

# Workshop on Ontology & Rich Semantics (26 Nov 2018)

Khoo Soo Guan, Christopher (Assoc Prof)    August 31, 2018 Workshop on Ontology & Rich Semantics (26 Nov 2018)2018-11-16T20:47:27+00:00

Workshop programme is [available](#)

Online registration is [available](#) (please register by Fri 16 Nov 2018 for catering planning)

The Workshop is funded by the [Centre for Liberal Arts and Social Sciences \(CLASS\)](#), College of Humanities, Arts, & Social Sciences, Nanyang Technological University.

## Workshop on Ontology and Rich Semantics: Frameworks and Application

- Date of event: Mon, 26th Nov. 2018, 9:15am to 5:00pm



Venue of Workshop

- Venue: Wee Kim Wee School of Communication & Information, Nanyang Technological University, Singapore  
31 Nanyang Link Singapore 637718  
(Executive Seminar Room, Rm CS02-19, Level 2)
- Organisers: Dr Chris Khoo, Dr Robert B Allen, & Dr Shigeo Sugimoto

- Target audience: academics, researchers, graduate students, librarians and knowledge organization professionals—who are doing research and professional work related to knowledge organization, ontology and metadata. Singapore and international participants are welcome
- Deadline for presentation proposals: Mon, 1st Oct 2018 (over)
- Registration: Free, but places are limited and prior registration is required. [Click here to register](#) by Fri 16 Nov 2018.

## **Description**

Explicit rich representations of domain knowledge, models, processes, historical development, and argumentation are important to support e-science, e-social science and digital humanities research. Ontologies define concept hierarchies and identify important types of relations between concepts as well as entities of interest. However, to support e-research as well as real applications, ontologies need to go beyond simple semantics to include representation of complex structures and systems, dynamic models and processes, flow of temporal events, as well as discourse structure, evolution of knowledge, argumentation and information across types of documents and media.

This workshop will explore frameworks for ontological representation with rich semantics, and applications in diverse knowledge domains such as systems analysis and design, molecular biology, clinical decision support systems, linguistics, law, history, biography, etc.

*The Workshop is funded by the [Centre for Liberal Arts and Social Sciences \(CLASS\)](#), College of Humanities, Arts, & Social Sciences, Nanyang Technological University.*

## **Call for Proposals**

Presentations of completed, ongoing or proposed research/project in ontology and rich semantics applications in a particular subject or application area are invited.

Authors of accepted presentations will be invited to submit a short paper (of 6-10 pages) for publication in a special issue of LIBRES e-journal (<https://www.libres-ejournal.info/>), after a round of review and revision.

Presentation proposal submission deadline: Mon, 1st Oct 2018 (over).

Please submit to Chris Khoo at [chriskhoo@pmail.ntu.edu.sg](mailto:chriskhoo@pmail.ntu.edu.sg)

---

## Workshop on Ontology and Rich Semantics: Frameworks and Application

### **PROGRAMME**

Venue: Wee Kim Wee School of Communication & Information, Nanyang Technological University

31 Nanyang Link, Singapore 637718

(Executive Seminar Room, Rm CS02-19, Level 2)

[Click here for online registration](#)

#### **9:15am-10:00am: Session 0 – Pre-Workshop Tutorial & Registration**

Introduction to ontology and knowledge representation

*Chris Khoo, Nanyang Technological University*

Registration

#### **10:00am-12:00pm: Session 1 – Keynotes**

- Keynote 1: Rich semantic representations for mechanisms  
*Robert B Allen, New York*
- Keynote 2: Metadata models for cultural/historical resources in non-conventional domains—pop-culture, intangible/dynamic contents and disaster records  
*Shigeo Sugimoto, University of Tsukuba*
- Keynote 3: The Open Multilingual Wordnet and teaching through tagging  
*Francis Bond and Luis Morgado da Costa, Nanyang Technological University*

#### **12:00pm-1:00pm: Lunch**

#### **1:00pm-2:45pm: Session 2 – Ontology and Research Data & Text**

1. Development and visualization of an enriched ontology to support reuse of social science quantitative data sets  
*Guangyuan (Schusie) Sun, Nanyang Technological University*
  2. Modelling argument and information structure in the introduction of sociology research papers  
*Wei-Ning Cheng, Nanyang Technological University*
  3. CausalChainNet as a resource of causal chains  
*Aliaksandr Huminski (Institute of High Performance Computing, A\*STAR), and Ng*
-

---

*Yan Bin (Cognitive Human-Like Empathetic & Explainable Machine-Learning, A\*STAR)*

4. Extracting semantic relations from a graphical representation of social science abstracts

*Chris Khoo, Nanyang Technological University*

### **2:45pm-3:15pm: Tea break**

### **3:15pm-5:00pm: Session 3 – Ontology and Digital Humanities**

1. Visual knowledge aggregation—from static to dynamic information systems  
*Andrea Nanetti, Nanyang Technological University*
2. OntoSenticNet: A commonsense ontology for sentiment analysis  
*Erik Cambria, Nanyang Technological University*
3. Coordinating and integrating faceted classification with rich semantic modeling  
*Robert B. Allen (New York), and Jaihyun Park (Syracuse University)*
4. Music Scores Web Ontology: Representing information on the music scores in the Jose Maceda Collection, based on the Arc2 Framework  
*Sonia M. Pascua, University of the Philippines at Diliman*

---

## **ABSTRACTS**

### **9:15am-10:00am: Session 0 – Pre-Workshop Tutorial & Registration**

#### **Introduction to ontology and knowledge representation**

*Chris Khoo, Nanyang Technological University*

Registration

### **10:00am-12:00pm: Session 1 – Keynotes**

#### **Keynote 1: Rich semantic representations for mechanisms**

*Robert B Allen, New York*

*Abstract.* It is proposed that “direct representation” knowledge bases should be developed to organize complex information. Because the text of science research articles is generally

---

---

highly structured and unambiguous, we propose that entire research articles could be implemented with rich semantics. Similarly, a community newspaper typically presents a relatively structured view of the events in that community. Thus, we have proposed the development of rich semantic community models to support the organization of information from historical newspapers. Our semantic models build on the Basic Formal Ontology (BFO) which is widely used in biomedicine. We have explored how BFO can be extended in scope and even used as a framework for object-oriented semantic modeling. Recently, we have focused on developing rich semantic structures for representing mechanisms, which are regular sequences of activities. Mechanisms are common in descriptions of human infrastructures, and they are fundamental to scientific research reports.

---

**Keynote 2: Metadata models for cultural/historical resources in non-conventional domains—pop-culture, intangible/dynamic contents and disaster records**

*Shigeo Sugimoto, Faculty of Library, Information and Media Science, University of Tsukuba*

*Abstract.* This talk reviews metadata models for describing cultural resources in non-conventional domains such as popular culture (e.g., Manga) and intangible/dynamic cultural content. Methods for preserving cultural resources and their metadata from natural disasters will also be discussed. The models are developed in a research project at the University of Tsukuba with the aim to enhance the usability of digital archives, by connecting cultural resources on the Web and those provided by memory institutions.

---

**Keynote 3: The Open Multilingual Wordnet and teaching through tagging**

*Francis Bond and Luis Morgado da Costa, Division of Linguistics and Multilingual Studies, Nanyang Technological University*

*Abstract.* The talk introduces the Open Multilingual Wordnet, a large lexical network of words grouped into concepts and linked by typed semantic relations. I discuss how the resource has evolved over time in size and complexity, and introduce some of the latest extensions. I further introduce an ongoing effort to enrich students' learning by involving them in sense tagging using the wordnets. The main goal is to lead students to discover how we can represent meaning and where the limits of our current theories lie. A

---

---

subsidiary goal is to create sense tagged corpora and an accompanying linked lexicon (in our case wordnets). I present the results of tagging several texts in multiple languages and present some ways in which the tagging process could be improved. Finally, I discuss what the students learned from this based on their project reports. The annotated corpora are available through the NTU multilingual wordnet (NTU-MC).

**12:00pm-1:00pm: Lunch**

**1:00pm-2:45pm: Session 2 – Ontology and Research Data & Text**

**Development and visualization of an enriched ontology to support reuse of social science quantitative data sets**

*Guangyuan (Schusie) Sun, WKW School of Communication & Information, Nanyang Technological University*

*Abstract.* Many universities are embarking on research data management initiatives, including building institutional data repositories to support the preservation and reuse of research data. An increasing number of social science data sets are publicly available in national data repositories (e.g., UK Data Archive) and in multi-institutional repositories (e.g., the Inter-university Consortium for Political and Social Research). However, it is challenging for researchers to reuse such data sets, that they had not collected themselves. Substantial intellectual effort is needed to understand and evaluate data sets for possible reuse.

This study seeks to identify the challenges researchers face in reusing social science quantitative data sets, and to develop an enriched ontology and visualization method to address the challenges. The talk will describe the enriched ontology that comprises three levels: Ontology Level, Conceptual Description Level, and Physical Description Level. The Conceptual Description Level and the Physical Description Level represent the structure and content of the data set files. The Ontology Level is derived from the questionnaires used to collect the data stored in the data set files, and assigns semantics to data set elements.

To evaluate the effectiveness of the enriched ontology in supporting reuse, a graphical visualization was designed and implemented using the data visualization software Cytoscape—to visualise data set information (i.e. variables and values), related questionnaire information, information from the associated research report (i.e. research concepts and research objectives/research questions/hypotheses), and associated ontology concepts. It is hoped that an implementation of this enriched ontology and its visualization

---

---

in existing data repositories will support social scientists in exploring data sets in data repositories and evaluate them for reuse.

---

## **Modelling argument and information structure in the introduction of sociology research papers**

*Wei-Ning Cheng, WKW School of Communication & Information, Nanyang Technological University*

*Abstract.* This study seeks to model the Argument Structure and Information Structure in research papers, in the fields of sociology, mechanical engineering and bioscience. Such models can be helpful for many purposes: for the teaching of academic writing to university students, to support automatic extraction and representation of information in research papers, and to inform the development of automatic text summarization and argumentation systems.

This talk will focus on the Argument and Information Structures found in the Introduction sections of 40 sociology research articles. In the study, an argument is analyzed into an argument claim supported by one or more supporting arguments. 27 types of argument claims (e.g., research gap, research objective, research result, and research contribution) have been identified, as well as 11 types of supporting arguments. The argument structure is modelled in two ways: as an argument chain (i.e. a sequence of argument claims), and as a graphical argument structure comprising a set of argument claims and support linked by directed edges, pointing from support to claim. Commonly occurring argument patterns have been derived for different types of sociology research: Investigative research, Development/evaluation research, Descriptive research, Historical analysis and Identification research.

Analysis of the Information Structure focuses on the content of the research objectives and research results, as well as on the supporting arguments that built up the research objectives/results. Thus, the analysis not only identifies the important types of information in the research objectives/results, but also how the information structure is incrementally built up along the argument chain. In a preliminary study, common information patterns were identified and represented as semantic frames. The following four semantic frames were developed: Research-relation frame, Comparison frame, Development/evaluation frame and Descriptive frame. Analysis of the Information Structure in a research paper involves instantiating and filling the “slots” in these semantic frames, which then

---

---

represent the information structure of the arguments, and can be seen as explaining the Argument Structure.

---

### **CausalChainNet as a resource of causal chains**

*Aliaksandr Huminski (Institute of High Performance Computing, A\*STAR), and Ng Y*

*Abstract.* An ontology was developed to catalog the 23 music scores in the Jose Maceda Collection (JMC), and to represent the information using the Resource Description Framework (RDF) and linked data methodologies, for publishing in the Linked Data Cloud. The Jose Maceda Collection is located at the UP Center for Ethnomusicology Library of the College of Music, University of the Philippines. It houses a collection *an Bin (Cognitive Human-Like Empathetic & Explainable Machine-Learning, A\*STAR)*

*Abstract.* This talk will present CausalChainNet as a resource of causal chains between events. The idea of CausalChainNet construction is to extract the causal chain, and not just a causal link between two events. CausalChainNet is considered a huge set of causal chains where each chain is a sequence of events related by causality: event-1 causes event-2 causes event-3 etc.

Traditional explicit syntactic patterns for the detection and extraction of causal relations like causal links, resultative constructions, conditionals, etc. cannot be used directly since they are focused on 2-member causal extraction. We need constructions where an effect in cause-effect relation can be taken, in turn, as a cause for the next step in chain causality. The following 2 linguistic patterns are used for causal link extraction:

V1+NP1+to/for+V2+NP2 (*stabbed the guy to kill him*)

V2+NP2+by+V1[ing]+NP1 (*kill the guy by stabbing him*)

where V1 is a verb representing a cause-event, and V2 is a verb representing the effect-event. NP1 and NP2 are noun phrases for objects representation. V2 can in turn represent a cause-event in the next step of the causality chain. The chain ends if the only pattern that can be applied contains an adjective to represent effect: V NP to become/be Adj (*buy a house to become independent*).

We are planning to use English Wikipedia to be applied on the patterns. CausalChainNet can be used in developing commonsense knowledge and reasoning resources. In particular, it can help when event prediction is needed, as in the case of script construction.

---

---

## **Extracting semantic relations from a graphical representation of social science abstracts**

*Chris Khoo, WKW School of Communication & Information, Nanyang Technological University*

*Abstract.* The study seeks to develop an information extraction and summarization method to extract useful information related to a topic (including research results) from a set of research abstracts, retrieved from a digital library or indexing/abstracting database. The method uses graphical pattern matching to identify semantic relations expressed in the text, and extracts the concepts that are linked to the topic concept by important semantic relations (e.g., the cause-effect relation). As terms and their synonyms that appear across multiple abstracts are mapped to the same concept, the method performs a kind of summarization to identify concepts that are frequently linked with the same semantic relation to the topic concept.

The approach taken is to parse the text into a dependency tree representation that represents the syntactic structure of the sentences. The resulting word-syntactic relation-word triples are imported into a graph database software, Neo4j (<https://neo4j.com/>), to perform graphical pattern matching to extract words that are commonly linked syntactically to the topic term, to infer semantic relations, and to generate a summary semantic representation by merging common graphical structures across sentences and documents.

The current focus of the study is on the types of graphical patterns that are useful for extracting important semantic relations. We shall also analyzing what kind of summary or overview of the topic can be obtained by applying this method to 500 abstracts related to the topic. The summary semantic representation derived can be used in the construction of an ontology for the topic.

**2:45pm-3:15pm: Tea break**

**3:15pm-5:00pm: Session 3 – Ontology and Digital Humanities**

## **Visual knowledge aggregation—from static to dynamic information systems**

*Andrea Nanetti, School of Art, Design and Media, Nanyang Technological University*

*Abstract.* This talk will present how Renaissance Italy world maps have been used to distill ontologies and design computer applications to visualise the information encoded in

---

---

historical cartography. Two case studies will be presented of how a theoretical approach and methods were developed to decode information from the 1457 World Map (Florence, Biblioteca Nazionale Centrale, Portolano 1) and the Fra Mauro World Map (Venice, Biblioteca Nazionale Marciana), and automatically link them to other historical as well as contemporary information via digital technologies. These methods can be applied as visual knowledge aggregators that will be helpful in some digital humanities research.

---

### **OntoSenticNet: A commonsense ontology for sentiment analysis**

*Erik Cambria, School of Computer Science & Engineering, Nanyang Technological University*

*Abstract.* OntoSenticNet is a commonsense ontology for sentiment analysis based on SenticNet, a semantic network of 100,000 concepts based on conceptual primitives. The key characteristics of OntoSenticNet are: (i) the definition of precise conceptual hierarchy and properties associating concepts and sentiment values; (ii) the support for connecting external information (e.g., word embedding, domain information, and different polarity representations) to each individual defined within the ontology; and (iii) the capability of associating each concept with annotations contained in external resources (e.g., documents and multimodal resources).

---

### **Coordinating and integrating faceted classification with rich semantic modeling**

*Robert B. Allen (New York), and Jaihyun Park (Syracuse University)*

*Abstract.* Faceted classifications define dimensions for the types of entities described. In effect, the facets provide an “ontological commitment”. We compare a faceted thesaurus, the Art and Architecture Thesaurus (AAT), with ontologies derived from the Basic Formal Ontology (BFO2), which is an upper (or formal) ontology widely used to describe entities in biomedicine. We consider how the AAT and BFO2-based ontologies could be coordinated and integrated into a Human Activity and Infrastructure Foundry (HAIF). To extend the AAT to enable this coordination and integration, we describe how a wider range of relationships among its terms could be introduced. Using these extensions, we explore richer modeling of topics from AAT that deal with Technology. Finally, we

---

---

consider how ontology-based frames and semantic role frames can be integrated to make rich semantic statements about changes in the world.

---

**Music Scores Web Ontology: Representing information on the music scores in the Jose Maceda Collection, based on the Arc2 Framework**

*Sonia M. Pascua, School of Library and Information Studies, University of the Philippines at Diliman* of sound documents comprising around 2,000 hours of recordings on reel-to-reel tapes and cassettes. This collection is the main holding of the UP Center for Ethnomusicology, founded in 1997 by José Maceda himself. Due to its universal significance, the collection was inscribed on the International Register of the Memory of the World Programme in 2007.

The ontology, called the Music Scores Web Ontology (MSWO), was developed by extending the Music Ontology as well as some of its extensions. The Music Ontology, an online community effort, provides a model for representing structured music-related data, starting from basic topics and building them into a larger framework. This allows other ontologies to be plugged on top of the Music Ontology namespaces and referenced as Music Ontology Modules. The ontology enables representation of data in “triples” through the use of RDF, Simple Protocol and RDF Query Language (SPARQL) for databasing using the Arc2 Framework, a PHP semantic tool.

---