

# Cognitively Mediated Research Discovery

A context-aware rich visualized knowledge graph co-created by humans and machines using a common language

Geeth de Mel      Dave Braines, Anna Thomas  
IBM Research, UK      IBM Emerging Technology, UK

Tien Pham      Will Dron  
Army Research Laboratory, USA      Raytheon BBN, USA

**Abstract**—We believe the language of choice, and the means by which humans and machines interact, and contribute knowledge, will play a key role in the success of future cognitive systems. In earlier research, we have investigated the potential role of a hybrid human-machine language for knowledge representation to serve as the basis for a highly agile and collaborative interface between human and machine agents. Inspired by the promise of rich semantics but motivated by the desire for simplicity, practicality, and applicability - especially the ability to engage with subject matter experts (SMEs) rather than technical specialists we developed a Controlled Natural Language (CNL) named Controlled English (CE). We have then applied CE in a wide variety of problem domains from decision-making to improving situation awareness among communities of interests, especially in military coalition environments wherein information is dynamic and complex. Our recent work has been focused on integrating natural language conversational capability to the core CE, thus enabling the human users to request information and, more importantly, contribute knowledge in a fully naturalistic manner. We have published experimental results, based on a series of crowd-sourcing collective intelligence games, which show that untrained human users can converse with such a system, learning to contribute meaningful information within a short timeframe. We will present details of this language, the principles against which it was created and highlight key differences to similar languages available.

**Keywords**—Controlled English; Human-Computer Interaction; Cognitive technologies; Science Library; Visualisation.

## I. SCIENCE LIBRARY

In this talk, we wish to focus on a particular recent use case that was developed and is publicly available. Specifically, this is the Science Library application which can be accessed at <http://nis-ita.org> and <http://sl.dais-ita.org> and has been built entirely using the CE language. The purpose of the Science Library is to provide a rich information exploration environment relating to academic publications, their impact and the social networks (authors and organisations) that lie behind these works. In the talk, we will focus on the key aspects of the approach we took including:

- the development of the Controlled English model in consultation with SMEs.
- the iterative approach enabling us to start extremely quickly with a few basic concepts and relationships.
- the role of logical inference rules to augment the basic data and infer many more rich relationships in the data.
- the ease of building complex interactive visualisations of the underlying information.
- the integration of a natural language chat capability to allow the user interface to be driven by textual means.



Figure 1: Science Library - Publication details

The architecture we have developed allows both human and machine agents to contribute knowledge into the environment through the creation of CE sentences. We will explain how machine agents can be used in such an environment to contribute additional knowledge based on their capabilities e.g., by using APIs on publication data (i.e., text, images, tables, and so forth) to extract relevant knowledge such as key research concepts, author networks, and so forth. We will also explain how conceptual and topical knowledge is identified using a combination of logical inference rules and similarity measures, along with techniques that can be used to further enhance the knowledge emerging from automated analysis of the data.



Figure 2: Science Library - Author narrative timeline

The approach we have taken brings together a number of important related capabilities which sit at the intersection of human-machine interaction. The CE language is used by both human and machine agents directly, minimizing the friction between them. The language is designed to be easy to parse by machine agents (e.g., in the same manner as XML or JSON)

but easy to read by human users. In order to reduce the burden on human users to produce knowledge based on the strict syntax of CE, our recent work has focused on providing a higher-level natural language conversational interface where users can write their information in natural language and simply read the CE interpretation from the system before deciding whether to approve the knowledge for ingestion. The result of CE ingestion is a knowledge graph with rich semantics (broadly comparable to typical semantic languages such as Web Ontology Language) with additional support for logical inference rules to allow premises and conclusions to be defined. The environment also contains rationale as a result of rule execution, capturing the exact premises that led to the conclusions in each case, and this rationale is also represented in the human-friendly manner.

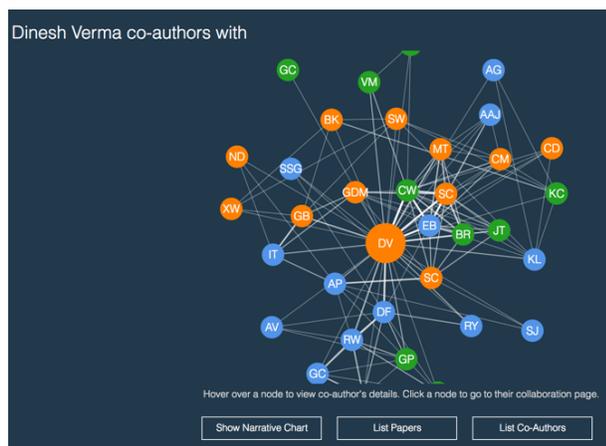


Figure 3: Science Library - Co-author social network

## II. RELATED WORK AND GETTING ACCESS

We will also discuss the applicability to other industries, technology areas and higher-level problem-solving capabilities, but the main focus will be to showcase the Science Library application and explain the approach, the solution and potential next steps.

All of the materials we cover are publically available and easily re-used. They can be found at <http://github.com/ce-store> including the main ce-store engine which is used to run the Controlled English model, rules and data. The user interface as well as the underlying data can also be found online in the same github repository.

### ACKNOWLEDGMENT

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Numbers W911NF-06-3-0001 and W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.