International Journal of Information Science and Management
Vol. 17, No. 2, 2019, 33-53

CivilOnto: An Ontology Based on Persian Articles Published in Civil Engineering Domain

Zahra Namvar
Ph.D. candidate in Knowledge and Information Sciences, Islamic Azad University, Science and Research Branch, Tehran, Iran.
Namvarzahra@gmail.com

Fatemeh Nooshinfard
Associate Prof. Department of Communication and Knowledge Sciences, Science and Research Branch, Islamic Azad University, Tehran, Iran.
Corresponding Author: f.nooshinfard@gmail.com

Fahimeh Babalhavaeji
Associate Prof. Department of Communication and Knowledge Sciences, Science and Research Branch, Islamic Azad University, Tehran, Iran.
f.babalhavaeji@gmail.com

Moluksadat Hosseini Beheshti
Assistant Prof. Research Department of Information Science, Iranian Research Institute for Information Science and Technology (IRANDOC).
beheshti@irandoc.ac.ir

Abstract

The present research aims to represent the terms and concepts of civil engineering and thus design its domain structure in the form of ontology. This is an applied research and the qualitative content analysis method with summative approach is used. The research population is 12309 published articles in the scientific journals (1983 to 2016), indexed in the full-text Persian articles database at Regional Information Center for Science and Technology (RICeST). Data collection was done via SQL queries from the RICeST database. The content analysis was used to create a conceptual model of the civil engineering domain and to explain the relationships and instances. To create the conceptual model, there were consulted with 10 Subject experts in civil engineering area. The process of ontology creation has been done through METHONTOLOGY methodology and protégé ontology management software (Beta 5 edition). The made-artifact of this research has 283 concepts, along with 62 object properties, 79 data type properties, 10 annotative properties, and 976 instances. In total, in the present research, 151 semantic properties were identified between the concepts. The results of this research can serve as a guide for the RICeST experts in order to identify the users' needs and facilitate the users' interaction with RICeST databases. This can also provide the users with access to more relevant resources. This ontology can also be served as a layer of semantic web in these databases.

Keywords: Ontology, Civil Engineering Ontology, Regional Information Center for Science and Technology, RICeST Databases, Persian Articles, CivilOnto.

Introduction

The emergence of information and communication technologies together with web environment access to the information exchange have led to a change in the information environments and, as a result, made changes in the knowledge representation tools. However, today the focus of modern information systems is moving from "data processing" towards "concept processing". In fact, a semantic concept, which has its own interpretations and exists in a context with other concepts, form the basis of these systems (Brank, Grobelnic and

Mladenic, 2005). In order to reduce problems such as semantic domain limitation, low flexibility of storage and retrieval systems, and also to increase the deduction capabilities in these systems, some movements have been done towards new knowledge representation systems. These movements, most often seen in the artificial intelligence, have got a different form with digital libraries and semantic web appearance (Sharif, 2006). They have led to concept-oriented information storage and retrieval systems, and have made the appearance of semantic tools such as ontologies. The Semantic web is a web of ontologies that allows the analysis of the domain knowledge by modeling the relevant concepts of this domain (Mezghanni & Gargouri, 2017). Ontology has been recommended as a knowledge representation method in semantic web. This semantic tool is used for representation of knowledge as concepts and roles. Concepts are the terms for indicating idea in a domain, and roles are the relations among the concepts (Lee, Rho & Lee, 2014). In fact, these concepts, roles and relationships form the structured knowledge of a specific domain (Nowroozi, 2015). Ontologies play a pervasive role in the modern knowledge-based systems as they constitute a powerful tool for supporting natural language processing, information retrieval, and text mining (Mezghanni & Gargouri, 2017). They are also used profoundly by experts in the databases and other computer systems to share thematic and specialized information of a domain (such as medicine, chemistry, Physics, etc.) which have led to the elimination of many shortcomings in the environments, such as databases, and creating them based on ontology (Shamsfard & Abdullahzadeh Barforoush, 2002). Therefore, today ontologies have multiple applications in databases and other information retrieval systems, such as understanding the user's information needs (Chen, 2009); conceptualization indexing (Mirzabeigi, 2012; Dan & Wang, 2006), understanding the subject area, query expansion (Andreou, 2005), semantic visualization (Mirzabeigi, 2012; Sanatjoo, 2006), the personalization of information retrieval system (Gauch, Chaffee & Pretschner, 2003), and the organization of retrieved documents based on precise semantic relationships (Chen, 2009). Therefore, it seems that these ontology applications can be used as one of the semantic web technologies in scientific articles' databases structure to provide better retrieval and higher precision. Regional Information Center for Science and Technology (RICeST), as one of the most important Iranian scientific centers, has different duties such as creation and maintenance of the databases. Indexed document in these databases can help one in tracing the ontology space, determine the concepts and their relationships. Lack of ontology in the RICeST databases has led to the lack of document integration and coherence, so that users have problems during document search and retrieval. Also, lack of ontology can cause users to have misunderstanding of the epistemic space of these databases domains.

In this regard, the present research intends to provide a conceptual structure of the scientific articles (published articles in the Iranian scientific journals) in civil engineering domain, which are indexed in RICeST full text Persian articles database in an ontology form. Therefore, this study intends to answer this research question: what and how the conceptual structures and semantic relations are formed between concepts of civil engineering based on the indexed articles in RICeST databases? Accordingly, this research seeks to provide domain ontology for conceptualizing civil engineering knowledge and determining the types of ontology relations between extracted concepts of its articles.

Literature Review

In the early 1990s, Gruber (1993) introduced the first attempts for designing and creating an ontology. With the emergence of the semantic web and the need for interaction between different retrieval systems, ontology and logic started to play an important role in the knowledge representation. Thus, ontology provides the content and logic provides the form, and the combination of these two elements provide machine reasoning and deduction. The semantics of information retrieval systems and the ontology's ability in semantics have led them to be one of the most popular and important research topics in various sciences such as knowledge engineering, natural language processing, artificial intelligence, information science, etc. hence, today we witness extensive research on ontologies. These researches can be divided into three sections: (1) researches focused on the transformation of thesauri into ontologies, investigation and enrichment of the existing relationships in ontologies (Soergel, Lauser, Liang, Fisseha, Keizer, & Katz, 2004; Huang, Chung & Kathleen, 2007; Khosravi & Vazifedoost ,2007; Hosseini Beheshti & Ejei, 2015; Ejei, Hosseini Beheshti, Rajabi & Ejehi, 2017; Chun & Wenlin, 2004; Nowroozi, Mirzabeigi & Sotudeh, 2018a); (2) Researches focused on the efficiency of thesauri and ontologies in information retrieval (Sanatjoo & Fathian, 2011; Nowroozi, Mirzabeigi & Sotudeh, 2018b); (3) Researches focused on the ontology creation based on a set of information sources and texts (Ahmadi, Osareh, Hosseini Beheshti, & Heidari, 2017; Mohammadi Ostani, Azargoon & Cheshmesohrabi, 2018; Zardari, 2016; Dumontier & Villanueva-Rosales, 2009; Teller, Keita, Roussey & Laurini, 2015; Aydin & Tecin, 1970; Elag & Goodall, 2013). Following are some research reviews based on the above-mentioned divisions.

Researches focused on the transformation of thesauri into ontologies, investigation and enrichment of the existing relationships in ontologies

Organizations such as FAO have been leading organizations in the transformation of thesauri into ontologies. The transformation of AGROVOC thesaurus into ontology by Soergel, Lauser, Liang, Fisseha, Keizer, & Katz, 2004 is one of the major projects of FAO, where they eliminated the ambiguity in the transformation process. In this research, the relations of AGROVOC thesaurus based on defined pattern and the rules of semantic relations in ontology have been investigated. Among the other organizations involved in the thesaurus transformation into ontology was the Chinese Information Institute of Ministry of Agriculture. In the first phase (2004), they have transformed part of this thesaurus into ontology, which included 898 Chinese concepts in the agricultural products field. In this project, Chun & Wenlin (2004) transformed the Chinese agricultural thesaurus into ontology. In the second phase (2006), this thesaurus was transformed completely into ontology.

In Persian language, there have been attempts to convert thesauri into ontologies and to investigate the existing relationships in ontologies. Among the leading scientific centers in Iran that have been designing and implementing projects in this field is the Iranian Research Institute for Information Science and Technology (IranDoc), National Library of Iran, and some universities. In this regard, Khosravi & Vazifedoost (2007) in a research re-engineered the LIS part of ASFA thesaurus for the ontology construction through ontology learning method. In this research, the existing relationships in ASFA thesaurus have been transformed into ontology relationships. Hosseini Beheshti & Ejei (2015) designed and implemented the basic sciences ontology based on the concepts and relationships in the related thesauri. In this project, they used the chemistry, physics, biology, geology and mathematics thesauri as a basis of ontology. This ontology is made in several steps eliminating the conflicts and overlaps of common concepts in the thesaurus, combining thesauri with each other, and creating a comprehensive thesaurus with conceptual designing of the ontology through METHONTOLOGY methodology and ontology implementation. In relation to researches on the ontology relationships, Ejei, Hosseini Beheshti, Rajabi & Ejehi (2017) examined and enriched the semantic relationships of basic sciences ontology. In this research, the enrichment of relationships was carried out in a semi-automated method. Ontology concepts were categorized. The semantic relationships between concepts were determined based on the related relationships in thesauri. Also, extractive semantic relationship patterns of upper-level ontology were added to ontology. In this research, approximately 70% of the related relationships was directly transformed into ontology semantic relationships.

Researches focused on information retrieval efficiency in the thesauri and ontologies

Fathian (2010) examined the efficiency of ontology capabilities in comparison with thesaurus. This designed ontology is the result of transforming the indexing domain concepts of ASFA thesaurus into ontology. This study showed that the ontology efficiency level in knowledge representation and concepts retrieval is more efficient than the ASFA thesaurus. On this basis, more effective semantic tools can be designed to represent knowledge and retrieve the concepts of subject areas in comparison to traditional tools such as thesauri. Nowroozi et al. (2018b), in a research, compared the efficiency of thesaurus and ontology in representation of the concepts and semantic relationships. In this research, they have based their work on ASIS&T web-based thesaurus and used the usability examination method. The results of this study showed that, from the viewpoint of indexers, ASIS&TOnto ontology is more efficient than the ASIS&T thesaurus in representing concepts and their relationships.

Researches focused on the creation of ontologies based on texts, resources and other information systems

In a research, Aydin & Tecin (1970) have attempted to construct the earthquake ontology using the METHONTOLOGY methodology. In this research, an upper- level ontology (Geo-ontology) was used and all the terms in the area of natural disasters, which were extracted from Izmir disaster and management information system, were conceptualized and implemented in protégé software.

Elag & Goodall (2013) also created the ontology of water resources systems due to the lack of a proper framework for modeling the water resources metadata. Their ontology is called Water Resources Component (WRC), which is based on the methodology of Uschold & Gruninger. In this ontology, the experts opinion have been used and five primary features of Gruber (1993), including clarity, coherence, extensibility, flexibility for merge, and the ability to overcome semantic heterogeneities, have been considered. The outcomes of this study are 18 super-classes to describe the four layers of resources, scientific affiliation, relationships, and techniques.

In another research, Dumontier & Villanueva-Rosales (2009) have discussed the ontology of pharmacogenetics domain using protégé software. It consists of five phases including defining the domain and ontology requirements, creating a primitive model of ontology based on essential concepts, mapping concepts in an upper-level ontology, identifying relationships between concepts, and adding other ontologies and more complicated descriptions to them. In the end, they believed ontologies and semantic web can be good tools for pharmacogenetics information representation. In relation to civil engineering, Teller, Keita, Roussey & Laurini (2015), in a research provided an urban engineering ontology. This research was carried out in two stages: (a) the construction of urban planning ontology through the integration of urban subjects with their descriptions and relationships; and (b) the development of software for urban planners in line with activity on this ontology semantic network. They also developed an ontology browser and editor in the research. They then developed the designed XML and used it to describe ontology.

Other studies have also had scientific endeavors to create ontologies based on published scientific texts in Persian language. Mohammadi Ostani, Azargoon & Cheshmesohrabi (2018), in a research, have explored the existing methodologies, designed scientometrics conceptual model and taken the steps of ontology construction as scientometric ontology. The used resources in this research to extract the required concepts for this ontology include books, articles, specialized glossaries, thesauri, dissertations and researches' projects of scientometrics in Persian language. Bermejo method together with domain analysis approach and methodology have been used in this research and were implemented in the protégé software. The results of this ontology generation are 11 major classes with 20 relationships and 100 individuals. Ahmadi, Osareh, Hosseini Beheshti, & Heidari (2017) modeled the conceptual knowledge in scientometrics by means of co-occurrence word analysis, C-value method, and protégé software. This ontology concept network construction has been carried out with 653 concepts. The ontology relationships used in this study were "is…a, part of, instance of" relations. The results indicated that most of the relationships between the concepts of the Iranian scientometrics domain were the non-classification relationships such as "part of & instance of". In other words, about 80% of the semantic relationships of the concepts in the Iranian scientometrics ontology are meronymy type.

Zardari (2016) presented a conceptual structure of LIS through ontology form. She applied a middle approach of ontology modeling and METHONTOLOGY methodology in this work. This ontology consists of 394 classes, 224 relationship types, 5633 instances, and 26490 axioms. A conceptual structure that constituted the explicit concepts in an official form was the ontology result.

An overview of the literature shows that a great part of their focus is to transform the thesauri to ontologies. Perhaps, reasons such as search capabilities and semantic information retrieval, economic and temporal issues are the causes for such an approach to traditional tools like thesaurus.

The review of the Persian literature indicates that the researchers and experts' efforts in LIS have led to the creation of a few specialized ontologies such as ASFAOnt, Iranian scientometrics ontology, LIS ontology, basic sciences' ontology and ASIS&T ontology; although there are comprehensive databases in other fields in Iran, less specialized ontologies have been created for these databases based on the existing data in them. The lack of such ontologies in information systems at centers whose mission is to provide research services to the scientific community, and the existing problems in information retrieval of these databases (e.g. RICeST civil engineering database) have led to the creation of specialized civil engineering ontology by researchers. According to the division in literature review, the current research and the CivilOnto ontology are grouped in the third division of these researches. The results of this research can be useful in the knowledge promotion and increasing the use of ontologies in civil engineering and the related projects. It also can facilitate communication between information systems, beneficiaries and professionals.

Methodology

The present research is an applied research and has used qualitative content analysis method with summative approach. Content analysis is a widely used qualitative research technique (Hsieh & Shannon, 2005). It is a family of systematic, rule-guided techniques used to analyze the informational contents of textual data. The object of qualitative content analysis can be all sort of recorded communication (transcripts of interviews, discourses, protocols of observations, video tapes, documents ...) (Mayring, 2004).

Researchers have used content analysis as a qualitative or quantitative method in their studies (Berelson, 1952). Qualitative content analysis is one of the numerous research methods used to analyze text data. It has three distinct approaches: conventional, directed, or summative. A summative content analysis involves counting and comparisons, usually of the keywords or the contents, followed by the interpretation of the underlying context. In this research, qualitative content analysis method with summative approach was used to identify the existing specialized concepts and their relations in civil engineering domain. So that the required data (12309 published articles (1983-2016) in 42 scientific journals of civil engineering) were extracted automatically from RICeST database by using SQL queries and then were manually processed in XCL files. In which, the title keywords were selected by authors and experts based on the articles' titles. Then, the ontology relationships were created and entered in the protege software. Finally, an overview of the concepts was created.

The METHONTOLOGY methodology is another method which has been used. This method has been used due to the purpose of this research and the features such as the type of extracted data, the continuous updating of the existing data in this research target database, the use of the content analysis method and also due to the ontology maintenance and development. This method belongs to the second generation of ontology-making techniques (Pinto, Staab & Tempich, 2004), which was developed by Fernandez- Lopez in 1997 in an ontology group at Technical University of Madrid. METHONTOLOGY is the most complete methodology for the creation of ontologies (especially, domain ontologies) and provides a more accurate description of each of the required activities in the ontology life-cycle (Brusa, Caliusco, & Chiotti 2006). This method has this feature to provide ontology reuse for its future development (Aydin & Tecin, 1970). METHONTOLOGY method has a life cycle for creating ontology. This life cycle is the relations among activities during the ontology development, which is also called Intra-dependency (Asosheh, Mehrasa, Khatibi & Khakshour Saadat, 2011). This life cycle exists in civil engineering ontology creation, and CivilOnto has been designed and created accordingly. This life cycle includes activities such as management (scheduling, control and quality assurance), development (specification, conceptualization, formalization, implementation and maintenance), and supporting activities (Knowledge Acquisition, integration, evaluation, documentation and configuration management).

Findings

Considering METHONTOLOGY method, following are the research findings and the explanations for each of the steps in CivilOnto ontology:

Stage I (management activities)

At this stage, the activities, arrangement of their implementation, timetable of their implementation, and the resources (people, software and hardware) should be planned (Fernandes Lopez, Gomez-Perez & Juristo, 1997). Upon CivilOnto ontology, the first step was to plan and determine the required resources and examine the data quality, as well as the required time to examine these resources and data. Therefore, at first, the journals (42 scientific journals) and articles (12309 articles) were determined and their time interval was also investigated. In the next step, titles and keywords of articles were extracted from RICeST database using SQL queries and were prepared to use in the various stages of ontology construction. So that, keywords[1], journals and articles title[2] queries were used in the RICeST software.

Stage II (developmental activities)

At this stage, the operations like specification, conceptualization, formalization, implementation, and maintenance have been done.

Specification

The goal of specification is to produce an informal, semi-formal or formal ontology of particular documents in a natural language. At this stage, the purpose of creating an ontology that includes target users, uses, ontology domain and the used resources should be determined (ibid). The purpose of this ontology is to represent the terms and concepts and to map the civil engineering domain structure in the form of an ontology. Its users are RICeST in-person and distant users. Its use is a tool for semantic and useful retrieval at RICeST engineering database. Its domain is civil engineering. Its resources are the published articles in the scientific journals of civil engineering (1983 to 2016).

Conceptualization

The next stage of building ontology are identified by conceptualization. By so doing, the knowledge domain of the conceptual model that describes the problems and strategies for determining the subject domain lexicon will be represented. Conceptualization consists of 11 stages: the creation of the terms list, concepts classification, relationships between concepts, concepts dictionary, the description of temporary binary relations, the description of the instances properties, the description of class properties, the description of the fixed values, the description of the formal axioms, the description of rules and instances. (Fernandes Lopez, Gomez-Perez & Juristo, 1997). The most important activity of METHONTOLOGY life cycle is the developmental activities and the conceptualization stage, which plays a decisive role in the continuation of ontology and has a strong relationship with the knowledge acquisition (Hosseini Beheshti & Ejei, 2015). Considering the importance of conceptualization, this study mostly focuses on the conceptualization stage because the semantic structure of civil engineering ontology will be achieved at this stage. The life cycle of conceptualization is expressed below.

Creating a list of Terms: The first step in conceptualization is the creation of terms list. These terms include concepts, instances, verbs, and their properties. In fact, this list is an aggregator of useful and potential domain knowledge of the ontology, and if well-prepared, many of the existing terms in the examined documents are identified in this way (Fernandes Lopez, Gomez-Perez & Juristo, 1997). This list introduces a set of existing terms in ontology, their definitions in natural language, and their synonyms and abbreviations (Asosheh, Mehrasa, Khatibi & Khakshour Saadat, 2011). List of terms is a lexical knowledge database, which is an important tool in the ontologies construction. In the present research, for the first stages (creating a list of terms), a list of terms containing the related terms in civil engineering was prepared and investigated from RICeST database by SQL queries. This list includes concepts, instances, and the properties of the concept characteristics, the relationships between concepts, their description in natural language, synonyms, and their abbreviated terms. In the early stages, this list may have a variety of terms that refer to the same concept, in which the authors, in consultation with subject experts, sought to identify synonyms and eliminate the overlaps.

Creating the Concepts Classification: Ontology is a taxonomic hierarchical structure based on subcategories, where the relationships among terms are richer and deeper than taxonomy (Hosseini Beheshti, 2014). In fact, at this stage, a classification of concepts was created which includes the classification of information entries that are hierarchical and are based on broader and narrower relationships in the actual world. In this classification, subclasses are split from the main node and divided from top to bottom, in the form of father-child relationships. This taxonomic hierarchical division method constitutes a semantic frame in ontologies (Dacunta, Obrst & Smith, 2003). The concepts classification output can be one or more classifications in which concepts are categorized at them (Asosheh, Mehrasa, Khatibi, & Khakshour Saadat, 2011). There is a list of general words or stop words in RICeST database which are extracted by computer man and were removed from XCL. Then, concepts are presented in an alphabetic order through a hierarchical structure. Indentations and different writing fonts are used for classes and subclasses distinction. The main classes are divided into subclasses with a "has a type" relationship. Then, subclasses are separated from super classes with the inverted relation “kind of", and are linked together with more indent. Also, concepts that are not accepted by experts are identified as underline, in contrast to which the accepted concept by expert is starred. Table 1 presents an example of the hierarchical structure of the basic concepts of the present ontology[3].

Table 1

Basic concepts of civil engineering ontology in a hierarchical structure

دوغاب ریزی*	**دوغاب سیمان**
دیوار(بخش سازه‌ای)	**دیوار(بخش سازه‌ای)**
دیوار برشی	دیوار برشی
دیوارک(ساختمان)	**دیوارک(ساختمان)**
زلزله*	**زمین لرزه**
زمین لغزه	**زمین لغزه**
ساخت بتنی*	**سازه بتنی**
بتن‌ریزی	بتن‌ریزی

بتن‌سازی	بتن‌سازی
بتن‌پاشی	بتن‌پاشی
قالب‌زنی لغزنده	قالب‌زنی لغزنده
سرباره*	<u>**سرباره کوره بلند**</u>
مصالح ساختمانی*	<u>**مصالح سازه ای**</u>
بلوک بتنی	بلوک بتنی

Establishing relationships between Concepts: The relationships must be constructed after concepts classification stage. In ontologies, a variety of relationships can be defined, most commonly known as the object's relationship (to express relationships between the instances), data type relationship (to create the relationships between instances) and the annotation relationship (to express the description of classes, instances, object relationships, and data type relationships). For each of these relationships, a sub-relation can also be defined (Sanatjoo & Fathian, 2012). In this study, based on the properties of concepts, the above three types of relationships have been used. 62 object relationships were used to express the relations between 976 instances of classes. Some of the object relationships are “constructed from; associate with; combination of; consist of; included services and …”. E.g., “Resalat tunnel” is an instance of “machine tunnel” which has object relationships with each other. One of the object relationship samples is “constructed from”, there is this relationship between “tunnel and structure” (Figure 1).

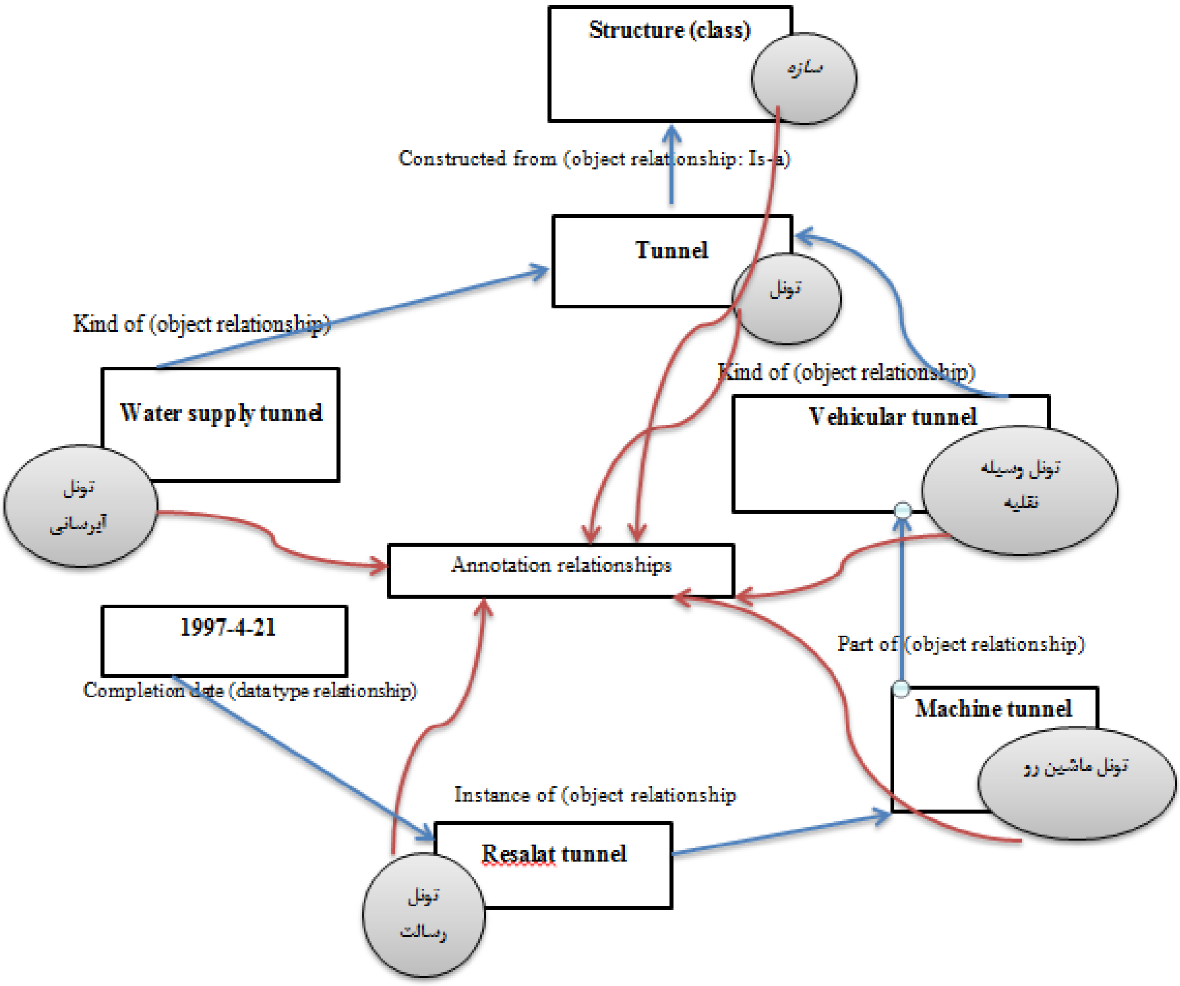

Figure 1. Examples of relationships between the CivilOnto concepts

There were 79 data types' relationships that link the data with their value. Some of the data type relationships are "invention date; completion date; rate of; unit of; start date; has/have dimensions; has/have branch and …" E.g., "1997-4-21 as completion date is a data type relationship (Figure 1). There were 10 types of annotation relationships[4], including definitions, translation to other languages, synonyms, notes, similar terms, etc. which are used for 283 concepts. E.g., the translations of tunnels in Figure 1 in Persian language are some kinds of annotation relationships (such as tunnel ab-resan or saze). The properties of concepts have an important role in the extraction of the relationships of ontologies. Authors extracted these relationships based on concepts properties through consultation with experts.

Another activity that can be done in this step, if necessary, is the creation of the binary relationships between concepts. These relationships are used in some cases and eliminate possible errors in the relationships between concepts and the related classes. Figure 2 shows an example of a binary relation diagram (such as the "kind of, use for and use of" relationship) between the concepts of civil engineering. E.g., the binary relationship between "cement and building materials" were created by "use for and use of" relationships. For creating these relationships, two classes "building materials as a domain" and "cement as a range" are needed (Figure 2). Binary relationships were extracted based on concepts features by authors with expert's consultation.

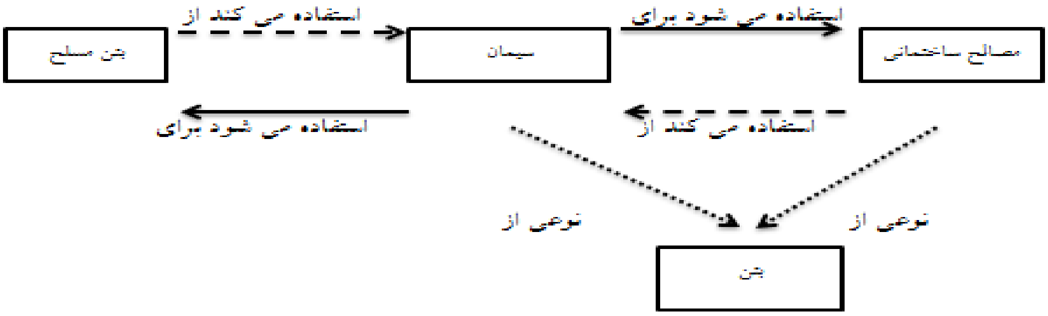

Figure 2. An example of the binary relationship among civil engineering concepts

The relationships between entities or concepts can be extracted from entities description according to expert help (Table 1).

Table 1

Extraction of the relationships between the entities

Relation	Range(instance)	Domain (class)	description
1. Is use for (binary relation) 2. Is use from (binary relation)	cement	building material	1. Cement is use for building material 2. building material is use from cement
Is kind of (binary relation & object relation)	concrete	building material	Concrete is kind of building material
Completion date (Data type relation)	1997-4-21	tunnel	1997-4-21 is completion date of Resalat tunnel
Is kind of (object relation)	foreshock	earthquake	Foreshock is kind of earthquake
Is scale (Data type relation)	earthquake intensity	earthquake	earthquake intensity is scale earthquake
known as (Annotation relation)	quake, tremor or temblor	earthquake	An earthquake also known as a quake, tremor or temblor

Creating the Concepts Dictionary: The relationships between these concepts, instances, instances properties and their classes are determined by constructing the concepts dictionary (Fernandes Lopez, Gomez-Perez & Juristo, 1997). In this research, based on the existing words bank in the RICeST database, a concept dictionary was created through SQL queries. After extraction of concept dictionary, each concept was considered as a class, and their properties, their instances, and the relationships between them were also identified and created. For example, in the “earthquake" class, "foreshock and aftershock" as the properties of this class, "bam earthquake" as an instance of this class, "earthquake intensity and its place" were considered as instance (bam earthquake) properties.

Description of the Temporary Binary Relationships: In fact, at this stage, all names, titles and concepts, members, reverse relationships and their mathematical properties must be identified. For example, the "kind of" relationship and its inverse relationship, "has kind", are considered for "damming" as a target concept, where the "concrete dam" is the source concept that can have transient mathematical properties and N cardinality (see Table 2).

Table 2
Temporary Binary Relationships

Relation	Source concept	Cardinality	Target concept	Mathematical properties	Inverse relation
Kind of	Concrete dam	N	Damming	Transient	has kind
Part of	Industry	N	Concrete industry	Transient	has part
Kind of	Form (concrete)	1	Concrete construction	Transient	has kind
Integrated with	Concrete association	N	Concrete association	Symmetry	---

Describing the Instance Properties: The existing instances in the concepts dictionary are described in detail at this stage. These concepts properties are different. For each instance, properties like name, related concepts, local properties of each concept, the value type, measurement unit, precision and the numerical values range, error values, minimum and maximum cardinality of a collection and ... are expressed. For example, for an “earthquake” concept, the time of occurrence is considered as the instance property; the "time " as type of the value, "earthquake" as the name of the concept, "Richter" as the measurement unit, and "0 and 1" can be also its cardinality (Table 3).

Table 3
Instance Properties

Instance Properties name	Concept name	Value type	measurement unit	cardinality
name	Seismologists	string	…	(1 & N)
Occurrence date	earthquake	date	Richter	(0 & 1)
Earthquake depth	earthquake	decimal	KM	(0 & 1)

Describing Details of the Class properties: At the seventh stage, a comprehensive description of the class's properties will be presented. The class’s properties actually describe the concepts and their values in the related class. The class property has no relation with subclasses and instances, and nothing inherits from them; it is independent and information such as name, concepts name, value and its type, measurement unit, its precision, etc. should be implemented for it. For example, in this research, for the "kind of " property in "beam & girder" class, the "centimeter" is a measurement unit, the “steel” is a value type, and "0 & 1" is cardinality (Table 4).

Table 4
Class properties

class properties name	concept	Value type	Measurement unit	cardinality
Kind of	beam and girder	concrete	CM	(1 & 1)
Kind of	beam and girder	steel	CM	(1 & 1)

Describing Fixed Values: If there were fix values in the previous stages (list of terms and concepts), ontologists must describe and present their details at this stage. The information that is usually given for fixed values is: name (e.g., light weight concrete), value (2800), value type (a number that can be a decimal and integer, such as 2800, which is an integer), unit measurement (kg/m^3) and other properties that fix values can be described through them (e.g., the maximum weight of light weight concrete) (Table 5).

Table 5
Fixed value of light and heavy concrete

name	value	Value type	Measurement unit	Presumable property
light concrete	2800	Integer	kg/m^3	maximum weight of light concrete
Heavy concrete	3500-6000	Integer	kg/m^3	maximum weight of Heavy concrete

Describing Formal Axioms: These axioms are used to specify limitations of ontologies. To describe these axioms, information such as the name (e.g., tunnels), natural language descriptions (e.g., any tunnel must have safe and welfare installations, such as standard entrance and exit, air conditioning and lighting), the logical expressions (e.g., "tunnels (A)"; "safety installations (B)"; "welfare facilities (C)" are used to express these axioms. The concepts such as "tunnels, safety, welfare, tunnel ventilation, tunnel lighting, tunnel entrance, tunnel exit"; the properties and contingency relationships related to these axioms (e.g., tunnel entrance - tunnel exit) and used variables (e.g., A, B, C) should be used to express these axioms. These axioms were used to add description to some of the classes and ontology extension. In this research, formal axioms were used by authors by the help of experts, in case of any necessity (Table 6).

Table 6
Sample of formal axiom of tunnel ontology

Axiom name	Tunnels
Natural language descriptions	any tunnel must have safe and welfare installations, such as standard entrance and exit, air conditioning and lighting
Logical expressions	tunnels (A); safety installations (B); welfare facilities (C)
concepts	tunnels, safety, welfare, tunnel ventilation, tunnel lighting, tunnel entrance, tunnel exit
contingency relationships	tunnel entrance; tunnel exit
variables	A, B, C

Rules Description: The required rules to create an METHONTOLOGY -based ontology should be identified and described in the “Table of Rules”. The cases identified in this table are approximately the same as the definitions in the formal axioms stage. The information should be described, such as name (e.g., the tunnel construction rule), natural language description (e.g., the safety of any constructed tunnel provided by the firefighting system), a phrase that describes the rule (e.g., tunnel safety, fire, firefighting system), concepts (e.g. tunnel, safety), properties (e.g., a variety of safety systems) and binary relationships and the used variables of (a) tunnel safety and (b) firefighting systems,. In METHONTOLOGY method, the “if <conditions> then <result>” pattern is used for rule description. This pattern is used for this research. The rules were created based on ontology concepts nature by authors through expert’s consultation and entered in protégé software. So that the concepts which had their related rules in natural language were separated from other concepts. Then, natural language description, main concepts related to rules, rules properties, binary relationships and the related variables were expressed for each rule (Table 7).

Table 7
Sample of ontology rule of tunnel safety

Rule name	the tunnel construction rule
natural language description	the safety of any constructed tunnel is provided by the firefighting system
phrase	Tunnel safety firefighting Firefighting system
concepts	Tunnel Safety
Related properties	Variety of safety systems
Binary relationships	...
variables	(a) tunnel safety (b) firefighting systems

Instances Description: The related instances that exist in concept dictionary and instances table should be defined. In the description and definition of the instances, the following information should be expressed: such as: the name of the instance (East Azerbaijan earthquake), the instance property, values and the values of different properties (such as the date of incidence with value 2012-8-11; incidence time with value 16:53; human fatalities number with value 306 person; earthquake magnitude with value 6.4; earthquake depth with value 17).

Formalization

The formalization stage begins after conceptualization stage. Ontology expression through representation language is done by formalization. In this research, protégé software was used to formalize existing ontology, which supports by web ontology languages and the resource description framework.

Implementation and Maintenance

Stages 4 and 5, respectively, include ontology implementation and maintenance. Therefore, choosing an appropriate tool for the implementation of ontology is an important decision, because this decision has a great impact on the ontology life cycle. In fact, tools used to implement ontology concepts should also be able to support, develop and maintain ontology. Protégé software has been used for this purpose because of the ability to implement and maintain the ontology of this research. Figures 2 and 3 show examples of civil engineering ontology for the concept of “building materials” in version 5 of this software. Maintenance activities can also be carried out during the development of ontologies in the form of ontologies updating and editing. The ontology maintenance in the present research was that if, according to the experts' opinions, the concepts needed to be edited and updated, they were applied by the researchers at this stage.

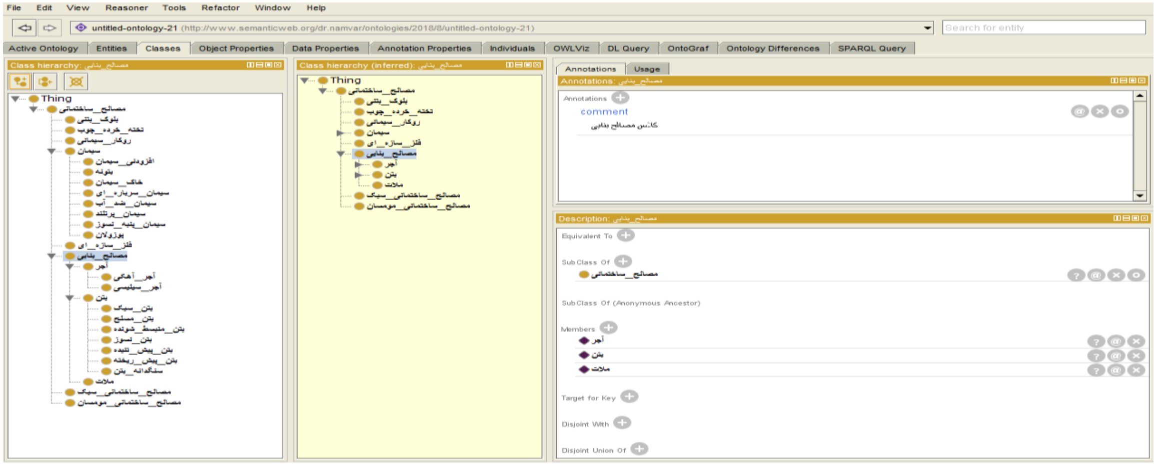

Figure 3. An Example of CivilOnto Ontology in Protégé Software

Figure 4 shows the conceptual structure of civil engineering in protégé software using OntoGraf. OntoGraf gives support for interactively navigating the relationships of your OWL ontologies. Various layouts are supported for automatically organizing the structure of your ontology.

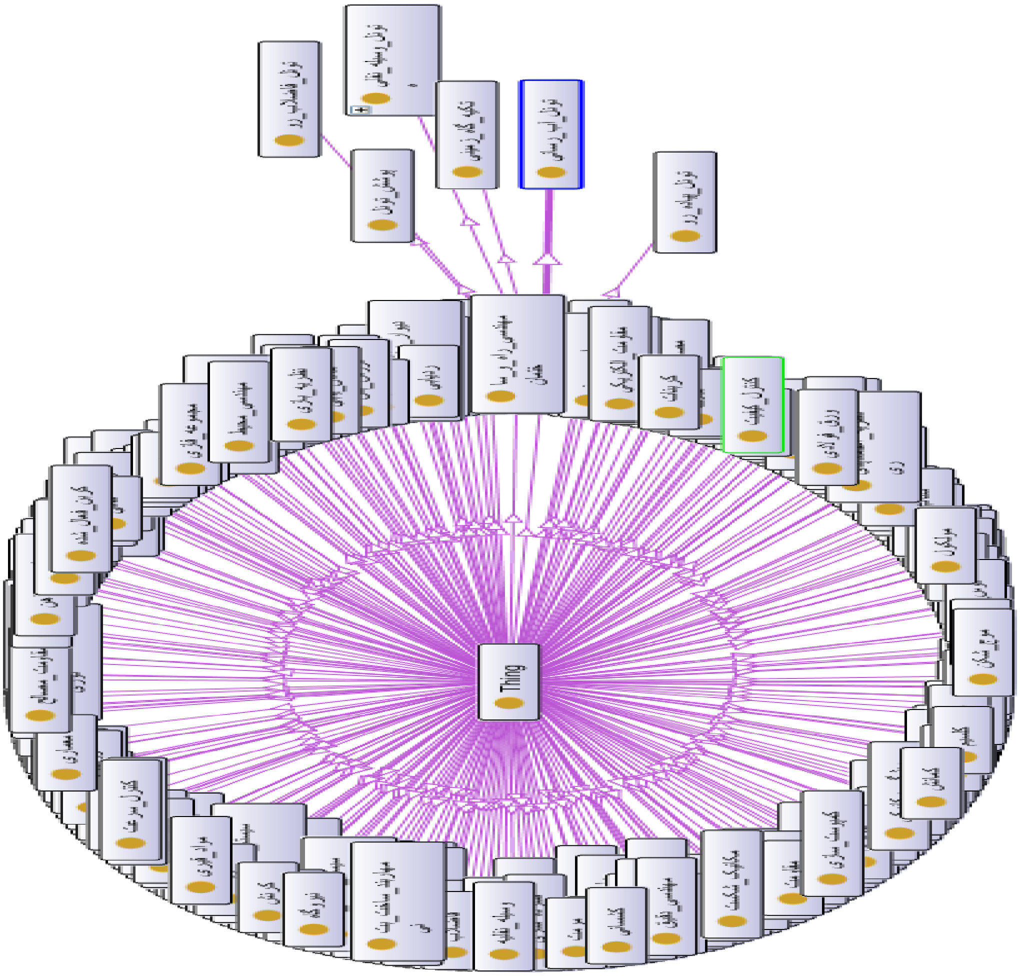

Figure 4. Conceptual structure of civil engineering using OntoGraf

Stage III: Supporting Activities

Supporting activities formed the sixth to tenth stages of ontology life cycle. Activities such as knowledge acquisition, integration, and evaluation are carried out during the ontology conceptualization stage. In fact, a large amount of knowledge is obtained at the primary stages of the ontology construction. The ontologies integration with each other led to the consolidation of the concepts and semantic relationships between them. Due to the lack of any other ontology in the civil engineering in Persian, integration in this research is not meaningful and it was not done. The ontology concepts evaluation can also be done in the conceptualization stage or during other stages by domain experts. In order to avoid possible errors, the process of ontology conceptualization and other stages must be evaluated. In the present research, 10 civil engineering experts were consulted at different stages, and in some cases, some editions were applied and then approved by them.

Documentation is another type of supporting activity. It seems that inadequate documentation is one of the main obstacles of effective knowledge sharing in the existing knowledge databases and ontologies. Therefore, all stages and assumptions that are made explicit must be recorded (Uschold & Gruninger, 1996). Ontologies can be reused if they are properly documented (Subhashini & Akilandeswari, 2011). All the performed stages in this research are recorded and documented.

Discussion and conclusion

In recent years, more attention has been paid to the use of thesaurus and controlled languages in ontologies (Namgoong & Kim, 2007; Bernstein & Kaufmann, 2006). The limited capability of subject headings and thesauri as tools for knowledge storage and retrieval in the precise presentation of concepts and the relationships between them can lead to the design of the ontologies (sanatjoo & Fathian, 2011). These constraints and attention to the semantic web have led to the concentration of researches on ontologies in different languages. Although there are researches in Persian language on the ontologies, such as ontology learning system called "Hasti" (Shamsfard, 2012) or the created ontologies in fields such as LIS (Zardari, 2016; Sanatjoo & Fathian, 2012; sanatjoo & Fathian, 2011; Khosravi & Vazifedoost ,2007; Nowruzi, 2015; Ahmadi, Osareh, Hosseini Beheshti, & Heidari, 2017; Mohammadi Ostani, Azargoon & Cheshmesohrabi, 2018; Basic Sciences (Hosseini Beheshti & Ejei, 2015), there is no ontology in the civil engineering domain .

The civil engineering ontology, entitled “CivilOnto”, is a domain ontology, for the design of which the ontology management tool (protégé) has been used. Terms have been extracted from titles and keywords of indexed articles in the RICeST database using SQL queries. Its conceptual model was also developed using the content analysis. The face and content validity of classes and subclasses of concepts was also reviewed and approved by 10 civil engineering experts. The METHONTOLOGY was used for CivilOnto creation. Ontology relationships are hierarchically arranged, and classes are divided into subclasses. The main classes with subclasses produce a tree of the relationships of their category members.

The results showed that the civil engineering conceptual model includes 283 concepts, along with 62 object properties, 79 data type properties, 10 annotation properties and 976 instances. Most of these properties are related to data type properties, and the least is also annotative properties. The increase in the number of properties was due to the fact that authors did not remove any properties from text sources and their actual mentions were based on the extracted data. The properties increases in data type in the present study indicate the attention to the entity property with its data values; in total, in the present research, 151 semantic properties were identified among the concepts.

The present research is aimed at creating a domain ontology. In this sense, it has similarity with researches by Zardari, 2016; Hosseini Beheshti & Ejei, 2015; Fathian & Sanatjoo, 2012; Aydin & Tecin, 1970; Ahmadi, Osareh, Hosseini Beheshti, & Heidari, 2017; Mohammadi Ostani, Azargoon & Cheshmesohrabi, 2018; Dumontier & Villanueva-Rosales, 2009; Elag & Goodall, 2013; Teller, Keita, Roussey & Laurini, 2015; and, in the sense of methodology and using the METHONTOLOGY, it has similarity with Zardari, 2016; Hosseini Beheshti & Ejei, 2015; and Aydin & Tecin, 1970. Also, in sense of using information texts and resources in ontology creation, it is in line with researches of Mohammadi Ostani, Azargoon & Cheshmesohrabi’s , 2018; Ahmadi, Osareh, Hosseini Beheshti, & Heidari, 2017; Dumontier & Villanueva-Rosales, 2009; Teller, Keita, Roussey & Laurini, 2015; Elag & Goodall, 2013; Aydin & Tecin, 1970; but, its method and data extraction methodology is different from these researches. In which, data extraction is done through SQL queries and the qualitative content analysis with summative approach was used for relations and properties analysis. It is while the mentioned researches have not used the above mentioned. The present research is similar to the mentioned researches in terms of using protégé software. This is the first civil engineering ontology, so it is impossible to compare the concepts, instances and properties of CivilOnto with these researches.

One of the usages of this research is to model it for other RICeST engineering databases, support Persian language in civil engineering, create semantic relations among indexed documents of RICeST and eliminate the heterogeneity between these documents. This can lead to semantic retrieval and receive more useful results from these databases. Also, due to RICeST database updating (including civil engineering section) and considering that this ontology can be expanded and developed (deletion, addition or edition) in the future, it can be served as a basis for ontology's completion in this domain.

The results of this research can serve as a guide for the RICeST experts in order to identify the users' need and facilitate the users' interaction with RICeST databases. This could also help users provide access to more relevant resources when searching in these rich databases. Therefore, it seems that the creation of such specialized ontologies as a strong support for these databases can improve databases knowledge representation and serve as a layer of semantic web in these databases. Using this ontology in RICeST databases (civil engineering section), can help eliminate the semantic levels errors when users search and retrieve content from these databases. This ontology can also be modeled for concepts organization and representing relationships among them in other RICeST databases, and could be an appropriate tool for mapping an ontology space for other specialized domain databases.

Endnotes

1. Select keyword from journal join article on JID = doc ID, join subject on JID = ID, where subject =' civil engineering'
2. Select title from journal join title on JID = ID ,join subject on JID = ID, where subject =' civil engineering'
3. A few examples of classes, instances and relationships have been mentioned due to the abundance of them in this ontology.
4. These relationships can be also used for object or data type relationships.

References

Ahmadi, H., Osareh, F., Hosseini Beheshti, M. S., & Heidari, GH. (2017). Designing semiautomatic system in ontology structure by to co-occurrence word analysis and c-value method (case study: The field of scientometrics of Iran). *Iranian Journal of Information Processing and Management*, 33(1), 185-216. [in Persian]

Andreou, A. (2005). *Ontologies and query expansion*. M.S. thesis, Computer engineering, School of Informatics, Edinburgh University.

Asosheh, A., Mehrasa, SH., Khatibi, T., & Khakshour Saadat, P. (2011). *Ontology: methodologies, language & tools*. Tehran: Tarbiat Modares University Press. [in Persian]

Aydin, C., & Tecin, V. (1970). Research on geo-ontologies at earthquake domain for disaster management on urban areas. In C. Ellul., S. Zlatanova & R. Laurini (Eds), *Urban and regional data management.* London: Taylor & Francis Group.

Berelson, B. (1952). Content analysis in communication research. Glencoe, IL: Free Press.

Bernstein, A., & Kaufmann, E., (2006). GINO – A guided input natural language ontology editor. In: Cruz I. et al. (Eds). *The semantic web* – ISWC 2006. Lecture notes in computer science, 4273. Berlin: Heidelberg. https://doi.org/10.1007/11926078_11

Brank, J., Grobelnic, M., & Mladenic, D. (2005). A survey of ontology evaluation techniques. In proceedings of the conference on data mining and data warehouses (SiKDD 2005). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi= 10.1.1.101.4788&rep=rep1&type=pdf. (Accessed May 4, 2017).

Brusa, g., Caliusco, M. L., & Chiotti, O. (2006). A process for building a domain ontology: an experience in developing a government budgetary ontology. Paper presented at proceedings of the Australian workshop on *advances in ontologies*, Hobart, Australia: Australian Computer Society, 72, 7-15.

Chen, M. (2009). *Exploring the use of ontological relations in information retrieval* .Posters presented at the iConference. Retrieved from https://core.ac.uk/download/ pdf/4824195.pdf

Chun, C., & Wenlin, L. (2004). From agricultural thesaurus to ontology. In 5th AOS workshop, April 27-29. Beijing, China. Retrieved frim http://www.fao.org/agris/aos/

Dacunta, m. c., Obrst, L. J., & Smith, K. T. (2003). *The semantic web: A guide to the future of XML web services and knowledge management*. Indianpolis: John Wiley.

Dan, W. U., & Wang, H. L. (2006). Role of ontology in information retrieval. *Journal of Electronic Science and Technology of China*, 4(2), 148-154.

Dumontier, M., & Villanueva-Rosales, N. (2009). Towards pharmacogenomics knowledge discovery with the semantic web. *Briefings in Bioinformatics*, 10 (2), 153-163.

Ejei, F., Hosseini Beheshti., M. S., Rajabi, T & Ejehi, Z. (2017). Enriching semantic relations of basic sciences ontology. *Knowledge Organization.* 44 (5), 325-318. Available at https://irandoc.ac.ir/sites/fa/files/attach/article/ko4420175b.pdf. [in Persian]

Elag, M. & Goodall, J. L. (2013). An ontology for component-based models of water resource systems. *Water Resources Research*. 49(8), 5077-5091.

Fathian, A. (2010). *The comparison of thesaurus and ontology efficiency in knowledge representation*. M.A thesis. Knowledge and information science, Faculty of education and psychology, Ferdowsi University. [in Persian]F

Fernandes Lopez, F., Gomez-Perez, A., & Juristo, N. (1997). Methontology: From ontological art towards ontological engineering. *In proceedings of the ontological engineering AAAI-97*, California: Stanford University. Retrieved from http://oa.upm.es/5484/1 /METHONTOLOGY_.pdf

Gauch, S., Chaffee, J., & Pretschner, A. (2003). Ontology-based personalized search and browsing. *Web Intelligence and Agent Systems: An international journal*, 1(3, 4), 219–234.

Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2), 199-220. https://doi.org/10.1006/knac.1993.1008

Hosseini Beheshti, M. S. (2014). Morphology, Terminology and Knowledge Engineering. Tehran: Irandoc: Chapar [In Persian]

Hosseini Beheshti, M. S., & Ejei, F. (2015). Designing and implementing basic sciences ontology based on concepts and relationships of relevant thesauri. *Iranian Journal of Information Processing and Management*, 30(3), 677-696. [in Persian]

Hsieh, H., Shannon, S. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15 (9), 1277-1288.

Huang C.R., Chung S.F., Ahrens, K. (2007). An Ontology-Based Exploration of Knowledge Systems for Metaphor. In Sharman R., Kishore R., Ramesh R. (Eds) *Ontologies: A handbook of principles, concepts and applications in information systems* (489-517). Integrated Series in Information Systems, vol 14. Springer, Boston, MA

Khosravi, F., & Vazifedoost, A. (2007). Creating a Persian ontology through thesaurus reengineering for organizing the Digital Library of the National Library of Iran. In International Conference on Libraries, Information and Society, ICoLIS (pp. 19-36). Petaling Jaya: Malaysia

Lee, K., Rho, S., & Lee, S. W. (2014). A method of extracting ontology module using concept relations for sharing knowledge in mobile cloud computing environment. *The Scientific World Journal*. http://dx.doi.org/10.1155/2014/382797

Mayring, P. (2004) Qualitative content analysis. InU. Flick, E. von Kardoff and I. Steinke (eds) *A Companion to Qualitative Research*. London: Sage.

Mezghanni, I. B., & Gargouri, F. (2017). Deriving ontological semantic relations between Arabic compound nouns concepts. *Journal of King Saud University –Computer and Information Sciences*, 29 (2), 212-228. http://doi.org/10.1016/j.jksuci.2017.03.001

Mirzabeigi, M. (2012). The role of ontology in information retrieval: Reviewing current research and representing a conceptual model. *Iranian Journal of Information processing and management*, 27(2), 839-855. [in Persian]

Mohammadi Ostani, M., Azargoon, M., & Cheshmesohrabi, M. (2018). Methodology of construction and design ontologies: A case study of scientometrics field. *Iranian journal of Information Processing and Management*. 33 (4), 1739-1822. [in Persian]

Namgoong, H & Kim, H. G. (2007). Ontology-based controlled natural language editor using CFG with lexical dependency. *In Proc. ISWC/ASWC, 4825, Lecture Notes in Computer Science*, K, Abrerer, et al (Eds), 353-366.

Nowroozi, M. (2015). *The comparison of thesaurus and ontology efficiency in semantic concepts and relations representation (case study: ASIS thesaurus and designed ontology*. M. A. thesis. Faculty of education and psychology, Shiraz University. [In Persian]

Nowroozi, M., Mirzabeigi, M., & Sotudeh, H. (2018a). Constructing an ontology based on a thesaurus: A case of ASIS&T Onto based on the ASIS&T web-based thesaurus. *The Electronic Library*, 36 (4), 750-764. http://doi.org/10.1108/el-02-2017-0037

Nowroozi, M., Mirzabeigi, M., & Sotudeh, H. (2018b). The comparison of thesaurus and ontology: Case of ASIS&T web-based thesaurus and designed ontology. Library Hi Tech, 36 (4), 665-684, DOI: https://doi.org/10.1108/LHT-03-2017-0060

Pinto, H.S., & Tempich, C. (2004). DILIGENT: Towards a fine-grained methodology for distributed, loosely-controlled and evolving engineering of ontologies. *In Artificial Intelligence (ECAI 2004).* Paper presented at the 16th European conference on artificial intelligence, 393-397, Spain: IOS press. Retrieved from http:// www.frontiersinai.com/ecai/ecai2004/ecai04/pdf/p0393.pdf. (Accessed September 25, 2017).

Sanatjoo, A. (2006). Necessity of the revision in thesauri structures: a review of thesauri inefficiency in the new information environment and ontology abilities in comparison with them. *National Studies on Librarianship and Information Organization*, 16(4), 79-92. [in Persian]

Sanatjoo, A., & Fathian, A. (2011). The comparison of thesaurus and ontology efficiency in knowledge representation. *Library and Information Science Research*, 1(1), 219-240. [in Persian]

Sanatjoo, A., & Fathian, A. (2012). Methodology of ontology designing, creating and implementation: approaches, languages and tools (case study: ASFAINT designing in the library and information science domain). *Library and Information Sciences*, 15(1), 113-142. [In Persian]

Shamsfard, M. (2012). *Creating semiautomatic of Persian WordNet: from word to word-ontology*. Paper presented at the proceeding of the *2th Computational Linguistics Seminar*. [In Persian]

Shamsfard, M., Abdullahzadeh Barforoush, A. (2002). Extracting conceptual knowledge from text: Using linguistic and semantic templates. *Advances in Cognitive Science*, 4(1), 48-66. [In Persian]

Sharif, A. (2006). A cognition of semantic relations in the web ontology. *Informology,* 4(1-2), 65-84, (In Persian).

Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., & Katz, S. (2004). Reengineering thesauri for new applications: The AGROVOC example. *Journal of digital information*. 4 (4), 1-23. Retrieved from https://journals.tdl.org/jodi/index.php/jodi /article/view/112/111.

Subhashini, R., & J, Akilandeswari. (2011). A survey on ontology construction methodologies. *International Journal of Enterprise Computing and Business Systems*. 1(1), 60–72.

Teller, J., Keita, A. K., Roussey, C & Laurini, R. (2015). *Urban ontologies for an improved communication in urban civil engineering projects*. Paper presented at the *COST Urban Civil Engineering Action C21 "TOWNTOLOGY".*

Uschold, M., & Gruninger, M. (1996). Ontologies: principles, methods and application. *Knowledge Engineering Review*, 11 (2), 93-136. https://doi.org/10.1017/ So269888900007797

Zardari, S. (2016). *Ontology engineering of knowledge and information science based on library and information encyclopedia*. Ph.D thesis. Faculty of education and psychology. Shahid Chamran University. [in Persian]