DR. MARIA VARGAS-VERA RESEARCH STATEMENT

From the 90s my primary interest was on obtaining a scientific understanding of the methods necessary in successfully building large programs and how to incorporate the results into automatic tools for large software projects. Success in this field would have a large impact on the software industry, as well as providing interesting scientific results about a process that is currently considered to be more art than science. There are various existing approaches to this problem. For example, there are specification languages that provide help in program development, however, none of these languages are easy to use, and they do not address important issues concerned with maintenance of programs under change of specifications. The strong restrictions imposed on the programmer currently mean that such methods, while important, are perhaps most useful for small but crucial programs. For large projects, Software Engineering (SE) provides methodologies for building programs. However, the methods are not automated and often seem to end up with the generation of so much paper work that there is a resistance to using them fully. Therefore, I was interested in approaches between those of formal methods and standard SE, in particular, the use of a Knowledge Based Systems (KBS) to store and use knowledge about a program in order to supply intelligent help to the programmer. The strength of such a system would relied on knowledge about the programming language and the associated programming practices used in program construction.

In my PhD work, I made significant steps towards such a system. I designed and implemented a system that automated the construction of programs by the repeated combination of simpler programs. In my system the knowledge about the program was represented by a “program history”. This history contains the initial control flow (schema) and the techniques that the user applied in the construction of the program. I was able to obtain this knowledge about the initial programs by exploiting the option of restricting the user to use of a specialized editor. This editor embodies knowledge of certain standard Prolog practices (techniques) to aid the program construction, and crucially for us we are able to record pertinent parts of the program development into the program history. This allowed us to obtain knowledge that would be very difficult to extract from just the program itself, but of course is likely to be known by the programmer. The system then used the program history knowledge in order to produce optimal combination of programs and do so with minimal supervision by the user.

During 1995-1997 my research was concerned with Fril++ an Object-oriented Fuzzy Logic Programming Language. This new programming language supports the modeling of fuzzy applications by integrating a fuzzy logic programming language with the object-oriented paradigm. Fril++ includes the concept of fuzzy objects as an explicit construction in the language. Fuzzy objects are naturally related with multiple inheritance, and so we allow both of these in the language. The multiple inheritance give us more flexibility, and in any case, the associated ambiguities are not so much of a problem in logic programming as they are in other paradigms. We also exploit the fact that information is expressed by the hierarchy itself. A Fril++ program can explicitly use the hierarchy by means of special constructions integrated into the language. Fril++

provides support for user decisions when there are several answers for a message by having a library of defuzzification methods. Finally, we concentrated on testing Fril++ by using it for real applications such as the object-oriented data browser.

Over the last ten years (2000-2010) the focus of my research has been in the use of knowledge in Natural Language Processing (ontology-based systems). In particular, the use of ontologies in Information Extraction (IE), semi-automatic populating ontologies, Ontology-Driven Question Answering Systems, Ontology Mapping and Information Retrieval. I conceived the vision and developed an Information Extraction System called MnM .
(Available for download as open source at http://kmi.open.ac.uk/projects/akt/MnM/). In this work I addressed the integration of Information Extraction (IE) and ontologies including its reasoning capabilities. In particular, using an ontology to aid the IE process, and using the IE results to help populate the ontology. The system performed IE by means of domain specific templates and the lightweight use of Natural Languages Processing techniques (NLP). The main goal of the system was to learn information from text by the use of templates and in this way to alleviate the main bottleneck in creating knowledge-base systems that is “the extraction of knowledge”. My domain of study or was news articles describing academic events. In summary, the main goal of my system was to classify an incoming article, obtain the relevant objects within the article, deduce the relationships between them, and to populate the ontology. Furthermore, the system aimed to do this with minimal help from the user. A published paper on MnM work had been widely cited in several research communities and boosted research on Information Extraction technologies (refers to Google Scholar for the number of citations).

Currently, I am working on Automated Question Answering (QA). Since Semantic web has becoming popular and also the need of services that could exploit the vast amount of information in the web. Therefore, there is a need for automated question answering systems. These kinds of systems should allow users to ask questions in everyday language and receive an answer quickly and with a context which allows user validate the answer. Current search engines can return ranked list of documents but they do not deliver answers to users. The question answering system should provide an answer in real time and it will be able to give a textual answer in a specific domain. Finally, my future research agenda comprises to build more intelligent Information Extraction (IE) Systems using knowledge encoded in ontologies and also to perform Question Answering across heterogeneous sources. In order to perform the latest we need to perform mappings between different ontological structures. Therefore, during the last six years (2005-2011), I have carried out research on ontology alignment (also called ontology mapping).