Chapter 1

INTRODUCTION

The American Heritage Dictionary defines "ontology" as "the branch of metaphysics that deals with the nature of being." Ontology, as a formal specification of a shared conceptualization [3], defines the thesaurus of the concept, relations between concepts and properties even relations between properties. Moreover, it describes axioms and individuals and relations between them, and provides sharing knowledge [4]. More specifically, it is an explicit specification of a conceptualization [5, 10].

The origin of ontologies in computer science can be referred back to 1991, in the context of the DARPA Knowledge Sharing Effort (Neches, Fikes, Finin, Gruber, Senator, & Swartout, 1991). The aim of this project was to devise new ways of constructing knowledge-based systems, so that the knowledge bases upon which they are based did not have to be constructed from scratch, but by assembling reusable components. This reuse applies both to static knowledge, which is modeled by means of ontologies, and dynamic problem-solving knowledge, which is modeled by means of problem solving methods [1].

Initial application of ontologies in computer science was that it acted as a means of providing the semantics for the semantic web i.e. web 2.0. By use of ontologies, the information retrieval could then be done based on the context (/meaning) rather than just simple string matching mechanisms.

In this work we provide an overview of what ontology is, describing the current trends, issues and problems in constructing them. We also propose an ontology development methodology *On-to-Methodology* that could be used as a standard model for ontology development tasks across various domains. We have automated the development process by implementing a tool for Ontology design process. We illustrate our methodology by

developing Ontology of Bikes. We then compare our methodology with other existing ontology development methodologies.

1.1 Motivation

The Ontologies created by different creators are different and inconsistent. There is a huge difference between different domains and currently there is no defined methodology for constructing Ontologies [12, 14, 19]. And inspite much being written in literature about Ontologies, the number and quality of actual, "non-toy" ontologies available on the Web today is remarkably low [19,20].

The development of ontologies generally consists of four main tasks:

- *Scope Definition*
  This activity involves establishment of scope of the ontology in terms of purpose that the ontology would serve. This is accompanied by the advantages that the ontology would provide after it has been developed.

- *Knowledge Acquisition*
  The target of this activity is to acquire the knowledge required for successful development of the ontology. This activity specifies the knowledge sources from where the knowledge would be compiled.

- *Design*
  The concepts identified during *Knowledge Acquisition* activity integrated to form an Ontology. The design specifies the architecture of the ontology and acts as an important step for successful formalization of the ontology.

- *Formalization*
  This activity deals with formalization of the ontology and represents it in a form so that it is understandable by the machines. The ontology is formalized using ontology languages and tools.

The development of ontologies is still a matter of craft skill rather than an understood engineering process. For the successful utilization of the potential of ontologies, it is required to formally define this practice.

1.2 Related Work

Fernández et. al. proposed "Methontology" and defined the ontology development process and the ontology life cycle [21]. Farooq et. al. proposed "Design of Ontology in Semantic Web Engineering Process and illustrated their method with the help on a case study [24]. Gaoyun et. al. identified the reasons for the inconsistencies in ontology construction and put forward a method for Consistent Ocean Ontology Construction [19]. Zhang et. al. proposed a methodology namely MADRE and illustrated their work by focusing on healthy housing [25]. Jia et. al. put forward "Automatic ontology construction approaches and its application on military intelligence" [27]. Chen et. al proposed Concept feature based ontology construction and maintenance which was backed up by a detailed framework [15]. A structured ontology construction by using data clustering and pattern tree mining was proposed by Yao-Tang et. al. [14].

These proposals seem to suffer from following drawbacks [19]

1. *Inconsistency of domain cognition*

   This kind inconsistency could be classified to two cases. One is about scope of domain that is people have no consistent definition to the scope of domain. The other is about contents of domain that is which contents belong to the domain people have their own opinions.

2. *Inconsistency of viewpoint*

   In many cases constructions which are restricted to the same scope and are built by domain experts with the same background knowledge still have essential dissimilarities. The reason is viewpoint. As we know one matter may

get contrary conclusions from different perspectives, so does ontology construction. Different viewpoints could also lead to inconsistent ontologies.

3. *Ontology construction is affected by subjective factors*

   From the definitions discussed in Section 1.1 above, it is clearly that ontology is category of subjective field, therefore ontology construction would be influenced by subjective factors heavily. Even in the same field with same viewpoint ontologies may be constructed various from one another due to creators' interests, tastes, preferences, ideas, abilities etc.

4. *Lexical inconsistency*

   Conceptualization is independent from vocabulary, but ontology is language-related. That is ontology should be expressed by a special formal language. Then there is another kind inconsistency that creators from different nationalities use their own language to build ontologies. Even the language is unified, the situation is still not improved. That is because all the languages existed cannot avoid the following two problems: "polysemy" and "thesaurus".

Martin Hepp [20] classifies the ontology-related tasks into two main groups: building or contributing to the development of ontologies and committing to a particular ontology. Committing to a given ontology, explicitly or implicitly, means agreeing that it properly represents the domain's conceptual elements.

So to solve the aforesaid problems with ontology engineering, we propose a methodology that is generic and comprehensive and narrows down the gap between constructing and committing to the ontologies.

1.3 Problem of the thesis

In this work we focus on proposing a method that is comprehensive, generic and makes it easy to commit and construct ontologies of various domains specifically transportation like bikes, cars, planes etc. We aim to produce documents similar to the documents in Software Engineering like Software Requirement Specification document known as IEEE std. 830. These documents will help the ontology to evolve. Also we wish to provide computer based support for these activities. The proposed methodology will have following features:

- *Better domain cognition*

  As listed in Section 1.3, there exists inconsistency in domain cognition. Our method handles this problem by adapting group oriented activities thus enabling experts and users to define the domain in better terms.
- *Unification of viewpoints*

  Inconsistency in viewpoints arises as different people have different perspectives and due to this contrary conclusions may be obtained. Our methodology resolves this problem by instrumenting the development process with evaluation activities after each phase of the development process. Also, the unification is achieved due to the adaptation of group oriented development activities.

- *Exhaustive documentation*

  On-To-Methodology provides an extensive and exhaustive documentation support thus making the development process more formal and helps in achieving traceability across phases. Documentation also simplifies the maintenance of the ontology by providing the backward traceability of various components of the ontology.

- *Extensive Verification & Validation*

  These two tasks of verification and validation are interleaved between various phases of the development process. This prevents introduction of bugs into the system right from the beginning of the development process and prevents faults till the maintenance phase.

So problem of the thesis is:

> **"Develop a comprehensive ontology methodology that is generic and construct a tool which can automate the construction process."**

1.4 Scope of the thesis

Our ontology construction methodology will have the following steps as shown in figure :

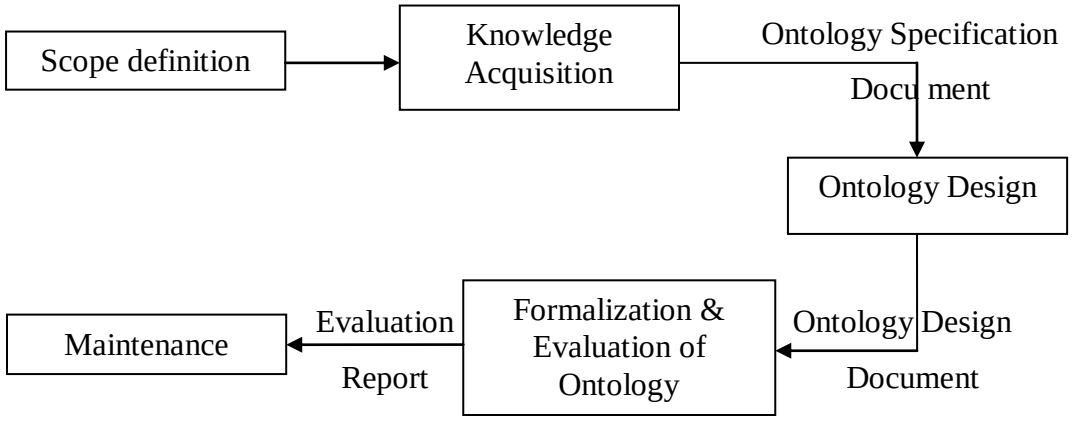

Figure 1: Steps involved in On-To-Methodology

Knowledge elicitation and representation plays an important role construction of Ontology. It is done through a formal group oriented technique based on FAST (Facilitated Application Specification Technique). We have designed set of forms to carry out these activities to formalize knowledge elicitation and then conceptual model of

the knowledge is built which is the base of ontology construction that will narrow down the gap between ontology building and commitment. Output of this activity is the Ontology Specification Document (OSD) which is similar to the Software Specification document in Software Engineering. Hence our knowledge is elicitated & represented in a formal manner.

For automating the ontology design process we present an algorithmic approach that will generate the ontology structure from the conceptual model of the knowledge. In this way we will automate the design process. This tool will be available online so that ontology engineers can access it from anywhere and anytime.

We shall apply On-To-Methodology to domain of Automobile, specifically Bikes. The intended ontology of Bikes will be representation of bikes in the domain of ontology. The purpose of this ontology is to provide information on bikes based on the criteria specified by the users.

- The ontology would hold information to answer queries of customers based on single (/combination of) parameter(s) which are Make, Engine capacity, Power, Price, Fuel tank capacity, Mileage, Brake type, Weight, Wheel type, Ignition and Number of gears.
- Bike manufacturing organizations can use this ontology to identify the bike configurations that are suitable for a particular market and can also use it to analyze current sales and make future predictions. This will guide them to plan their production & inventory.
- This ontology can prove to be beneficial for bike retailers as they can use it to plan their inventory and analyze their sales.

The advantage that this ontology would provide is its capability to answer the queries of the customers across a large information base of different bikes, based on multiple search criteria with complex inter-relations.

Broadly, the scope of this work can be summarized as:

- Develop a detailed process for knowledge elicitation to build the conceptual model with comprehensive documentation support. The output is the Ontology Specification Document (OSD).
- Develop an algorithm that will automate the ontology design activity.
- Develop an ontology of Bikes using the proposed methodology.
- Evaluate our methodology with other existing methodology including Methontology, Methodology by Farooq et. al., Methodology by Gaoyun et. al., MADRE, approach by Jia et. al, Concept feature based Ontology Construction and Maintenance and Structured Ontology Construction by using data clustering & Maintenance.
- Develop a tool that will support our methodology.

1.5 Thesis Statement and Outline

This aim of this dissertation is to propose a well defined methodology for construction of ontology.

The rest of the thesis is organized as follows:

In chapter 2, we review the concept of methodology for creation of ontologies and also describe the current state of the art in ontology engineering.

Chapter 3 provides the proposed methodology for ontology construction.

In chapter 4, we illustrate our methodology by constructing an Ontology of Bikes.

In chapter 5, we evaluate the Ontology of Bikes for various competency questions and compare our approach with other existent methodologies.

Chapter 6 concludes the thesis and proposes future work.

Chapter 7 gives the list of references that I have gone through during my research.