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

3 Background

3.1 Formal Thesauri

From a linguistic perspective, a thesaurus is a collection of terms connected through lexical relationships such as synonym, antonym, and meronym. *Terms* are the basic semantic units used for conveying concepts in a thesaurus. A *concept* is "a unit of thought, formed by mentally combining some or all of the characteristics of a concrete or abstract, real or imaginary object" [19]. Concepts exist in the mind as abstract entities independent of terms used to represent them. The International Organization for Standardization (ISO) defines thesaurus as " the vocabulary of a controlled indexing language, formally organized with the aim of stating explicitly the relationships between the concepts" [20]. The American National Standards Institute (ANSI) [19] defines thesaurus as "a controlled vocabulary arranged in a known order and structure so that equivalence, homographic, hierarchical, and associative relationships among terms are displayed clearly and identified by standardized relationship indicators that are employed reciprocally." A thesaurus can be regarded as a *lightweight ontology* that connects various terms and concepts through a selected set of lexical relations. A thesaurus makes as few claims as possible about the world being modeled. This allows ontology developers to adopt the concepts within a shared thesaurus and specialize and instantiate the concepts as needed to create more sophisticated knowledge constructs designed for answering arbitrary queries in a certain domain. WordNet [21] is an example of a linguistic thesaurus developed for English terms.

The practices of integrating thesauri with information retrieval systems date back at least to 1960s [22, 23]; and they gradually evolved from mere lexical resources towards powerful instruments for conceptual representation and knowledge organization and retrieval [24]. By indexing data and documents with a thesaurus, information retrieval systems are able to achieve improved performance. There exist several formal thesauri in different areas of applications such as National Agricultural Library Agricultural Thesaurus [25], Medical Subject Heading [26], National Cancer Institute (NCI) thesaurus [27], and GEMET (General Multilingual Environmental Thesaurus) [28] developed to support automated information retrieval in different application domains. Most of the larger thesauri, such as the NCI thesaurus, are created over several years and extended by a group of developers with expertise in different subject matters. In the engineering domain, there are few formal thesauri that are specifically designed for information retrieval and knowledge organization. The first thesaurus in engineering is the Thesaurus of Engineering and Scientific Terms (TEST) developed by the U.S. government in the 60s with the objective of indexing technical documents. However, TEST never gained the expected widespread acceptance in the engineering community particularly in the private sector. One possible reason for lack of adaptation of controlled vocabulary in engineering is the dominance of the symbolic and mathematical representations over the linguistic representation. Also, engineering knowledge repositories are typically isolated entities developed on the basis of a company-specific and proprietary vocabulary as opposed to being parts of a community-wide knowledge base. However, as product development practices are becoming increasingly knowledge-intensive, collaborative, and distributed, the need for standard semantic models for exchanging

Corresponding Author: Farhad Ameri

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

product information and organizing engineering knowledge is becoming more evident. One of the emerging applications of thesauri in the engineering domain is to support information retrieval in engineering design through indexing the informal engineering documents such as design logbooks [29]. Development of ManuTerms is a step toward systemic knowledge capture and organization in the manufacturing domain. The thesaurus developed in this work supports the conventional applications of thesauri such as document indexing and concept organization. One novel application introduced in this paper is to use a thesaurus as a support for ontological conceptualization.

3.2 Simple Knowledge Organization System (SKOS)

SKOS (Simple Knowledge Organization System) [11, 30] is used for representing the content and structure of the thesaurus in this work. SKOS is a W3C recommendation for formal thesaurus representation and is built upon the Resource Description Framework (RDF) and RDF Schema (RDFS) and enables publication of controlled vocabularies on the Semantic Web as an RDF graph. SKOS thesauri are *concept-based*, as opposed to *term-based*. In a term-based thesaurus, terms are directly connected together by the means of semantic relationships. However, in a concept-based thesaurus, semantic connection is at a concept level. In SKOS formalism, terms are the lexical labels for the concepts[2]. An advantage of a concept-based thesaurus is that it allows conceptualization to be performed at the semantic level. A SKOS thesaurus, like any other concept-based thesaurus, has a three-level structure: (a) conceptual level, where concepts are identified and their hierarchical and associative interrelationships established; (b) terminological correspondence level, where terms or labels (preferred or alternative) are associated to their respective concepts; and (c) lexical level where lexical relationships are defined to interconnect the terms. The conceptual nature of SKOS is particularly useful in ontology development as it urges the developers to draw a distinction between terms and concepts and build a sound conceptual understanding of the domain of discourse. SKOS Core Vocabulary is a set of RDF properties and RDFS classes. `skos:Concept` is one of the core classes in SKOS used to define an atomic conceptual resource. Each concept in SKOS has exactly one *preferred label* (`skos:prefLabel`) and can have multiple *alternative labels* (`skos:altLabel`). These labels are used to associate terms with concepts. For example, the metal casting process in which the mold is made of ceramic represents a concept that is typically referred to as `Ceramic Mold Casting` in the metal casting community but `Ceramic Molding` is also used interchangeably to point to the same concept. However, since `Ceramic Mold Casting` is used more frequently, it is used as preferred label while the `Ceramic Molding` is regarded as the alternative label for this concept. Nevertheless, the choice of the preferred label can be arbitrary and it does not impact the semantics of the concept. The *broader*, or more generic, concept of ceramic mold casting is `Expendable Mold Casting` and the concepts such as `Shaw Process` and `Unicast Process` are the *narrower* concepts; meaning that they are more specialized forms of the ceramic mold casting process. `Ceramic Mold Casting` is related to concepts such as `Stamping Die` and `Extrusion Die` since it is the process that is used for producing these products. Technically, all terms in ManuTerms can be indirectly related to one another since they all belong to the same domain. Therefore, in connecting concepts through `skos:related` relationship,

---

[2] It should be noted that semantic relationships include lexical relationships as well. Equivalency, hierarchy, and associativity are the main types of semantic relationships that are often used in ontologies.

Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

more direct relationships are taken into account. *Broader*, *narrower*, and *related* are the semantic relations used in any SKOS thesaurus. Also, each SKOS concept can have a *definition* provided in a natural language.

Figure 1 shows the broader and narrower concepts for `Ceramic Mold Casting [30]`.

Figure 1: Narrower and broader concepts for `Ceramic Mold Casting`

4 Ontological Conceptualization Methodology

In this section the proposed process for ontological conceptualization in the manufacturing domain is introduced. Figure 2 shows the overall view of the proposed steps. Thesaurus development is the first phase of ontological conceptualization in this method. The second phase deals with evaluation of the thesaurus with respect to its coverage and semantic correctness. The last phase concerns with selection of the concepts to be included in the ontology.

Figure 2: Ontological conceptualization process

4.1 Phase I: Thesaurus Development

The main steps in developing and extending ManuTerms include *term extraction*, *concept identification*, *internal concept linking*, and *external concept linking*. Pool Party Thesaurus Manager[3] (PPT) is employed as the thesaurus development tool in this work. PPT provides a web-based frontend for collaborative thesaurus development and management and offers different functionalities such as automated term extraction from a given URL or document, connection to other concepts available on Linked Open Data (LOD) [4], and automated population of the thesaurus from external datasets such as DBpedia[5].

4.1.1 Term extraction

This step involves extracting terms that point to an aspect of manufacturing capability knowledge. The extracted terms will be used for answering the competency questions (CQs) identified for the target ontology. Competency questions are the questions that the ontology should be able to answer. Competency questions are developed by the domain experts and are specific to the application envisioned for the ontology. The interested readers are referred to [8] for a more in-depth discussion on how to formulate competency questions for an ontology since it is not within the scope of this paper. Table 1 shows some of the competency questions for the MSDL ontology.

Table 1: Examples of competency questions for ManuTerms and its target ontology

[3] http://www.poolparty.biz/portfolio-item/poolparty-thesaurus-server/
[4] http://linkeddata.org/
[5] http://dbpedia.org/About

Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

Terms collected at this stage are treated as free concepts that are added as encountered to the thesaurus and may be converted into controlled concepts in the later stages. The main resources used for term extraction include online supplier profiles, web-based knowledge portals, DBpedia (the structured extension of Wikipedia), engineering documents, and technical handbooks. DBpedia datasets were created as a result of a crowd-sourced community effort to extracting structured information from Wikipedia and making the information available on the Web as a component of LOD. DBpedia uses RDF triples to represent the extracted information, and therefore, it is compatible with the syntax of SKOS. DBpedia datasets were used early in the term extraction step to create a *seed thesaurus*. The seed thesaurus is a list of entry terms (not preferred terms) that may be used as preferred terms in the thesaurus. Through importing the *Metalworking* dataset in DBpedia, its tree structure was replicated automatically into ManuTerms with the aid of the import feature in PPT. However, since DBpedia datasets, by definition, are domain-independent and generic, not all the imported concepts in Metalworking were deemed relevant to the scope and purpose of ManuTerms. Therefore, to arrive at a more refined seed thesaurus, it was necessary to prune the imported tree and eliminate the irrelevant or redundant concepts. Once the seed thesaurus was generated and trimmed, it was further populated by handpicking more terms from other resources mentioned earlier. In particular, the online profiles of manufacturing suppliers were explored extensively because the technical terms used by the suppliers for describing their technological capabilities directly address different features of manufacturing capabilities. PPT provides a functionality for importing documents and webpages into the thesaurus development environment, tagging terms within the text that point to concepts relevant to ManuTerms, and automatically adding the tagged terms to the thesaurus. Figure 3 and Figure 4 shows a *tagging event* created in PPT in which the capability narrative of a manufacturing company is imported from the online profile of the company and the relevant terms are tagged. The terms that already exist in the thesaurus are automatically highlighted in dark gray in both the original text (Figure 3) and the term cloud (Figure 4). The user then manually tags the relevant terms that can be used as alternative label for existing ManuTerms concept or point to new ManuTerms concepts. The terms that are tagged manually are highlighted in light gray. Similarly, manual tagging can be done either in the original text or in the term cloud. Term cloud includes terms which are automatically generated by PPT from the import based on a reference corpus in PPT's backend). The reference corpus used in term extraction in this project is a generic corpus. However, in PPT it is possible to create a more customized corpus based on a large collection of documents that have manufacturing relevance. A customized reference corpus will result in a more refined term cloud to be used in tagging events. Once the tagging event is saved, the terms that are tagged manually are registered as temporary concepts under a category called *Free Concepts* with the tagged terms being used as the *preferred labels* (`skos:prefLabel`) for the generated concepts. Free concepts are not part of the thesaurus yet and need to be approved by the domain experts[6] before being formally added to the main body of the thesaurus and treated as *Controlled Concepts*. For example, `Vertical CNC` is a term that is manually tagged in the text shown in Figure 3. Once the event is saved, a new concept, identified by `Vertical CNC` as its preferred label, is added to the thesaurus under Free Concepts. `Vertical CNC` technically is eligible for being added to the thesaurus

[6] Domain experts in the context of current work refers to the individual who has in-depth knowledge of various manufacturing processes and their associated equipment.

Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

as a concept since it pertains to a particular category of manufacturing resources that have implications with respect to the manufacturing capability.

Figure 3: Tagged terms in the original text in a PPT's tagging event

Figure 4: Tagged terms in the term cloud in a PPT's tagging event

Term tagging guidelines:

Below some guidelines are provided for more efficient term tagging based on the experience gained through developing the pilot thesaurus. These guidelines are particularly valid for collecting the terms that pertain to the manufacturing capability knowledge. For different domains, different sets of guidelines should be developed. Term tagging guidelines are closely related to the competency questions of the target ontology.

- Tag all unique terms for manufacturing processes (related to CQ8). Suppliers tend to use different terms for referring to the same process. For example `Blanchard Grinding` and `Rotary Surface Grinding` are the alternative terms used for the same type of abrasive machining process.
- Tag all the terms that refer to a special form of manufacturing process such as `Forging Machining`, `Heavy Duty Machining`, and `High Performance Machining`. Such terms usually have a broader meaning beyond a specific manufacturing process and contain information about the process properties and types of parts the process can generally produce.
- Tag all the terms that suppliers use for describing their product (related to CQ5). It is an important aspect of process capability. Example terms include `pump`, `valve`, and `gear`. Search for the terms that refer to categories of parts that have similar attributes such as `Custom precision machined parts` or `Screw machine parts`.
- Tag all the terms for engineering materials. Also, look for the terms that refer to a subgroup of engineering materials such as `Aerospace Materials` or `Exotic Materials` (related to CQ1).
- Tag all the terms that describe a physical resource such as a machine tool or a piece of equipment (related to CQ4). Examples include `Pallet Changer`, `Bar Feeder`, `Twin Pallet Horizontal Milling Machine`, or `Chucking Machine`.
- Tag all the terms that refer to properties of manufacturing processes such as `Precision`, `Surface Roughness`, and `Wall Thickness`.
- Tag all the terms related to the industries a supplier serves such as `Automotive`, `Medical`, `Aerospace`, `Oil and Gas`, and `Alternative Energy`.
- Tagging of the semantically equivalent terms is acceptable and will not cause redundancy at this stage. These terms can be used as alternative labels for a new or an existing concept.

Concept analysis procedures for extracting terms from product/service categories, product/service capability descriptions, and term definitions as described in [31] are also useful for term tagging.

Corresponding Author: Farhad Ameri

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

4.1.2 Concept Identification

In this stage, the tagged terms that are saved as free concepts are converted into controlled concepts and placed under the appropriate *concept scheme*. A SKOS concept scheme can be viewed as an aggregation of several SKOS concepts that belong to the same category semantically. For example, *Process* is an example of a concept scheme because it is the highest level of abstraction for different manufacturing processes. Recall that ontology capture entails identifying the relevant concepts within the domain of discourse. In the same vein, thesaurus concept identification refers to conversion of a term or a group of terms that pertain to the domain of interest (in this case manufacturing capability modeling domain) into a SKOS concept. If a term represents a unique concept that can describe a certain dimension of manufacturing capability, it is eligible to be treated as a controlled concept in ManuTerms. The term can represent a physical resource (such as CNC machine), a process parameter (such as spindle speed), an abstract process (such as solidification process), an industry (such as aerospace, automotive) and the like.

ManuTerms is composed of 16 *concept schemes*: `General Term`, `Facility`, `Feature`, `Hardware`, `Industry`, `Material`, `Metric`, `Model`, `Phenomenon`, `Process Input`, `Process`, `Product`, `Production`, `Service`, `Software`, and `Supply`. The concept schemes do not contribute to the semantics of the terms within the thesaurus and they only serve as header concepts. Therefore, they can be selected arbitrarily as long as they don't overlap with the existing concept schemes. A *middle-out* approach was used for concept identification as opposed to top-down or bottom-up approaches. In the middle-out approach, concepts are identified as they are encountered and deemed relevant to the domain of interest; and generic or specific relationships are established between identified concepts—rather than trying to identify the most generic concept first in the top-down approach or the most specific concept first in the bottom-up approach. For example, `Mechanical Machining` was among the first concepts introduced under the `Process` concept scheme since it was considered to be a relevant concept in the manufacturing domain. `Mechanical Machining` was later generalized to arrive at more abstract concepts such as `Subtraction Process` followed by identification of more specific forms of mechanical machining, such as `Abrasive Machining`, `Single-point Cutting`, and `Multi-point Cutting`. The middle-out approach cannot be enforced during the tagging events because terms are tagged in arbitrary order during such events. However, in converting free concepts to controlled concepts, the middle-out approach can be employed. The middle-out approach creates a more balanced thesaurus in terms of the level of detail incorporated in the concept hierarchy [13]. Furthermore, since many of the higher-level concepts are gleaned from the lower-level ones, as opposed to being enforced from the top, the resulting thesaurus will be more stable structurally. The concepts in the first level below the concept scheme are referred to as *top concepts* in PPT. During concept identification, the identified concepts are classified in the taxonomy of ManuTerms concepts. In this way, the broader/narrower relationships between the concepts are defined explicitly in this step. Also at this stage, the natural language definition of the concept is added to the thesaurus to supply a complete explanation of the intended meaning of the concept. The definition is identified through the `skos:definition` property. `skos:definition` is a sub-property of `skos:note` that is provided for general documentation purpose. Other sub-properties of `skos:note` such as `skos:scopeNote`, `skos:historyNote`, and `skos:editorialNote` can be used at this stage for annotation and providing

Corresponding Author: Farhad Ameri

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

human-readable information for the new concepts added to the thesaurus including the sources of textual definitions.

4.1.3 Internal concept linking

Internal concept linking is conducted after the broader/narrower relationships are established. Internal concept linking entails connecting various concepts within the thesaurus through `skos:related` property. For example, `Hole Making Process` is related to `Drill Press Machine` and `Twist Drill Bit` because they are considered to be enablers of the hole making process. In general, all manufacturing processes under `Process` concept scheme should be linked to some corresponding physical resources under `Hardware > Machine` or `Equipment`. Services are typically linked to core processes, materials, and products that are relevant to the service. Only *direct relations* between concepts are made explicit. Since `skos:related` can be considered a transitive property, several indirect relationships will be established through this property. For example, `Gear Manufacturing Service` is linked to `Gear` (under `Product > Machinery > Driving Element`). Also, `Gear Manufacturing Service` is linked to the related processes such as `Gear Cutting` and `Hobbing`. Since `Hobbing` is the narrower concept for `Milling`, direct linking of the `Milling` process to `Gear Manufacturing Service` will cause semantic redundancy. Since `Milling` is located at the higher levels of the process taxonomy, it has more abstract (hence, less informative). Therefore, linking `Gear Manufacturing Service` to `Milling` adds little formal semantics to `Gear Manufacturing Service`. As a general recommendation, lower-level concepts are preferred candidates for internal linking as they are more information-dense. Based on the same reasoning, `Assembly Service` is linked to processes such as `Mechanical Welding, Rivet Joining, Snap Fit Joining, Threaded Fastener Joining` (as opposed to being linked to more abstract processes such as `Mechanical Joining` or `Consolidation Process`).

4.1.4 External concept linking

External linking entails connecting ManuTerms concepts to the concepts in other PPT projects or the concepts available in various datasets in the LOD. There are five types of possible external linking as described below:

Exact Matching Concept (`skos:exactMatch`): Is used for linking two concepts in two different datasets that have equivalent meaning with a high degree of confidence such that they can be used interchangeably. For example, the exact-matching concept for `ManuTerms:Bar Stock` is `dbpedia:Billet`.

Close Matching Concept (`skos:closeMatch`): Is used for linking two concepts in two different datasets that are sufficiently similar in meaning that they can be used interchangeably. `skos:closeMatch` is not a transitive property whereas `skos:exactMatch` is transitive.

Broader Matching Concepts (`skos:broadMatch`): Is used for linking a concept to its broader concept in a different dataset or PPT project. As an example, `dbpedia:Plating` is a broader match for `ManuTerms:Electroplating`.

Narrower Matching Concept (`skos:narrowMatch`): Is used for linking a concept to its narrower concepts in a different dataset or PPT project. `dbpedia:lost-wax casting` is a narrower matching concept for `ManuTerms:Expendable Mold Casting`, for example.

Corresponding Author: Farhad Ameri

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

Related Matching Concept (`skos:relatedMatch`): Is used for linking a concept to its related concept in a different dataset or PPT project. For example, `Freebase:Metalworking` is a related-matching concept for `ManuTerms:Mechanical Subtraction`.

Figure 5 shows the concept diagram for `Plaster Mold Casting`. As can be seen in this figure, the alternative labels for plaster mold casting are `Rubber Plaster Molding` and `Plaster Molding`. The broader concept is `Expendable Mold Casting` and the narrower concepts are `Antioch Process` and `Prototype Metal Casting`. Since plaster mold casting is primarily used for casting parts made of `Aluminum`, `Copper`, `Magnesium`, and `Zinc`, these concepts are considered to be related to plaster mold casting. The external exact-matching concepts on LOD are `dbpedia:Plaster Molding` and `FreeBase:Plaster Mold Casting`. The natural language definition of `Plaster Mold Casting` will be particularly helpful when developing the axiomatic definition of this process, in an OWL ontology for instance.

Figure 5: The concept diagram for `ManuTerms: Plaster Mold Casting`

4.2 Phase II: Thesaurus Evaluation

Thesaurus evaluation typically deals with *verification* of the thesaurus with respect to satisfying the predetermined requirements and also *validation* with respect to its ability to deliver its intended purpose. There is little research on quantitative methods for thesaurus evaluation and most of the existing methods are informal and qualitative based on subjective evaluation by domain experts or focus groups. In this work, a heuristic is proposed to quantitatively measure the quality of the thesaurus. This measure can be used as a method for thesaurus *verification*. The following questions are used for guiding the thesaurus evaluation process:

- Does the thesaurus adequately cover the relevant terms the domain of interest?
- Are the semantic relationships (broader, narrower, related) adequately captured?
- Are the alternative terms adequately captured?

Twenty online supplier profiles from contract machining sector were sampled to participate in an experiment for thesaurus verification. The objective of this experiment was to assess the ability of the thesaurus in meeting its requirements. It is an *indirect* assessment in a sense that it compares the conceptual model of ManuTerms with that of a human expert. A direct approach will assess the thesaurus with respect to its ability to meet the individual requirements. Each profile was indexed manually by a domain expert through identifying the terms related to the core capabilities of the supplier. The expert was asked to select up to ten terms per profile. The indexed terms may or may not appear explicitly in the profile. For example, a profile may be indexed by the term `Prototyping Service` only because the supplier owns 3D printing technology. The expert did not have access to the list of terms and concepts in ManuTerms. The same sample of supplier profiles was indexed automatically using ManuTerms as supported by the PoolParty Extractor (PPX) tool. Table 2 shows the indexed terms, generated both manually and automatically, for an example supplier profile. The degree of match between the index list generated by the expert and the index list generated by PPX is

Corresponding Author: Farhad Ameri

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

used as an indirect measure of quality of ManuTerms. A higher degree of match with expert judgment suggests a higher level of requirement satisfaction. It is acknowledged that to reduce subjectivity, more domain experts should be included in thesaurus verification process.

Table 2: Indexed terms for the example supplier

Index Match Ratio (IMR) is proposed for measuring the degree of match between manual and automated indexing and it ranges from 0 to 1. IMR for each supplier profile is calculated using the following equation:

$$IMR = \frac{I_{em}+I_{alt}+w_1 \times I_{n-b}+ w_2 \times I_{rel}}{I_{total}} \quad (1)$$

In this equation, I_{em} is the number of exact matches between manual and automated indexing, I_{alt} is the number of terms matched as alternative terms (such as `Inspection Service` matched with `Quality Control Service`), $I_{n\text{-}b}$ is the number of terms matched as narrower or broader terms (such as `EDM` matched with `Wire EDM`) and I_{rel} is the number of terms matched as related terms. I_{total} is the number of indexed terms identified by the expert. w1 and w2 (both equal to or less than one) are the weighting coefficients assigned to $I_{n\text{-}b}$ and I_{rel} respectively to specify their relative importance in different circumstances. In most situations, the number of terms returned by the extractor tool is higher than I_{total}. The IMR equation generally measures the recall rate[7] and does not take into account those terms that appear in the PPX's list but not in the expert's list. This characteristic, however, satisfies the evaluation questions, which are concerned with the thesaurus coverage. The average value of IMR for the sample of twenty suppliers in this experiment was 0.71, which demonstrates a reasonable level of match between expert's list and PPX's list. IMR is only a relative value and doesn't convey much meaning in an absolute term. Average IMR can be improved as more terms and semantic relations are added to the thesaurus.

4.3 Phase III: Converting Thesaurus Concepts to Ontology Classes

All controlled concepts in ManuTerms can potentially be converted into MSDL classes. However, given the burden of defining and maintaining classes in a heavily axiomatic ontology such as MSDL, classes should be added to the ontology only if they refer to the important concepts in the domain. Such concepts are typically generic enough to be useable across multiple ontologies in the manufacturing domain. One way to assess the importance of the collected terms and concepts is through domain experts' evaluation. However, as the thesaurus grows in size, assessment by domain experts becomes subjective and not scalable. In this work, a quantitative technique for concept screening and scoring is proposed that uses two metrics for measuring the importance of a concept: 1) Concept Frequency (CF) and 2) Concept Connectivity (CC). *Concept Frequency* is mainly used for

[7] Recall here refers to the number of concepts in the expert list that are indexed automatically as well.

Corresponding Author: Farhad Ameri

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

concept screening to reduce a large number of concepts to a smaller set of more important and generic concepts while *Concept Connectivity* is used for further refining the concepts that pass through concept screening. The connectivity score is determined based on the number of relationships a concepts has.

Concept Frequency: CF measures the number of occurrences of a concept in a reference text. The purpose of calculating CF for a given concept is to provide a measure of importance of the concept. The reference text is not necessarily the same text used during concept extraction. To provide meaningful evaluation of the importance of the concepts, the reference text should be fairly large in terms of the count of words and highly relevant to the domain knowledge. To evaluate the relevance of the reference text, a metric called *Concept Return Ratio* (CRR) is introduced. For a given natural language text, CRR is calculated using the following equation:

$$CRR = \frac{N_{ec}}{N_{ec} + N_{et}} \quad (2)$$

Where N_{ec} is the number of concepts in the thesaurus present in the reference text (either directly or through alternative labels) and N_{et} is the number of terms in the reference text indexed by PPX. More precisely, N_{et} represents the number of terms or phrases found in the PPT corpus [8] and the reference text that are not mapped to any ManuTerms concept. The denominator in this fraction is the normalization factor which cancels the impact of the size of the document. The reference texts with higher manufacturing relevance have higher CRR . A Java-based concept extractor (matcher) tool was developed to automatically calculate the CRR for any input reference text.

Three sets of reference text were created by aggregation of the profiles of 100 manufacturing suppliers in three different areas, namely, 5-axis machining, CNC milling, and metal casting. The calculated CRR for the three reference text files corresponding to 5-axis machining, CNC milling, and metal casting are 0.19, 0.17 and 0.13 respectively. The lower CRR for the casting reference text reflects the fact the metal casting reference text is less relevant to the thesaurus (domain of interest) compared to the 5-axis machining and CNC milling reference texts. This corresponds to the fact that the set of machining-related concepts in ManuTerms is richer compared to the set of casting-related concepts. After evaluating CRR of sample documents using the PPT's generic corpus, it was determined that a document with a CRR greater than or equal to 0.1 is sufficiently related to manufacturing domain to be used for concept evaluation. It should be noted that the threshold is generally project-specific as the value is dependent on the corpus used. Concept frequency (CF) is calculated for concepts extracted from each reference text passing the minimum CRR threshold. Table 3 shows the concept frequency of the top twenty concepts returned by the extractor tool from the three reference text files.

[8] In this work PPT corpus was not the same as the reference text. PPT corpus, in its default setting, is a generic text without any manufacturing significance. CRR measure can be improved if a manufacturing-related PPT corpus is available.

Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

Table 3: Top twenty concepts extracted from three text files corresponding to 5-axis machining, CNC Milling, and casting.

Concept Connectivity: CC refers to the number of internal or external concepts connected to a given concept. The reasoning behind selecting connectivity as a measure of importance a concept is that more important and frequently used concepts have higher number of links to other concepts in the domain. Concept connections can be made through different types of semantic relations including `skos:related`, `skos:exactMatch`, `skos:closeMatch`, and `skos:relatedMatch`. Except for the `skos:related` relation that connects a given concept to other concepts within the same thesaurus, other semantic relations connect the concept to the external concepts in different PPT projects or LOD datasets related to the domain of interest. The connectivity score for the concept *i* is calculated using the following equation:

$$CC_i = \mu N_{re\text{-}i} + \nu (N_{em\text{-}i} + N_{cm\text{-}i} + N_{rm\text{-}i}) \quad (3)$$

Where $N_{re\text{-}i}$ is the number of `skos:related` relations, $N_{em\text{-}i}$ is the number of `skos:exactMatch` relations, $N_{cm\text{-}i}$ is the number of `skos:closeMatch` relations, and $N_{rm\text{-}i}$ is the number of `skos:relatedMatch` relations for the concept *i*. μ and ν are the weighing factors used for varying the importance of the external or internal links in different circumstances. In this work, $\mu = 1$ and $\nu = 2$ is used to assign more weight to the links that connect a ManuTerms concept to external concepts. External links are deemed more important since they suggest that the concept is already in use in other concept models and is validated by a wider range of knowledge users. The normalized concept connectivity score of the *i*th concept in the thesaurus is calculated as:

$$CC^n_i = CC_i / CC_a \quad (4)$$

where :

$$CC_a = \mu N_{re\text{-}a} + \nu (N_{em\text{-}a} + N_{cm\text{-}a} + N_{rm\text{-}a}) \quad (5)$$

CC_a is the average connectivity score for the thesaurus calculated based on the average number of internal and external links for a concept in ManuTerms. It should be noted that CC^n_i is not a constant value and it changes as the thesaurus evolves in time. The changing nature of CC^n_i does not pose a problem for the proposed method since this relative measure is calculated at the same timestamp for all concepts.

The CC score for the hypothetical concept shown in Figure 6 is 11 (3+4×2) before normalization. CC_a = 5 is used for normalization since on average a concept in ManuTerm has three internal links and one external link (3+1×2=5). A java-based tool was developed for automated calculation of the connectivity score of ManuTerms concepts. Table 4 shows the CC scores calculated for some of the concepts in ManuTerms.

Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

Figure 6: Concept Connectivity (CC) score is calculated based on the number of external and internal links for each concept.

Table 4: Concept Connectivity (CC) score for some of the most connected concepts in ManuTerms

Using the provided concept evaluation metrics and the developed tools for score calculation, it is possible to dynamically evaluate the concepts with respect to connectivity and frequency as the thesaurus evolves in time. Figure 7 shows the steps for converting ManuTerms concepts into MSDL classes. This process is triggered whenever a specified number of new concepts (N_c, e.g., 100) are added to ManuTerms. In the first step, concept frequency score is calculated for all concepts in the thesaurus. A number of concepts whose CF scores are above a certain level are added to the set of candidate concepts to be converted into MSDL classes. The concepts with connectivity score greater than the average connectivity score (CC_a) are selected eventually as the qualified concepts. It should be noted that the average connectivity score (CC_a) needs to be constantly updated since its value changes as the size and structure of the thesaurus changes.

Both CF and CC change as the domain expands; therefore, concepts that were deemed irrelevant to MSDL earlier can become important later on. Because of such behavior, the proposed measurements are well-suited for ontology development in large and evolving domains. The concepts that are selected at each iteration are then converted into OWL classes and the necessary properties and axioms are defined for them in the Protégé environment. ManuTerms concepts and their associated MSDL classes will be then connected through their Uniform Resource Identifiers (URIs). This is the last step of the proposed methodology for ontological conceptualization guided by a thesaurus.

Figure 7: The process of introducing ManuTerms concept to MSDL based on their connectivity and frequency score

5 Conclusion

In this paper a methodology was introduced for evolutionary ontological conceptualization in the manufacturing domain. A thesaurus-based approach was used in this work supported by the syntax and semantics of Simple Knowledge Organization System (SKOS). SKOS can support the creation of semantically rich meta-models and, at the same time, it is simple enough for domain experts to understand and use for thesaurus development. The developed thesaurus serves as an interface between informal engineering terms and formal ontological classes. One notable advantage of using SKOS for thesaurus representation is that the conceptual model is formalized as it is being developed by the knowledge engineers. The steps in the proposed ontological conceptualization methodology include thesaurus development, thesaurus evaluation, and concept conversion. The proposed methodology ensures that ontology building is mainly conducted at the semantic level rather than the symbol level or the implementation level. Three principles can be summarized as best practices for the thesaurus development: 1) Use the middle-out approach to concept identification; 2) Assign the broader and narrower relationships before the related relationship; and 3) Use the lowest level concept possible when linking concepts with the related relationship.

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

Although the proposed procedure is developed in the context of manufacturing ontology development, the underlying logic, tools, and metrics can be applied to develop any domain ontology. The developed thesaurus can serve as a standalone lightweight ontology and be shared as a dataset in LOD such that its concepts can be reused by other semantic models and controlled vocabularies. Also, because of its minimal ontological commitment, ManuTerms can support knowledge sharing among disperse agents that are implemented using different ontologies.

One possible avenue for further research in this area is using ManuTerms for search and information retrieval. In a previous research, an OWL-based search engine was developed for manufacturing supplier discovery purpose [32]. However, given the complexities associated with using heavily axiomatic ontologies, OWL-based search engines might impose unnecessary computational burden on the search system. Therefore, less complex information models such as SKOS thesauri can provide a viable solution for more efficient search and retrieval of engineering information.

This paper reported the early findings of an ongoing research aimed at developing formal methods for collaborative ontology development. For continuous enrichment and consensus-based validation of ManuTerms, a crowdsourcing approach will be explored in the future to enable decentralized and collaborative thesaurus development by a network of domain experts and end users. The long-term objective is to develop and implement a methodology, supported by Semantic Web technology suite, that enables all the stakeholders, including suppliers, software vendors, and standard development organizations to participate in creation, validation, and extension of reference service ontologies in the manufacturing domain.

Acknowledgement
This work was partially funded by National Institute of Standards and Technology (NIST) Award 60NANB11D108 with Texas State University at San Marcos.

Disclaimer:
Certain commercial software products are identified in this paper. These products were used only for demonstration purposes. This use does not imply approval or endorsement by the authors or their organization, nor does it imply these products are necessarily the best available for the purpose.

References

[1] Li, Z., and Ramani, K., 2007, "Ontology-Based Design Information Extraction and Retrieval," Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM, 21(2), pp. 137-154.
[2] Seung-Cheol, Y., Patil, L., and Dutta, D., 2010, "Function Semantic Representation (Fsr): A Rule-Based Ontology for Product Functions," Journal of Computing and Information Science in Engineering, 10(3), pp. 031001 (8 pp.).
[3] Yan, Y., Dong, Y., Zhibin, J., and Lixin, T., 2008, "Ontology-Based Semantic Models for Supply Chain Management," International Journal of Advanced Manufacturing Technology, 37(11-12), pp. 1250-60.

Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

[4] Martinez Lastra, J. L., and Delamer, I. M., 2008, "Ontologies for Production Automation," Tiergartenstrasse 17, Heidelberg, D-69121, Germany, 4891 LNCS, pp. 276-289.
[5] Pinto, H. S., and Martins, J. P., 2004, "Ontologies: How Can They Be Built?," Knowledge and Information Systems, 6(4), pp. 441-64.
[6] Neuhaus, F., and Vizedom, A., 2013, "Ontology Summit 2013 Communiqué:Towards Ontology Evaluation across the Life Cycle," Gaithersburg, MD.
[7] Gruber, T. R., 1993, "A Translation Approach to Portable Ontology Specification," Knowledge Acquisition, 5(2), pp. 199-220.
[8] Grüninger, M., and Fox, M. S., 1995, "Methodology for the Design and Evaluation of Ontologies," Proc. Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, Canada.
[9] Ahmed, S., Sanghee, K., and Wallace, K. M., 2007, "A Methodology for Creating Ontologies for Engineering Design," Transactions of the ASME. Journal of Computing and Information Science in Engineering, 7(2), pp. 132-40.
[10] De Nicola, A., Missikoff, M., and Navigli, R., 2009, "A Software Engineering Approach to Ontology Building," Information Systems, 34(2), pp. 258-75.
[11] W3c - World Wide Web Consortium, 2009, Skos Simple Knowledge Organization System Reference, http://www.w3.org/TR/skos-reference/
[12] Ameri, F., and Dutta, D., 2006, "An Upper Ontology for Manufacturing Service Description," ASME Conference Proceedings, 2006(42578), pp. 651-661.
[13] Uschold, M., and King, M., 1995, "Towards a Methodology for Building Ontologies," Proc. Workshop on Basic Ontological Issues in Knowledge Sharing - IJCAI, Montrela Canada.
[14] Fernández López, M., Gómez Pérez, A., and Juristo, N., 1997, "Methontology: From Ontological Art Towards Ontological Engineering," Proc. Symposium on Ontological Engineering of AAAI, Stanford University, California pp. 33-40.
[15] Li, Z., Raskin, V., and Ramani, K., 2007, "A Methodology of Engineering Ontology Development for Information Retrieval," Proc. International Conference on Engineering Design, Paris, France, DS 42.
[16] Staab, S., Studer, R., Schnurr, H. P., and Sure, Y., 2001, "Knowledge Processes and Ontologies," IEEE Intelligent Systems, 16(1), pp. 26-34.
[17] Pinto, H. S., Staab, S., and Tempich, C., 2004, Ecai 2004: 16th European Conference on Artificial Intelligence, Proceedings.
[18] Gomez-Perez, A., and Suarez-Figueroa, M. C., 2009, "Scenarios for Building Ontology Networks within the Neon Methodology," Redondo Beach, CA, United states, pp. 183-184.
[19] 2005, "Guidelines for the Construction,Format, and Management of Monolingual Controlled Vocabularies " Technical Report No. ANSI/NISO Z39.19-2005, ANSI American National Standards Institute New York, NY.
[20] 1986, "Guidelines for the Establishment and Development of Monolingual Thesauri," Technical Report No. ISO 2788:1986 nternational Organization for Standardization Geneve.

Corresponding Author: Farhad Ameri

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

[21] Kolte, S. G., and Bhirud, S. G., 2009, "Wordnet: A Knowledge Source for Word Sense Disambiguation," International Journal of Recent Trends in Engineering, 2(4), pp. 213-17.
[22] Sparck Jones, K., and Needham, R. M., 1968, "Automatic Term Classifications and Retrieval (Cambridge Language Research Unit, England)," Information Storage and Retrieval, 4(2), pp. 91-100.
[23] 1969, "Faa Thesaurus of Technical Descriptors," Federal Aviation Administration, Department of Transportation.
[24] Pastor-Sanchez, J. A., Mendez, F. J. M., and Rodriguez-Munoz, J. V., 2009, "Advantages of Thesaurus Representation Using the Simple Knowledge Organization System (Skos) Compared with Proposed Alternatives," Information Research-an International Electronic Journal, 14(4), pp. 31.
[25] 2011, Agricultural Thesaurus and Glossary Home Page, http://agclass.nal.usda.gov/
[26] 2011, "Medical Subject Heading."
[27] National Cancer Institute (Nci) Thesaurus, http://ncit.nci.nih.gov/
[28] 2011, "Gemet Thesaurus - Eionet."
[29] Yang, M. C., Wood, W. H., Iii, and Cutkosky, M. R., 2005, "Design Information Retrieval: A Thesauri-Based Approach for Reuse of Informal Design Information," Engineering with Computers, 21(2), pp. 177-92.
[30] Todd, R. H., Allen, D. K., and Alting, L., 1994, Manufacturing Processes Reference Guide, Industrial Press Inc, New York.
[31] Shin, J., Kulvatunyou, B. S., Lee, Y., and Ivezic, N., 2013, "Concept Analysis to Enrich Manufacturing Service Capability Models," Atlanta, Georgia.
[32] Ameri, F., and Patil, L., 2010, "Digital Manufacturing Market: A Semantic Web-Based Framework for Agile Supply Chain Deployment," Journal of Intelligent Manufacturing, 10.1007/s10845-010-0495-z), pp. 1-16.

Corresponding Author: Farhad Ameri

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

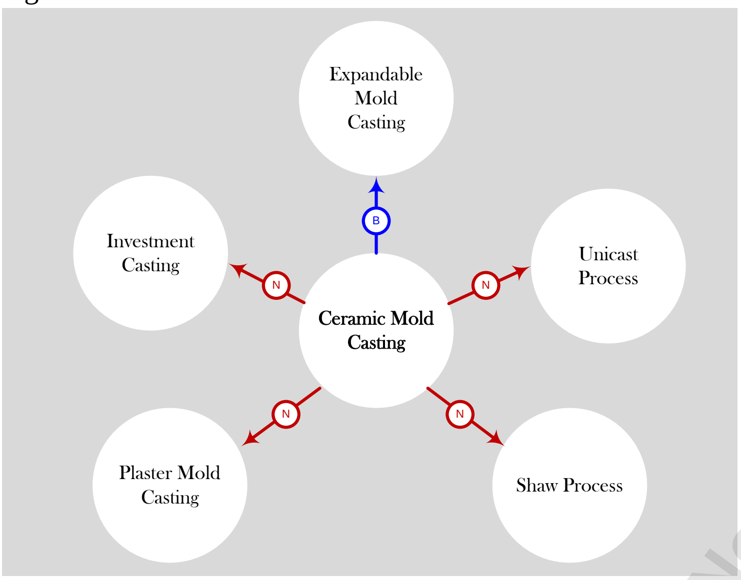
List of Figures:

List of Tables:

Corresponding Author: Farhad Ameri

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

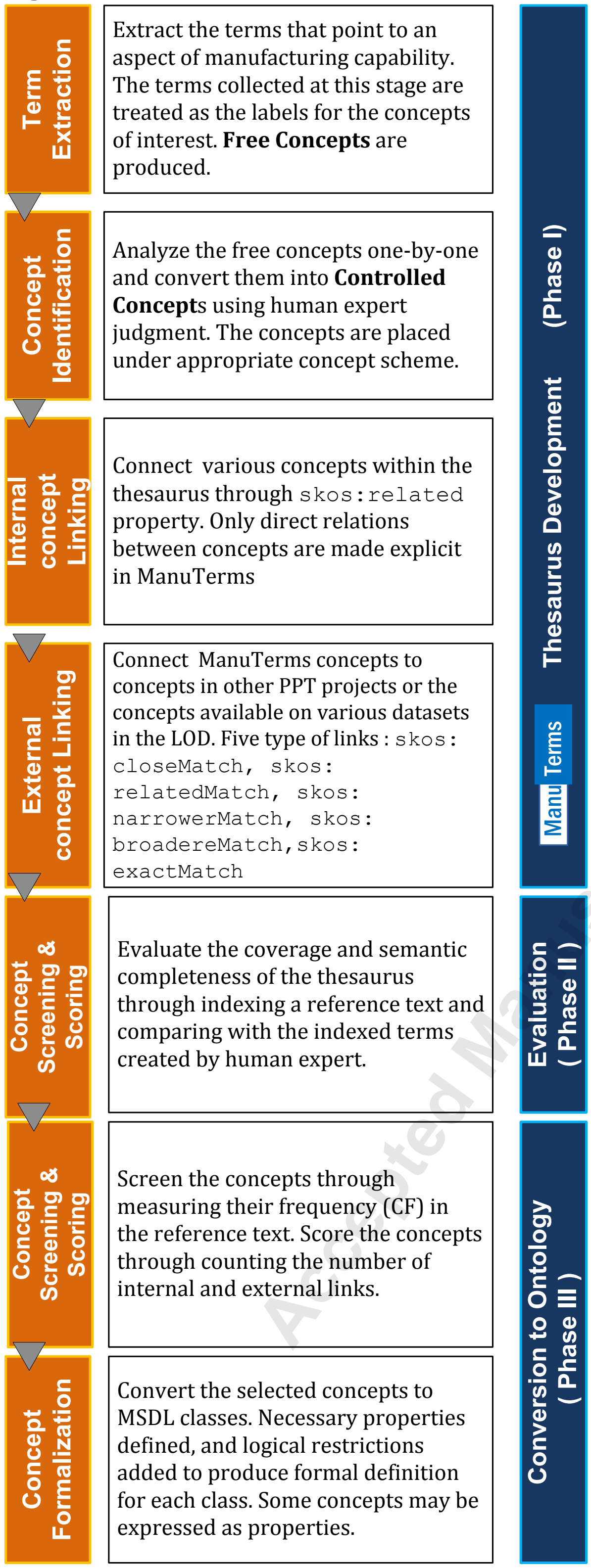
Figure 1



Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

Figure 2



Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

Figure 3

Standard vertical CNC machining for custom made parts or build assemblies. Includes 5-axis vertical machining centers for machining complex shapes, undercuts difficult angles, twin pallet vertical CNC precision machining centers with low end torque capabilities of 35 lb-ft with speed ranges of 60 to 8000 rpm. Also includes optional magnetic spindle installation, ultra precision ballscrews, bearings, micro switches. Materials include aluminum, plastics, stainless steel, copper, titanium cut to precision using CADCAM software.

Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

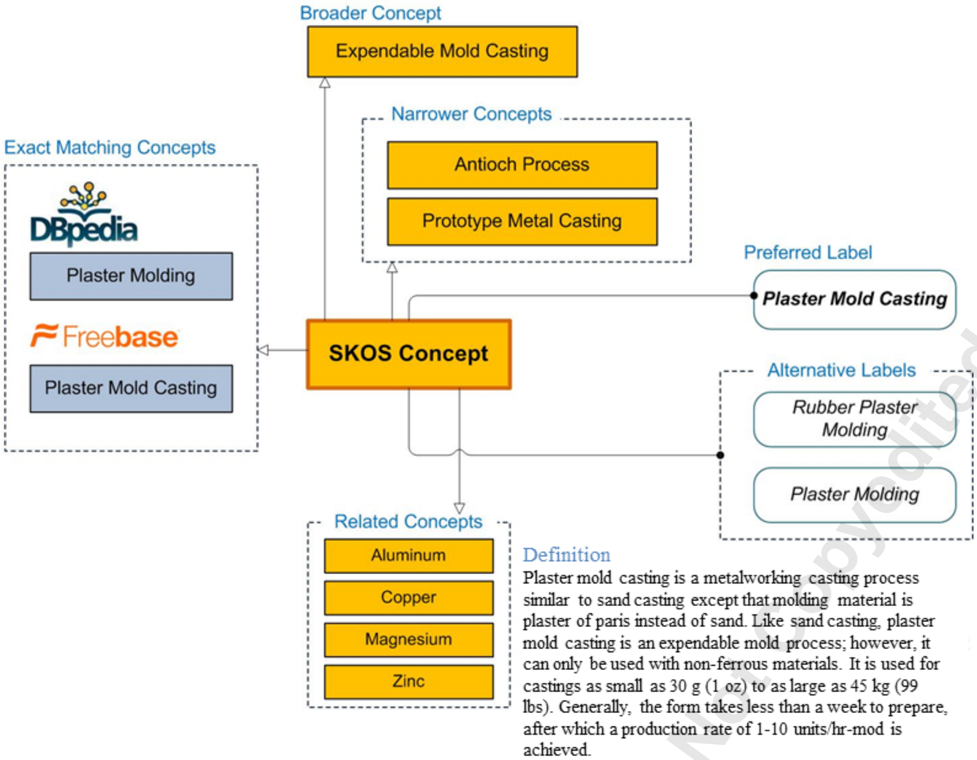
Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

Figure 4

Corresponding Author: Farhad Ameri

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

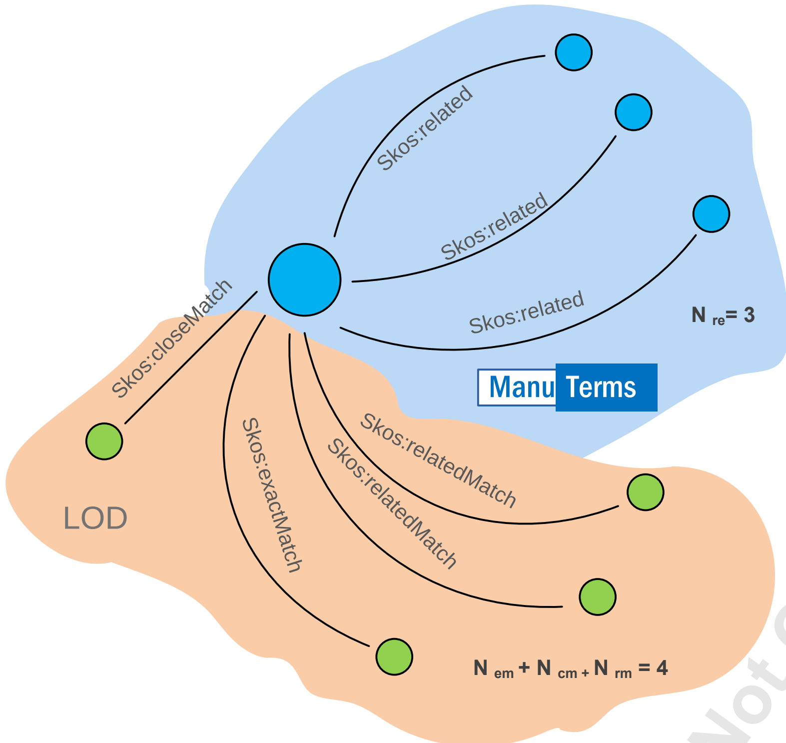
Figure 5



Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

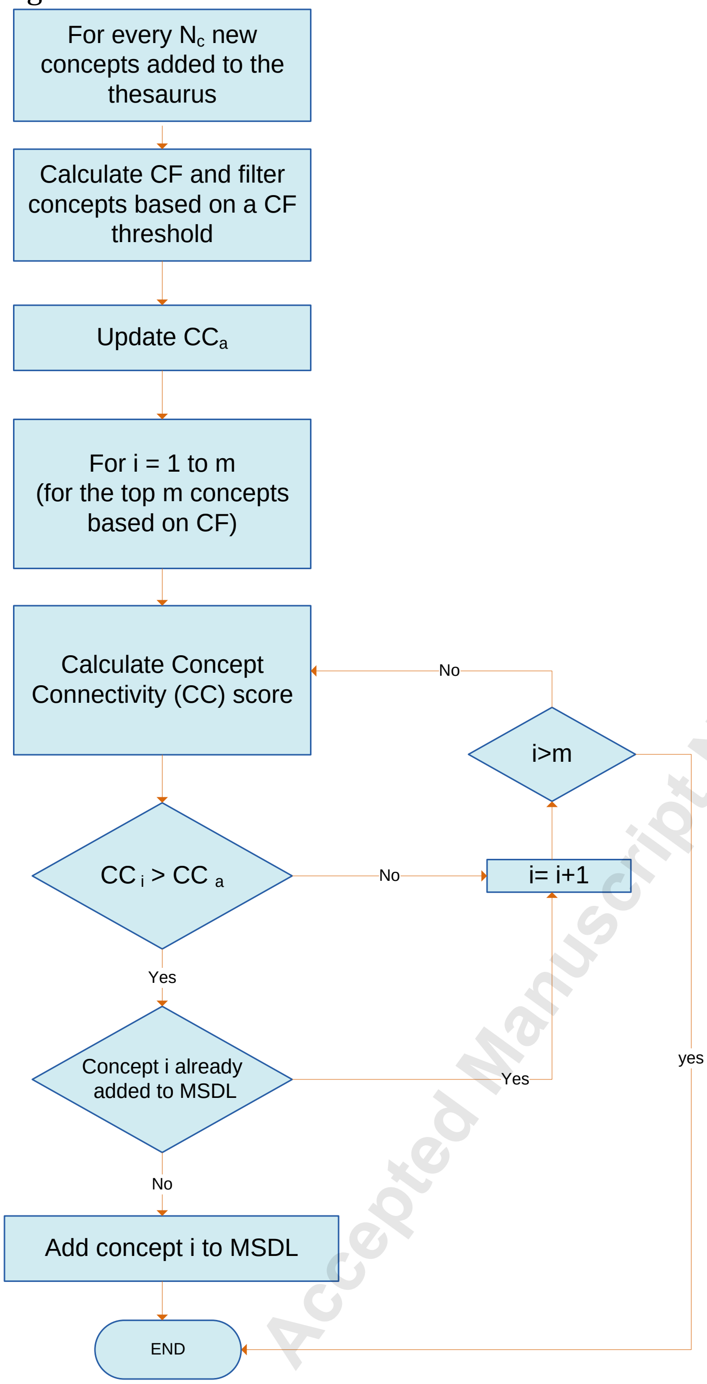
Figure 6



Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

Figure 7



Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

Table 1

CQ_1	What are the materials the supplier can process?
CQ_2	What are the geometries the supplier can accommodate?
CQ_3	What is the main expertise of the supplier?
CQ_4	What are the types of machinery and equipment the supplier owns?
CQ_5	What are the types of products and industries the supplier specializes in?
CQ_6	What are the tolerances the supplier can accommodate?
CQ_7	What is the typical production volume of the supplier?
CQ_8	What are the processing services the supplier can offer?

Corresponding Author: Farhad Ameri

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

Table 2

Capability narrative from the online profile	Manual Indexing	Automated Indexing	Match Type
Our company is an ISO-compliant, full-service precision machine shop capable of supplying prototype quantities to production quantities. Utilizing the most modern machining technologies coupled with the latest innovations in CNC multifunctional turning center, vertical machining centers, programming software and tooling allows us to optimize product quality, increase production, and deliver to our customers with exceptional results. We supply the highest quality precision machined parts on time, every time. All job routers, materials, outside services and inspection procedures are 100% traceable and tracked throughout the entire machining process. Services supplied but not limited to CNC Milling, CNC Turning, Gun Drilling, Grinding (Cylindrical / Surface), Honing, EDM, Heat Treating, Water Jet, Dynamic Balancing. Assembly service utilizes the same tracking as our machining process. With dedicated assembly/staging areas, we can supply small to large, and simplistic to complex assemblies and subassemblies. We are flexible to our customers' needs and can include complete testing with traceability as needed. Testing ranges from performance to visual ND. Leak test includes pressure, vacuum and hydrostatic (the profile is shortened).	Grinding Turning CNC Machining Heat Treating Honing Precision Machining	Grinding Turning CNC Machining Heat Treating Honing Precision Machining	Exact Match I_{em}= 6
	Inspection Service	Quality Control Service	Alternative label match I_{alt}=1
	EDM Milling	Wire EDM Machining Service	Broader / narrower match I_{n-b}=1
	Assembly Fabrication	Assembly Services	Related match I_{rel}=2
	Low-volume production High-volume production	Horizontal Turning CNC Turning CMM Service Reverse Engineering Service CMM Service Metal Stamping	No match
	w_1= .5 and w_2=.5 and I_{total} = 13 : IMR= 0.65		

Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

Table3

5-Axis Machining	
Nec	**182**
Nec + Net	**957**
CRR	**0.19**
Concept	***CF***
Mechanical Subtraction	282
Machining Service	179
CNC Machining	168
5-Axis Machining	110
Custom Manufacturing Service	107
Milling machine	81
Precision Machining	80
Milling	79
Precision Part	78
Turning	71
Horizontal Turning	63
Assembly Service	63
Steel	62
ISO 9001	56
Grinding	52
Multiaxis Machining	51
Aluminum	37
Electrical Discharge Machining	36
Polymer	31
Consolidation Process	30

CNC Milling	
Nec	**174**
Nec + Net	**1,023**
CRR	**0.17**
Concept	***CF***
Machining Service	342
CNC Machining	244
Milling machine	217
Milling	215
Mechanical Subtraction	148
Custom Manufacturing Service	122
Turning	85
Steel	80
Assembly Service	77
Horizontal Turning	76
Precision Machining	73
Precision Part	68
Polymer	47
Aluminum	43
Consolidation Process	40
Grinding	40
Drilling	40
Hole Making	40
ISO 9001	38
Stainless Steel	34

Casting	
Nec	**167**
Nec + Net	**1,284**
CRR	**0.13**
Concept	***CF***
Casting Service	195
Cast Workpiece	147
Custom Manufacturing Service	111
Aluminum	60
Casting	48
Mechanical Subtraction	42
Die	41
Steel	35
Iron	27
Zinc	26
Prototyping	25
Metal	25
Assembly Service	25
ISO 9001	25
Molding	21
Prototyping Service	20
Precision Machining	20
Consolidation Process	18
Bronze	18
Precision Part	18

Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138

Journal of Computing and Information Science in Engineering. Received July 24, 2013;
Accepted manuscript posted May 2, 2014. doi:10.1115/1.4027582

Table 4

Concept	Internal links	External links	CC	CC^n
Machining	19	5	29	5.8
Casting	17	5	27	5.4
Lathe	15	5	25	5
CNC Machine	13	4	21	5.2
Drill	10	3	16	3.2
Machine Tool	9	3	15	3
Precision Machining	7	3	13	0.6

Corresponding Author: Farhad Ameri
Paper #: JCISE-13-1138