# Towards an ontology for automatic scientific discovery *

Tezira Wanyana[1] and Deshendran Moodley[1,2]

[1] University of Cape Town, Cape Town, South Africa
[2] Center for Artificial Intelligence Research (CAIR), South Africa
{twanyana, deshen}@cs.uct.ac.za

**Abstract.** While some attempts have been made to automate the scientific discovery process in specific domains, these approaches have limited support for formal representation and reasoning about observations and phenomena. This research aims to create a generic formal ontology to support an intelligent agent for observation induced knowledge discovery.

**Keywords:** Agents · ontologies · Automatic Hypothesis Generation.



**Introduction:** One of the goals of intelligent agents is to learn and adapt to a dynamic environment. An agent typically takes in observations from its environment, identifies anomalous observations, i.e. unexpected observations, and determines whether the anomaly is indicative of a new phenomena or a change in the environment. If this is the case the agent's goal is to generate and evaluate a hypothesis as an attempt to explain the underlying causal mechanism for this phenomenon. A first step towards designing such agents is to settle on a formal language or ontology for representing and reasoning about hypotheses. In this research, we explore the requirements for such an ontology.

**Existing Approaches:** Some attempts have been made to formalize the representation of hypotheses using ontologies, e.g. the Robot Scientist[3] uses LABORS (LABoratory Ontology for Robot Scientists) and the DISK system[2] uses the DISK ontology. An attempt is made in [4] to link research statements to associated probabilities using the HELO ontology. There are other hypothesis representation models analysed in [1]. In this analysis, only the DISK ontology attempts to cater for most of the aspects except hypothesis classification which checked if a taxonomy of hypothesis statements is supported. The DISK ontology and the other ontologies are not based on phenomena-triggered hypothesis generation and hence do not represent some of the key hypothesis elements of hypothesis generation and evaluation. For example, the phenomena that triggered the hypothesis and its detection mechanism. However, some of the elements presented and lessons learned will be used to design a formal representation for hypothesis generation and evaluation.

* Supported by Center for Artificial Intelligence Research (CAIR) and Hasso-Plattner-Institut (HPI)

**Table 1.** Summary of the core elements represented in previous ontologies

| **Element** | **LABORS** | **DISK** | **HELO** |
|---|---|---|---|
| Phenomena detection mechanism | No, hypotheses are from background knowledge | No, initial hypothesis is provided by the user | No |
| Triggering phenomenon | No | Yes, in form of evidence for revised hypotheses | No |
| Hypothesis Statement Representation | Predicates | RDF Triples | Predicates |
| Hypothesis Qualifier | No | Yes(confidence level) | Yes(Probability) |
| Hypothesis appraisal mechanism and unsuccessful hypotheses | No | No | No |

**A Hypothesis Ontology; Core Requirements:** Hypotheses and their semantic meaning have to be consistently and precisely represented to aid reusability and reproducibility [1]. We suggest that the following top level elements as the core requirements for the representation: 1) The Hypothesis statement: an assertion of the explanation of the underlying causal mechanism of the phenomenon. 2) The hypothesis Qualifier: the probability value that represents the agent's belief of the extent to which the hypothesis explains the observed phenomenon. 3) Triggering Phenomenon: the phenomenon for which the hypothesis was generated. 4) The Provenance Record: This consists of the phenomenon detection mechanism, the qualifier threshold used in hypothesis selection and the hypothesis appraisal mechanism used in selecting the most plausible hypotheses. 5)Unsuccessful Hypotheses: These are the competing alternatives that are unsuccessful. Table 1 shows some of the required elements and which hypothesis representation ontology has catered for them.

**Conclusion:** In conclusion, we have presented some of the core elements towards a generic formal ontology for automatically generating hypotheses to explain new phenomena in some environment.

## References

1. Garijo, D., Gil, Y., Ratnakar, V.: The disk hypothesis ontology: Capturing hypothesis evolution for automated discovery. In: K-CAP Workshops. pp. 40–46 (2017)
2. Gil, Y., Garijo, D., Ratnakar, V., Mayani, R., Adusumilli, R., Boyce, H., Mallick, P.: Automated hypothesis testing with large scientific data repositories. In: Proceedings of the Fourth Annual Conference on Advances in Cognitive Systems (ACS). pp. 1–6 (2016)
3. King, R.D., Rowland, J., Aubrey, W., Liakata, M., Markham, M., Soldatova, L.N., Whelan, K.E., Clare, A., Young, M., Sparkes, A., et al.: The robot scientist adam. Computer **42**(8), 46–54 (2009)

4. Soldatova, L.N., Rzhetsky, A., De Grave, K., King, R.D.: Representation of probabilistic scientific knowledge. In: Journal of biomedical semantics. vol. 4, p. S7. BioMed Central (2013)