# A Reasoning Mechanism for Corpus-Based Knowledge Representation

Shabnam Surjitsingh Ivkovic
Department of Computer Science
University of Waterloo, Ontario, Canada.
sivkovic@uwaterloo.ca

Anne Banks Pidduck
Department of Computer Science
University of Waterloo, Ontario, Canada.
apidduck@uwaterloo.ca

**Abstract**

Corpus-based knowledge representation is based on the idea that statistical patterns in a bank of domain models can be leveraged to create novel tools for knowledge representation. We have observed that a statistics based reasoning mechanism added to the existing model will facilitate functions such as hypothesis testing and partial predictability. Our research suggests the methodology of *corpus-based knowledge representation with reasoning* to assist the finding of semantic relations between models, and then reasoning with those relations towards recognizing patterns and testing hypotheses. Towards this goal, we propose statistics that facilitate reasoning. These statistics are based on a Basis of Reasoning model. This paper explains this model and the statistics for reasoning.

## 1 Introduction

Computing is gradually gravitating towards a symbiotic design that encompasses a variety of cross-connected domains. For example, what people shop, how they use their credit cards, what bills they pay, what institutions are they affiliated to, etc., all combined can pretty much build a profile of a person in particular. Most aspects that computing is applied to can be represented by a domain model. Considering the fact that these aspects are gradually being integrated into one whole, it would seem prudent to find common, underlying characteristics of these domain models.

One of the methodologies that suggest an approach to accomplish the above is that of corpus-based knowledge representation. In this method, semantic relations are established between different elements of a corpus of domain models, also called a domain bank. The semantic relations arise from statistics that are run over the corpus. [AlHa03] Our research suggests that this method can be enhanced if some measure of reasoning, based on statistical analyses, is added to it. This paper details the suggested Basis of Reasoning model and the additional statistics that we adopt to facilitate reasoning.

## 2 Basis of Reasoning Model

There is a distinctive link between semantic relationships among elements of the domain and the reasoning that is carried out within the domain. That is: reasoning is done using the elements and the semantic relationships that are identified between them. Statistical methods can be used to describe phenomena, or to look for patterns of relationships among two or more variables. We contribute a set of statistics to facilitate reasoning in the original model of corpus-based knowledge representation. These statistics are based on the Basis of Reasoning model. The Basis of Reasoning model is developed on the lines of the Domains of Reasoning Model [BrGa87], which elaborates on the method in which the domain should be modeled to represent it comprehensively and completely enough. The Domains of Reasoning model is built on the assumption that reasoning can be analyzed and then effected if the knowledge processes in a given domain are distinctly recognized. This model is used to categorize the components that contribute to reasoning and represent the attributes of elements and interactions within that domain. In this model, the domain is essentially considered to be a complex informal chaos. Our interpretation of Gaines' Domains of Reasoning model segregates this given domain into distinct worlds generated as under. Figure 1, below, is our interpretation of the model and has been adapted from [BrGa87].


**Figure. 1: Domains of Reasoning model [Adapted]**

The worlds are described as follows:

*Axiology [World 0]:* Within the domain, distinct knowledge processes and the elements involved therein are identified. Only useful processes that are based on utilitarian truth and pragmatic inference are identified. It is called the World of Axiology since the processes identified are ascribed a truth value.

*Epistemology [World 1]:* The elements of a domain must convey their manifestation of physical reality, i.e. which elements encompass what amount of physical reality. It is termed the World of Epistemology since the elements are characterized by their value in modeling the physical world and setting limits to the manifestation that must be represented.

*Psychology [World 2]:* The elements of the domain share relationships between one another. Some are explicit and evident, some are implicit, while others are latent or have yet to be formed. This is the World of Psychology since it is formative of the relationships that are exhibited between the elements of the domain.

*Ontology [World 3]:* The elements within the domain that are affiliated by some common underlying characteristics. These are justified in terms of criteria that determine that they belong together. They have an established nature of being and a system that seeks to find associations within them. This is called the World of Ontology, since the nature of the objects clusters them in associative groups, thereby forming an ontology for the domain.

The Worlds 1,2 and 3 must be dealt with in conjunction to one another. At the same time, the summation obtained must be given a truth value using the World of Axiology.

Our perception of the descriptions of the Worlds, as adapted from [BrGa87], is thus defined as the following:

*Axiology* - The study of the nature of values and value judgments. [Dict1]

*Epistemology* - The theory or science of the method or grounds of knowledge especially with reference to its limits and validity. [Dict2]

*Psychology* - The science that deals with mental processes and behavior. [Dict1]

*Ontology* - That department of the science of metaphysics, which investigates and explains the nature and essential properties and relations of all beings, as such, or the principles and causes of being. [Dict2]

When a domain is distinctly identified by the components that make up a world, it is easier to objectify the relationships between elements of the domain and the inferences that arise out of them. The characteristics and attributes of every identified element and its association to other elements can be indicated more methodically. Once such preliminaries are in place, reasoning can be effected by leveraging them. We have taken this division of Worlds as the basis for the spectrum of statistics that are needed for the reasoning required to enhance the base model. We call this the Basis of Reasoning Model. Figure 2, below, displays the 'Levels of Reasoning' established for the identification of the additional statistics for reasoning.

*Level 0:* This represents the domain that must be represented.

*Level 1:* The World of Axiology, now called Concept Value, stands for the goodness of a concept or the value of an object.

*Level 2A:* The World of Epistemology, now called Domain Frame, stands for nature, origin and limits of the knowledge contained in the domain.

*Level 2B:* The World of Psychology, now called Object Relationships, stands for the behavioral characteristics for objects as individuals or groups, i.e. subsets.

*Level 2C:* The World of Ontology, now called Object Nature, stands for the nature of being of the objects as individuals or groups, i.e. subsets.


**Figure 2: The Basis of Reasoning Model**

The three components of Level 2 – 2A, 2B, and 2C – work in conjunction to define the nature, origin and limits of the knowledge contained in the domain in terms of the behavioral characteristics and the nature of being of the objects, as individuals or groups, within the domain. Level 1 gives a goodness value, a weight of belief to the components of Level 2. Level 1 and Level 2 are used in conjunction to establish statistics for reasoning, such as Extraction of Roles of the individual objects or groups, Object Connectivity as cause or effect with respect to a given instance, and Object Collocations and Associativity. The information acquired from the statistics of the 3 components in level 2: Domain Frame, Object Relationships, Object Nature leads to the epistemic concept of degree of belief or confidence, which is level 1: Concept Value. This level determines whether the statistics fall in the confidence intervals of the hypothesis that is being tested, where a confidence interval of the hypothesis is the range of values of the degree of belief or confidence for which the hypothesis will be correct. These statistics, based on the model above, are the instruments to facilitate reasoning within the corpus-based representation model. The enhanced model is now able to address those applications that are based on reasoning that arises from observing a sufficient number of regularities.

## References

[AlHa03] Halevy A., Madhavan J. 2003. Corpus-Based Knowledge Representation. Proceedings of the International Joint Conference on Artificial Intelligence. Invited paper for IJCAI-03

[BrGa87] Gaines B. (1987). Logical Foundations for Knowledge Representation in Intelligent Systems. International Symposium on Methodologies for Intelligent Systems. Proceedings of the ACM SIGART international symposium on Methodologies for intelligent systems. Pgs: 366 – 380

[Dict1] The American Heritage® Dictionary of the English Language, Fourth Edition, © 2000, by Houghton Mifflin Company.

[Dict2] Webster's Revised Unabridged Dictionary, © 1996, 1998 MICRA, Inc.