

ICININFO

Developing Ontology for the University Archives: The Domain of Technological Education

Daphne Kyriaki-Manessi*, Markos Dendrinou

Technological Educational Institute of Athens, Department of Library Science and Information Systems, Agiou Spyridonos Str., 12210 Aegaleo, Athens, Greece

Abstract

The present study, is the outcome of the project: “Technological Educational Institute of Athens Library: Development of Digital Services” funded by the EU Digital Plan 2009-2013 programme. One of the programme’s mandates was the digital organization of the university archives within the collections of the institutional repository. Key element of the process was to secure subject access to archival data. This obligation led us to explore the potentials of ontologies. The domain of technological education was selected, as our institution is technological in nature and our archival material corresponded to the aforementioned subject area. Similar existing ontologies were examined and ideas on classes, subclasses and properties were taken into account. However, the complex structures of these ontologies with multiple calls on other existing controlled vocabularies did not accommodate our needs. As a result, a customized structure was designed and classes, subclasses and properties were defined. Subdomain analytical description and visualization are also part of this paper. Finally, discussion on ontology development issues and further research follow.

© 2014 Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Selection and peer-review under responsibility of the 3rd International Conference on Integrated Information.

Keywords: *ontologie; technological education; domain; institutional repositories; university archives*

* Corresponding author. Tel.: +30-210-538-5202; fax: +30-210-538-5254.

E-mail address: dkmanessi@teiath.gr

1. Introduction

1.1. The project

The present study, is the outcome of the project: “Technological Educational Institute of Athens Library: Development of Digital Services” funded by the EU Digital Plan 2009-2013 programme. One of the programme’s mandates was the digital organization of the university archives within the collections of the institutional repository. The project aimed at creating an institutional repository housing the academic community’s research and educational output. This includes research papers, patents, student theses, teaching material, etc. In addition, the Institutional Repository is directly connected with the Institute’s following related projects:

- a. “*Open Academic Courses Project*” aiming at creating a pool of open courses on the institute’s e-class platform. Data and metadata of educational material have taken full advantage of the systems’ interoperability.
- b. “*Evaluation of the TEI of A*”, an ongoing evaluation procedure, according to EU standards and under the Guidance of the Quality Assurance Agency. IR’s metadata feed the professors’ evaluation profiles.

Within this framework the content of the IR is of importance. Communities of the Repository correspond to Faculties and Departments and collections are of thematic nature.

The IR was set in the DSpace platform, Dublin Core was the standard for the description of items, metadata were to be bilingual and subject access was based on providing broad subject categories and keywords following the rules of the Subject Headings issued by the Greek National Library (National Library of Greece, 2003); whereas there was no relevant term the LCSH (LCSH, 2013) was to be used.

The development of the Institute’s archives in a digital environment was also a mandate of the project. For this purpose, a distinct collection was formed at the IR’s “community” level. It should be noted, that this served as a starting point for acquiring, locating and collecting the actual physical material.

1.2. The Institute’s archives: concept and structure

The Technological Educational Institute of Athens, is the largest and oldest technological institute in Greece, and it marks the development of technological institutes in Greece along with the history of technological education in the country. A significant amount of paper clippings, reports, photos, videos, white papers, blueprints, administrative papers, student guides, state gazettes, legislation, maps, awards, designs of the degrees and other official papers, etc were accumulated over the years, but no formal archives were ever set up. Some departmental collections, housed old laboratory equipment, such as old optometric instruments, or the nurses’ old uniforms and kits. Student works were also included, especially those of the faculty of Fine Arts. We decided that these should all become part of the Institute’s archival collection, and their metadata to become part of the IR’s collection.

A customized form of Dublin Core was used for the metadata of the archival items, and relevant mapping with EAD (EAD, 2012) was created, taking into account, the collections’ compatibility with other archival projects. In regards to subject access to the archival collection, the issue revealed several problems, as materials were diverse in focus, but the main subject area was “technological education”; in itself its course and development as a distinct part of higher education. The discussion on developing an ontology on “technological education” led us to a twofold project: a. Defining the terms and building the ontology, and b. Connecting the ontology to the DSpace platform so as to accommodate the semantic metadata of items and facilitate searching.

This paper presents the process of building of the ontology within the domain of “technological education”, and its final outcome.

2. Literature review

Literature review was conducted in two areas: a. In relevant controlled vocabularies such as thesauri that cover the domain of higher education, and b. In existing ontologies that also cover higher education.

Initially ERIC Thesaurus (ERIC, 2012) was examined, and then relevant terminology especially that regarding “higher education” was extracted. ERIC covers in detail all elements of higher education, its widespread use and specialization on education issues was important. Areas that needed verification and comparison with the nature of “Technological education” within Higher education in Europe and in Greece in particular were also examined. The same holds for the LCCS (LCCS, 2012). In detail the terminology included in EUROVOC (EUROVOC, 2013) the official thesaurus of the European Union was taken into account. Terms related to the administrative part of Higher education were also gathered, while there was an emphasis on the European aspect of it. Along the same lines, but much narrower in scope and in terminology, is the European Education thesaurus (EET, 2012) which offered some terms related to the job market and employment opportunities in connection to higher education.

In addition, the existing Greek controlled vocabularies were examined these being: a. The Subject Headings of the National Library of Greece (National Library of Greece, 2003) and the Thesaurus of the National Documentation Centre (National Documentation Centre, 2005) and the Union Catalogue of the Greek Academic Libraries (Hellenic Academic Libraries Link, 2012). It is worth noting, that the terminology gathered from these vocabularies, however limited, was far more accurate and relevant to Greek “technological education” issues.

Within the area of ontologies, we located two ontologies in the broader area of “higher education”. We determined that some of their classes and subclasses, could be used in building the more specific domain of “technological education”. The most important existing ontology in the domain of higher level education is VIVO. The domain of this ontology is networking of scientists and researchers, including university structure (faculty member, student, department, division, university, library, information resources, working papers, research proposals, facility, room, building, faculty administrative position, teacher role, reviewer role, researcher role, editor role etc.).

VIVO ontology makes use of available classes of other ontologies, and complements them with additional classes and properties relative to the structure and function of an educational institute. The ontologies called and used by VIVO are among others foaf (foaf, 2013), bibo (bibo, 2013), event (event, 2013), c4o (c4o, 2013), skos (skos, 2013) whose classes are interrelated with the classes introduced and defined in the frame of VIVO itself. VIVO ontology is both structured based (modeling the social networks of scientists) and domain based (modeling the expertise of scientists). Figure 1 presents an extract of VIVO classes that we considered importing in our ontology.



Fig. 1. An extract of the class hierarchy of VIVO (calling BIBO).

The other existing ontology was designed and implemented especially for the analytical description of a higher education institute is HERO (Higher Education Reference Ontology). HERO includes location issues (campus, laboratory), research work (projects, publications, seminars), faculty issues (departments, semesters, conferred degrees, faculty members, academic staff, committees, administrative staff, support staff, technical staff, graduate students, undergraduate students, postgraduate students).

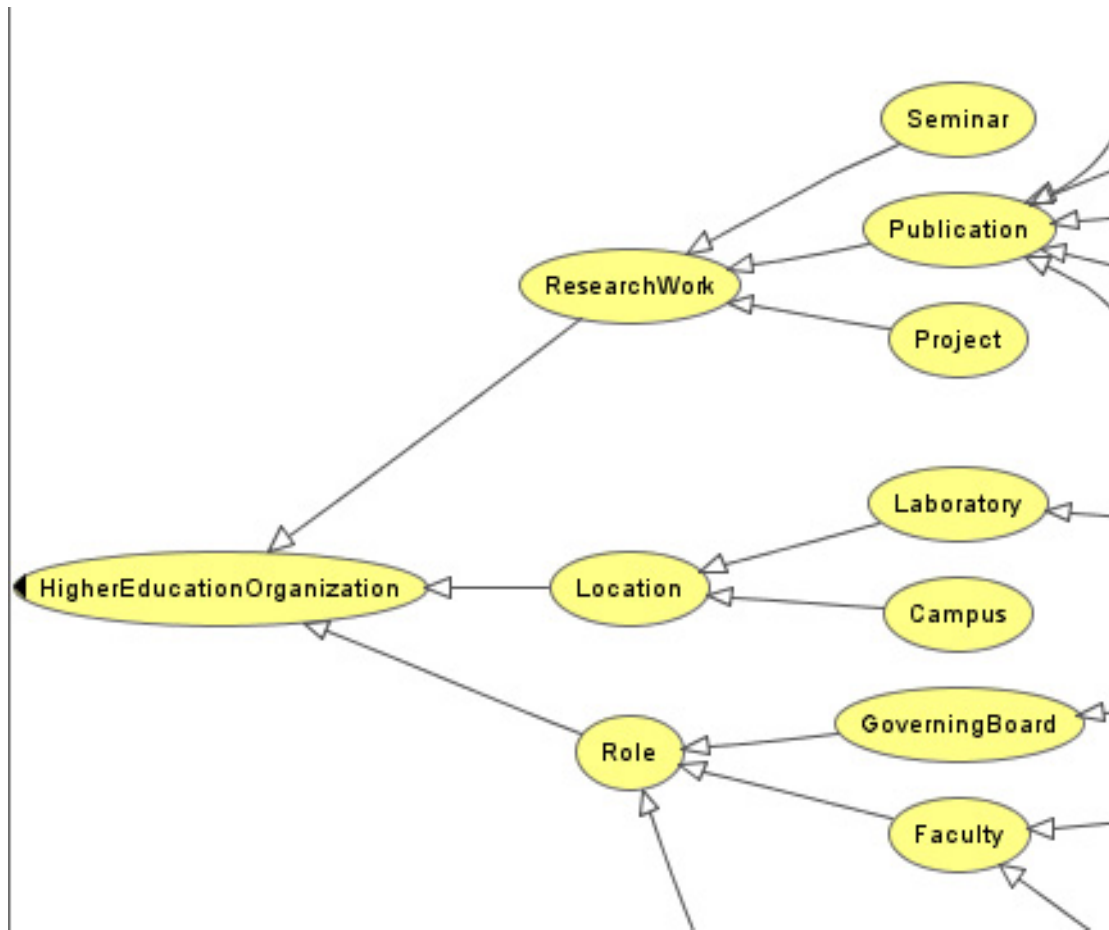


Fig. 2. An extract of the class hierarchy of HERO

However, the complex structures of these ontologies, with multiple calls on other existing controlled vocabularies, did not accommodate our needs. As a result, a customized structure was defined and classes, subclasses and properties were defined. For the development of the ontology on Technological Education presented here the open source software Protégé 3.4.8 (Stanford Center for Biomedical Informatics Research, 2013) was used.

3. Ontology structure and domain analysis

3.1. Classes, properties and individuals for the description of technological education

Ontology is the knowledge scheme that offers the greatest possibilities concerning the definition of relations among individuals of different classes. While taxonomies include only BT/ NT (broader term/ narrower term) relations, thesauri include also part-of, synonym, antonym and RT (related term) relations, ontologies allow designers to define any customized relation between different type of objects (classes).

Class is a concept; i.e. a category including objects of the same type. Property, in OWL/ Protégé environment, is a relation defined between objects of two different classes (object property) or a characteristic of a class taking

values of a certain type (strings, integers, decimals, dates etc (datatype property). Individual, in OWL/ Protégé environment, is an object of a certain class.

In the ontology concerning technological education, defined in the present paper, the superclass/ subclass relation is not restricted only in BT/ NT relation but is extended to include relation also. The hierarchical scheme of classes and subclasses of the ontology under consideration is expressed through the domain concept. The Domain Concept subsumes five main concepts/ classes: areas of activity, human resources, infrastructure, legislation, institutions.

- a. Areas of activity subsume academic programs and continuing education.

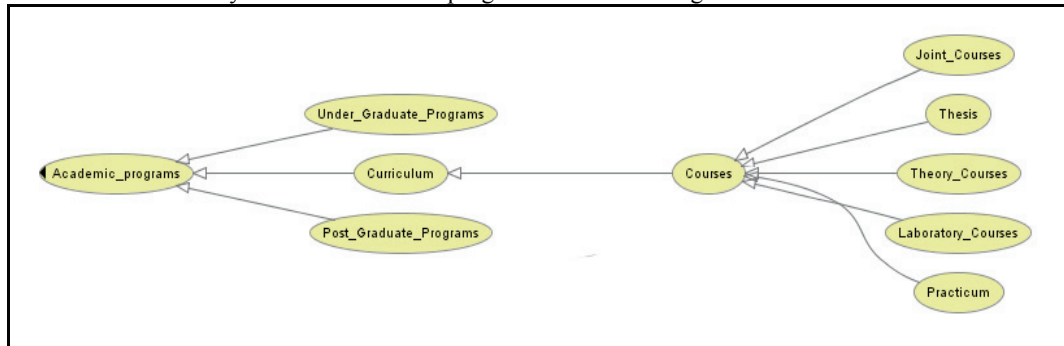


Fig. 3. Class hierarchy with root: Academic Programs (Technological Education Ontology)

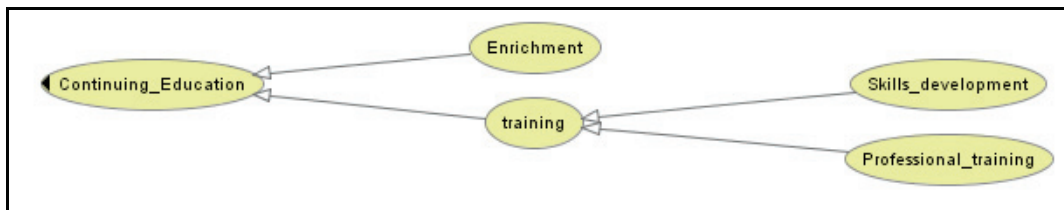


Fig. 4. Class hierarchy with root: Continuing Education (Technological Education Ontology)

- b. Human Resources include administrative staff, faculty members, post doctorate researchers, researchers, students and visiting scholars.
- c. Infrastructure includes Academic Libraries (Digital Libraries, Institutional Repositories) and Campuses.
- d. Legislation includes Campus Sanctity and Higher Education Legislation.
- e. Institutions include Higher Education Institutions, Research Institutes and Student Elections.



Fig. 5. Class hierarchy with root: Higher Education Institutions (Technological Education Ontology)

Apart from the basic class Domain Concept, the following general concepts have been defined in parallel to this one: Educational Philosophy, Educational Policy, Educational Programs, Higher Education, Interuniversity cooperation, Professional Education, Quality of Education, Research Programs and Teaching.

3.2. Subdomain analytical description and visualization

Each class may have one or more subclasses either related or disjoint. Classes and subclasses are characterized by properties having a range within a domain. This is applied on the individuals (presented as instances) defined within the domain. In fact, this ability of protégé software is extending traditional thesaurus functions in establishing relationships between terms to all possible functionalities of terminology selected. In this respect, the implementation of the ontology in protégé has given the opportunity to create a far more flexible system in spite of the fact that interrelationships are simpler based on one to one correlation.

The triplets of Domain class, Object properties and Range class, are presented in the following figures 6-10.

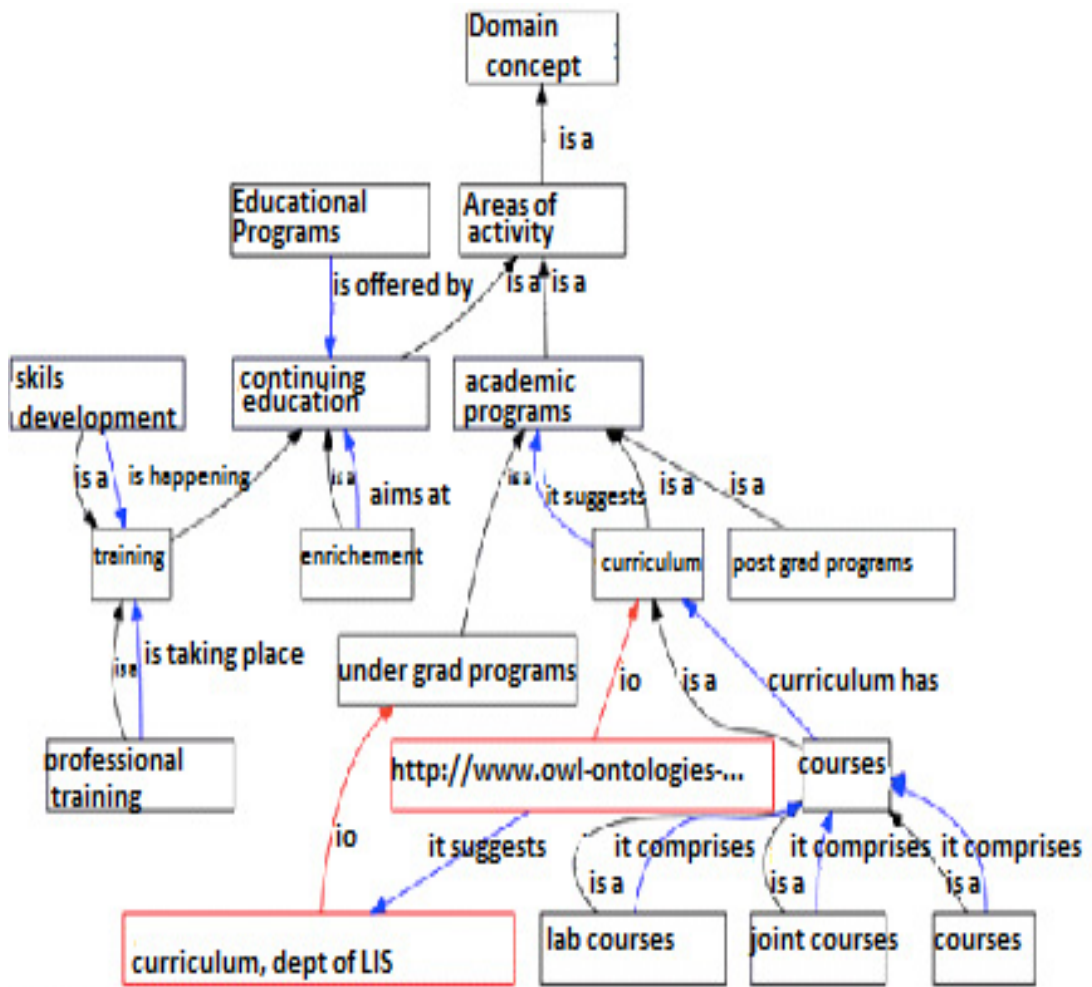


Fig.6. Subdomain "Areas of activity": Subclasses, properties, instances

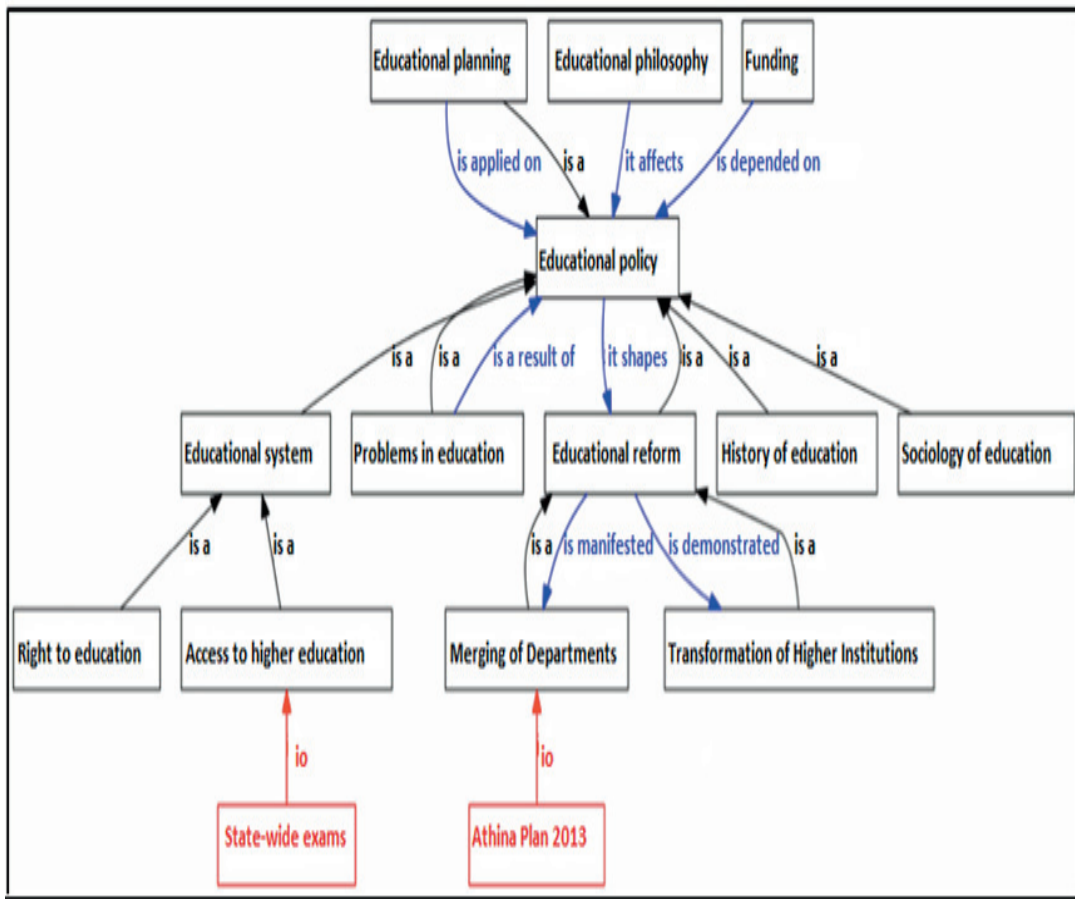


Fig. 7. Subdomain "Educational Policy": Subclasses, properties, instances

The object properties connecting objects of certain classes and the individuals (instances) of them are given at the figures below for a number of selected subdomains (classes).

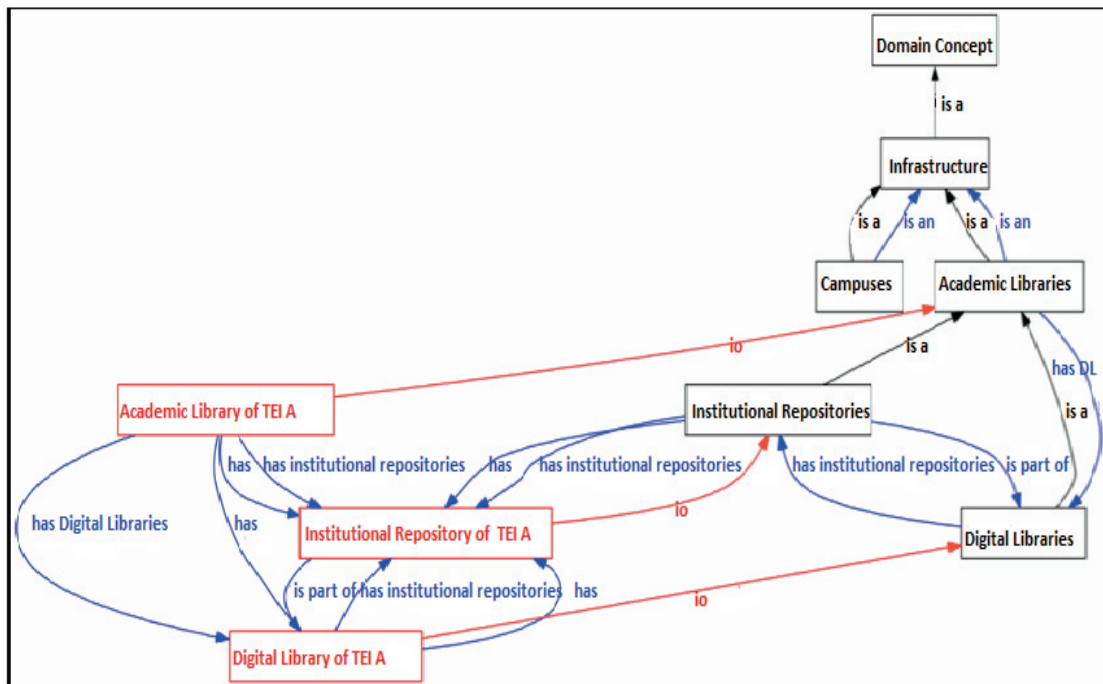


Fig. 8. Domain: Academic Libraries. Properties and individuals (Technological Education Ontology)

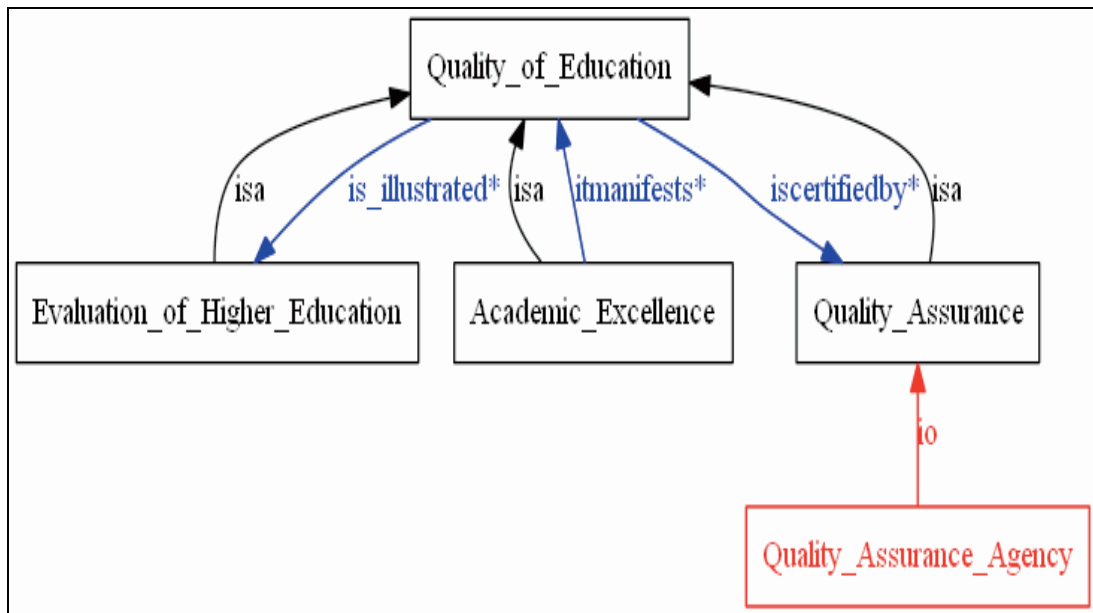


Fig. 9. Domain: Quality of Education. Properties and individuals (Technological Education Ontology)

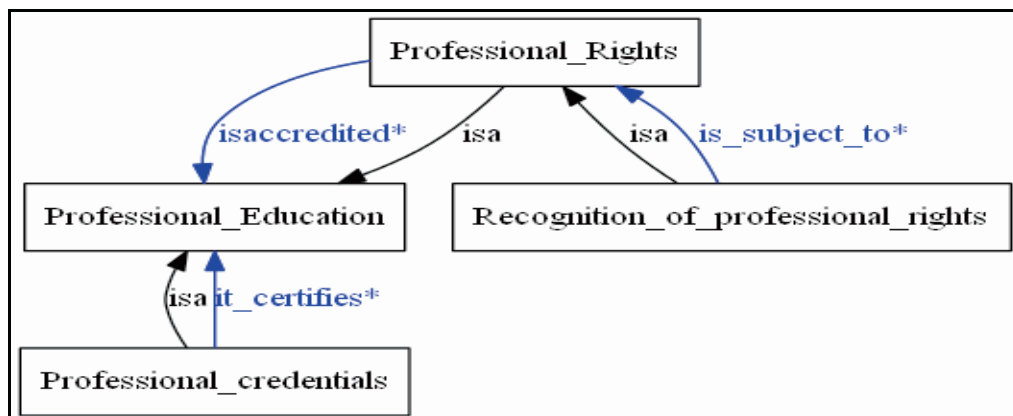


Fig. 10. Domain: Professional Education. Properties and individuals (Technological Education Ontology)

4. Significance of the work and further research

The ontology on “Technological education” was built in order to serve the subject organization and access to the archival collection of a higher education institution of technological character. It was designed to serve the aforementioned collection and its users, these being academics, sociologists, education policy makers, education historians, students as well as the academic community of the Institute along with the broad public.

In addition, the present ontology can be used in whole or as apart by other owl ontologies. In this sense the work adds to the already existing pool of ontologies and contributes to the weave of the semantic network.

As it is already mentioned, the project was twofold a. defining the terms and building the ontology, which is the present work and b. connecting the ontology to the DSpace platform so as to accommodate the semantic metadata of items and facilitate searching. This second objective is currently under development and a relevant report will follow at later date.

References

- (2013). *Bibo*, <http://bibliontology.com/> (accessed 1 July 2013).
- (2013). *C4O, the Citation Counting and Context Characterization Ontology*, <http://purl.org/spar/c4o/> (accessed 1 July 2013).
- (2013). *Event*, <http://motools.sourceforge.net/event/event.html> (accessed 1 July 2013).
- (2013). *Foaf*, <http://xmlns.com/foaf/spec/> (accessed 1 July 2013).
- (2013). *SKOS (Simple Knowledge Organization System)*, <http://www.w3.org/2004/02/skos/core#Concept> (accessed 1 July 2013).
- Educational Resources Information Centre (2012). *ERIC thesaurus* http://www.eric.ed.gov/ERICWebPortal/resources/html/thesaurus/about_thesaurus.html (accessed 1 July 2013).
- Encoded Archival Description (2012) <http://www.loc.gov/ead/> (accessed 1 July 2013).
- Europa (2012). *The European Education Thesaurus Eyridice*, http://eacea.ec.europa.eu/education/eurydice/documents/tese/pdf/teseen_005_alphabetic.pdf (accessed 1 July 2013).
- Europa (2013). *Eurovoc multilingual thesaurus of the European Union*, <http://eurovoc.europa.eu/drupal/?q=el> (accessed 1 July 2013).
- Hellenic Academic Libraries Link (2009). *Union Catalogue of Greek Academic Libraries*, http://www.unioncatalog.gr/ucportal/index.php?option=com_content&view=article&id=300&Itemid=28&lang=el (accessed 1 July 2013).
- Library of Congress (2012). *Library of Congress Classification Scheme (LCCS)*, <http://www.loc.gov/catdir/cpso/lcco/> (accessed 1 July 2013).
- Library of Congress (2013). *Library of Congress Subject Headings*, <http://www.loc.gov/aba/cataloging/subject/weeklylists/> (accessed 1 July 2013).
- National Documentation Centre (2005). *Thesaurus*, <http://www.ekt.gr/products/thesaurus/> (accessed 1 July 2013).
- National Library of Greece (2003). *Authority files*, <http://www.nlg.gr/www/el/node/56> (accessed 1 July 2013).
- Stanford Center for Biomedical Informatics Research (2013). *Protégé 3.4.8*, <http://protege.stanford.edu/download/download.html> (accessed 1 July 2013).