Ontology for Research in Artificial Intelligence

Abebe Teferi
HiLCoE, Computer Science Programme, Ethiopia
abebeteferi.teferi@gmail.com

Tibebe Beshah
HiLCoE, Ethiopia
School of Information Science, Addis Ababa University, Ethiopia
tibebe.beshah@gmail.com

Abstract

An ontology is a vocabulary that describes the concepts and relationships in a particular domain serving as a knowledge representation. As there is no ontology for research in Artificial Intelligence (AI), this paper attempts to create an ontology for research in AI using OWL based Protégé 4, and then implements the ontology by building a website using HTML, CSS and JavaScript for the front end and PHP and MySQL for the backend. Input from experts is integrated into the ontology. The ontology has the research areas under AI as subclasses of the class "Research in Artificial Intelligence" and other fields that are related to AI as classes. These classes of research area have further subclasses such as "Software", "Tools," "Models," "Algorithms," "Data", "Hardware" and "Papers" relevant to the research areas. A class like "Place" where research in AI is conducted is included in the ontology. Most of these classes and subclasses are annotated with an explanation for the class. Object Properties and Data Properties are used to create relationships and axioms.

The website is designed to help those conducting research in Artificial Intelligence and is based on the ontology in that the classes and subclasses are menus and links on the website that lead to their own page. A user can register and then log in and update the website based on the ontology.

The ontology is evaluated and it well describes the domain of research in AI. The ontology can be imported for interoperability to develop and extend the ontology, for instance, to an ontology of AI, instead of just an ontology for research in AI, where education, jobs in AI and the like would be included in the ontology.

Keywords: Artificial Intelligence; Machine Learning; Computer Vision; Information Retrieval; Intelligent Interaction; Natural Language Processing

1. Introduction

Artificial Intelligence is the intelligence of machines. In Computer Science, an "intelligent" machine is an agent capable of reasoning that perceives its environment and takes actions that maximize its chances of success for some goal [2]. Artificial intelligence is a science of intelligence as it involves testable hypothesis underlying processes that are necessary for intelligent behavior. It is the study of mimicking perception, cognition, and action through using machines.

The main goal of AI research is to increase our understanding of perceptual, reasoning, learning, linguistic and creative methods. This understanding is no doubt helpful in the design and construction of useful tools in science, industry, and culture. The major problems of AI research include reasoning, knowledge, planning, learning, natural language processing, perception and the ability to move and manipulate objects. General intelligence is among the field's long-term goals. Approaches include statistical methods, computational intelligence, soft computing (e.g., machine learning), and traditional symbolic AI. Numerous tools are used in AI, including versions of search and mathematical optimization, logic, methods based on probability and economics. The AI field is a combination of Computer Science, Mathematics, Psychology, Linguistics, Philosophy, Neuroscience and Artificial Psychology [2].

In Computer Science, an ontology is a formal naming and definition of the types, properties, and

interrelationships of the entities that exist in a particular domain. An ontology classifies the variables needed for some set of computations and creates the relationships between them. The field of artificial intelligence creates ontologies to limit complication and to categorize information. The ontology can then be used for problem solving [1].

Ontologies have been used to develop a common understanding in the field of medicine, e-tourism, construction, plans and planning information, workflow and management system endeavors, for defining and specifying products, for retrieving objects and invoking operations on objects across a network. However, to the best of our knowledge, there is no ontology developed for the field of research in Artificial Intelligence. As a result, there is no single agreed upon point of reference for research in AI, where most can agree upon as a set of concepts and relationships in the field, which can be used to develop software or websites for research in AI.

AI, as an emerging field that aims at making technological related products more intelligent and more useful to humans, would need a site of its own to aid those conducting research and to find all the information needed for those interested in the field. Particularly, a specific site for research in this area is relevant to facilitate research by providing all the tools needed in one site. The website built for this paper is constructed based on the ontology and helps those planning to conduct research in the field of AI, mainly serving as a place where they can access research materials, such as papers, books, tools, software, models, data, algorithms, and hardware in a specific research area and even guide them in choosing a topic by providing them with a list of current and future research topics or areas. Therefore, by first developing an ontology for research in the field, the ontology is implemented by building a site for the field.

This research is a combination of quantitative and design science research. As mainly a design science research, after identifying a gap on previous literature, solution design is presented. The quantitative part of the research comes in as primarily the focus is on getting input from experts in the field in Ethiopia on the applicability of the ontology for research in AI through interviews and questionnaires and the input of experts by posting the ontology on the Web-Protégé website at Stanford University where users can edit the ontology and provide their comments.

Then an ontology for research in AI is proposed and implemented by constructing a web based system using WampSever 5, particularly HTML, CSS3, JavaScript for the front end and MySQL and PHP 5 to connect to the database and store and access data from the database. Finally, the developed ontology is evaluated by experts. The ontology can be updated by users on the website and on the Stanford University website, and hence the ontology goes through a constant wave of updating.

2. Related Work

There are quite a few papers on ontologies, although none on the domain of research in Artificial Intelligence. One of the papers, for example, is Cardoso's paper that is about developing an owl ontology for e-tourism. The paper uses OWL Full to develop the e-tourism ontology [3].

The ontology consists of concepts relations, instances, and axioms. According to the paper, "a concept represents a set or class of entities within the tourism domain". In the e-tourism ontology, "what," "where," "when," and "tourist" are the main classes. The class "Squash," "Paintball" and "Golf" are cited as examples of subclasses of the class "What." The author applies "complementOF" element to imply that tourists from European Union are not tourists from the non-European Union countries. The author describes senior tourists by taking the intersection of the class of tourism with the anonymous class of people that are senior citizens.

As far as object property is concerned, "hasActivity" relates the class "Where" with the class "What". As far as data property is concerned, the year a tourist was born is specified using a positive

integer value. Further, using cardinality restriction, the property #visitLocal of the class SiteSeeingPackage may have a maximum cardinality of 10 because it is assumed that a site seeing package ought not to have more than ten places to visit.

Machado's [4] work is the representation of domain knowledge in ontologies that can be taken advantage of to improve the current knowledge about a disease and to smooth the progress of the diagnosis and prognosis processes. The method the author used to validate is to first create a semantic model that stands for and integrates the heterogeneous sources of data necessary for the representation of a disease and its prognosis process through taking advantage of semantic web technologies and existing ontologies. Secondly, it is to develop a methodology that utilizes the knowledge represented in existing ontologies to ameliorate the results of knowledge exploration methods obtained with translational medicine datasets. The contribution of the paper is first a methodology for the creation of a semantic model in the OWL language portraying a genetic disease. This methodology is composed by numerous steps such as the identification of the data to represent, the definition of the hierarchical relations between the data elements, and the labeling of external resources to use. Secondly, a semantic model of the disease hypertrophic cardiomyophathy (used as case study) is developed. No public semantic representation existed before this paper.

The above ontologies are on e-tourism and medicine. There are other similar ontologies, but none in the domain of research in artificial intelligence. This paper bridges the gap by proposing an ontology for the domain of research in AI.

3. The Proposed Solution

In this Section, an ontology for research in Artificial Intelligence is proposed. The ontology is drawn upon literature mainly after studying an ontology of Pizzas by Matthew Horridge in his tutorial [5] and an ontology of a family in a tutorial for Protégé [6]. Protégé is an ontology development tool, based on OWL2, and has been used to build the current ontology. The classes are drawn from reviewing the information on the websites of various Artificial Intelligence research institutions and universities such as Microsoft, Facebook, Machine Intelligence Research Institute, Massachusetts Institute of Technology, University of Washington, and University of Michigan.

The properties, the relationships, the axioms are intuitively designed, again by drawing upon the ontology of pizzas and that of a family.

3.1 The Ontology

3.1.1 Domain and Scope of the Ontology

The domain of the ontology is research in Artificial Intelligence. Artificial Intelligence is the field of computer science that deals with expert reasoning to make machines as intelligent as humans. Machines are usually fed programs according to which they operate. But with AI, programs and machines will be able to reason, make their own decisions and act intelligently to be more useful to humans.

3.1.2 Important Terms in the Ontology

Important terms in the ontology, as compiled from the websites of institutions in Artificial Intelligence are Machine Learning, Expert Systems, Robotics, Natural Language Processing, Computational Neuroscience, Computational Biology, Computer Vision, Speech and Text Recognition, Fuzzy logic, Intelligent Interaction, Brain-Computer Interfacing, Neural Networks, and Automated Planning and Control.

3.1.3 Defining the Classes and Class Hierarchy

At this level, the top-down approach is used where the most general concepts and classes are defined first and the subclasses are defined afterwards.

Main classes:

- Places (places where research in AI is being conducted)
- Research in Artificial Intelligence

Subclasses:

- Current Advances in Research
- Areas of Future Research
- Automated Planning and Control [7]
  - Robotic control
    - papers
  - Use of planning and control models in computational neuroscience
    - papers
  - Automated planning under uncertainty
    - papers
- Brain-Computer Interfacing [7]
  - Computational Neuroscience
    - papers
  - Computational Biology
    - papers
- Computer Vision [7, 8]
  - Object Recognition
    - papers
  - Scene Understanding
    - papers
  - Human Activity Recognition
    - papers
  - Active Vision
    - papers
  - Grouping and Figure-Ground
    - papers
  - Visual Data Mining
    - papers
- Ethics
  - papers
- Information Retrieval [15]
  - Collaborative filtering
    - papers
  - Information extraction
    - papers
  - Image and video search
    - papers
  - Intelligent information systems
    - papers
- Intelligent Interaction [7]
  - Activity recognition
    - papers
  - Applications of machine learning and preference elicitation in UIs
    - papers
  - Crowd-sourcing, community sourcing and mixed AI/human computation
    - papers
  - Decision-theoretic interfaces and agent interaction
    - papers
  - Interactive machine learning and active learning
    - papers
  - Programming by demonstration
    - papers
  - Smart sensing and embedded devices
    - papers
- Machine Learning [7, 8]
  - Knowledge representation and reasoning
    - First order probabilistic logics
      - papers
    - Symbolic algebra
      - papers
  - Expert Systems
    - papers
  - Fuzzy logic
    - papers
  - Learning and Probabilistic Inference
    - Graphical models
      - papers
    - Kernel methods
      - papers
    - Nonparametric Bayesian methods
      - papers
    - Reinforcement learning
      - papers
    - Problem solving, decisions, and games
      - papers
- Natural Language Processing [15]
  - Speech and Text Recognition

    - papers
  - Parsing
    - papers
  - Information Extraction
    - papers
  - Machine Translation
    - papers
  - Context Modeling
    - papers
  - Dialog Systems
    - papers
- Robotics: subclasses [7, 8]
  - Mechanism design, sensors
    - papers
  - Computer vision,
    - papers
  - Robot learning
    - papers
  - Bayesian state estimation
    - papers
  - Control theory
    - papers
  - Numerical optimization
    - papers
  - Biomechanics
    - papers
  - Neural control of movement
    - papers
  - Computational neuroscience
    - papers
  - Brain-machine interfaces
    - papers
  - Natural language instruction
    - papers
  - Physics-based animation
    - papers

All the above classes are annotated with descriptions and have individual members.

3.1.4 Defining the Properties of the Classes

Based on the Horridge's pizza ontology, besides to classes, an ontology needs to have object and data properties and these properties are used to form restrictions and axioms that apply to the classes [5].

a. Object Properties

Object properties are used to link classes and as a result the individuals of the classes [5]. The following are the revised object properties in the ontology:

- Conducts_A_Research: This property would link a place to the type of artificial intelligence research conducted. Its domain is Place and Range Research in Artificial Intelligence.
- Has Papers: This would link a place like MIT - Massachusetts Institute of Technology - to the papers in AI it has. This property is a subclass of Conduct_A_Research. If a place conducts a research in some AI topic/s then it should have papers in the field. Its domain is Place and Range Research in Artificial Intelligence.
- Is_Paper_In: This would link a paper to a research area. Its domain is All Papers and Range Research in Artificial Intelligence.
- Research_Conducted_At: This would link a research area to the place it is conducted at and is the inverse of the object Property Conducts_A_Research. Its domain is Research in Artificial Intelligence and range is Place.

b. Data Property

Data properties are used to link individuals to certain data values [5]. One data property is Level_Of_Advance _in_Research. This is to indicate whether the members of the class "Current Advances in Research," are "High Level Research" (if an integer number above or equal to 5 is assigned to the research) or "Low Level Research" (if an integer number below 5 is assigned to the research).

"High_Level_Research" and "Low_Level_ Research" are defined classes serving as axioms, a current advanced research being categorized as one of these depending on the integer value it is assigned.

The goal of this data property is to classify the current advances as being really advanced, or a benchmark research that has a big impact in the field of AI or an advance that is not quite as revolutionary based on the opinions of the experts.

3.1.5 Defining the Facets of the Properties

As a subclass of the class Research in Artificial Intelligence is an existential restriction "Conducts_A_Research some Research in Artificial Intelligence" meaning all the research conducted at the Place, like Google, Microsoft, Facebook, MIT fall under one of the subclasses of Research in Artificial Intelligence. As a subclass of Place, there are two existential restrictions:

- "Has_Papers some Place" meaning A place where research in AI is done should have some papers in the field of artificial intelligence.
- "Research Conducted At min 1 Place": A research is conducted at a minimum of one place.

3.1.6 Data Collection and Analysis

In order to review the ontology developed and to get input from others, especially from scientists, a questionnaire consisting of nine questions was distributed to five experts in the field from Addis Ababa University and from HiLCoE School of Computer Science and Technology. The questionnaires consisted of whether the scientists agree with the proposed classes for research in artificial intelligence ontology, with the object, and data properties and if they have any additional suggestions and comments.

The answers to the questions from the five experts were not all the same but they have some aspects in common. They all commented that the class "Ethics" should be redefined. Some suggested additional classes and some mentioned that related fields to research areas to artificial intelligence should be included with their relationship to research in AI. Some proposed additional object property, "Is_Related_To," to show the relationships of other fields to AI, and commented a data property that helps classify researches as theoretical or advanced should be included. All their suggestions have been used to modify and extend the previously designed ontology.

The related fields to research in AI, drawn from literature, are Business Intelligence, Neurology, Cognitive Science, Data Mining, Linguistics, and Mathematics. The class "Ethics" has been redefined as "Ethics of Artificial Intelligence." Additional subclasses to the class "Research in Artificial Intelligence" and additional members have been added. The subclasses "Software," "Tools," "Models," "Algorithms," "Hardware", and "Data" have been added under each subclass of each research area in AI to include relevant software, tools, models, etc. Further, the class Places has been modified with additional subclasses such as "Address", "People," "Projects," "Publications," "Research Areas." Additional classes such as "Big Data Analytics," "Decision Support Systems," "Cognitive Architectures and Computational Cognitive Science", "Assistive Technology," "Multi-Agent Economic Systems", "Books," "Papers" subclasses under each major field to categorize the papers written in that field, "All_papers_in_Research_Artificial_Intelligence" which has as its subclasses the papers category of each field.

In addition, there is an additional object property "Is-Related-To" that relates the additional fields such as Business Intelligence, Neurology, Cognitive Science, Data Mining, Linguistics and Mathematics to the research areas in AI. For example, brain computer interface is labeled as "Is_Related_To some Neurology" while "Information Retrieval" is labeled as "Is_Related_To some Data Mining."

The object property "Has Papers" is no more a subclass of Conduct_A_Research. It stands on its own and it would link a place to the papers in AI as before. If a place conducts a research in some AI topic/s then it should have papers in the field.

One of the existential restrictions for the class "Place" has been changed to Conducts A Research in some research in artificial intelligence, meaning a place, like Massachusetts Institute of Technology, Google, or Microsoft, has to conduct some research in artificial intelligence to be included in the ontology under the class "Place." The existential restriction on "Research in Artificial Intelligence" is changed to "Research Conducted at some place" meaning the subclasses or the research areas are conducted at some place.

An additional data property "Level of Theory" is also added based on which the defined classes "Theoretical" and "Applied" are created. These classes gain members based on the integer value assigned to "Level of Theory" for the individual of the class "Current Advances in Research", an integer value above five labeling the individual as a member of the defined class "Theoretical", and an integer value below four labeling the individual as a member of the defined class "Applied".

3.2 Proof of Concept

A website is developed using HTML, CSS3, JavaScript for the front end and PHP and MySQL for the backend with a title "Research in Artificial Intelligence". It is aimed for those planning to conduct a research in the field. It has seven navigating links at the top right, below the title of the web page: Home, Places where Research in AI is Conducted, Future Research Areas, Current Advances in Research, Papers, Books, and Contact Us. The above, excluding Contact Us, are the classes of the ontology. In the left corner are links for the research areas of artificial intelligence, as per the subclasses of the class Research in Artificial Intelligence and links for other areas of fields. There are also links for other related areas of research in AI. These links both at the top and bottom lead to their own pages with full description and Java script based moving related pictures. Users can search for any entity of the web site. They can also register and log in to update the website.

4. Discussion

There are various evaluation techniques of ontologies as there is no agreed standard form of evaluation [9]. For the evaluation of the current ontology, seven questions were formulated drawn from the evaluation techniques of Brank and Hartman [10, 11] to be reviewed by five experts in the field from Addis Ababa University and from HiLCoE School of Computer Science and Technology. The evaluation questions are based on two formal quantitative evaluation techniques - evaluation of the result, the implementation as to its relation to the ontology, and evaluating the ontology on how well it has covered the domain through assessing the different levels of the ontology. According to the evaluation, the majority of the research areas in artificial intelligence have been included and the other classes are found to be quite relevant. As to the object and data properties, they apply to the majority of their instances and hence are relevant.

5. Conclusion and Future Work

Ontology is a proper naming and explanation of the types, properties, and interrelationships of the entities that exist in a particular domain. In review of previous literature, ontology for e-tourism and for the field of medicine have been developed. However, there is no ontology developed for research in AI. In this paper, ontology for research in AI is developed. In the ontology, research in AI was divided into subclasses of the major research areas with each subclass having further subclasses. To create relationship between the classes and between the classes and the domain, object properties and data properties were created to ensure that a class is within the scope of the domain, to establish relationship with other classes, and to create defined classes.

A knowledge base is an ontology and a set of individual instances of classes. The ontology developed is a knowledge base adding to the knowledge base of research in AI through proper

naming and explanation of the types, properties, and interrelationships of the entities that exist in the field of research in AI.

The ontology is a back-end means to support the use of the ontology for specific group of users. In this paper, the specific group of users are those interested in pursuing a research in AI. A website that is based on the modeled ontology is built by integrating the distributed information about research in AI from various sites. Users can register and update the website besides to being able to update the ontology on the Web-Protégé site of Stanford University. The basic idea of an ontology based website is to provide all the links associated with an entity. For instance, when a user clicks on "Place where research in AI is conducted," s/he will find publications, projects, people, research areas and when s/he clicks on a research area, s/he will find most of the links associated with a research area. Overall, the ontology can be used to build software for a reference point for research in AI.

A further research area would be to develop an ontology for AI, instead of just for research in the field, which would include jobs in AI, education in AI, integrating AI at the work-place for companies for increasing productivity, and other areas of AI.

References

[1] "Ontologies (Information Science)," Wikipedia, the free encyclopedia, white paper, 2016.retreived from https://en.wikipedia.org/wiki/Data_integration, Last accessed on January 15, 2017.

[2] "Artificial Intelligence," Wikipedia, the free encyclopedia, white paper, May 2016, retrieved from https://en.wikipedia.org/wiki/Data_integration, Last accessed on February 15, 2017.

[3] J. Cardoso, "Developing an OWL Ontology for E-Tourism," Dept. Math. and Eng., University of Madeira, Funchal, Portugal, 2001.

[4] C. M. Machado, "Knowledge Representation for Data Integration and Exploration in Translational Medicine," Dept. Inform. Sci., Univesidade de Lisboa, 2013.

[5] M. Horridge, "A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools," Ed. 1.3, The University of Manchester, white paper, 2011, retrieved from http://mowl-power.cs.man.ac.uk/protégéowltutorial/resources/Protégé OWLTutorialP4_v1_3.pdf, Last accessed on April 30, 2017.

[6] "Introduction to Ontologies with Protégé", white paper, 2010, retrieved from https://wiki.csc.calpoly.edu/OntologyTutorial/wiki/IntroductionToOntologiesWithProtégé, Last accessed on April 30, 2017.

[7] "Artificial Intelligence: Computer Science and Engineering Department: Graduate Research," white paper, 2016, retrieved from https://www.cs.washington.edu/research/ai, Last accessed on February 15, 2017.

[8] "Artificial Intelligence: Computer Science and Engineering Department: Graduate Research," white paper, 2016, retrieved from http://www.eecs.umich.edu/eecs/research/area.html?r_id=2, Last accessed on March 15, 2017.

[9] Marta Sabou et al," Learning Domain Ontologies for Web Service Descriptions: an Experiment in Bioinformatics, "Dept. of Computer Science. University of Amsterdam, white paper, 2005, retrieved from https://pdfs.semanticscholar.org/3f90/2fadec5c4b6baaece3a25c474f62c12dbac5.pdf, Last accessed on April 30, 2017.

[10] Janez Brank et al, "A Survey of Ontology Evaluation Techniques," Department of Knowledge Technologies, Jozef Stefan Institute, Ljubljana, Slovenia, white paper, 2010, retrieved from http://ai.ia.agh.edu.pl/wiki/_media/pl:miw:2009:brankevaluationsikdd2005.pdf, Last accessed on April 30, 2017.

[11] Jens Hartman et al, "Methods of Ontology Evaluation", EU-IST Network of Excellence, white paper, January 2005, retrieved from ttp://www.starlab.vub.ac.be/research/projects/knowledgeweb/KWeb-Del-1.2.3-Revised-v1.3.1.pdf, Last accessed on April 30, 2017.