Chapter XI

Natural Language Intelligences:
The Virtual or Digital Aristotle

Introduction

Of all possible intelligent NL applications and semantic artifacts, a special value is today ascribed to building the question answering systems (Q&A) with broad and wide ontological learning (Onto Query Project, 2004), classified as open-domain Q&A knowledge systems [Question Answering, From Wikipedia, 2006]. This line of research is considered as upgrading of a traditional keyword query processing in database systems, as endowing the Web search engines with answering deduction capacities. Ideally, such a general-purpose Q&A agent should be able to cover questions (matters, subjects, topics, issues, themes) from any branch of knowledge and domain of interest by giving answers to any meaningful questions, like the Digital Aristotle, "an application that will encompass much of the world's scientific knowledge and be capable of answering novel questions and advanced problem-solving" (Project Halo, 2004). The trade name of the Digital Aristotle was inspired by the scholar mostly admired for the depth and width of his perception, whose mind spread over ontology, physics, logics, epistemology, biology, zoology, medicine, psychology, literary theory, politics, and art.

The reasons why the project Halo failed may be very illustrative and instructive. To construct such a knowing-all, versatile computing entity, Paul Allen (the owner of Vulcan Inc. funding the project and the initiator of the idea of the DA) projected to reach the pinnacle of knowledge technology in several separate stages. The completed first phase was intended to test the validity of the state-of-the-art KR&R systems for the real world knowledge domains, such as inorganic chemistry.

Quite unusual, but as a positive result of three competing teams, Cycorp Inc., the SRI International, and Ontoprise GmbH, there was presented the negative findings making a whole "taxonomy of types of failures": knowledge modeling errors, non-expressive languages, faulty reasoning, poor coverage, inadequate answer explanation, poor scalability, evaluation insufficiency, meta-knowledge and meta-reasoning deficiencies (Friedland & Allen, 2004). The second 30-month phase was aimed to remove these drawbacks by hiring *world class teams* in knowledge representation and reasoning, knowledge acquisition, intelligent interfaces, natural language understanding, and system integration. As a matter of fact, the DA project is just the follow-up of past DARPA projects to develop advanced knowledge systems and tools, Knowledge Sharing Effort, High Performance Knowledge Base, and Rapid Knowledge Formation. The would-be powerful reasoning application, the DA appears as a new attempt to create a large versatile knowledge-based system, now as empowered with question answering and explanation generation capacities, with the old means, formal logic languages having little to do with the real world knowledge formulation.

Additionally, many experimental investigations in information engineering increasingly indicate that the major difficulty in making automated question answering is not in knowledge representation languages but rather in having a top ontology organizing a knowledge domain and a lexicon. In other words, the DA system should not be seen just a catchy trade name but rather as the speaking title suggesting a classical ontological pedigree deeply rooted in its progenitor's universal system of entity types. For, by general consent, the *real* Aristotle was the first who sought to create a systematic universal ontology as the study of primary kinds of being forming the highest level of things.

The project Halo lesson teaches us that that to become a general query system, or the Virtual Aristotle Machine (VAM), it should appear as an ontology-based NL question-answering system with the world knowledge and vast lexicon able to real (causal) reasoning and explanation in any domain area. Accordingly, the NL application should be designed as a universal language machine made up of knowledge bases and natural language resources organized by standard ontology and thoroughly processed by onto-logical reasoning engine for formulating novel knowledge, principles, rules, and facts in natural language forms. Ideally, its reasoning engine is planned to possess the most sought-after capacities:

- To identify and process top constructs and relationships in any knowledge area
- To resolve complex problems revealing their central causes and reasons
- To generate questions and answer explanations in natural language
- To interact with the human being exchanging information and thoughts on every lively topic
- To predict possible events, consequences, effects, or outcomes brought about by human actions or natural processes

The VAM is expected to critically contribute into transforming the Syntactic HTML Web into the Semantic Ontology Web as the next generation of the Internet's World Wide Web. Emphasizing the principal distinctions of the Virtual Aristotle from the Digital Aristotle as a traditional KR&R system, we further will study the way how to create the VAM as a know-all natural language machine with an open domain Q&A facility.