Semantic applications for ontological information theory

Amany K. Alnahdi
Computer Science Department, King Abdulaziz University, Jeddah, Saudi Arabia
E-mail: akalnahdi@kau.edu.sa

Abstract. Information theory is a field that has applications in several scientific domains. This field, which was initially proposed by Claude Shannon, is interdisciplinary; it combines probability, electrical engineering, information engineering, computer science, and statistics. Representing knowledge in information theory provides scientific scholars a resourceful reference for the filed. By learning the precise terminology used in information theory, researchers can discover documents in the domain. In addition, they can communicate knowledge and ideas by using the accurate jargon. A scientific ontology representing knowledge corpus can be implemented digitally by articulating the concepts in the domain of information theory using semantic graphs. These scientific semantic artifacts are valuable with Web 3 technologies. They facilitate searching documents online by providing precise related results. In addition, it is expected that the semantic knowledge to be applicable in Web service applications based on information theory as Web services are software components that communicate information. This work presents an ontology for the Information theory. The ontology conceptualizes the knowledge of information domain by identifying the main concepts, principles, equations, metrics, and history of information theory. The ontology composed of classes, instances, and properties is built using protege, an ontology editor, and is stored in OWL format. Educational applications for the built ontology are indicated.

1. Introduction
Information theory is a legacy field that have contemporary applications such as physics, electrical engineering, text analysis, economics, and computing. The usefulness of this field makes it beneficial to build resources that formulate the concepts in the domain and to define the hierarchical representations of these concepts. These hierarchal representations organize knowledge.

Semantic representations can support the Web with knowledge structures. In addition, semantic components have effect on many Web service applications. Retrieving information based on semantic infrastructure has impact on the accuracy of the search results. Therefore, using metadata can improve computing applications by supporting the syntactic layer. Semantic applications include educational applications in computing. Semantic computation surpasses syntactic computing where communicating between software modules is effective and based on meaningful contextual information. In addition, describing Web services semantically would make it accessible in Universal Description, Discovery, and Integration (UDDI) registries.

This research work proposes ontology for information theory and presents its semantic applications in the domain. This paper is organized as follows. Section 2 presents some of the related work. Section 3 explains the method for building the ontology and presents a case study for evaluating the ontology and discussing its semantic applications in different domains. Section 4 presents the discussion for the case study results. Section 5 concludes the paper and indicates some of future research directions.
2. Related work
In [1], the author addressed the difficulty of incorporating semantics in the information theory due to natural language and proposed theory of semantics. The author of [2] addressed the beneficial use of semantics in the information theory by analyzing information theory. In [3], the author indicated the comprehensive dimensionality of information namely syntactic, semantic, and pragmatic. The author of [4] studied the use of semantic information methods in machine learning algorithms, while in [5] the use for semantic information for resolving multi-labeling problem in machine learning was introduced.

Semantics and context exchange gained interest in smart cities applications. In [6], authors presented a document communication framework in smart cities for global information transformation based on semantic concepts. In [7], authors presented an approach for cross-context tabular documents for semantic cross-domain interactions. Authors of [8] proposed an approach for Machine Natural Language (MNL) smart contracts that provides agreement based on context. In [9], authors indicated that role of artificial intelligence that will play in processing abundant data transferred in smart cities.

3. Methodology
This section explains the process of building the ontology and presents a case study for evaluating and discussing the semantic applications of the built Information Theory Ontology. Computerized ontologies are artifacts that conceptualize knowledge in domains in a digital file format. These semantic artifacts present the knowledge in a graph form rather than structured tabular files. The flexibility of graph presentation combines the applicability of database and also provides flexibility of representing and querying data in standard format.

3.1. Ontology building
To build the ontology, proté gé [10], an ontology editor, was used. The concepts of the domain of information theory were enumerated, classified, and edited in a computational ontological form. Main concepts such as principles, equations used for quantifying information, and metrics are included in the built ontology. The ontology is demonstrated in 3 forms. As can be seen in Figure 1, XML tags describe the ontology classes. Class hierarchical form of the ontology using proté gé editor is shown in Figure 2. Figure 3 depicts the interactive visualization of the ontology using WebVOWL [11]. The resultant ontology consists of 55 classes.

Figure 1. Information theory ontology as XML tags.
Figure 2. Information theory ontology in protégé.
3.2. Case study

To evaluate the ontology, interviews with academic educators and scholars in scientific fields were conducted. As the ontology term is philosophical and is not common to scientific researchers and digitized ontologies are new to some users, conducting qualitative interview provided a dialogue evaluation by browsing the ontology through online video conference sessions. The interviews started by an introduction explaining the built information theory ontology. The ontology was presented in both protégé [10] editor and the visualization library WebVOWL [11] during the interview. Dialogue interviews were progressed by prompting the interviewee to give their feedback about the ontology comprehensiveness and classification accuracy.

In addition, the interview incorporated a valuable discussion about possible semantic applicability in the different scientific domains. Participants’ demographic information such as specialty and education level is described and the rating for the ontology is reported as can be seen in Table 1.

Figure 3. Interactive visualization for the information theory ontology using WebVOWL.
Table 1. Information theory ontology evaluation responses.

| Participant | Demographic Information | Interview Time (min) | Ontology Evaluation |
|-------------|-------------------------|----------------------|---------------------|
|             | Major                   | Education Level      | Comprehensiveness [1-5] | Classification Accuracy [1-5] |
| P1          | Biochemistry            | Masters              | 15                  | 5                     | 5                     |
| P2          | Physics                 | PhD                  | 10                  | 4                     | 4                     |
| P3          | Business Administration | Bachelor             | 10                  | 3                     | 4                     |
| P4          | Statistics              | Bachelor             | 10                  | 5                     | 5                     |
| P5          | Business Administration | Bachelor             | 20                  | 5                     | 4                     |
| P6          | Information technology  | Bachelor             | 20                  | 4                     | 4                     |

| Ontology Evaluation Average Score | 4.33 | 4.33 |

Qualitative participants’ feedback about the ontology and its educational applicability of the ontology were reported as follows. The ontology consists of numerous classes and it has detailed classes and examples as it presents intersections with other scientific fields. In the bio-informatics domain there are educational computer programs for drawing chemical structures. In pure chemistry and formal dynamics, equations compute quantified values. The entropy concept is used in biochemistry as it computes the randomness of a specific molecule to expect the interaction between several molecules. In addition, there are quantified values that are used to estimate 3D sizes and concentration for chemical component to estimate their measures in the space. Many quantifiers and equations are well defined to describe chemical concepts such as protons and electrons. The proposed ontology is generic and provides broad information about information theory domain.

The ontology consists of good classification. There should be application class to demonstrate the applicability of information theory. Many education practices are focusing reinforcement for applications of theory and training in the educational curriculum. These applications are used in pure sciences and theory. In the physics field, equations are well defined.

Such ontology is good for academia as it can work as scientific dictionary. It is good to know the synonyms of a term. Also, it is beneficial to use the appropriate and accurate term in the context. Naive translations can cause confusion in using terms in a specific context. A student in any major has to know what concepts are commonly used in a specific domain. It is better to present these concepts available in a digital form instead of using a paper dictionary or even finding these concepts within articles or books. Using interactive visualization and images for the summarized form of the ontology makes it easier to learn and memorize the knowledge in the domain.

There are many statistical applications as in the statistics domain that quantifies information to compute the probability of occurrence, success, or failure of an event. During the COVID-19 pandemic time, the statistical health indicators are crucial and are used in hospitals. Depicting information in a knowledge graphical form facilitates the process of remembering facts as some learning styles are based on visualizations. Creating software programs in the domain of information theory would be of beneficial use.

Information theory would help academic researchers whose research is based on information to quantify those concepts. Statistical figures in economical domain are essential. The built information theory provides a classified presentation for concepts; this organized structure instead of random acquisition of information will affect in presenting research in sequenced and organized matter. In addition, a researcher can precisely know the branches of each knowledge field. Ontologies provide
concise knowledge acquisition because searching on the Internet can cause confusion due to many non-relevant results to search query.

4. Discussion
Domain information theory ontology that conceptualizes knowledge in information theory specific domains is expected to be of interest to researchers in specialized fields. Information theory has several applications in different fields. For example, entropy concept is common in chemistry, physics, and computing. The publicity of such foundational concepts has an indication that other information concepts might have definitions and applicability in other fields.

As quantifying information provides a mean for decision making and classification, conceptualizing knowledge of information theory is expected to be beneficial for researchers when reading, searching, and finding documents in the domain. Quantifying information provides mean for reducing the uncertainty about the value of a specific concept. This measurement aspect enables comparing values. In addition, having the ontology articulated in a digitized form is mandatory in the education process for the digital erra specifically during pandemic time where education and research have shifted to an online theme.

Semantic artifacts available in graphical forms provide flexibility in expressing domain knowledge. This flexibility enables scalability of representing information by adding more concepts and new concepts. In addition, formulating domain ontologies provides standardization for information communication which in turn would expedite information sharing. Expediting information sharing has its impact in knowledge production in research.

The built Information Theory Ontology has its applications in scientific fields such as statistics, physics, economy, and biochemistry. While some of the legacy equations in scientific fields are well defined and formulated, new notional aspects can also be formulated in numerical form by thoughtful consideration to the theories in the domain of information theory. Enthusiastic researchers who are willing to discover new scientific facts can associate their research theories by measurement techniques which have solid foundation in the information theory.

Using the ontology has also a beneficial pedagogical application in teaching information theory. This reference resource is constructive for both instructors and educators. Instructors who read and review the ontology are enlightened by a classified form of knowledge in the domain of information theory which will help in elevating the linguistic aspects in the education process. On the other hand, educators can learn the precise terms in the domain by browsing the ontology. This learning process is expected to raise knowledge acquisition by facilitating the reading and the comprehension of educational text materials.

5. Conclusion and future research directions
Semantic applications for Information theory are of interest to scientific research community. This work proposes Information Theory Ontology. The built ontology conceptualizes knowledge in the domain of information theory. The ontology is built in a Web Ontology Language (OWL) format, a standard form of data presentation which enables processing, sharing, and communicating data in a graph form in the Web. The built ontology is considered a knowledge reference for concepts in the domain. The ontology is extensible and can be used to annotate semantic documents and services in the domain. The open-world assumption which used for knowledge representation in semantic artifacts signifies the possibilities for discovering, creating, and incorporating new concepts to the presented ontology.

In addition, this work discusses semantic applications for the Information Theory Ontology. Merging semantic philosophical domain with quantifying domain is expected to have valuable effect in scientific research. Quantifying semantic information reduces the uncertainty of natural language components which is expected to have a valuable impact in several scientific domains including text analysis. Moreover, new concepts that do not have quantified equations to explain them can be formulated by contemplating the ontology. Additionally, visualizing ontology is of interest to visual learners which provides summarized pictorial form of the knowledge in the domain.
Representing the ontology in a standardized form namely OWL enables the using this ontology by different semantic software applications based on processing ontology data. Also, it can be used for annotating documents which helps in retrieving relevant documents by semantic search. Having the ontology in a digital format increases the accessibility for this semantic artifact and documents annotated by the ontology. The extensibility feature of the ontology enables creating detailed domain ontologies and adding new concepts.

Acknowledgments
I would like to express my gratitude to the ontology evaluators for their time and their participation. Their feedback enlightened semantic applications of using the built Information Theory Ontology and using the ontology in several scientific fields.

“This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institute of Health.”

References
[1] Zadeh LA. A key issue of semantics of information. In: 2016 IEEE 15th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC). IEEE; 2016. p. 1–1.
[2] Isaac AM. The semantics latent in Shannon information. The British Journal for the Philosophy of Science. 2019;70(1):103–125.
[3] Zhong Y. A theory of semantic information. China communications. 2017;14(1):1–17.
[4] Lu C. From Shannon’s Channel to Semantic Channel via New Bayes’ Formulas for Machine Learning. arXiv preprint arXiv:180308979. 2018;
[5] Lu C. Semantic information G theory and Logical Bayesian Inference for machine learning. Information. 2019;10(8):261.
[6] Qin P, Guo J. A novel machine natural language mediation for semantic document exchange in smart city. Future Generation Computer Systems. 2020;102:810–826.
[7] Yang S, Wei R. Semantic Interoperability Through a Novel Cross-Context Tabular Document Representation Approach for Smart Cities. IEEE Access. 2020;8:70676–70692.
[8] Qin P, Guo J, Shen B, Hu Q. Towards Self-automatable and Unambiguous Smart Contracts: Machine Natural Language. In: International Conference on e-Business Engineering. Springer; 2019. p. 479–491.
[9] Sanchez-Anguix V, Chao K-M, Novais P, Boissier O, Julian V. Social and intelligent applications for future cities: Current advances. Elsevier; 2020..
[10] Musen MA. The protégé project: a look back and a look forward. AI matters. 2015;1(4):4–12.
[11] WebVOWL [Internet]. Available from: http://vowl.visualdataweb.org/webvowl.html