

Semantic technologies for the automation of research in biomedicine



Ross D. King

School of Computer Science, University of Manchester, Manchester, United Kingdom

Dr Ross King moved to the School of Computer Science, University of Manchester, in February 2012. Before that, he was at the University of Wales, Aberystwyth, for fifteen years. His first degree was in Microbiology, but he also has a M.Sc. and Ph.D. in Computer Science. The research achievement he is most proud of is developing Robot Scientists, a physical implementation of the task of Scientific Discovery in a microbiology laboratory, representing the merging of increasingly automated and remotely controllable laboratory equipment and knowledge discovery techniques from Artificial Intelligence.

The use of computers is changing the way that science is described and reported. Scientific knowledge is best expressed in formal logical languages with associated probabilities. Only formal languages provide sufficient semantic clarity to ensure reproducibility and the free exchange of scientific knowledge. Despite the advantages of logic, most scientific knowledge is expressed only in natural languages. This is now changing through developments such as the Semantic Web and ontologies.

A Robot Scientist is a physically implemented robotic system that applies techniques from artificial intelligence to execute cycles of automated scientific experimentation. A Robot Scientist can automatically execute cycles of hypothesis formation, selection of efficient experiments to discriminate between hypotheses, execution of experiments using laboratory automation equipment, and analysis of results. We have developed the Robot Scientists Adam (functional genomics), and Eve (drug design).

Robot Scientists provide excellent test-beds for the development of methodologies for formalising science. Using them it is possible to completely capture and digitally curate all aspects of the scientific process. We attempted to record and formalise Adam's experiments. For the core organization of this we used the ontology of scientific experiments EXPO. This ontology formalizes generic knowledge about experiments. We then developed LABORS, a customized version of EXPO. Application of LABORS produces experimental descriptions in the logic-programming language Datalog. In the course of its investigations, Adam observed 6,657,024 optical density (OD595nm) measurements (forming 26,495 growth curves). These data are held in a MySQL relational database. Use of LABORS resulted in a formalization of the scientific argumentation involving over 10,000 different research units (segments of experimental research). This has a nested treelike structure, 10 levels deep, that logically connects the experimental observations to the experimental metadata. This structure resembles the trace of a computer program and takes up 366 Mbytes. Making such experimental structures explicit renders scientific research more comprehensible, reproducible, and reusable.

My vision of the future is a collaboration between human and Robot Scientists will produce better science than either can alone, and the scientific knowledge produced will be primarily expressed in logic with associated probabilities and published using the Semantic Web.